Interpreting patterns in transcriptomes and ancient DNA from Next Generation Sequencing data

Part I: Endogenous retroviruses and human genetic diseases

Part II: Using large-scale ancient animal DNA datasets to address fundamental evolutionary questions

A thesis submitted for the
Degree of Doctor of Philosophy



Trinity Term 2019

Audrey Tseng Lin
Green Templeton College

Abstract

A goal of biological research is to understand the implications of the evolutionary process. Efforts using direct analysis of nucleotide sequences have had a tremendous impact in achieving this goal. The revolution in palaeogenomics and transcriptomics research over the past decade is inextricably intertwined with the power of next generation sequencing platforms and the ability to target and sequence short DNA molecules at a high resolution. In palaeogenomics, archaeological specimens that were previously unsuitable for DNA analyses due to extensive postmortem sequence damage can now be used successfully as source material. The increasing availability of ancient DNA allows for greater precision and accuracy in the molecular dating of phylogenetic trees, and provides invaluable insights into the timing, location, mechanisms, and dynamics of important evolutionary processes such as domestication. In gene expression data, the sensitivity of RNA-seq allows for the potential detection of alternative splice isoforms of transcripts as well as rare and context-specific transcripts, such as those seen in complex tumour environments. In this thesis, I use bioinformatic techniques to 1) evaluate the expression of endogenous retrovirus-encoded putative oncogenes *np9* and *rec* in RNA-seq data, and find that these gene products are upregulated in many cancers, and in specific stages of early embryonic development; 2) find that *np9* and *rec* are positively and negatively correlated with the upregulation and downregulation of many cancer stem cell and epithelial-mesenchymal transition biomarkers; 3) utilise ancient and modern animal mitochondrial genomes to find evidence of time-dependence evolutionary rate estimates, showing that the degree of this effect varies among parts of the mitochondrial genome and with the estimation method being used; 4) reconstruct early animal domestication processes of pigs in Europe.

Acknowledgements

I'd first like to extend my thanks to my dear friends and colleagues in Zoology and the Medawar for the helpful discussions that greatly improved my thesis, and for their support. Thank you to the members Palaeovirology group, past and present, including Pakorn Aiewsakun, Amr Aswad, Samuel Campbell, Nico Kist, Fabricia Nascimento, Emilia Skirmuntt; and the former Applied Evolutionary Virology group – Timokratis Karamitros, Tara Hurst, Charlie Ince, Cindy Santander, and Tetyana Vasylyeva. Without these folks and the knowledge and insight from Aris Katzourakis and Gkikas Magiorkinis, all of the hard work of Part I of my thesis would not exist. Many thanks also to Emanuele Marchi and Adrian Smith for the helpful discussions during my transfer which greatly improved my first data chapter.

I would most like to thank my supervisor Greger Larson, for welcoming me into the group so warmly and for his bottomless support, passion, and enthusiasm over the past two and a half years. I am most thankful for Greger and Laurent Frantz for having confidence in me and involving me in such amazing projects I never imagined I'd be a part of. My experience working on the piggy paper, a huge and richly collaborative undertaking, is one of the highlights of my DPhil. My time in Oxford has also been greatly enriched by discussions and interactions with the members of Palaeo-BARN, past and present – including but not limited to Joel Alves, Lisa Briggs, Antony Dimopoulos, Evan Irving-Pease, Alex Jamieson, Ophélie Lebrasseur, and Hannah Ryan. Thank you also for being such fun company, especially at conferences.

Thank you to my co-supervisor Tim Coulson for being so supportive and for ensuring that the rest of the DPhil goes as smoothly as possible. Thank you to Simon Ho for being a wonderful source of help and knowledge for my time-

dependency chapter. Thank you to Louis du Plessis for his tremendous patience, for all the inspiring discussions, and for being one of my favourite human beings.

Finally, I'd like to thank Alison Stenton, Marella de Bruijn, Chris Sauer, and Paul Giangrande from Green Templeton College for their patience, kindness, empathy, and unwavering support during my entire time as a student here at Oxford.

LLAP.

Dedication

I dedicate this brick of work to my dad (1949-2015), my cat Vasily (2011-), and Spock (2230-2285; 2285-2263).

> "You could tell by the way [she] talked, though, that [she] had gone to school a long time. That was probably what was wrong with [her]."
> — John Kennedy Toole, **A Confederacy of Dunces**

Contents

ABSTRACT **2**

ACKNOWLEDGEMENTS **3**

DEDICATION **5**

CONTENTS **6**

STATEMENT OF CONTRIBUTION AND ASSOCIATED PUBLICATIONS **15**

GLOSSARY **21**

PART I: ENDOGENOUS RETROVIRUSES AND HUMAN GENETIC DISEASES 21

PART II: USING LARGE-SCALE ANCIENT ANIMAL DNA DATASETS TO ANSWER FUNDAMENTAL EVOLUTIONARY QUESTIONS 24

PART I: ENDOGENOUS RETROVIRUSES AND HUMAN GENETIC DISEASES **26**

CHAPTER 1, INTRODUCTION **27**

"ENDOGENOUS RETROVIRUSES AND HUMAN GENETIC DISEASES" **27**

ABSTRACT 28

1. INTRODUCTION 29

2. RETROVIRUSES AND CONSEQUENCES OF ENDOGENISATION 29

2a. Retrovirus biology *29*

2b. Retroviral endogenisation *31*

2c. Deactivating mutations and deletions *31*

2d. Epigenetic silencing *32*

2e. ERV expansion through host co-option *32*

2f. ERV classification *34*

2g. HERV-L, H, W and FRD *35*

2h. MaLRs *35*

2i. ERV9 *35*

2j. HERV-K (HML-2) *36*

Figure 1.1. Genome schematics of endogenous retrovirus and mammalian apparent LTR retrotransposon 37

3. HERVS AND HOST IMMUNITY 37

Figure 1.2. Model of HERV-K (HML-2) in human innate immune function. 40

Table 1.1. HERVs and host immunity 42

4. HERVS AND CANCER 43

4a. HERV-K and cancer *44*

4b. MaLRs and cancer *45*

4c. Np9 and Rec *46*

Table 1.2 HERVs and cancer 47

5. HERVS AND AUTOIMMUNE DISEASE 48

5a. Rheumatoid arthritis, lupus, and psoriasis *48*

5b. Amyotrophic lateral sclerosis *49*

Figure 1.3. Model of HERV-K expression and amyotrophic lateral sclerosis pathogenesis 51

5c. Multiple sclerosis *51*

Figure 1.4. Model HERV-W Env and multiple sclerosis pathogenesis 52

Table 1.3 HERVs in autoimmune disease 54

6. HERVS AND NEUROPSYCHIATRIC DISORDERS 56

6a. Schizophrenia and bipolar disorder *56*

6b. Alcoholism *58*

Table 1.4 HERVs and neuropsychiatric disorders 58

8. KEY CONCEPTS 62

10. FURTHER READING LIST 63

CHAPTER 2, "TARGETING ENDOGENOUS RETROVIRUS GENE TRANSCRIPTION IN HUMAN CANCERS." 65

Abstract *67*

1. Introduction *68*

2. Materials and Methods *71*

2a. cDNA reference sequences 71

Figure 2.1. Workflow describing method in measuring expression of *np9* and *rec* in RNA-seq data 72

2b. Alignment *73*

2c. Filtering for np9 and rec-specific aligned reads *73*

Figure 2.2. Sequenced cDNA reads from primary tumours mapped to HK2 targets as viewed on Integrative Genome Browser (IGV) 74

2d. Quantifying depth of coverage of HK2 *75*

2e. Normalisation of HK2 transcripts by actb transcripts *75*

2f. Analysis of publicly available raw data *75*

Table 2.1. Summary of RNA-seq data from solid primary tumour/non-diseased tissue 76

2g. Non-diseased reference tissue selection *77*

Table 2.2. Summary of groups of non-diseased reference tissue samples used for analysis 78

2h. Calculation of relative difference in expression between np9 and rec in cancers compared to non-diseased tissue reference group *78*

2i. Statistical Analysis *79*

3. Results *80*

3a. Describing HK2 expression in different categories of non-diseased tissue *80*

Figure 2.3. Describing HK2 expression in different categories of non-diseased tissue 81

Table 2.3. Relative expression of HK2 in different non-diseased tissue group categories compared to non-diseased tissue reference (NT) 81

3b. HK2 expression in early embryogenesis and stem cell lines *83*

Figure 2.4. Expression of HK2 targets in human preimplantation embryos and embryonic stem cells 84

3c. HK2 expression in non-diseased tissues and cancer samples *85*

Table 2.4. Relative expression of HK2 targets in cancers compared to NT 85

a) *np9* 85

Figure 2.5. Expression of HK2 genes in non-diseased reference tissue group (NT) compared to 15 different cancers 90

4. Discussion *91*

5. Acknowledgements *96*

6. Funding Statement *96*

7. Ethics Approval and Consent to Participate *96*

8. Supplement *97*

Figure S2.1. Partial schematics of HERV-K (HML-2) proviruses encoded gene products 98

Table S2.1. HERV-K (HML-2) reference sequences used 98

S2.2 Supplementary spreadsheet 99

CHAPTER 3, "HERV-K HML-2 TRANSCRIPTION IN DIVERSE CANCERS IS RELATED WITH CANCER STEM CELL AND EPITHELIAL-MESENCHYMAL TRANSITION MARKERS." **127**

ABSTRACT 129

1. INTRODUCTION 130

2. MATERIALS AND METHODS 132

2.1 Data Analysed *132*

Table 3.1. Summary of The Cancer Genome Atlas (TCGA) RNA-seq data 133

2.2 Testicular germ cell cancer as cancer stem cell reference *133*

2.3 Analysis of EMT and CSC genes from TCGA datasets *134*

Table 3.2. Lists of gene groups analysed 134

2.4 Calculation of transcriptional distance of HK2 genes from transcriptional profile of all EMT/CSC genes *135*

2.5 Statistical Analysis *137*

3. Results *138*

3.1 HK2 expression is high in cells with CSC/EMT-like transcriptional profiles *138*

Figure 3.1. HK2 upregulation in cancers in comparison to epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) gene expression signature 139

3.2 Independent correlates of HK2 expression and CSC/EMT gene expression *139*

Table 3.3. Chi-square contingency tables showing significant correlations between different functional gene categories and transcription of *rec* and *np9* 140

Figure 3.2. Breakdown of different gene categories that *np9* and *rec* are significantly correlated with 143

Table 3.4. *np9* and *rec* correlates with CSC/EMT genes 143

4. Discussion *148*

Figure 3.4. Model of *np9* and *rec* transcriptional activity during epithelial-mesenchymal transition (EMT) 152

Figure 3.5. Host gene function categories significantly correlated with HK2 gene expression 154

5. Acknowledgements *154*

6. Funding Statement *154*

7. Ethics Approval and Consent to Participate *154*

8. Supplement *155*

Table S3. EMT and CSC upregulated and downregulated genes 155

CHAPTER 4, "DISCUSSION AND CONCLUSIONS CONCERNING PART I" **159**

1. CONCERNING CHAPTER 1: ENDOGENOUS RETROVIRUSES AND HUMAN GENETIC DISEASES 159

2. CONCERNING CHAPTER 2: TARGETING ENDOGENOUS RETROVIRUS GENE TRANSCRIPTION IN HUMAN CANCERS 160

Limitations and future directions *161*

3. CONCERNING CHAPTER 3: HERV-K HML-2 TRANSCRIPTION IN DIVERSE CANCERS IS RELATED WITH CANCER STEM CELL AND EPITHELIAL-MESENCHYMAL TRANSITION MARKERS 163

Limitations and future directions *165*

4. CONCERNING CHAPTER 9 (APPENDIX): HUMAN ENDOGENOUS RETROVIRUS-K HML-2 INTEGRATION WITHIN *RASGRF2* IS ASSOCIATED WITH INTRAVENOUS DRUG ABUSE AND MODULATES TRANSCRIPTION IN A CELL-LINE MODEL 166

Limitations and future directions *167*

5. CONCLUSIONS 168

PART II: USING LARGE-SCALE ANCIENT ANIMAL DNA DATASETS TO ADDRESS FUNDAMENTAL EVOLUTIONARY QUESTIONS **171**

CHAPTER 5, INTRODUCTION **172**

"ANCIENT DNA FOR EVOLUTION-FOCUSED ENQUIRY" **172**

1. RATES OF EVOLUTION 173

1a. The molecular clock *173*

1b. Evolutionary rate variation *174*

1c. Time-dependent molecular evolution *175*

Figure 5.1. A schematic of the declining rate curve of time-dependent molecular evolution on a log-log chart 176

Figure 5.2. Time-dependent patterns in evolutionary rate estimates have been observed in a variety of taxa 177

1d. Calibrating the molecular clock *177*

Figure 5.3. A schematic demonstrating the value of increasing the sample size of dated samples in a molecular phylogeny 179

2. ANIMAL DOMESTICATION 180

2a. The concept of domestication *180*

2b. Early domestication *181*

2c. Using ancient DNA to study domestication *183*

2d. Mitochondrial genomes in ancient DNA studies *186*

2e. Nuclear genomes in ancient DNA studies *186*

3. ORIGINS, EXPANSION, AND GENETICS OF ANIMAL DOMESTICATION 187

3a. Case study: Pigs *187*

3b. Case study: Goats *188*

3c. Case study: Horses *189*

3d. Case study: Dogs *191*

4. Conclusions *194*

CHAPTER 6, "TIME-DEPENDENT ESTIMATES OF EVOLUTIONARY RATES IN ANCIENT ANIMAL MITOGENOMES." 196

ABSTRACT 198

1. INTRODUCTION 199

2. MATERIALS AND METHODS 203

2.1 Data collection and organisation *203*

Figure 6.1 Plot showing the temporal distribution of all ancient and modern samples used for analysis 205

Table 6.1. Number of samples in each taxon per each sampling window (in years) 207

2.2 Date-randomisation tests *208*

2.3 Rapid dating methods *210*

2.4 Bayesian inference *212*

3. RESULTS 213

3.1 Date-randomisation test results *213*

Table 6.2 Date-randomisation test results for full mitogenomes of 7 taxa 214

3.2 Difference in rate estimates between the sampling windows *215*

3.3 Difference in rate estimates among partitions and methods *216*

Figure 6.2 Substitution rate estimates for full mitogenomes from seven taxa with seven different sampling time windows 220

Figure 6.3 Substitution rate estimates for D-loop generated from seven taxa with seven different sampling time windows 222
Figure 6.4 Substitution rate estimates for 1st codon sites (chicken, goat, horse) or 1st+2nd codon sites (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows 225
Figure 6.5 Substitution rate estimates from 2nd codon sites (chicken, goat, horse) or 1st+2nd codon sites (bison, dog, pig, sheep) from seven taxa with seven different samping time windows 227
Figure 6.6 Substitution rate estimates from 3rd codon sites from seven taxa with seven different sampling time windows 230
Figure 6.7 Substitution rate estimates for rRNA (chicken, goat, horse) or all RNA (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows 232
Figure 6.8 Substitution rate estimates for tRNA (chicken, goat, horse) or all RNA (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows 235
Figure 6.9 Summary of substitution rate estimates from the D-loop (a) and full mitogenome (b) generated by Treedater for seven taxa with seven different sampling time windows, scaled by the generation time for each taxon. 236
Figure 6.10 Summary of t2 rate estimates that are lower than t1 rate estimates 238
Figure 6.11 Regression of mean rate estimates from all mitogenome partitions generated from Treedater 239
Figure 6.12 Summary of bison t2 rate estimates of all mitogenome partitions from BEAST2, Treedater, and Treetime 240
Figure 6.13 RAxML tree of chicken dataset with red jungle fowl as outgroup 241
4. DISCUSSION 242
4.1 Purifying selection *243*
4.2 Ancient DNA and tip-calibration *244*
4.3 Postmortem damage *246*
4.4 Bayesian vs. rapid dating methods *249*
5. CONCLUSIONS 251
6. SUPPLEMENT 252

CHAPTER 7, "ANCIENT PIGS REVEAL A NEAR COMPLETE GENOMIC TURNOVER FOLLOWING THEIR INTRODUCTION TO EUROPE" 253

CHAPTER 8, "DISCUSSION AND CONCLUSIONS CONCERNING PART II" 259

1. NEXT STEPS FOR CHAPTER 6 259

2. DISCUSSION: RETRIEVAL OF ANCIENT DNA 262

4. DISCUSSION: RECONSTRUCTING PATTERNS OF PAST POPULATION MOVEMENT 264

5. DISCUSSION: RECONSTRUCTION OF SPECIATION HISTORY 265

6. DISCUSSION: HISTORICAL VIRUS EPIDEMIOLOGY 266

8. CONCLUSIONS 272

APPENDIX 275

CHAPTER 9, "HUMAN ENDOGENOUS RETROVIRUS-K HML-2 INTEGRATION WITHIN RASGRF2 IS ASSOCIATED WITH INTRAVENOUS DRUG ABUSE AND MODULATES TRANSCRIPTION IN A CELL-LINE MODEL" 275

CHAPTER 10, "DIRE WOLVES WERE THE LAST OF AN ANCIENT NEW WORLD CANID LINEAGE" 277

CHAPTER 11, "SPECIALISED SLEDGE DOGS AIDED INUIT DISPERSAL ACROSS THE NORTH AMERICAN ARCTIC" 279

BIBLIOGRAPHY 281

Statement of Contribution and Associated Publications

The work presented in this thesis includes research that has been published in journals, papers in various stages of the publication process, and papers that are currently in preparation. Here I present a brief summary of my contributions to each chapter, and those of others. Asterisks (*) in authorships of papers denote equal contributions.

Chapter 1

Audrey T. Lin, *and Gkikas Magiorkinis. "Role of Endogenous Retroviruses in Human Genetic Diseases." eLS (2016): 1-17. doi:10.1002/9780470015902.a0026711*

This chapter is wholly my own work. I wrote the manuscript and Gkikas Magiorkinis supervised the project and provided editorial assistance.

Chapter 2

Audrey T. Lin, *Cindy G. Santander, Emanuele Marchi, Timokratis Karamitros, Aris Katzourakis, and Gkikas Magiorkinis. "Targeting endogenous retrovirus gene transcription in human cancers." bioRxiv (2018): 449686. doi: https://doi.org/10.1101/449686*

I wrote the manuscript and performed all the computational work and analyses. The bioinformatic method development was done with assistance from the other coauthors. Gkikas Magiorkinis supervised the project and provided editorial assistance.

Chapter 3

Audrey T. Lin, *Cindy G. Santander, Fabricia F. Nascimento, Emanuele Marchi, Timokratis Karamitros, Aris Katzourakis, and Gkikas Magiorkinis. "HERV-K HML-2 transcription in diverse cancers is related with cancer stem cell and epithelial-mesenchymal transition markers." bioRxiv (2018): 451997. doi: https://doi.org/10.1101/451997*

I wrote the manuscript and performed all the computational work and analyses. Gkikas Magiorkinis supervised the project and provided editorial assistance.

Chapter 4

Audrey T. Lin. *"Discussions and Conclusions concerning part I."*

This chapter is wholly my own work. I wrote the chapter and Gkikas Magiorkinis supervised the projects that I write about. Tim Coulson advised and provided editorial assistance.

Chapter 5

Audrey T. Lin. *"Ancient DNA for evolution-focused enquiry."*

This chapter is wholly my own work. I wrote the chapter and Greger Larson supervised the projects that I wrote about. Simon Ho, Tim Coulson, and Greger Larson advised and provided editorial assistance.

Chapter 6

Audrey T. Lin, *Laurent A.F. Frantz, Louis du Plessis, Dan Bradley, Ophélie Lebrasseur, Ivica Međugorac, Ludovic Orlando, Beth Shapiro, Simon Y.W. Ho, and Greger Larson. "Time-dependent estimates of evolutionary rates in ancient animal genomes."*

This chapter is wholly my own work, with input from Simon Ho, Greger Larson, Laurent Frantz, and Louis du Plessis. Greger Larson and Simon Ho

supervised the project. I performed all the computational work and analyses. I wrote the chapter and Simon Ho and Laurent Frantz provided editorial assistance.

Chapter 7

Laurent A. F. Frantz, James Haile*, **Audrey T. Lin***, Amelie Scheu, Christina Geörg, Norbert Benecke, Michelle Alexander, Anna Linderholm, Victoria E. Mullin, Kevin G. Daly, Vincent M. Battista, Max Price, Kurt J. Gron, Panoraia Alexandri, Rose-Marie Arbogast, Benjamin Arbuckle, Adrian Bălăşescu, Ross Barnett, László Bartosiewicz, Gennady Baryshnikov, Clive Bonsall, Dušan Borić, Adina Boroneanț, Jelena Bulatović, Canan Çakirlar, José-Miguel Carretero, John Chapman, Mike Church, Richard Crooijmans, Bea De Cupere, Cleia Detry, Vesna Dimitrijevic, Valentin Dumitraşcu, Louis du Plessis, Ceiridwen J. Edwards, Cevdet Merih Erek, Aslı Erim-Özdoğan, Anton Ervynck, Domenico Fulgione, Mihai Gligor, Anders Götherström, Lionel Gourichon, Martien A.M. Groenen, Daniel Helmer, Hitomi Hongo, Liora K. Horwitz, Evan K. Irving-Pease, Ophélie Lebrasseur, Joséphine Lesur, Caroline Malone, Ninna Manaseryan, Arkadiusz Marciniak, Holley Martlew, Marjan Mashkour, Roger Matthews, Giedre Motuzaite Matuzeviciute, Sepideh Maziar, Erik Meijaard, Tom McGovern, Hendrik-Jan Megens, Rebecca Miller, Azadeh Fatemeh Mohaseb, Jörg Orschiedt, David Orton, Anastasia Papathanasiou, Mike Parker Pearson, Ron Pinhasi, Darko Radmanovic, François-Xavier Ricaut, Mike Richards, Richard Sabin, Lucia Sarti, Wolfram Schier, Shiva Sheikhi, Elisabeth Stephan, John R. Stewart, Simon Stoddart, Antonio Tagliacozzo, Nenad Tasić, Katerina Trantalidou, Anne Tresset, Cristina Valdiosera, Youri van den Hurk, Sophie Van Poucke, Jean-Denis Vigne, Alexander Yanevich, Andrea Zeeb-Lanz, Alexandros Triantafyllidis, M. Thomas P. Gilbert, Jörg Schibler, Peter Rowley-Conwy, Melinda Zeder, Joris Peters, Thomas Cucchi, Daniel G. Bradley, Keith Dobney, Joachim Burger, Allowen Evin, Linus Girdland-Flink, and Greger Larson. "Ancient pigs reveal a near complete genomic turnover following their introduction to Europe." Proceedings of the National Academy of Sciences of the United States of America 116, no. 35 (2019):17231-17238. doi:10.1073/pnas.1901169116*
*denotes equal contribution

In this paper, I organised and compiled the database used for the analysis (Table S1). I created the maps seen in Fig 1 and Fig S6 demonstrating the biogeographical boundaries of these Asian, Near Eastern, European, and Y2

haplogroups. Maps were also drawn demonstrating the overall frequencies of these haplogroups in domestic pigs at various archaeological sites/locations between 8,000 BP and 5,100 BP; 5,099 BP and 180 BP; and in modern domestic pigs. I collated information for and wrote Archaeological Site Descriptions (Supplementary Information, pp. 2-50). I created the bar plots seen in Fig S5 that show the number of domestic pig and wild boar samples that possess different haplotypes. I also helped write the overall manuscript that was published in PNAS.

Chapter 8

***Audrey T. Lin**. "Discussions and Conclusions concerning part II."*

This chapter is wholly my own work. I wrote the chapter and Louis du Plessis, Tim Coulson, and Greger Larson advised and provided editorial assistance.

Chapter 9 (Appendix)

*Timokratis Karamitros, Tara Hurst, Emanuele Marchi, Eirini Karamichalie, Urania Georgopoulou, Andreas Mentis, Joey Riepsaame, **Audrey Lin**, Dimitrios Paraskevis, Angelos Hatzakis, John McLauchlan, Aris Katzourakis, and Gkikas Magiorkinis. "Human Endogenous Retrovirus-K HML-2 integration within RASGRF2 is associated with intravenous drug abuse and modulates transcription in a cell-line model." Proceedings of the National Academy of Sciences 115, no. 41 (2018): 10434-10439. doi:10.1073/pnas.1811940115*

In this paper, I designed a set of primers targeting a exon-exon junction in the human RASGRF2 gene to be used in PCR tests. In addition, I also performed the PCR screening tests on the Greek cohort (n=202) of HIV-infected people who inject drugs (PWID) and non-PWID samples. All other laboratory work and bioinformatic analyses were performed by the other authors.

Chapter 10 (Appendix)

Angela R. Perri, Kieren J. Mitchell[2], Alice Mouton, Sandra Alvarez Carretero, Ardern Hulme-Beaman, James Haile, Alexandra Jamieson, Julie Meachen, ***Audrey T. Lin****, Blaine W. Schubert, Carly Ameen, Pere Bover, Selina Brace, Alberto Carmagnini, Christian Carøe, Jose A. Samaniego Castruita, James C. Chatters, Keith Dobney, Mario Dos Reis, Allowen Evin, Philippe Gaubert, Shyam Gopalakrishnan, Graham Gower, Holly Heinger, Kristofer Helgen, Josh Kapp, Anna Linderholm, Andrew T. Ozga, Samantha Presslee, Alexander Salis, Nedda F. Saremi, Colin Shew, Katherine Skerry, Mary Thompson, Matthew J. Collins, Mikkel-Holger S. Sinding, M. Thomas P. Gilbert, Anne C. Stone, Beth Shapiro, Blaue Van Valkenburg, Robert K. Wayne, Greger Larson, Laurent A. F. Frantz, and Alan Cooper. "Dire wolves were the last of an ancient New World canid lineage." In review.*

In this paper, I built a phylogenetic tree using the COL1A1 and COL1A2 amino acid sequence generated by MS/MS analysis from a dire wolf and other Carnivora species, whose sequences were publically available. The COL1A1 and COL1A2 gene sequences from all other modern canid were also downloaded from Genbank: gray fox, andean fox, black backed jackal, side striped jackal, african wild dog, dhole, ethiopian wolf, golden wolf, wolf, and yellowstone wolf. I extracted the amino acid sequences from the canid DNA sequences and then translated the coding sequences. I concatenated the amino acid sequences, which were then masked, and built a phylogenetic tree. I originally used BEAST2 but in the end, for the manuscript, a phylogenetic tree was generated using MrBayes. The result of this work can be seen in Figure S12.

Chapter 11 (Appendix)

Carly Ameen, Tatiana Feuerborn, Sarah K. Brown, Anna Linderholm, Ardern Hulme-Beamanb, Ophélie Lebrasseur, Mikkel-Holger S. Sinding, Zachary T. Lounsberry, ***Audrey T. Lin****, Martin Appelt, Lutz Bachmann, Kate Britton, Matthew Betts, John Darwent , Rune Dietz, Merete Fredholm, Shyam Gopalakrishnan, Olga I. Goriunova, Bjarne Grønnow, James Haile, Jón Hallsteinn Hallsson, Ramona Harrison, Mads Peter Heide Jørgensen,*

Rick Knecht, Robert J. Losey, Edouard Masson-MacLean, Thomas H. McGovern, Ellen McManus-Fry, Morten Meldgaard, Åslaug Midta, Madonna L. Moss, Iurii G. Nikitin, Tatiana Nomokonova, Albína Hulda Pálsdóttir, Angela Perri, Aleksandr N. Popov, Lisa Rankin, Joshua D. Reuther, Mikhail Sabline, Anne Lisbeth Schmidt, Scott Shirar, Konrad Smirarowski, Christian Sonne, Mary C. Stiner, Mitya Vasyukov, Gro Birgit Ween, Sanne Eline Wennerberg, Catherine F. West, Øystein Wiig, James Woollett, Love Dalén, Anders J. Hansen, M Thomas P Gilbert, Benjamin Sacks, Laurent Frantz, Greger Larson, Keith Dobney, Christyann M. Darwent, and Allowen Evin. "Specialised sledge dogs aided Inuit dispersal across the North American Arctic." Proceedings of the Royal Society B 286 (2019): 20191929. doi:10.1098/rspb.2019.1929

In this paper, I partitioned the mitochondrial genomes for the BEAST2 analysis. I also selected the parameters and made the xml file for the Bayesian skyline analysis in BEAST2. The effective population size was inferred from the Bayesian Skyline plots (Figure S3 & S9) generated from these BEAST2 analyses.

The Bayesian tree dates the origin of the so-called X-clade to ~5,300 years BP. The result of this work can be seen in Figure S7-S8.

Glossary

Part I: Endogenous retroviruses and human genetic diseases

Cancer stem cell (CSC): Within a tumour, there is a small subset of biologically distinct and rare "tumour-initiating" cells with attributes that parallel the functional properties of normal stem cells. CSCs have been proposed to be the source of all malignant cells in a primary tumour, and their presence also gives rise to metastatic growth. Following chemotherapy-induced remission, the failure to eliminate CSCs may result in relapse and a drug-resistant phenotype of the tumour regrowth.

Enhancer: sequences of DNA that contain binding sites for transcription factors that regulate gene expression. The enhancer may activate transcription independently of their distance, location, or orientation of the gene it affects.

Epithelial-to-mesenchymal transition (EMT): occurs during normal embryonic development, tissue regeneration, organ fibrosis, and wound healing. During the dynamic process of EMT, epithelial cells convert to a mesenchymal phenotype. It is also associated with tumour progression, specifically during metastatic expansion, where the generation of tumour cells with stem cell properties (CSCs) also play a major role in the resistance to anticancer drugs.

Exaptation: when a specific trait is used for a function that is different from its original adaptive function. For example, the incorporation of part of the ERV into the host genome to serve a new function that benefits the host, such as the co-option of retroviral *env* for its immunosuppressive and fusion properties in placentogenesis.

Exogenous retrovirus (XRV): a retrovirus with the ability to form infectious virus particles by exiting the cell where it was assembled and infect a new cell.

Exonisation: ERV sequences have been functionally incorporated into the protein-coding regions of host genes.

Human endogenous retrovirus (HERV): endogenous retroviruses found in human genomes because of retroviruses endogenising the germline of a common ancestor. For example, HERV-L, the oldest HERV family in the human genome infected the common ancestor of mammals approximately 100 mya.

HERV-K (HML-2) / HK2: the most recently acquired ERV in the human genome, and the only known group of HERVs that not only have full-length proviruses with intact open reading frames, but are insertionally polymorphic in the human population. Depending on the presence or absence of a 292-bp deletion in the pol-env junction, can either encode for *np9* or *rec*.

Inflammasome: a component of the innate immune response that induces an inflammatory cascade and activated when infectious agents bind to innate immune sensors and receptors. Inflammasomes are implicated in many inflammatory diseases.

Np9: a 9 kDa protein with currently unknown function in ERV replication. It is a product of a 292 bp deletion in *env*, resulting in a different splice donor that produces an *np9* transcript instead of *rec*. Only HERV-K (HML-2) proviruses with the deletion (termed type I) encode for Np9.

Pluripotent: a transient property of embryonic cells during early embryogenesis, specifically as the inner mass cells in a blastocyst. Pluripotent cells can give rise to all three embryonic germ layers (endoderm, ectoderm, and the mesoderm), and subsequently differentiate into all of the

cell types of the adult body, with the exception of extra-embryonic tissues such as the placenta.

Pluripotency factors: Oct3/4, Sox2, Nanog, KLF4

Provirus: a DNA retrovirus genome that has integrated into the DNA of the host.

Rec: a 14.5 kDa functional homolog of accessory proteins Rev and Rex, encoded by HIV-1 and T-cell leukaemia virus type 1 (HTLV-1) that is presumed to regulate viral gene expression by transporting viral mRNA from the nucleus into the cytoplasm. Only HERV-K (HML-2) proviruses that do not have a 292-bp deletion in *env* (termed type II proviruses) encode for Rec.

Retrotransposon: mobile genetic elements that amplify within the genome via a "copy and paste" mechanism that involves reverse transcription of an RNA intermediate followed by integration into the genome.

Solo LTR: short for Solitary Long Terminal Repeat. An integrated endogenous retrovirus that had been deleted during homologous recombination between the LTRs, resulting in the excision of the provirus body and one of the two LTRs.

Superantigen (SAg): a class of antigens that causes non-specific T-cell activation followed by release of cytokines.

Syncytiotrophoblast: the epithelial surface of the placenta composed of fused multinuclear fetal cells that functions in transporting nutrients, oxygen, and waste between the developing embryo and the mother.

Totipotent: cells that can differentiate into all of the cell types in an adult body, including the extra-embryonic tissue such as the placenta. In mammals, the zygote is the ultimate totipotent cell. During embryogenesis, the embryonic cells progressively lose totipotency as they develop into one of

the three embryonic germ layers (endoderm, ectoderm, and the mesoderm) in the blastocyst.

Part II: Using Large-scale ancient animal DNA datasets to answer fundamental evolutionary questions

Heterochronous data: datasets where the sequences have distinct ages. Examples include serially sampled viruses and bacteria from a single patient, and ancient DNA datasets.

Molecular clock: Originally introduced in the 1960s by Zuckerkandl and Pauling, it is the hypothesis that differences in amino acids or nucleic acids accumulate at a constant rate, so there is a linear relationship between the degree of sequence divergence between species and the time since their last common ancestor. Therefore, if the rate of amino acid change is known, the times of evolutionary divergence between species could be estimated.

Mutation rate: the instataneous rate of change of a nucleotide in the genome.

Pedigree rate: an estimate of the mutation rate, assessed by calculating the number of nucleotide changes observed over a known number of reproductive events (based on a known geneology of individuals and generation time).

Purifying selection: also referred to as negative selection. The selective removal of deleterious mutations.

Saturation: the result of multiple substitutions at the same site in a sequence, or identical substitutions in a different sequence, so that the apparent sequence divergence rate is lower than the actual divergence that has occurred.

Substitution rate: the rate at which mutations are fixed in the population. Because most nucleotide changes that appear within a population are eventually eliminated (by purifying selection or by genetic drift), there will be fewer observed changes per unit of time. Consequently, the substitution rate will always be slower than the mutation rate, with the exception of perfectly neutral conditions.

Tip-dating calibration: A phylogenetic tree that has been calibrated by the ages assigned to the tips of the tree. The tips must contain sufficient temporal information for populations to be characterised as "measurably evolving". "Measurably evolving" populations exhibit detectable amounts of nucleotide changes among DNA sequences sampled at different points in time.

Part I: Endogenous retroviruses and human genetic diseases

Chapter 1, Introduction

"Endogenous retroviruses and human genetic diseases"

The Introduction in Part I of my thesis will have the text of the review article that was published in eLS in 2016 (doi:10.1002/9780470015902.a0026711). Following the review article, the next 2 chapters and the PNAS appendix paper are case studies of endogenous retroviruses serving a role in human genetic diseases, specifically in cancers (Chapters 2 and 3) and addiction (Chapter 9, Appendix).

Statement of Contribution

I wrote the manuscript with assistance from Gkikas Magiorkinis.

Author list and affiliations

Authors

Audrey T. Lin[1] and Gkikas Magiorkinis[2]*

* Corresponding Author: audrey.lin@gtc.ox.ac.uk

Affiliations

1. Department of Zoology, University of Oxford, Oxford, UK
2. Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Athens, Greece

Abstract

It is now increasingly realised that many repetitive elements such as endogenous retroviruses (ERVs), remnants of ancient retroviral infections, can contribute to the host's genetic and transcriptional identity in complex ways. There is growing evidence that specific ERVs are not genomic "junk," but have biological significance in humans. Some ERVs have been "co-opted" for human biological processes, and the nucleic acids and proteins of these ERVs are able to modify human gene expression and regulatory networks in a cell and tissue-specific manner, with implications for healthy development, immune function, and susceptibility to diseases. Ongoing research explores the possibility that aberrant expression of HERVs would contribute to the pathogenesis of a number of inflammatory disorders, including cancers, multiple sclerosis, amyotrophic lateral sclerosis, and schizophrenia.

1. Introduction

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections and make up 5-8% of the DNA. Many of these ERVs have entered the germline more than 30 million years ago where they amplified and proliferated in the genome. These ancient elements are mostly inactive due to mutations or deletions, or are epigenetically repressed. Some of the youngest insertions belonging to the HERV-K (HML-2) family are still intact, with open reading frames encoding for viral proteins that interact with and potentially impact cellular signalling pathways. The presence of ERVs is a double-edged sword for the host: there is growing evidence that some ERVs have biological significance by impacting gene expression and gene regulatory networks in a cell and tissue-specific manner, with implications for development and immune function. However, aberrant activation of ERVs has also been associated with a number of diseases. As a result, the presence of ERVs requires a balance in the host due to its ability to contribute both deleterious and beneficial effects, and this homeostasis appears to have been shaped during evolution.

In this article, we will discuss what is known about ERV biology, and explore the various roles of ERVs and the general mechanisms by which these ancient elements may impact the host genome gene function, and other cellular processes in the context of human health and disease.

2. Retroviruses and consequences of endogenisation

2a. Retrovirus biology

A retrovirus is an encapsidated enveloped virus with a genome of two positive-sense RNA strands—the 5' to 3' polarity of the RNA genome may

directly be translated into protein by the host cell. However, the life cycle of the retrovirus is distinct from many other RNA viruses in that the genome is reverse-transcribed into DNA using reverse transcriptase, an enzyme packaged by the retrovirus, and a tRNA primer. The retroviral DNA then integrates into the host genome as a provirus. The provirus then is transcribed into genomic RNA and messenger RNA, and viral gene products are translated and processed, resulting in the budding of infectious virions from the host cell surface. A canonical replication-competent retrovirus has a 7-11 kb genome encoding for genes *gag*, *pol*, and *env*, with long terminal repeats (LTRs) that flank the 5' and 3 ends of the provirus. *gag* encodes for matrix and capsid structural proteins, *pol* encodes for reverse transcriptase, integrase, and protease, enzymes integral to retroviral replication, and *env* encodes for viral surface glycoproteins necessary for the infection. These envelope proteins allow the virion to bind to the cell surface receptors of the host and fuse with the plasma membrane.

Within the LTRs are unique untranslated U3 and U5 regions separated by a repeat segment (R) and basal promoter motifs for RNA polymerase II, such as TATA box coupled with GC/GT box that specify transcription initiation. Polyadenylation signals, core promoter motifs, and enhancer elements are also present in LTRs, with the ability to recruit a number of transcription factors, co-activators, and transcriptional mediators to affect the expression of adjacent host gene(s), both upstream and downstream of their insertion site. Many ERV LTRs in the mammalian genome drive the transcription of non-coding RNAs (ncRNAs) from ERV insertions in introns, intergenic regions, and the 3' region of protein-coding genes.

2b. Retroviral endogenisation

Retroviruses usually infect somatic cells, but occasionally may infect a sperm or egg cell and enter the germline where the provirus would be inherited. These inherited proviruses are referred to as endogenous retroviruses (ERVs). Once ERVs enter the germline, they may proliferate either by re-infecting germ cells as exogenous retroviruses (XRVs) or through intracellular retrotransposition within the genome of the germ cell. There are two mechanisms of retrotransposition: in *cis*, where the virus uses its own encoded proteins for intracellular mobilization; or in *trans*, where the proteins for mobility are hijacked from another ERV, retrotransposon or XRV within the same cell. Retrotransposition does not require a functional env gene for movement.

The presence of ERVs can be detrimental to the host, due to their ability to cause insertional mutagenesis in the host genome by disrupting host gene regulation. ERVs can also promote non-allelic homologous recombination, which may result in duplication, deletion, inversion, and translocation events. Because of this, active ERV insertions that cause strong adverse effects or disease are likely to be selected against by host defence mechanisms that have co-evolved to suppress the proliferation of ERVs. Pathogenic ERVs that are highly deleterious to the host are therefore unlikely to reach high allelic frequencies and remain in the host population.

2c. Deactivating mutations and deletions

The majority of ERVs are defective, due to mutations that result in frameshifts and premature stop codons. The majority of deactivating mutations that ERVs accumulate appear to be random, occurring after endogenisation by genetic drift. However, ERV inactivation can also occur in

the presence of selection pressure, such as host-derived restriction. For example, apolipoprotein B mRNA-editing complex 3 (APOBEC3)-mediated cytidine deamination results in hypermutation of ERV sequences, thereby inhibiting the retroviral life cycle. Another method of restricting ERV amplification is by homologous recombination between the two viral LTRs, excising an LTR and the body of the provirus, leaving behind a solitary LTR termed "solo LTR". Solo LTRs usually outnumber full-length proviruses by tenfold in all species.

2d. Epigenetic silencing

The persistent threat to genomic integrity posed by the amplification of ERVs and other transposable elements has resulted in the co-evolution of highly adaptable host defences, most notably seen in germ cells and early embryonic cells. Heritable global silencing of ERVs and other autonomous and non-autonomous retrotransposons can occur by epigenetic modification that causes changes in histone modification or DNA methylation. It is possible that that the functional role of DNA methylation was originally a host defence mechanism against ERVs. In the mammalian genome, a methyl group is targeted to a cytosine at position 5 of CpG dinucleotides by DNA methyltransferases (DNMTs). Transcriptional repression by DNA methylation results in the direct inhibition of transcription factor binding, as well as the recruitment of methyl-CpG-binding proteins.

2e. ERV expansion through host co-option

However, if the presence of the ERV is neutral or even slightly beneficial for host fitness, the ERV frequency within the genome may increase within populations. The exaptation of ERV sequences by the host has been

demonstrated in several well-known examples of physiological importance. For example, *env* sequences from the HERV-FRD and HERV-W families have been co-opted into mammalian *syncytin* genes for placental development, likely due to the cellular fusogenic properties of retroviral envelope gene sequences that give rise to syncytiotrophoblast formation or the immunosuppressive properties of *env* that assist in preventing fetal rejection by the mother's immune system. There is also some evidence suggesting that members of the HERV-K family may also play a role in the normal physiology of the placenta during pregnancy. It has been reported that HERV-K Env is expressed in the inner layer of trophoblasts, specifically the villous and extravillous cytotrophoblast cells of the placenta, along with elevated levels of the immunosuppressive cytokine IL-10. Thus, it is possible that HERV-K may have similar functional capabilities in the placenta as Syncytin-1 (encoded by HERV-W *env*) and Syncytin-2 (encoded by HERV-FRD *env*) [1].

ERV LTR sequences may be co-opted by the host for use of alternative promoters, allowing for alternative splicing that yield differential functional effects, such as regulation of tissue-specific expression patterns, and protein isoforms with distinctive functional activities. An example is demonstrated by [2] who reported the presence of an ERV-9 LTR with strong promoter activity integrated upstream of *TP63*, a homolog of the tumour suppressor *TP53*. Differential splicing of the p63 transcript due to the presence of the LTR results in a unique p63 isoform, where the 5' end of the processed incorporates exons containing the LTR. This p63 isoform is specifically and highly expressed in healthy human testis, particularly in spermatogenic precursors. The authors proposed a model whereby the unique testis-specific p63 isoform acts as a tumour suppressor by protecting the genomic integrity of male germ cells. In response to DNA damage, the germ-line encoded p63 isoform was found to be activated by cleavage of caspase, inducing apoptosis, and downregulating p63 in human testicular cancer cell lines and tissues.

ERV LTRs have also been appropriated by the host in starch digestion. The promoter region of the salivary α-amylase genes are derived from two ERV integrations that date approximately 40 million years ago. α-amylase provides the first step in dietary starch and glycogen digestion in the metabolism of glucose, the principal carbon and energy source of mammalian cells. The presence of the LTR results in expression of specific α-amylase in the salivary glands, and the removal of the LTR results in reversion to the ancestral pancreas-specific expression.

2f. ERV classification

There is currently no standard classification system of ERVs, and most ERVs have not been assigned to retroviral genera by the International Taxonomy of Viruses (ICTV). In general, phylogenetic classification divides XRVs into three classes: class I, gammaretroviruses; class II, betaretroviruses; class III, spumaviruses, and ERVs are loosely distributed into all three classes. The historical standard of ERV classification and nomenclature is based on the amino acid carried by the tRNA primer complementary to the primer binding site (PBS) of the retroviral genome. For example, many members of the HERV-K group contain a PBS complementary to lysine-tRNA, and many HERV-H members have a histidine-tRNA PBS. However, this classification method is not exact: many unrelated proviruses may use the same tRNA primer and many phylogenetically related proviruses have different PBS sequences. Currently, HERVs are classified into 31-40 families, each family representing ancient germline infections by different XRVs that had endogenised. We have only highlighted a few HERV families here.

2g. HERV-L, H, W and FRD

HERV-L, the oldest HERV family in the human genome infected the common ancestor of mammals approximately 100 mya [3]. The human genome is dominated by the HERV-H family, where ~88% of all ERVs are comprised of HERV-H insertions that occurred within the last 30 million years, prior to the split of the Old World monkeys and the New World monkeys.

A HERV-W and a HERV-FRD retrovirus also infected a common ancestor of New and Old World monkeys, and are notable due to the host co-option of *env* for the formation of *syncytin*-1 and *syncytin*-2, respectively.

2h. MaLRs

MaLRs (Mammalian apparent LTR retrotransposons) are an ancient class of retrotransposons that entered the germline of eutherian animals approximately 80-100 mya [4]. They comprise of approximately 3.65% of the human genome, with ~240,000 copies. In contrast to the structure of typical retroviruses, with the genes *gag*, *pol*, and *env* flanked by LTRs, MaLRs have 5' and 3' LTRs that surround a 1.35kb ORF lacking any detectable homology with known retroviral genes. MaLRs do not contain a transcription initiation site, primer binding site, nor encode for reverse transcriptase and integrase [4]. It is thought that MaLRs amplify within the host genome by non-autonomous retrotransposition. MaLR sequences have been incorporated into the protein-coding regions of a number of genes ("exonisation").

2i. ERV9

ERV9 is one of the youngest HERV families, with 75% of its elements that had entered the primate germline between 6-15 mya, with high proliferative

activity around 6-8 mya, ceasing approximately during the divergence of the human and chimpanzee lineages, possibly due to the rapid spread of virus restriction alleles in the host populations, suggesting that the proliferation of ERV9 loci was highly deleterious to host fitness [5].

2j. HERV-K (HML-2)

The HML-2 subfamily of HERV-K is the most recently acquired ERV in the human genome, having entered the germline as recently as 150,000 years ago, after the emergence of anatomically modern humans. Consequently, HERV-K (HML-2) are the only known group of HERVs that not only have full-length proviruses with intact ORFs, but are insertionally polymorphic in the human population. There is some evidence suggesting polymorphisms in HERV-H integrations, but the details remain unclear [6][7]. So far, approximately 1,000 HERV-K (HML-2) integrations have been identified in the human reference genome, with 11 polymorphic full-length proviruses out of 91 currently known [6]. It is possible that there are HERV-K (HML-2) proviruses capable of movement. However, no HERV-K (HML-2) provirus is known to produce fully infectious virions, although it has been shown that viral infectivity could be restored by recombination or complementation between loci [8].

All the HERV families with the potential exception of HERV-K (HML-2) and HERV-H are no longer active in the genome, silenced by a number of mechanisms. Although infectious ERVs have been observed in a number of mammals, including mice, cats, sheep, and koala bears, active ERVs have not yet been observed in humans. See **Figure 1.1** for a comparative schematic of HERV-K (HML-2) and its associated transcripts, and a MaLR genome.

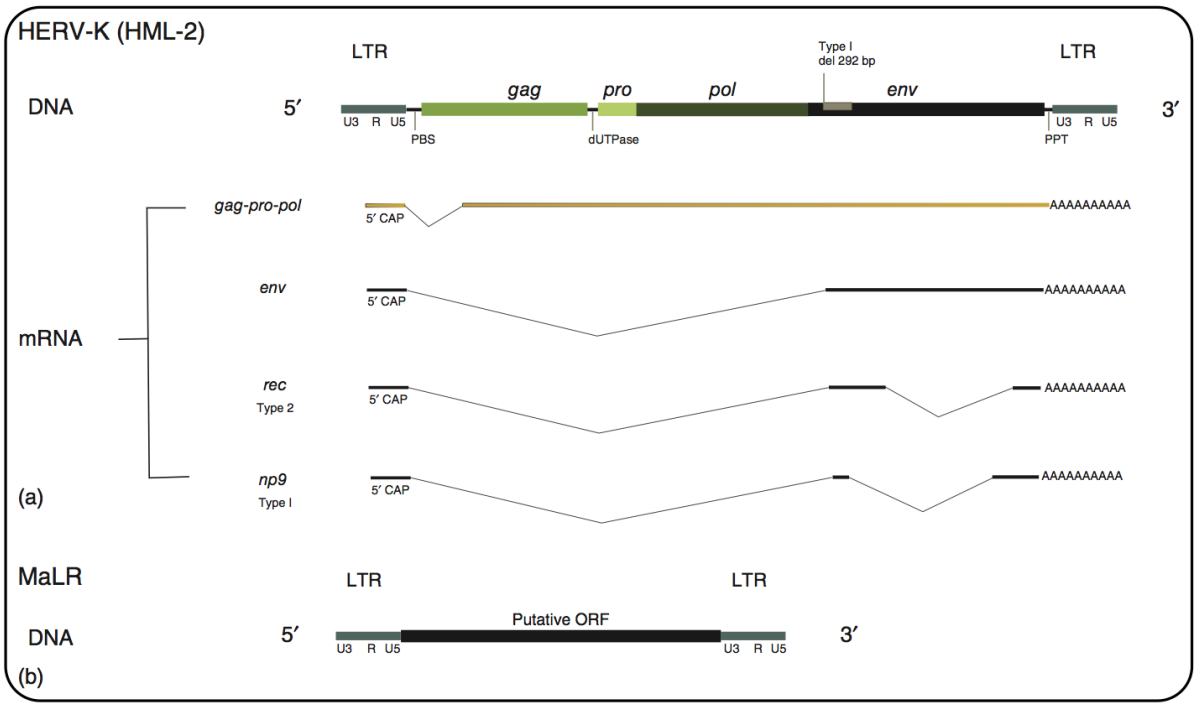


Figure 1.1. Genome schematics of endogenous retrovirus and mammalian apparent LTR retrotransposon

(a.) Genomic organisation of a HERV-K (HML-2) provirus and its 5' capped and polyadenylated transcripts. The coding sequences of the retroviral genes *gag*, *pro*, *pol*, and *env* are flanked by long terminal repeats (LTRs) at the 5' and 3' ends. Each LTR are composed of untranslated 3 (U3), repeat (R), and untranslated 5 (U5) regions. The locations of the primer binding site (PBS), dUTPase, and polypurine tract (PPT) are also indicated. Compared to Type II HERV-K (HML-2) endogenous retroviruses, Type I HERV-K (HML-2) endogenous retroviruses have a 292 base pair deletion in the junction of *pol* and *env*, resulting in a different splice donor that produces an *np9* transcript instead of *rec*. **(b.)** Generalised genomic organisation of a mammalian apparent LTR retrotransposon (MaLR). The 5' and 3' LTRs of retroviral origin are flank a putative open reading frame (ORF) region that do not code for any known retroviral proteins.

---

3. HERVs and host immunity

It is an interesting paradox how HERVs may contribute to autoinflammatory and autoimmune diseases. Because HERVs are part of the host genome, they should be recognised and established by the host immune system as “self”-

antigens during lymphocyte development. Unfortunately tolerance of self-antigens can be imperfect, and the loss of self-tolerance results in the immune system being activated by self-antigens, leading to autoimmune diseases. HERVs may activate innate immune pathways by several mechanisms. HERV nucleic acids or proteins often resembles the pathogen-associated molecular patterns (PAMPs) of exogenous viruses, which are recognised by pattern recognition receptors (PRRs) that can either be cytosolic or membrane-bound. The binding of PAMPs to PRRs activates a complex signalling cascade resulting in the expression of genes encoding for proinflammatory effector molecules such as cytokines, chemokines, and type I interferons (e.g. IFN-α and IFN-β) that induces a general antiviral state as well as activating adaptive immunity.

[9] proposed a model where coevolution of ERVs and host had shaped the host's DNA-sensing immune response. The authors suggest that low levels of ERV transcription is regulated by Trex1, a DNA repair enzyme with 3' to 5' exonuclease activity that degrades endogenous DNA in the cytoplasm. In mice, it was observed that TLRs activated in dendritic cells result in a signalling cascade leading to the expression of type I interferon and NF-κB, followed by the upregulation of TREX1. The authors postulate that the detection threshold of cytoplasmic DNA sensors must be above the normal burden of ERV transcription in order to prevent inappropriate activation and autoimmunity.

Most recently, it was found that the primate-specific MER41, a family of ERVs that had endogenised approximately ~45-60 mya in a primate ancestor, had an impact on the gene regulatory networks induced by proinflammatory cytokine interferon-g (INFG). MER41 LTRs act as cis-regulatory binding sites for the important transcription factors inteferon regulatory factors (IRFs) and

signal transducer and activator of transcription (STATs), following activation by IFN signalling. CRISPR-Cas9 deletion of the LTRs in CD44+ macrophages in HeLa cell lines impaired the expression of the adjacent IFN-induced immune response genes, which includes the activation of the AIM2 inflammasome [10].

It also appears that HERVs play a role in priming and activating B cell responses in a T cell-independent pathway against specific pathogens. As seen in **Figure 1.2**, ERV RNA activates B cells via two different innate immunity pathways: the retroviral RNA either triggers the intracellular nucleic acid sensor retinoic acid-inducible gene-1 (RIG-I) and melanoma differentiation-associated protein 5 (MDA5) and its downstream adaptor mitochondrial antiviral-signalling protein (MAVS); or the retroviral RNA is reverse-transcribed triggering a different cytosolic nucleic acid sensor called cyclic GMP-AMP synthase (cGAS), in which downstream effects include the activation of the stimulator of interferon genes (STING). The activation of either sensing pathway results in the priming and proliferation of antigen-specific extrafollicular B cells that quickly differentiate into IgM and IgG antibody-secreting plasma cells [11]. Both RIG-I-MAVS and cGAS-STING pathways are involved in the sensing of retroviruses such as HIV-1.The ERV-primed B cell response within the spleen appears to be restricted to specific TI-2 antigens, namely capsular polysaccharides from the bacteria *Streptococcus pneumoniae* and Pneumovax 23, a commercial vaccine against *S. pneumoniae*.

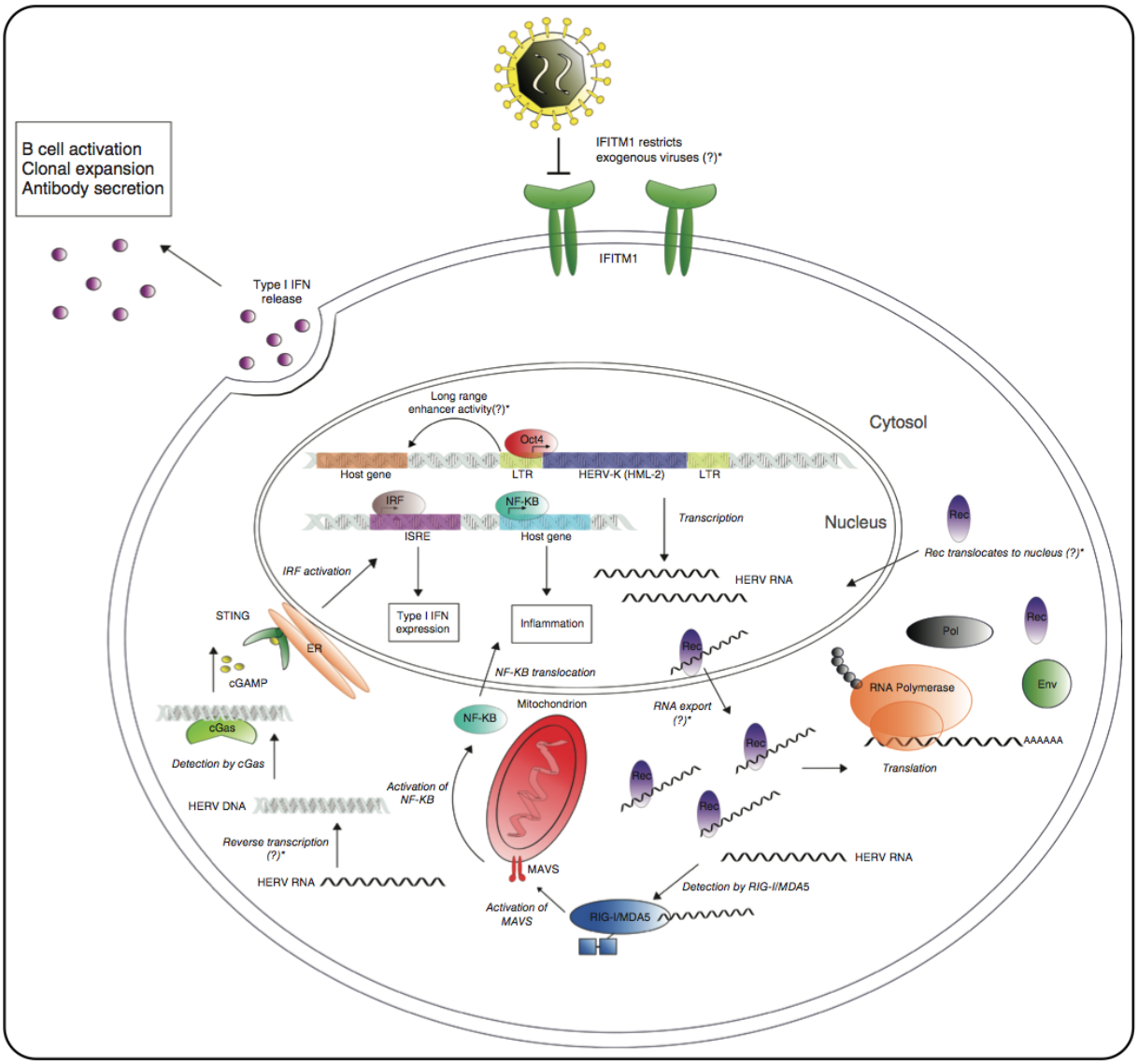


Figure 1.2. Model of HERV-K (HML-2) in human innate immune function.

The transcription factor Oct4 binds to the LTR of a type II HERV-K (HML-2) provirus, inducing its transcription. The LTR may also have long-range enhancer activity on neighbouring host genes. HERV mRNA is processed and translocated to the cytosol for translation. The accessory protein Rec, a product of differential splicing of env transcript is presumed to translocate into the nucleus and facilitate in the export of HERV RNAs into the cytoplasm. The presence of HERV-K (HML-2) RNA in the cytosol is detected by the cytosolic immune receptors retinoic acid-inducible gene-1 (RIG-I) and melanoma differentiation-associated protein 5 (MDA5), which then activates downstream immune signalling processes. The activation of mitochondrial

antiviral-signalling protein (MAVS) results in the activation of nuclear factor Kappa-B (NF-KB), a transcription factor that transalocates to the nucleus where it induces the expression of proinflammatory factors. HERV-K (HML-2) mRNA also may be reverse-transcribed to DNA, which is detected by another cytosolic immune sensor cyclic GMP-AMP synthase (cGAS), resulting in the production of cGMP that activates stimulator of interferon genes (STING). STING induction leads to the activation of interferon regulatory factors (IRFs) that bind interferon-stimulated response elements (ISREs) and express Type I interferons (IFNs). The release of Type I IFNs lead to the activation and clonal expansion of B cells, and secretion of antibodies from plasma cells. The HERV-associated immune response may also result in the expression of Interferon-induced transmembrane proteins (IFITMs) such as IFITM1 that are shown to restrict exogenous viruses. This model depicts findings from [12] and [11].

* indicates proposed mechanisms.

The antigen-sensing repertoire of T cells may also be shaped by the expression of immunoregulatory HERVs. Excessive and aberrant primary T cell activation that results in deletion of the stimulated T cells can be triggered by interaction with microbial superantigens presented by major histocompatibility complex (MHC) class II molecules on antigen-presenting immune cells. In viruses, superantigens are typically glycosylated proteins produced in the infected cells. LMP-2A, the latent membrane protein of Epstein-Barr virus (EBV) has been shown to transactivate the expression of *env* encoded by HERV-K18, with superantigen activity [13] potentially resulting in consequences for people chronically infected with EBV. The interactions between exogenous viruses that cause chronic infections such as EBV and the chronic “genomic infection” of ERVs encoding for immunoregulatory proteins may have great implications in shaping the antigen-sensing repertoire of T cells and contributing to autoinflammation.

Another potential mechanism for autoimmune disease is through HERV-W Env protein, which is able to induce the innate immune response by

interacting with specific PRRs such as Toll-like receptor 4 (TLR4), thereby promoting a type 1 helper T cell (TH1)-like response, which stimulates the release of proinflammatory cytokines IL-6, IL-1ß, and TNF-α [14]. A summary of specific HERV properties in host innate immunity can be seen in **Table 1.1**.

Table 1.1. HERVs and host immunity

HERV family	Gene	Expressed product type	Location of expression	Other findings	Ref.
HERV-K (HML-2)	*rec*	Protein	Human embryonic stem cells, human induced pluripotent stem cells	Transports HERV RNA into cytoplasm to be translated, inducing antiviral mechanisms including expression of *IFITM1*	[12]
HERV-W	*env*/*syncytin -1*	Protein	Inner layer of trophoblast	Part of syncytiotrophoblast formation, placental fusion, prevention of fetal rejection	[14]
HERV-K	*env*	Protein	Inner layer of trophoblast	Possible similar function as syncytins	[1]
Unspecified HERVs	Unspecified HERVs	RNA	Spleen	Primes and activates B cell responses in a T cell-independent pathway by binding to either RIG-I or cGAS	[11]
HERV-K (HML-2)	*env*	RNA	Plasma, PBMCs from HIV-1 infected individuals	Upregulated transcription of Env, but no virions were expressed	[15]
HERV family	**Function**	**Affected host gene**	**Location of expression**	**Other findings**	**Ref.**
ERV9	Promoter	*TP63*	Testis	Possible tumour suppressor in response to DNA damage	[2]

MER41	Enhancer	Adjacent IFN-inducible immune response genes	CD44+ macrophages in HeLa cell lines	Regulate innate immune response by providing binding site for IRFs and STATs	[10]
HERV-K (HML-2) (LTR5HS)	Enhancer	HERV-K (HML-2) provirus and possible adjacent host genes	Human embryonic stem cells, induced pluripotent stem cells	Binding site for Oct4, transcription of HERV RNA, inducing antiviral mechanisms including expression of *IFITM1*	[12]

Note: RIG-I—*retinoic acid-inducible gene-1;* cGAS—*cyclic GMP-AMP synthase*; IFN—interferon; IRF—*interferon regulatory factors*' STATs—*signal transducer and activator of transcription*

4. HERVs and cancer

Because of the genomic burden placed upon the host due to the likelihood of ERV insertions to deleteriously affect host gene expression, it is not surprising that ERVs near exons, gene promoter regions, or intronic splice sites are subject to strong negative selection. ERVs and TEs are found in both orientations, and are found in both intergenic and intronic regions. There is a strong bias of intronic ERVs in the antisense orientation, with respect to the gene it inserted into, perhaps because of the increased likelihood of sense-oriented insertions disrupting host gene expression.

There are a number of mechanisms by which exogenous retroviruses may affect host gene expression. The enhancer elements found within an LTR inserted by a neighbouring host gene may stimulate transcription from the host promoter, or provide alternative promoter or splice donor and acceptors, resulting in stabilization of RNA transcripts or read-through transcription. In addition, retroviruses may contribute to tumourigenesis by activating cellular oncogenes or mutating cellular tumour suppressor genes, resulting in aberrant gene expression that promotes a cascade of uncontrolled cellular proliferation. Although these tumourigenic mechanisms by ERVs have been

observed in other host organisms, namely in mice, there is yet no evidence of HERVs contributing to cancer pathogenesis.

4a. HERV-K and cancer

Because many HERV-K (HML-2) insertions contain intact open reading frames, and are also shown to be insertionally polymorphic, there is a great deal of interest in the involvement of this particular HERV family in disease pathogenesis. HERV-K (HML-2) is expressed in a number of human diseases, including melanoma [16], germ cell tumours [17], lymphoma, breast cancer [18], leukaemia [19], ovarian cancer [20] and viral infections such as human immunodeficiency virus type I (HIV-1) [15].

Antibody responses specific to HERV proteins such as HERV-K Env have been observed in cancer patients, and has been reported to be correlated with poor prognosis of patient survival, such as in mid-stage melanoma [18]. Within the U3 region of the HERV-K LTR, there are several predicted oestrogen, progesterone, and androgen response elements. *env* transcription has been shown to be upregulated up to 10-fold in breast cancer cell lines that had been treated with estradiol followed by progesterone. Moreover, HERV-K reverse transcriptase is overexpressed in breast cancer cell lines that have been treated with the estradiol/progesterone combination, suggesting that these steroid hormones indeed have a regulatory effect on HERV transcription [21].

Because Env protein is also capable of inducing the response of anti-inflammatory cytokines such as IL-10, it is possible that this is one of the mechanisms by which HERVs may protect the tumour it is upregulated in. Elevated IL-10 release has been observed during normal human pregnancy as well as in metastatic melanoma tumours [22][23]. The transmembrane

subunit of Env confers the immunosuppressivity due to its function in the fusion of host and virion cellular membranes. Env protein expressed from HERV-K (HML-2) also appears to have the ability to suppress activation of T-cells by modulating dendritic cells [24].

4b. MaLRs and cancer

It has been demonstrated that the de-repression of previously silenced retroviral LTRs in cells can drive inappropriate expression in a human disease, specifically in the development of Hodgkin's lymphoma. The oncogenic development of Hodgkin's lymphoma requires aberrant autocrine signalling by macrophage growth factor and the oncogenic gene products colony-stimulating factor 1 receptor (*CSF1R*) binding to its ligand colony-stimulating factor 1 (CSF1). In the absence of CSF1/CSF1R signalling, lymphoma cells undergo apoptosis. Specifically, the expression of *CSF1R* is not directed from its canonical promoter, but rather a MaLR of the *THE1B* family, located 6.2 kb upstream of the promoter. In normal B cells, the MaLR LTR is methylated and *CSF1R* is not expressed, through an indirect mechanism involving the transcriptional repressor CBFA2T3. However, in Hodgkin's lymphoma cells, the demethylated LTR acts as a promoter for aberrant transcription of *CSF1R* by recruiting nuclear factor-κB (NF-κB), indirectly leading to the inhibition of CBFA2T3. The activation of CSF1R transcription results in an autocrine proliferative loop allowing for lymphoma growth. The demethylation of *THE1B* LTRs found specifically Hodgkin's lymphoma cells allowing for *CSF1R* transcription contrasts strikingly with CpG-methylated *THE1B* LTRs in non-Hodgkin's lymphoma cells, peripheral blood mononuclear cells, and cells of the myeloid lineage, where *CSF1R* is silenced [25].

4c. Np9 and Rec

Some HERV-K (HML-2) proviruses encode for accessory proteins called Np9 and Rec that have been found to interact with cellular pathways linked to cancer [26][27]. The production of the putative oncogenes Np9 and Rec depend on alternative splicing as a result of the presence or absence of a 292 base pair deletion in *env*, resulting in a different splice donor that produces an *np9* transcript instead of *rec*. Proviruses that do not have this deletion (termed type II proviruses) encode for Rec, and proviruses with the deletion (termed type I) encode for Np9.

Rec is a 14.5 kDa functional homolog of accessory proteins Rev and Rex, encoded by HIV-1 and T-cell leukaemia virus type 1 (HTLV-1) that is presumed to regulate viral gene expression by transporting viral mRNA from the nucleus into the cytoplasm [12]. See Figure 1.2 for the putative role of Rec in transport of HERV-K (HML-2) transcripts. In mice, the expression of Rec was found to interfere with germ cell development and also cause carcinoma [28].

Although Np9 and Rec both are alternatively spliced from HERV-K (HML-2) *env* mRNA, they share only the first 14 amino acid residues of the N-terminus [26]. Np9 is a 9 kDa protein with currently unknown function in ERV replication, however it does appear to have tumourigenic activity. Np9 is preferentially expressed in transformed cells and also interacts with LNX, an important part of the Notch signal transduction pathway, which is implicated in regulating breast and prostate cancer proliferation [27]. It was also found that the silencing of Np9 inhibits the growth of myeloid and lymphoblastic leukaemic cells, and Np9 expression significantly promotes the growth of leukaemic cells, both *in vitro* and *in vivo*. Np9 was found to activate the ERK, Akt and Notch1 pathways, and upregulate ß-catenin, promoting survival and growth of leukaemia stem/progenitor cells [19]. In addition, it

was found in one study that 70% of all patients with chronic lymphocytic leukaemia expressed *np9* mRNA five times higher compared to healthy controls [29].

Both Np9 and Rec interact with promyelocytic leukaemia zinc finger (PLZF) tumour suppressor, a chromatin remodeller and transcriptional repressor implicated in spermatogonial stem cell renewal and cancer. This interaction between Np9, Rec, and PLZF leads to the de-repression of the c-*myc* proto-oncogene, c-Myc overexpression, altered expression of genes regulated by c-Myc, and increased cell proliferation and survival [26]. *np9* and *rec* transcripts are also found in human embryonic stem cells and human induced pluripotent stem cell lines, and are associated with pluripotency maintenance [30]. A summary of specific HERV properties in human cancers can be seen in **Table 1.2**.

Table 1.2 HERVs and cancer

HERV family	Expressed HERV product	Expressed product type	Location of expression	Disease associations	Other findings	Ref.
HERV-K (HML-2)	Np9	Protein	Testicular germ cell tumour	Breast cancer, prostate cancer	Interacts with LNX, ligand of Numb protein X	[27]
HERV-K (HML-2)	Np9	Protein	Myeloid and lymphoblastic leukaemic cell lines	Leukaemia	Activates ERK, Akt, Notch1 pathways, upregulate ß-catenin	[19]
HERV-K (HML-2)	Np9, Rec	Protein	Testicular germ cell tumour	Testicular cancer	Interacts with PLZF tumour suppressor	[26]
HERV-K (HML-2)	Env	Protein	Blood serum from melanoma patients	Mid-stage melanoma	Antibody-specific response, correlated with poor prognosis	[18]

HERV-K (HML-2)	Env	Protein	Placenta choriocarcinoma cell lines	Choriocarcinoma	Modulate dendritic cell activity	[24]
HERV family	**LTR function**	**Affected host gene**	**Location of expression**	**Associated disease**	**Other findings**	**Ref.**
MaLR	Promoter	*CSF1R*	B cells	Hodgkin's lymphoma	Recruits NF-κB, inhibits CBFA2T3, activates CSF1R transcription	[25]
HERV-K	Promoter	Estrogen, progesterone response elements	Breast cancer cell lines	Breast cancer	When cell lines treated with estradiol followed by progesterone, HERV-K RT is overexpressed	[21]

Note: ERK - Extracellular signal-regulated kinases, PLZF - promyelocytic leukaemia zinc finger tumour suppressor, RT – reverse transcriptase

5. HERVs and autoimmune disease

5a. Rheumatoid arthritis, lupus, and psoriasis

Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation of the synovial membrane, resulting in loss of function and destruction of the bone and cartilage in affected joints, pain, and disability. As with many autoimmune diseases, the aetiology of RA pathogenesis is largely unclear.

Elevated antibody response against HERV-K proteins has been observed in rheumatoid arthritis patients [31]. HERV-K (HML-2) transcripts of both *rec* and *np9* have been detected in synovial cells from both normal controls and patients with RA, but there was no significant difference observed in the expression between the two disease states [32]. HERV-K18 which has been

reported to have superantigen activity, was also found to have elevated expression in the peripheral blood of juvenile RA patients, but not in patients with paediatric systemic lupus erythematosus (SLE) [33]. However, it was reported in one study that there is a significant correlation observed in HERV-K113 insertions in patients with RA and SLE compared to healthy controls in the Polish population [34].

There is also some evidence indicating links between HERV-K expression and psoriasis, a chronic inflammatory disease of the skin. A deoxyuridine triphosphate nucleotidohydrolase (dUTPase) encoded by a HERV-K was found to induce NF-kB activation via TLR-2, triggering the release of Th1 and Th17 cytokines and chemokines associated with the formation of psoriasis plaques in dendritic cells [35].

5b. Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS), the most common motor neuron disease of adulthood is characterised by the progressive wasting of muscles associated with limb movement, speech, swallowing, and breathing. The peak age in onset of the motor neuron disease occurs in the mid 50s, and death typically occurs within 3 years following the appearance of symptoms. Only 10-15% of ALS patients have the familial form of the disease, which is inherited as an autosomal dominant trait. What causes the sporadic form of the disease in the majority of ALS patients is unknown.

Links have been observed between the expression of HERV-K and ALS, building upon the hypothesis that Env protein may be neurotoxic. Elevated reverse transcriptase activity has been observed in blood and cerebrospinal fluid of ALS patients. HERV-K *gag* sequences from chromosomal locus 7q34 have been found upregulated in the brain tissue of ALS patients compared to

non-ALS patient controls. Furthermore, it was also found that HERV-K (HML-3) *pol* sequences from the adjacent chromosomal locus 7q36.1 was uniquely expressed in ALS brain, with no expression detected in non-ALS brain. The 7q36.1 region is associated with motor neuron disease (MND), but contributing genes to the MND phenotype has not yet been identified. HERV-K *pol* transcripts are also strongly positively correlated with the expression of TAR DNA Binding Protein 43 (TDP-43) transcripts in ALS brain [36]. TDP-43, encoded by *TARDBP*, is a multi-functional protein involved in many cellular processes, including mRNA binding, splicing, stability, and transport, ssDNA binding, biogenesis of microRNA, cell cycle regulation and apoptosis. The gene was originally identified as encoding for a protein that binds to the trans-activation response (TAR) element in HIV-1 DNA, repressing viral gene expression [37]. The pathological form of TDP-43 is aberrantly ubiquitylated, hyperphosphorylated, and degraded. In the neuronal tissues of individuals with diseases associated with pathological TDP-43, abnormal localization of the protein to the cytosol results in the formation of insoluble aggregates. Pathological TDP-43 is one of the primary biomarkers for ALS and other motor neuron diseases associated with the frontal and/or the temporal lobes of the brain. As seen in **Figure 1.3**, [38] propose a model in which neuroinflammation activates expression of proinflammatory cytokines such as tumour necrosis factor-α (TNF-α) and nuclear factor κ-B (NF-κB) which promotes expression of TDP-43 and HERV-K, including *env* and reverse transcriptase (transcribed from *pol*). In addition, TDP-43 was also found to bind the LTR of HERV-K, activating transcription of HERV-K, which further stimulates a neuroinflammatory state, resulting in the neurodegenerative symptoms of ALS.

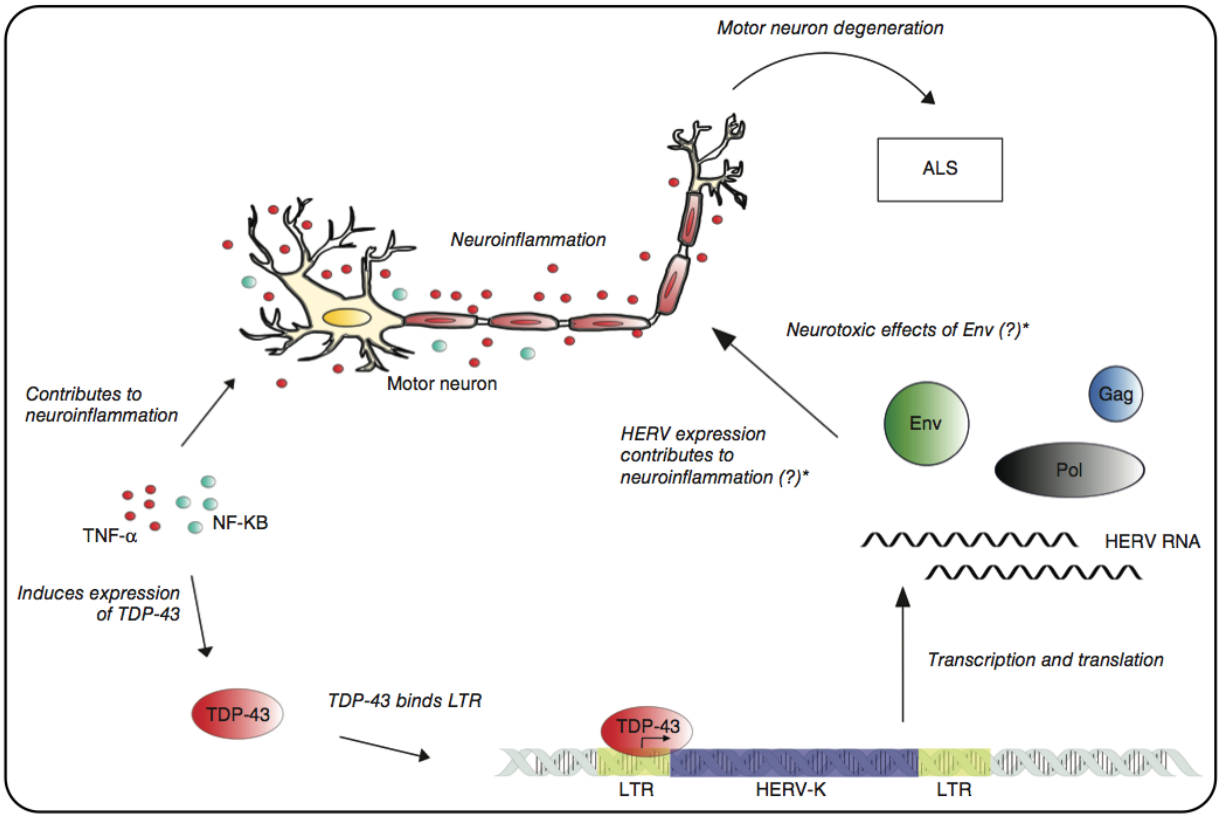


Figure 1.3. Model of HERV-K expression and amyotrophic lateral sclerosis pathogenesis

Motor neurons are susceptible to inflammation due to a combination of genetic predisposition to ALS and environmental factors. Proinflammatory factors TNF-α and NF-KB are expressed, inducing the expression of TAR DNA Binding Protein 43 (TDP-43) that binds the LTR of HERV-K, resulting in its expression. *env* expression has been shown experimentally in transgenic mice be be neurotoxic. The induced expression of HERV-K may stimulate a neuroinflammatory state, resulting motor neuron degeneration and the symptoms of ALS. This model depicts findings from [38].

5c. Multiple sclerosis

Multiple sclerosis (MS) is a neurodegenerative disease characterized by inflammation of the central nervous system (CNS) and disruption of axonal signalling. The inflammation results in the destruction of oligodendrocytes, the neural cells that make up the myelin sheaths around the spinal cord and

axons in the brain, leading to disability and brain atrophy in a progressive decline of autonomic, motor, and neurocognitive function.

Members of three HERV families are often detected in multiple sclerosis patients, potentially acting as a component of the inflammatory cascade that results in MS pathogenesis. HERV-W *env*, which encodes for *syncytin-1*, located on chromosome 7q21.2, appears to have proinflammatory effects. It has been found to be upregulated in the brain of MS patients, specifically in astrocytes, activated parenchymal microglia, and perivascular macrophages in both acute and chronic demyelinated lesions. As seen in **Figure 1.4**, the overexpression of syncytin in astrocytes, activated macrophages and T-lymphocytes results in an inflammatory state that is toxic to oligodendrocytes: high amounts of cytotoxic factors, reactive oxygen radicals (ROS) and the proinflammatory cytokine interleukin-1ß (IL-1ß) appear to contribute to demyelination and the destruction of oligodendrocytes in MS patients [39].

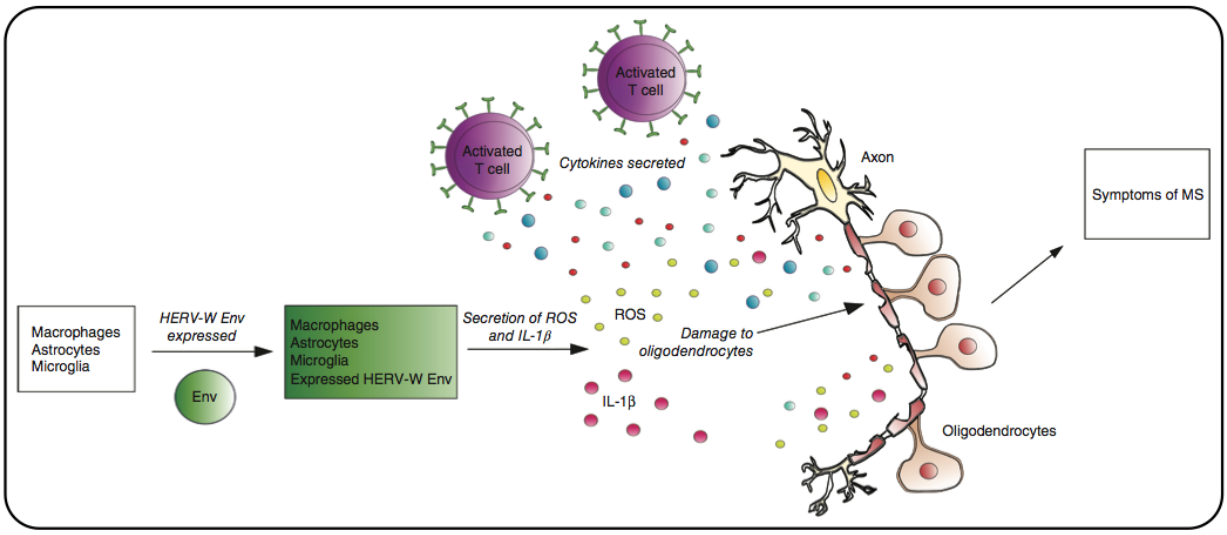


Figure 1.4. Model HERV-W Env and multiple sclerosis pathogenesis

Macrophages and activated T cells from blood migrate to the site of neuroinflammation, where they secrete proinflammatory cytokines.

Astrocytes, microglia, and macrophages have increased expression of HERV-W Env, which induces secretion of factors that are toxic to oligodendrocytes. Reactive oxygen species (ROS) and cytokines such as interleukin-1ß (IL-1ß) damage oligodendrocytes and may result in the demyelination of the myelin sheaths that wrap around the spinal cord and axons in the brain, resulting in symptoms seen in multiple sclerosis. This model depicts findings from [39].

* indicates proposed mechanisms.

The most intensely studied HERV associated with MS belongs to another member of the HERV-W family that is also known in MS literature as multiple sclerosis-associated retrovirus (MSRV).

Upregulation of HERV-W in the brains of MS patients as well as increased copy number of HERV-W in the blood of MS patients have been associated with poorer prognosis [40]. It is also possible that HERV-W may work in conjunction with other inflammatory factors or exogenous viruses in MS progression. For example, it was observed that HERV-W expression is enhanced by co-infection with herpes simplex virus type 1 (HSV-1), human herpesvirus-6A (HHV-6A), and influenza. HERV-W *pol* and *env* RNA and protein also appears to stimulate the production of proinflammatory cytokines. Specifically, the surface subunit of HERV-W Env protein was found to interact with TLR4 and its coreceptor CD14 and stimulate the expression of TNF-α, IL-1ß, and IL-6 [14].

The expression of HERV-H is also reported in several studies to be associated with MS [41]. HERV-H has been found to be upregulated in the lymphocytes of MS patients, and a proliferative cell-mediated immune response has been observed against HERV-H-encoded peptides of Gag and Env, in combination with HSV-1 and HHV-6A antigens [42]. Finally, a member of the HERV-K family, specifically HERV-K18 *env* has been reported to be significantly associated with MS in a case-control study [43]. A summary of specific HERV properties in autoimmune disease can be seen in **Table 1.3**.

Table 1.3 HERVs in autoimmune disease

HERV family	Target gene	Expressed product type	Location of expression	Associated disease	Other findings	Ref.
HERV-K10	*env*	Protein	Blood serum	Systemic lupus erythematosus	Elevated antibody response against HERV Env	[31]
HERV-K18	*env*	Protein	Peripheral blood	Rheumatoid arthritis	Elevated expression	[33]
HERV-K	Reverse transcriptase (*pol*)	Protein	Cortical brain tissue	ALS	Colocalises with TBP-43 protein in brain	[36]
HERV-K	*pol*	RNA	Brain tissue	ALS	Strong correlation with TBP-43 transcription	[36]
HERV-W, ERV9	*pol*	RNA	Brain tissue (frontal cortex)	ALS, schizophrenia	Upregulation of HERV-W, downregulation of ERV9	[44]
HERV-K	dUTPase	Protein	Human primary cells, keratinocytes	Psoriasis	Induce NF-kB activation via TLR2, resulting in release of Th1 and Th17 cytokine/chemokine associated with psoriasis plaques in dendritic cells	[35]
HERV-W	*env*/*syncytin-1*	Protein	Brain tissue (astrocytes, activated parenchymal microglia, perivascular macrophages), specifically frontal cortex	Multiple sclerosis	Binds TLR4 which stimulates proinflammatory cytokine release	[14]

HERV-W	*env*/*syncytin-1*	Protein	Brain tissue (astrocytes, activated parenchymal microglia, perivascular macrophages), specifically frontal cortex	Multiple sclerosis	Overexpression promotes proinflammatory state involving cytokine and ROS production that is toxic to oligodendrocytes	[39]
HERV-H	*gag*, *env*	Protein	PBMCs	Multiple sclerosis	Proliferative cell-mediated immune response against HERV-H peptides in conjunction with HSV-1 and HHV-6A antigens	[42]
HERV-H	*env*	Protein	Blood	Multiple sclerosis	Elevated antibody response against Env peptides	[41]
HERV-K18	*env*	Protein	Lymphoblastoid cell lines	Epstein-Barr virus (EBV infection)	Possible superantigen activity, is transactivated by EBV-enncoded protein LMP-2A	[13]
HERV family	**LTR function**	**Affected host gene**	**Location of expression**	**Associated disease**	**Other findings**	**Ref.**
HERV-K	Promoter	HERV-K	Brain tissue	ALS	Expressed when bound by TDP-43, further promoting neuroinflammation	[38]
HERV family	**Target gene**	**Method**	**Location of expression**	**Associated disease**	**Other findings**	**Ref.**
HERV-W	*pol*	Copy number count	Blood	Multiple sclerosis	Increased copy number in blood is associated with poorer prognosis	[40]
HERV-K18	*env*	*env* SNP genotyping	Blood	Multiple sclerosis	HERV-K18 env genotype associated with increased risk of MS	[43]

Note: ALS—a*myotrophic lateral sclerosis*; TLR2—Toll-like receptor-2' Th1—type 1 helper T-cells; Th17—type 17 helper T-cells; TRL4—Toll-like receptor-4; ROS—reactive oxygen species; PBMCs—peripheral blood mononuclear cells

---

6. HERVs and neuropsychiatric disorders

6a. Schizophrenia and bipolar disorder

Schizophrenia is a heritable psychiatric disorder that emerges in either late adolescence or early adulthood. The pathological mechanisms that underlie the genetic risk factors of schizophrenia are unknown, but numerous genetic markers strongly associated with the disease are found in the major histocompatibility complex (MHC) locus that spans over 3.6 kb of chromosome 6 [45]. The MHC locus is divided into three regions, classes I, II, and III, that are densely packed with immunoregulatory genes, including those encoding for human leukocyte antigen (HLA) molecules. Recent evidence indicates that the gene for complement factor C4, a component of the classical complement cascade is strongly associated with risk of schizophrenia. The complement system is an integral part of innate immunity that enhances the ability of phagocytes and antibodies to eliminate the presence of pathogens.

C4 is found within the class III region and exists as two functionally distinct isotypes (*C4A* and *C4B*) that vary from person to person in copy number and structure. *C3* and *C4* are found in both long and short genomic forms, determined by the presence or absence of a HERV integration within an intron. In humans, C4 expressed by neurons localizes to axons, dendrites, and synapses, where it is then secreted, and decreased numbers of synapses have been observed in schizophrenia patients. C4 expression is elevated in the brains of schizophrenia patients, and the copy number of *C4A* and *C4B* is positively correlated with expression of the gene, and the presence of the

HERV also increases the ratio of *C4A*/*C4B* expression. The higher the expression levels of *C4A* and *C4B* in the brain, the greater the risk of schizophrenia, suggesting that the HERV integration in *C4A* and *C4B* may play a role in the synapse loss observed in brains of schizophrenia patients [46].

The expression of members of specific HERV families are also associated with schizophrenia. In three expression studies on postmortem brains with primers targeting *pol*, increased expression of HERV-W was found in frontal cortex, and increased HERV-K10 expression from frontal cortex and prefrontal cortex brains of schizophrenia and bipolar disorder patients . Increased expression of HERV-K (HML-2) *gag* was found in the prefrontal cortex region of schizophrenia patients. In the same brain samples, no significant difference in *env* transcription from HERV-W, HERV-FRD, and HERV-K was observed. ERV9 *pol* transcription was increased in the prefrontal cortex of schizophrenic brain, however ERV9 *pol* transcription was significantly decreased in the frontal cortex. Increased transcription of *pol* from HERV-W, ERV9, and HERV-FRD was also found in the cerebrospinal fluid of schizophrenia patients [44]. Increased HERV-W, ERV9 expression of both *gag* and *pol* transcripts were also observed in blood from schizophrenia patients [47]. However, there are contradicting observations of HERV-W *env* expression in blood from schizophrenia patients [48].

HERV-W Gag and Env protein have been detected from the blood serum of patients with schizophrenia, and increased reverse transcriptase activity observed in the blood serum and postmortem brain of schizophrenia patients. Env and Gag protein in serum of schizophrenia patients are also positively correlated with C-reactive protein (CRP), a PRR with levels that are elevated during inflammation [49].

Finally, a HERV-K (HML-2) element located upstream of *PRODH* acts as a tissue-specific enhancer for the gene's expression [50]. *PRODH* encodes for a mitochondrial enzyme proline dehydrogenase (oxidase), which catalyses the first step in proline metabolism vital for healthy central nervous system (CNS) function. Mutations in *PRODH* was found to be associated with several neuropsychiatric disorders, including schizophrenia [51].

6b. Alcoholism

Ponomarev and colleagues [52] had the first study suggesting that LTR-containing TEs may have a functional role in alcoholism and addiction behavior. Hypothesizing that neuroadaptations and neuropathology that contribute to alcoholism are partly mediated by alcohol-induced epigenetic changes in gene expression, transcriptomes gathered from the post-mortem brains of 17 alcoholic humans and 15 matched controls were examined. It was found that several LTRs were upregulated in the human alcoholic brain samples. It was also found that DNA hypomethylation in the alcoholic brain is associated with the activation of ERV expression, specifically members of HERV-L (*MLT2A1*), MaLR (*THE1B*), and LTR8 (which maps to the HUERS-P3 ERV belonging to the ERV1 superfamily). A summary of specific HERV properties in neuropsychiatric disorders can be seen in **Table 1.4**.

Table 1.4 HERVs and neuropsychiatric disorders

HERV family	Target gene	Expressed product type	Location of expression	Associated disease	Other findings	Ref.
HERV-FRD	*pol*	RNA	Cerebrospinal fluid	Schizophrenia	Upregulated transcription	[44]
HERV-W	*pol*	RNA	Cerebrospinal fluid	Schizophrenia	Upregulated transcription	[44]

ERV9	*pol*	RNA	Cerebrospinal fluid	Schizophrenia	Upregulated transcription	[44]
HERV-W, ERV9	*pol*	RNA	Brain tissue (frontal cortex)	Schizophrenia	Upregulation of HERV-W, downregulation of ERV9	[44]
HERV-W, ERV9, HERV-FRD	*pol*	RNA	Cerebrospinal fluid	Schizophrenia	Upregulation of HERV-W, ERV9, HERV-FRD	[44]
HERV-W	*env*	RNA	Blood (PBMCs)	Schizophrenia, Bipolar disorder	Upregulation of HERV-W Env	[48]
HERV-W	*gag*, *env*	Protein	Blood serum	Schizophrenia	Gag and Env positively correlated with CRP	[49]
HERV-W	Reverse transcriptase (*pol*)	RNA → DNA	Blood serum, postmortem brain	Schizophrenia	Increased RT activity	[49]
HERV-L (*MLT2A1*), MaLR (*THE1B*), LTR8	LTR	RNA	Brain tissue	Alcoholism	DNA hypomethylation associated with HERV activation	[52]
HERV family	**HERV type**	**Affected host gene**	**Location of expression**	**Associated disease**	**Other findings**	**Ref.**
Not specified	Not specified (~17 kb sequence)	*C4A*, *C4B*	Brain tissue	Schizophrenia	HERV integration in *C4A* or *C4B* intron increases expression ratio of *C4A/C4B,* which increases schizophrenia risk	[46]
HERV family	**LTR function**	**Affected host gene**	**Location of expression**	**Associated Disease**	**Other findings**	**Ref.**

HERV-K (HML-2)	Enhancer	*PRODH*	Brain tissue (hippocampus)	Schizophrenia	Binds SOX2, *PRODH* catalyses first step in proline metabolism for healthy CNS function; mutations in *PRODH* associated with schizophrenia	[50]

Note: CRP—C-reactive protein; PBMCs—peripheral blood mononuclear cells; RT—reverse transcriptase

7. Conclusions

We have only now begun to understand the complexities that HERV integrations have on human cellular physiology. These HERVs are sometimes beneficial: LTRs are used as a genomic reservoir for promoter and enhancer sequences; viral nucleic acids are used for activating specific components of the innate immune response; the fusogenic properties of the HERV envelope protein is used in early placentogenesis. Yet the presence of these ancient genomic elements often involves trade-offs regarding host fitness. A more sensitive innate immune response to exogenous retroviruses would better protect the host against infection and resulting pathologies, but also may predispose the host in inappropriate detection of HERV nucleic acids and proteins, thereby potentially promoting autoimmune disease. However, although links between HERV expression many diseases have been established, there is no clear evidence that they are involved in the disease pathogenesis. For example, whether the presence of HERV-K (HML-2) RNA and its gene products are merely byproducts of immunological dysfunction and inflammation or if they play an aetiological role in disease development, especially human cancer, is a hotly debated topic. Because copies of HERVs are ubiquitous in the human genome, and most are present in all individuals due to fixation, their presence is not considered unique in disease. It is

difficult to establish the causality or contribution of the HERV to disease, especially compared to single aetiologic agents applicable to Koch's Postulates. Therefore, the investigation of the potential role HERVs in disease pathogenesis requires careful detection of viral expression in the disease of study.

In addition, although there is clear evidence that some HERV LTRs have been exapted resulting in significant changes in host gene expression, as exemplified by the LTR-derived promoter in the expression of salivary α-amylase, it has been argued that most HERV LTRs are minor, alternative promoters that exert a subtle overall effect on host gene expression. Instead, these LTRs would induce expression patterns similar to that of the native host promoters, suggesting that selective pressures resulting in LTR exaptations as host promoters are rare occurrences. However, most recent example of the enhancer activity of MER41 that regulates the IFN-inducible immune response indicates that the regulatory roles of ERVs in the human genome on important physiological processes are only being uncovered and elucidated.

There are also further implications to consider regarding HERV polymorphisms. Most integrations are fixed in the human population (that is, every individual has them), however evolutionarily young HERVs like HERV-K (HML-2) are not fixed. The presence or absence of an HERV at a specific genomic location, whether or not they are full-length proviruses or solo LTRs, or whether there are mutations within the viral sequence from either single nucleotide polymorphisms (SNPs) or larger insertions/deletions may impact the host in noticeable ways. Inappropriate expression of domesticated HERVs (that are tightly regulated by epigenetic modification in healthy tissues) may contribute to the pathogenesis of various human diseases that have multifactorial causes. These activated HERVs, such as derepressed LTRs and aberrantly expressed RNAs and proteins from both HERV and host are

capable of perturbing host gene expression, and activate innate immune sensors, thus promoting autoinflammation. The current research that parses how these ancient relics in our genomes are worthwhile in the context of human health and disease, but in doing so, we are are also expanding our knowledge on the evolution of the human genome.

8. Key concepts

1. Some HERVs have been “co-opted” for human biological processes such as placentogenesis, innate immunity pathways, anticancer mechanisms, and central nervous system function.

2. ERVs provide a genetic reservoir for promoters, transcription start sites, enhancers, and alternative splice donors and acceptors for regulating host gene expression.

3. HERV-K (HML-2), the youngest HERV in the human genome, is polymorphic in the human population, meaning that not everyone has the same HERV integration at the same chromosomal position.

4. Np9 and Rec, accessory proteins encoded by HERV-K (HML-2) are associated with specific human cancers in several ways: their expression is upregulated, they interact with specific cancer biomarkers, and they impact signalling pathways that are implicated with cancer.

5. Specific HERV LTRs help regulate the innate immune response by acting as a reservoir of interferon-inducible enhancers.

6. HERV transcripts play a role in priming and activating B cell responses in a T cell-independent pathway against specific pathogens.

7. Ongoing research explores the possibility that aberrant expression of HERVs would contribute to the pathogenesis of a number of inflammatory disorders, including cancers, multiple sclerosis, amyotrophic lateral sclerosis, and schizophrenia.

10. Further Reading list

1. Stoye JP. (2012) Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature Reviews Microbiology* **10**(6): 395-406.

2. Feschotte C, and Gilbert C. (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics* **13**(4): 283-296.

3. Magiorkinis G, Belshaw R, Katzourakis A. (2013) 'There and back again': revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **368**(1626): 20120504.

4. Hurst TP, and Magiorkinis G. (2015) Activation of the innate immune response by endogenous retroviruses. *Journal of General Virology* **96**(6): 1207-1218.

5. Manghera M, and Douville RN. (2013) Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors. *Retrovirology* **10**(1): 16.

6. Young GR, Stoye JP, and Kassiotis G. (2013). Are human endogenous retroviruses pathogenic? An approach to testing the hypothesis. *Bioessays* **35**(9):794-803.

7. Downey RF, Sullivan FJ, Wang□Johanning F, *et al.* (2015). Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice?. *International Journal of Cancer* **137**(6): 1249-1257.

8. Cohen CJ, Lock WM, and Mager DL. (2009). Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* **448**(2): 105-114.

9. Jern P, and Coffin JM. (2008) Effects of retroviruses on host genome function. *Annual Review of Genetics* **42**: 709-732.

10. Meisler MH, and Ting CN. (1993) The remarkable evolutionary history of the human amylase genes. *Critical Reviews in Oral Biology & Medicine* **4**(3): 503-509.

Chapter 2, "Targeting endogenous retrovirus gene transcription in human cancers."

This work has been submitted to bioRxiv in 2018. **doi:** https://doi.org/10.1101/449686

Statement of Contribution

Research design

I developed the overall project with GM and AK.

Data

I curated the publically available and controlled-access data. The diseased and non-diseased cancer data was obtained through The Cancer Genome Atlas Project and the Illumina Body Map 2.0 Project. The Cancer Genome Atlas Project data was used according to the Data Use Certification as stipulated by the TCGA Data Access Committee and dbGaP authorisation (https://dbgap.ncbi.nlm.nih.gov/aa/). I acquired permission to use controlled-access dbGAP data under projects #57108 and #7621.

Method development

I designed and the developed the method for quantifying the HERV-K (HML-2) target transcripts with CGS, TK, AK, and GM. The creation of the cDNA reference sequences were done with TK.

Analysis

I performed all of the RNA-seq analyses, including alignment of the downloaded raw files, filtering, quantification, and normalisation of transcripts. I performed all of the statistical analyses under the supervision of GM.

Manuscript

I wrote the manuscript with additional input from CGS, TK, AK, and GM. I generated all the tables and figures. I curated and edited the information presented in the Supplement.

Author list and affiliations

Authors

Audrey T. Lin[1], Cindy G. Santander[1], Emanuele Marchi[2], Timokratis Karamitros[1,3], Aris Katzourakis[1], Gkikas Magiorkinis[1,4]**

* Corresponding Authors: audrey.lin@gtc.ox.ac.uk and gmagi@med.uoa.gr

Affiliations

1. Department of Zoology, University of Oxford, Oxford, UK
2. Nuffield Department of Medicine, University of Oxford, Oxford, UK
3. Public Health Laboratories, Department of Microbiology, Hellenic Pasteur Institute, Athens, Greece
4. Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Athens, Greece

Abstract

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections that make up to 8% of the human genome. Although these elements are mostly fragmented and inactive, many proviruses belonging to the HERV-K (HML-2) family, have intact open reading frames, some encoding for accessory genes called *np9* and *rec* that interact with oncogenic pathways. It has been shown in previous studies that the transient expression of ERVs in both stem cells and cancers results in aberrant self-renewal and uncontrolled proliferation. This paper presents a novel and a highly specific pipeline that is able to capture and measure transcription of *np9* and *rec* encoded by proviruses that share great sequence similarity, and are transcribed at very low levels. Using our novel methodology we find that HERV-K (HML-2)-encoded genes *np9* and *rec* are upregulated in a number of cancers that have not been described previously. We also show that *np9* and *rec* are specifically expressed in the 8 and 16-cell stage in human preimplantation embryos.

1. Introduction

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections that make up about 5-8% of the human genome. These ancient elements are mostly inactive due to mutations or deletions, or are epigenetically repressed. Many proviruses belonging to HERV-K (HML-2) (shortened to HK2 here) family, the only lineage still proliferating in the genome after the human-chimpanzee split, have intact open reading frames, some encoding for accessory proteins called Np9 and Rec that have been found to interact with cellular pathways linked to cancer [19][26][27]. Type I proviruses have a 292 base pair deletion in *env* that leads to a different splice donor, producing an *np9* transcript instead of *rec*, which is encoded by type II proviruses [53][54] (**Fig. S1**). Rec is a functional homolog of accessory proteins Rev and Rex, encoded by HIV-1 and leukaemia virus type 1 (HTLV-1) respectively. Rev/Rex/Rec functions in regulating viral gene expression by transporting viral mRNA from the nucleus into the cytosol [55][12]. Rec protein was found to interfere with germ cell development and also cause carcinoma in mice [28].

It is unclear what function *np9* has in virus replication, if any, but it does appear to have oncogenic activity. Np9 protein can activate the ERK, Akt and Notch1 pathways, and upregulate ß-catenin, promoting survival and growth of leukaemia stem/progenitor cells [19]. Np9 protein is also preferentially expressed in transformed cells and is known to interact with LNX, an important part of the Notch signal transduction pathway, implicated in regulating breast and prostate cancer proliferation [54][27]. Its expression significantly promotes the growth of leukaemia cells, and when silenced, the growth of myeloid and lymphoblastic leukaemic cells is inhibited. Both Np9 and Rec proteins interact with promyelocytic leukaemia zinc finger (PLZF) tumour suppressor, a chromatin remodeller and transcriptional repressor that is implicated in spermatogonial stem cell renewal and cancer. The Np9-

Rec-PLZF interaction results in the de-repression of the c-*myc* proto-oncogene, c-Myc overexpression, altered expression of genes regulated by c-Myc, and increased cell survival and proliferation [26]. Np9 protein is also reported to bind and regulate ubiquitin ligase MDM2, which represses the activity of p53 tumour suppressor [56]. Transcripts of *np9* and *rec* are also found in human embryonic stem cells and human induced pluripotent stem cell lines, and are associated with maintenance of pluripotency [30][57].

With recent advances in high-throughput mRNA sequencing (RNA-seq), it is possible in a single assay to find novel genes and transcripts as well as measure transcription [58][59]. Small RNA-seq experiments are capable of producing vast volumes of data: current sequencers can generate more than 500 gigabases of raw sequencing reads per run [60]. The sensitivity of RNA-seq allows for the potential detection of alternative splice isoforms of transcripts as well as rare and cell and context-specific transcripts. Moreover, because the number of reads that are produced from an RNA transcript is a function of the transcript abundance, the depth of read coverage, or read density, can be used as a proxy to measure gene expression [61][62][60].

However, there may be inherent biases in the RNA-seq methodology that must be taken into consideration. In general, most sequencing platforms produce short sequence reads that are assembled computationally to generate full-length transcripts. Even for fairly well-characterised transcripts, such as from protein-coding genes, ambiguity arises from the expression of transcriptional variants that often share exons.

Given the sensitivity of RNA-seq, it is likely that HERV transcripts can be detected, even if they are expressed in low abundance. However, because HERV transcripts are often longer than the lengths of reads generated by most sequencing platforms, and that HERVs are often greater than 97% identical to one another [6], there are many uncertainties arising from

analysing HERVs in RNA-seq data. Due to the high similarity of sequences from repetitive regions, it is difficult to reconstruct these transcripts with accuracy. Problems with library preparation as well as GC content of the transcript also produce biases [63]. Inconsistent coverage can result from biological stochasticity – transcripts can be expressed at different levels and at different lengths depending on cellular conditions [64]. As a result, measuring the expression of the alternatively spliced *np9* and *rec* transcripts with confidence can be challenging.

To the best of our knowledge, there is currently no standardised computational method for inferring expression of repetitive elements from RNA-seq data. We designed a pipeline that enabled us to measure the cDNA transcripts of *np9* and *rec* in mapping RNA-seq reads to a reference transcript and only counting the reads spanning the splice junction. This ensures the measurement of these HERV alternative splice products with high sequence similarity with great specificity. By using our novel methodology of processing and analysis of publicly available RNA-seq data, we will show overexpression of the HK2 targets in the cancer types that are known in the literature to overexpress HERVs. The cancers we had expected to see HERV overexpression in comparison to non-diseased tissue include breast cancer, germ cell tumours, skin melanoma, lymphoma, ovarian cancer, and prostate cancer.

In addition, it had been reported that during early human embryogenesis, HK2 is transcribed only during the 8-cell, 16-cell (morula), in early blastocyst outgrowth stages, as seen in early passages of human embryonic stem cell (hESC) lines [12][65]. In order to determine whether *np9* and *rec* follow these patterns of expression, published single cell RNA-seq data [65] was used to measure the transcription of *rec* and *np9* in human preimplantation embryos.

Using RNA-seq data from The Cancer Genome Atlas and other repositories, we observed high transcription of *np9* and *rec* in testicular germ cell cancer, and cancers of the eye, skin, liver, lung, and thymus, as well as during the 8- and 16-stage cell stages of human embryogenesis.

2. Materials and Methods

2a. cDNA reference sequences

The 225 bp *np9* cDNA sequence (Genbank accession number JA719353.1) was used as a reference for the RNA-seq reads to map against. Because there are no other known isoforms of *np9*, no consensus sequence was created. However, if the splice junction is shorter than the 48 bp-long RNA-seq read, the read will not map properly even if it spans the junction, because the read would be longer than the *np9* sequence leading up to the splice junction (48 bp vs 44 bp long, respectively) (**Fig. 2.1**).

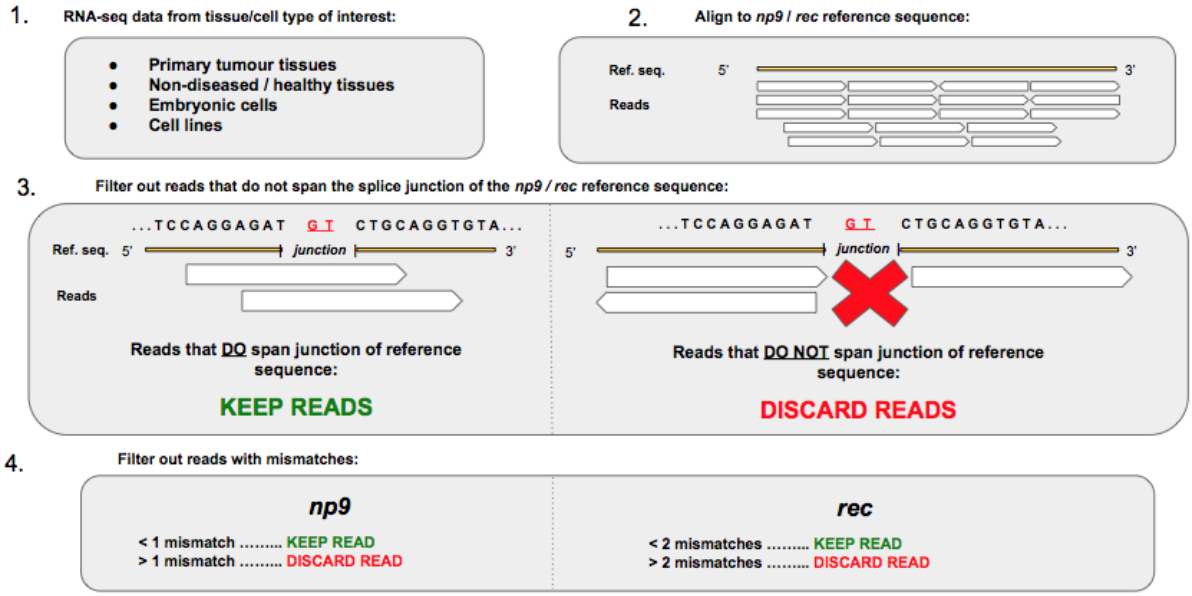


Figure 2.1. Workflow describing method in measuring expression of *np9* and *rec* in RNA-seq data

1. Download RNA-seq files from the tissue/cell type of interest. This can include transcriptional sequence data from primary tumour tissues, non-diseased or healthy tissues, embryonic cells, and cell lines. **2.** Align raw RNA-seq data to the *np9* or *rec* reference sequence (**Table S2.1**). **3.** Filter out reads that do not span the splice junctions of the *np9*/*rec* reference sequence, keeping only the reads that do span the respective splice junctions of *np9* and *rec*. **4.** Filter out the reads with sequence mismatches. For *np9*, if there's more than 1 mismatch in these reads, discard the reads; for *rec,* if there's more than 2 mismatches in these reads, discard the reads.

To capture these reads that would otherwise be lost, 52 bp from the 3' end of HERV-K (HML-2.HOM) *pol* (accession number AF074086.2) was added to the *np9* cDNA reference, forming a 274 bp sequence, allowing the full RNA-seq read to map to the *np9* reference (**Table S2.1**).

All *rec*, or complete ORF (cORF) as reported on Genbank sequences were downloaded from Genbank. Because multiple isoforms of *rec* transcript are known, including a 414 bp isoform with a 58 bp deletion containing the splice donor [66], sequences of the longer 471 bp isoform were downloaded and aligned, and a consensus sequence was created. All RNA-seq reads were mapped to this *rec* consensus sequence (**Table S2.1**).

pol transcript from HK2 proviruses and HERV-K113 5' LTR (accession number NC_022518.1) transcript were used to quantify the background transcription level of HK2 loci in both solid primary tumour and solid non-diseased tissue samples. Full type 1 and type 2 HK2 provirus sequences were downloaded [6], aligned, and a consensus sequence created. A 532 bp fragment from the 5' end of the *pol* consensus sequence was used to map all RNA-seq reads to (**Table S2.1**). All consensus sequences were created in Geneious version 8.1 [67].

2b. Alignment

Because of the polymorphic nature of HK2 integrations, it is unknown how many proviruses express *np9* and *rec* in the genome, from one individual to the next. Therefore, transcription levels within each sample were standardised by selecting a "housekeeping" or "reference" gene to normalise HK2 activity with. *actb* (accession number NM_001101.3) was used as a reference gene in all the samples analysed. The RNA-seq reads were aligned with bowtie2 v.2.2.5 [68] to all cDNA reference sequences in end-to-end mode, using default settings. All aligned reads in the resulting BAM file with a mapping quality of less than 20 were filtered out using Samtools version 1.2 [69], which was then followed by sorting and indexing with Picard Tools version 2.1.1 (Broad Institute), for visualization on Integrative Genome Browser (IGV) (Broad Institute).

2c. Filtering for *np9* and *rec*-specific aligned reads

The high sequence identity between different HK2 proviruses necessitates stringent filtering to ensure that the reads mapped to the *np9* and *rec* reference sequences indeed belonged to expressed *np9* and *rec* transcripts. For this reason, the junction between the splice donor and the splice acceptor site was the critical target in capturing *np9* and *rec* transcripts within each given sample. In order to ensure high specificity when capturing alternative splice products of the same gene, within the filtered BAM file, all reads that did not span the splice junction of *np9* and *rec* within 40 bp in either direction were discarded. Also discarded were reads mapped to *np9* reference sequence with >1 mismatch, and >2 mismatches for *rec* (**Fig. 2.1, 2.2**). The mismatch thresholds differ to allow for capture of transcripts expressed from isoforms of *rec*. A table with a list of the raw reads mapped to *np9* and *rec* from all samples are attached at the end of this chapter in the Supplement.

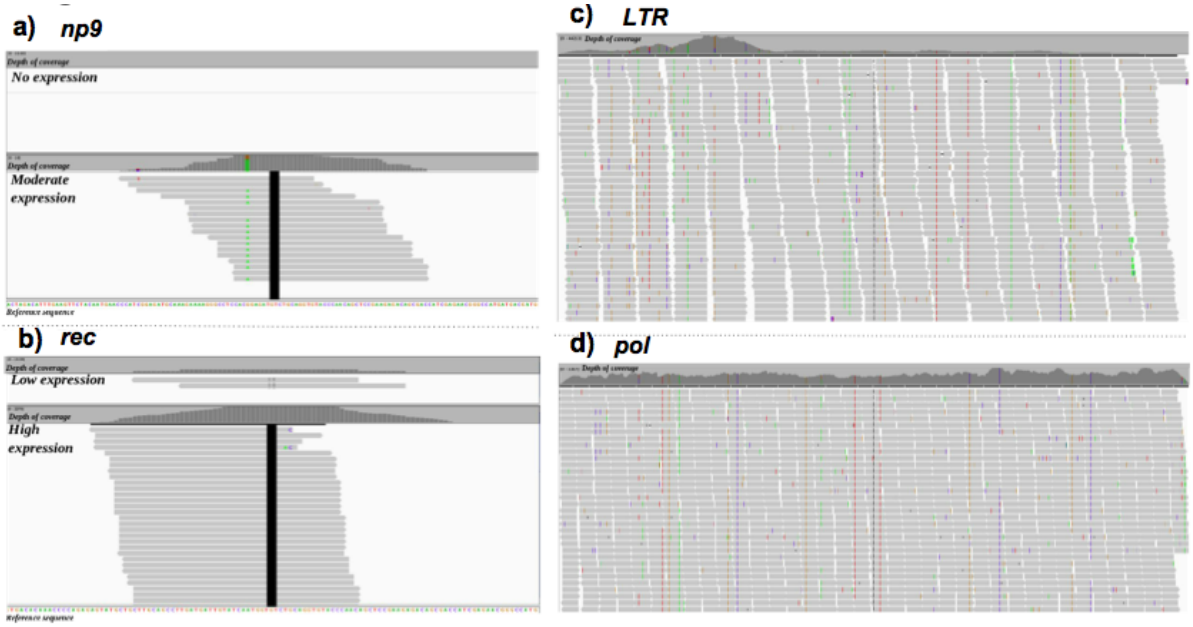


Figure 2.2. Sequenced cDNA reads from primary tumours mapped to HK2 targets as viewed on Integrative Genome Browser (IGV)

The horizontal shaded boxes display the depth of coverage of mapped reads. Thick vertical black bars indicate the splice junction of *np9* and *rec.* A segment of the reference sequence is displayed at the bottom of each panel. **(a.)** RNA-seq reads from lung adenocarcinoma (LUAD) primary tumour samples from two different cancer patients mapped to *np9* reference sequence. The top panel shows a patient with no *np9* expression. The bottom panel shows a patient with moderate *np9* expression. **(b.)** RNA-seq reads from testicular germ cell tumour (TGCT) primary tumour samples from two different cancer patients mapped to *rec* reference sequence. The top panel shows a patient with low *rec* expression, with only 2 reads that map perfectly to the reference. The bottom panel shows a patient with high expression of *rec*. **(c.)** RNA-seq reads from testicular germ cell tumour (TGCT) primary tumour sample mapped to HERV-K113 LTR reference sequence. LTR is highly expressed, most likely due to lack of specificity. **(d.)** RNA-seq reads from testicular germ cell tumour (TGCT) primary tumour sample mapped to HK2 *pol* reference sequence. *pol* is highly expressed, most likely due to lack of specificity.

2d. Quantifying depth of coverage of HK2

The average depth of coverage of the mapped reads was quantified using the GenomeAnalysis TK package (GATK) version 3.6.0 [70] and Picard Tools version 2.1.1 (Broad Institute). For the other HK2 cDNA targets *pol* and LTR, as well as for the normalisation gene *actb*, the stringent filtering and processing necessary for capturing *np9* and *rec* transcripts was not necessary, and we used the average read count across the entire reference sequence. Because of this, the depth of coverage of *pol* and LTR is not directly comparable to the depth of coverage observed for *np9* and *rec* expression, for each given sample (**Fig. 2.2**). For these genes the BAM file with reads of a MAPQ >20 is directly input into Picard Tools for sorting and indexing.

2e. Normalisation of HK2 transcripts by actb transcripts

The read counts generated by GATK DepthOfCoverage were analysed in R version 3.3.1 [71]. Normalised expression of the HK2 transcripts (*np9*, *rec*, LTR, and *pol*) from the samples was calculated by dividing the average depth of coverage of the expressed HK2 by the average depth of coverage of the expressed reference gene (*actb*). However, because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

2f. Analysis of publicly available raw data

Raw FASTQ files of unaligned Illumina RNA-seq reads sequenced from 305 primary solid tumour tissues and blood-derived non-diseased tissues from different individuals were downloaded from Cancer Genomics Hub (data now available at Genomics Data Commons). Cancers that had been reported in the literature to be associated with HK2 expression were selected for study.

The other cancer types were arbitrarily selected, and a total of 15 different cancers were analysed.

RNA-seq data sequenced from single cells of oocytes, zygotes, 2-, 4-, 8-cell embryos, morula, late blastocysts, and also early passages of human embryonic stem cell lines (71 files in total) were downloaded from Yan *et al.* [65] (**Table 2.1**). The full list of all of the files downloaded, including the accession and ID numbers are listed in the Supplement.

Table 2.1. Summary of RNA-seq data from solid primary tumour/non-diseased tissue

Tissue Source	Tissue Type
Breast invasive carcinoma (BRCA)	Primary solid tumour, solid tissue normal (NT)
Cervical squamous cell carcinoma/endocervical carcinoma (CESC)	Primary solid tumour, solid tissue normal (NT)
Cholangiocarcinoma (CHOL)	Primary solid tumour, solid tissue normal (NT)
Colon adenocarcinoma (COAD)	Primary solid tumour, solid tissue normal (NT)
Lymphoid neoplasm large B cell lymphoma (LBCL)	Primary solid tumour
Liver hepatocellular carcinoma (LIHC)	Primary solid tumour, solid tissue normal
Lung adenocarcinoma (LUAD)	Primary solid tumour, solid tissue normal (NT)
Lung squamous cell carcinoma (LUSC)	Primary solid tumour, solid tissue normal (NT)
Ovarian serous cystadenocarcinoma (OV)	Primary solid tumour
Pancreatic adenocarcinoma (PAAD)	Primary solid tumour, solid tissue normal
Prostate adenocarcinoma (PRAD)	Primary solid tumour, solid tissue normal (NT)
Skin cutaneous melanoma (SKCM)	Primary solid tumour, solid tissue normal (NT)
Testicular germ cell tumour (TGCT)	Primary solid tumour

Thyroid carcinoma (THCA)	Primary solid tumour, solid tissue normal (NT)
Uveal melanoma (UVM)	Primary solid tumour
Ovary*	Solid tissue normal (NT)
Testis*	Solid tissue normal (NT)

Note: All except for marked * were sequenced at the University of North Carolina Lineberger Comprehensive Cancer Centre and come from The Cancer Genome Atlas project. Those marked with * were sequenced at the European Bioinformatics Institute and come from the Illumina Human Body Map 2.0 Project.

2g. Non-diseased reference tissue selection

We first wanted to evaluate what the background expression of HK2 genes *np9* and *rec* are in normal, non-tumour tissues. Normal matched controls are not available for the majority of the tumour samples in TCGA so we decided to use a selection of random normal non-tumour tissues that are provided by TCGA and Illumina Human Body Map 2.0. The non-diseased tissue (NT) reference group is a mix of different healthy solid tissues, the majority of which are matched control tissue samples downloaded from TCGA. These tissue types include the cervix, bile duct, colorectum, liver, lung, ovary, pancreas, prostate, testis, and thymus. To ensure that there is no significant bias in the non-diseased reference tissue samples selected for this comparison group, we also measured the expression of other groups of non-cancerous tissues from TCGA. These other groups **(Table 2.2)** consist of non-diseased breast tissue samples (BRCA_NT), tissues that came from organs that adenocarcinomas are derived from (AD_NT – bile duct, colorectum, lung, pancreas, prostate), tissues that come from organs that squamous cell carcinomas are derived from (SQ_NT – cervix, lung), and tissues that come from organs that cancer types are derived from that do not fit in the other

categories (OT_NT – liver, thymus). All accession/ID numbers of the samples used are in the Supplement.

Table 2.2. Summary of groups of non-diseased reference tissue samples used for analysis

Category	Tissue source
NT (non-diseased tissue reference group)	Bile duct (*n* = 2), cervix (*n* = 3), colorectum (*n* = 1), liver (*n* = 4), lung (*n* = 2), ovary (*n* = 1), pancreas (*n* = 1), prostate (*n* = 1), testis (*n* = 1), thymus (*n* = 2)
BRCA_NT (breast tissue only)	Breast (*n* = 18)
AD_NT (organs that adenocarcinomas are derived from)	Bile duct (*n* = 5), colorectum (*n* = 5), lung (*n* = 9), pancreas (*n* = 4), prostate (*n* = 5)
SQ_NT (organs that squamous cell carcinomas are derived from)	Cervix (*n* = 3), lung (*n* = 5)
OT_NT (organs that other cancer types are derived from)	Liver (*n* = 10), thymus (*n* = 5)

2h. Calculation of relative difference in expression between *np9* and *rec* in cancers compared to non-diseased tissue reference group

In order to determine the difference in relative expression of *np9* and *rec* in the different cancers compared to the non-diseased tissue reference group, the average read count of the HK2 gene of each sample was divided by the average read count of *actb* within the same sample. The average normalised read count of HK2 gene/*actb* of each sample is then calculated for each tissue group. Finally, the ratio of the average relative expression value of the HK2

gene of each cancer type divided by the average relative expression value of the non-diseased tissue reference group (NT) is calculated.

2i. Statistical Analysis

The Wilcoxon signed rank-sum test was used to compare the relative expression (normalised within each sample using *actb*) of the HK2 genes from each of the unmatched samples within each cancer type compared to the non-diseased tissue reference group (NT). The Wilcoxon signed rank-sum test was also used in comparing the relative expression of the same HK2 genes within the NT group to other non-diseased tissue categories (BRCA_NT, AD_NT, SQ_NT, OT_NT). Results with a $p < 0.05$ are considered significant. We also corrected for multiple comparisons using the Bonferroni method and report p-values < 0.0033 as significant.

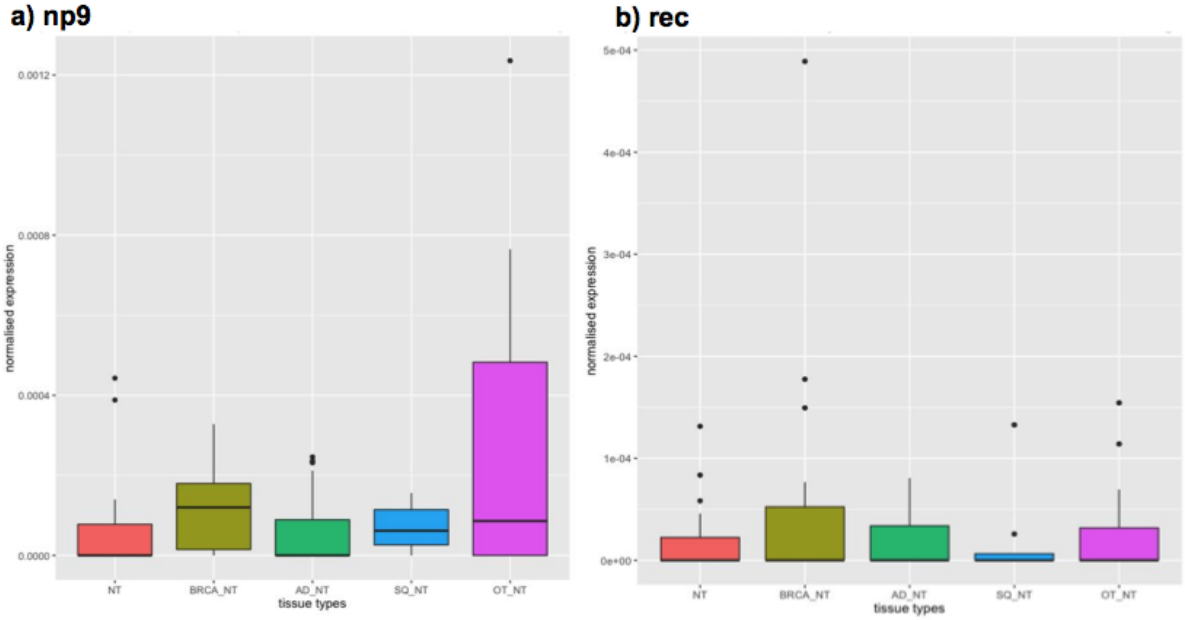
For visualisation, the normalised values were transformed in order to account for samples with read counts of 0: $\log((\text{normalised read count})10{,}000+1)_{10}$).

For all statistical methods, p-value < 0.05 is considered statistically significant. All statistical analyses were performed with R (version 3.3.1) in the integrated development environment (IDE) "Rstudio" (version 1.1.383).

3. Results

3a. Describing HK2 expression in different categories of non-diseased tissue

To ensure that there is no significant bias in the non-diseased tissue samples selected for this comparison group, we also measured the expression of other groups of non-cancerous tissues from TCGA and found that there was no significant difference between the expression of the HK2 targets in non-diseased tissue reference group (NT) and other non-diseased tissue types. *np9* and *rec* in the non-diseased tissue reference group (NT) is not significantly different compared to the expression of the same HK2 genes in breast tissues (BRCA_NT), tissues from organs that adenocarcinomas are derived from (AD_NT), tissues from organs that squamous cell carcinomas are derived from (SQ_NT), and tissues from organs with cancer types that do not fit in the other categories (OT_NT) (**Fig. 2.3**, **Table 2.3**).



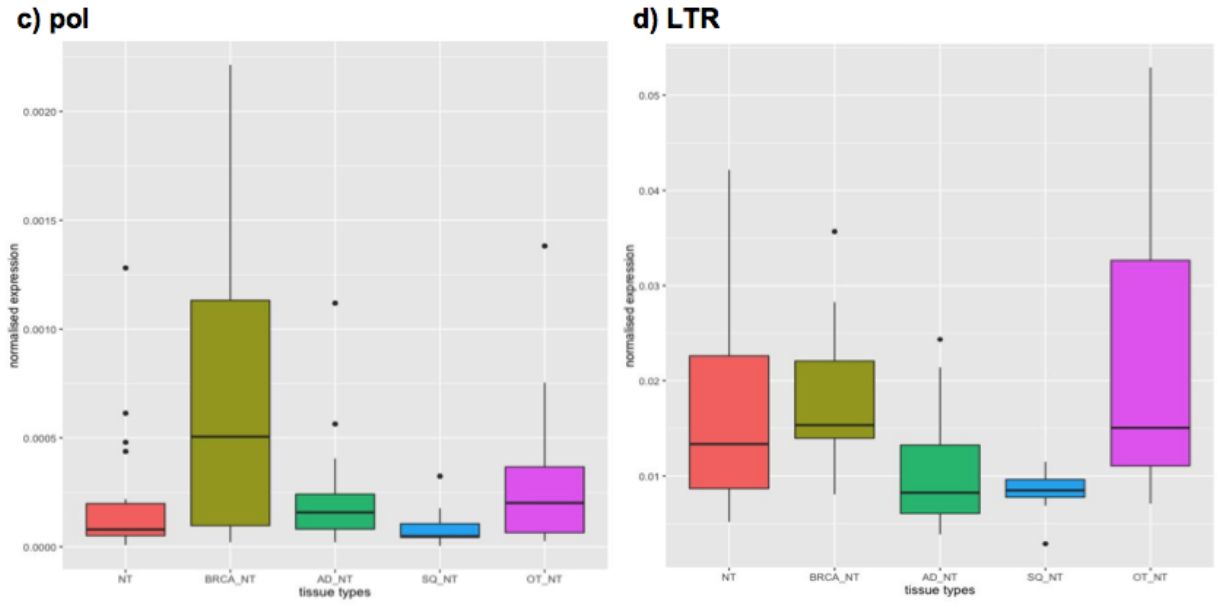


Figure 2.3. Describing HK2 expression in different categories of non-diseased tissue

x-axis indicates non-diseased tissue group types (from left to right): NT – pan-non-diseased tissue reference which includes tissues from bile duct, cervix, colon, liver, lung, ovary, pancreas, prostate, testis, and thymus (n = 18), BRCA_NT – breast tissue (n = 18), AD_NT – tissues from bile duct, colorectum, lung, pancreas, and prostate (n = 28), SQ _NT – tissues from cervix and lung (n = 8), OT_NT – tissues from liver and thymus (n = 15); y-axis indicates HERV transcription normalised by transcription of *actb*. Black horizontal lines indicate the mean. Non-diseased reference tissue (NT) group is placed on the far left in all box and whisker plots, for clarity. Shown is the relative expression of **(a.)** *np9*; **(b.)** *rec*; **(c.)** HK2 *pol*; **(d.)** HERVK-113 LTR across the different categories of non-diseased tissues.
Note: Because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

Table 2.3. Relative expression of HK2 in different non-diseased tissue group categories compared to non-diseased tissue reference (NT)

a)

np9			*rec*		
Category	**W test statistic**	**p-value**	**Category**	**W test statistic**	**p-value**
BRCA_NT	219	0.0635	BRCA_NT	181	0.496
AD_NT	247	0.9109	AD_NT	240	0.751
SQ_NT	90	0.306	SQ_NT	68	0.812
OT_NT	169	0.1925	OT_NT	133	0.937

b)

pol			**LTR**		
Category	**W test statistic**	**p-value**	**Category**	**W test statistic**	**p-value**
BRCA_NT	245	0.007874 *	BRCA_NT	190	0.389
AD_NT	307	0.2224	AD_NT	142	0.0127 *
SQ_NT	54	0.3383	SQ_NT	36	0.0457 *
OT_NT	162	0.3428	OT_NT	160	0.381

(a.) The expression of *np9* and *rec* and the expression of **(b.)** HK2 *pol* and HERVK-113 LTR in samples within the NT reference group in comparison to other non-diseased tissue groups. NT – pan-non-diseased tissue reference which includes tissues from bile duct, cervix, colon, liver, lung, ovary, pancreas, prostate, testis, and thymus (n = 18), BRCA_NT – breast tissue (n = 18), AD_NT – tissues from bile duct, colorectum, lung, pancreas, and prostate (n = 28), SQ _NT – tissues from cervix and lung (n = 8), OT_NT – tissues from liver and thymus (n = 15)
* indicates $p < 0.05$

Note: Because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

---

np9 and *rec* expression from BRCA_NT, AD_NT, SQ_NT, and OT_NT compared to NT was not statistically significantly different ($p > 0.05$). However, a significant difference in relative expression of *pol* was observed in samples from the BRCA_NT group compared to NT ($p = 0.007874$). The expression of LTR was also significantly different within the NT samples compared to the AD_NT group ($p = 0.01265$) and SQ_NT ($p = 0.045699$). Because there appears to be no significant difference in expression levels of *rec* and *np9* between the different NT groups, we reasoned that the samples used in NT would be adequate to use as a non-diseased reference group when comparing expression profiles to different cancers.

3b. HK2 expression in early embryogenesis and stem cell lines

Transcription of *np9* and *rec* is activated at, and is highest at the 8-cell stage of early embryogenesis, but *rec* expression is overall silenced during the following developmental stages (**Fig. 2.4b**). Interestingly, *np9* is transcriptionally active (but downregulated) in morula and in the early passages of the hESC lines (**Fig. 2.4a**), and both *pol* and HERV-K113 LTR show upregulation in the late blastocyst that is not seen in *np9* and *rec* (**Fig. 2.4c-d**). This discrepancy is because because the transcription of *pol* and LTR most likely come from multiple HK2 elements, and therefore these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

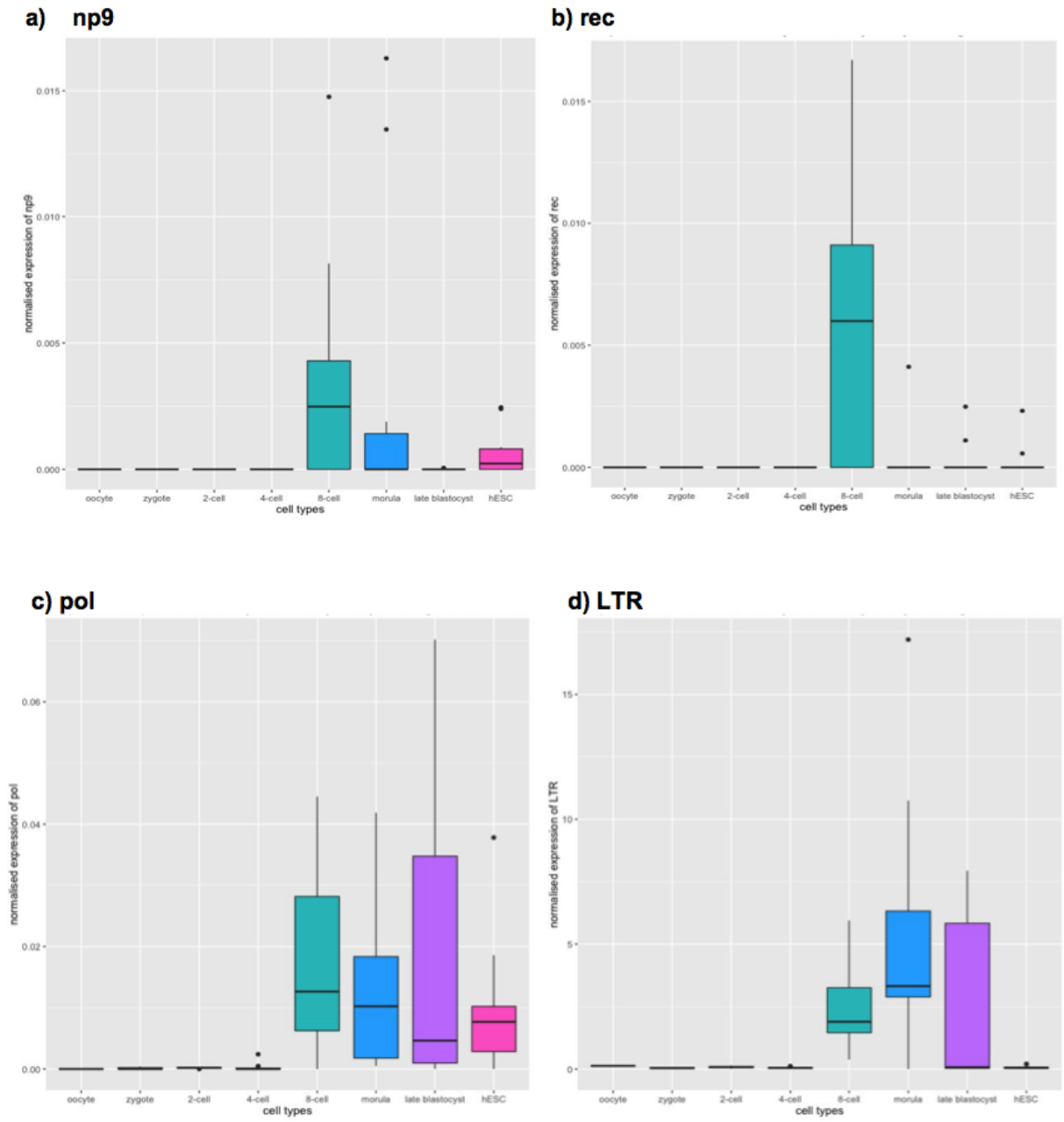


Figure 2.4. Expression of HK2 targets in human preimplantation embryos and embryonic stem cells

x-axis indicates cell types (from left to right): Oocyte (n = 3), zygote (n = 3), 2-cell (n = 4), 4-cell (n = 12), 8-cell (n = 17), morula (n = 10), late blastocyst (n = 10), hESC (n = 13); y-axis indicates HERV transcription normalised by transcription of *actb*. Black horizontal lines indicate the median. Data are from (15). Expression of (a.) *np9*; (b.) *rec*; (c.) HK2 *pol*; (d.) HERVK-113 LTR

in the different stages of human preimplantation embryos and human embryonic stem cell lines.
Note: Because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

3c. HK2 expression in non-diseased tissues and cancer samples

When we compared the expression of the HK2 genes *np9* and *rec* in non-diseased tissue reference and the 15 different cancers, we found that there is a general trend of HERV gene upregulation across the cancers, with some cancers that show significant overexpression **(Table 2.4)**.

Table 2.4. Relative expression of HK2 targets in cancers compared to NT

a) *np9*

Cancer type	§∑*np9*(cancer)/*np9*(NT)	p-value
BRCA	2.9	0.0731
CESC	3.9	0.0764
CHOL	4.9	$8.68 \cdot 10^{-3}$ *
COAD	1.2	0.837
LBCL	1.5	0.102
LIHC	5.7	$5.76 \cdot 10^{-3}$ *
LUAD	4.2	$7.11 \cdot 10^{-4}$ **

LUSC	0.9	0.370
OV	3.0	0.0531
PAAD	1.9	0.176
PRAD	0.7	0.679
SKCM	2.9	0.0598
TGCT	55.5	$9.83 \cdot 10^{-7}$ **
THCA	1.8	0.0433 *
UVM	12.9	0.0112 *

b) *rec*

Cancer type	§∑*rec*(cancer)/*rec*(NT)	p-value
BRCA	0.7	0.439
CESC	0.6	0.459
CHOL	1.2	0.986
COAD	0.6	0.247
LBCL	0.4	0.743
LIHC	3.0	0.533
LUAD	16.6	0.217
LUSC	0.9	0.879

OV	1.0	0.931
PAAD	0.4	0.632
PRAD	0.7	0.281
SKCM	43.7	0.0411 *
TGCT	83.0	$1.53 \cdot 10^{-6}$ **
THCA	3.1	0.0923
UVM	1469.5	$5.81 \cdot 10^{-7}$ **

c) *pol*

Cancer type	§∑*pol*(cancer)/*pol*(NT)	p-value
BRCA	19.6	$1.34 \cdot 10^{-6}$ **
CESC	1.2	0.521
CHOL	2.6	$2.07 \cdot 10^{-3}$ **
COAD	20.8	$1.48 \cdot 10^{-3}$ **
LBCL	0.7	0.965
LIHC	2.1	0.0476 *
LUAD	2.1	0.0383 *
LUSC	0.9	0.346
OV	2.2	0.0160 *

PAAD	2.6	$8.20 \cdot 10^{-4}$ **
PRAD	5.8	$2.86 \cdot 10^{-3}$ **
SKCM	9.5	$4.30 \cdot 10^{-4}$ **
TGCT	7.9	$2.21 \cdot 10^{-8}$ **
THCA	1.8	0.0188 *
UVM	23.5	$3.41 \cdot 10^{-7}$ **

d) LTR

Cancer type	§$\sum LTR$(cancer)/LTR(NT)	p-value
BRCA	6.0	$1.65 \cdot 10^{-3}$ **
CESC	2.4	0.0266 *
CHOL	1.6	0.0412 *
COAD	3.6	0.196
LBCL	0.6	0.0205 *
LIHC	2.1	$4.99 \cdot 10^{-4}$ **
LUAD	0.8	0.497
LUSC	1.6	0.0930
OV	2.2	0.0146 *
PAAD	0.8	0.534

PRAD	2.2	0.0305
SKCM	2.1	0.443
TGCT	10.4	$3.07 \cdot 10^{-7}$ **
THCA	1.4	0.0990
UVM	13.3	$1.15 \cdot 10^{-5}$ **

Relative expression of **(a.)** *np9*; **(b.)** *rec*; **(c.)** *pol*; and **(d.)** LTR in cancer compared to non-diseased reference tissue.
§ is the sum of the ratio of the average read count of the HERV gene divided by the average read count of *actb* of each sample in each tissue type (cancer or non-diseased reference tissue group).
* $p < 0.05$
** $p < 0.0033$
Note: Because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

---

np9 is significantly upregulated by ~5x in cholangiocarcinoma ($p = 0.009$), almost 5.5x higher expression in liver hepatocellular carcinoma ($p = 0.006$), 4x higher expression in lung adenocarcinoma ($p = 0.001$), ~2x higher expression in thyroid carcinoma ($p = 0.043$), and ~13x higher expression in uveal melanoma ($p = 0.011$). *np9* was highly upregulated in testicular germ cell cancer, with 55.5x fold change in overexpression compared to the non-diseased tissue group testicular germ cell cancer ($p = 9.83 \cdot 10^{-7}$) **(Fig. 2.5a)**. On the other hand, *rec* is significantly upregulated in 3 cancers. In skin cutaneous melanoma, *rec* is overexpressed by ~44x compared to the expression of *rec* in the non-diseased tissue group ($p = 0.041$), overexpressed by 83x in testicular germ cell cancer ($p = 1.53 \cdot 10^{-6}$), and is highly upregulated in uveal melanoma by more than a thousandfold ($p = 5.81 \cdot 10^{-7}$) (**Fig. 2.5b**).

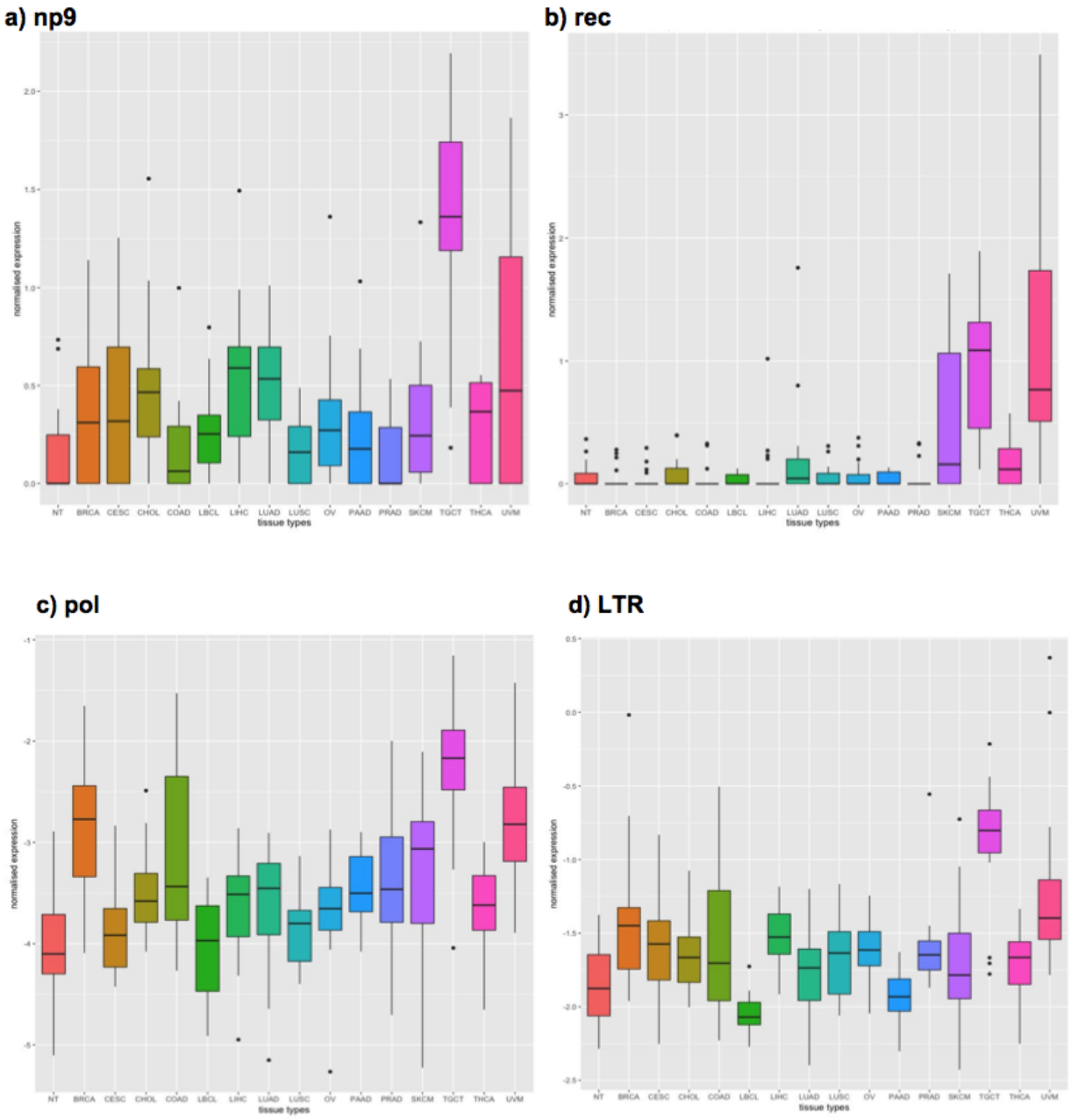


Figure 2.5. Expression of HK2 genes in non-diseased reference tissue group (NT) compared to 15 different cancers

x-axis indicates tissue types (from left to right): NT – normal tissue reference group (n = 18); BRCA – breast invasive carcinoma (n = 20); CESC – cervical squamous cell carcinoma/endocervical adenocarcinoma (n = 18); CHOL – cholangiocarcinoma (n = 20); COAD – colorectal adenocarcinoma (n = 20); LUSC – lung squamous cell carcinoma (n = 20); OV – ovarian serous cystadenocarcinoma (n = 20); PAAD – pancreatic adenocarcinoma (n = 20);

PRAD – prostate adenocarcinoma (n = 20); SKCM – skin cutaneous melanoma (n = 24); TGCT – testicular germ cell cancer (n = 20); THCA – thyroid carcinoma (n = 20); UVM – uveal melanoma (n = 20). y-axis indicates HK2 transcription normalised by transcription of *actb*, on a scale transformation of $\log_{10}$(normalised read count)10,000+1). Black horizontal lines indicate the median. Normal tissue (NT) group is placed on the far left for clarity. Expression of (a.) *np9*; (b.) *rec*; HK2 *pol*; and (d.) HERVK-113 LTR in NT compared to 15 different cancers.
Note: Because the transcription of *pol* and LTR are likely to come from multiple HK2 elements, these results should not be seen as directly comparable to the transcription of *np9* and *rec*.

---

4. Discussion

Retroviruses have the ability to contribute to tumourigenesis by activating cellular oncogenes or mutating cellular tumour suppressor genes, resulting in aberrant gene expression that promotes a cascade of uncontrolled cellular proliferation [72][73]. Although these tumourigenic mechanisms by ERVs have been observed in other host organisms, namely in mice, there is yet no evidence of ERVs directly contributing to cancer pathogenesis in humans.

In comparison to non-diseased tissue, we had expected to see HK2 target overexpression in breast cancer, germ cell tumours, skin melanoma, lymphoma, prostate cancer, liver cancer, and prostate cancer. We also surveyed other cancer types that, to our knowledge, have not been examined for HK2 expression. In addition, because the HK2-encoded accessory genes *np9* and *rec* may be considered as biomarkers to specific cancers, due to their association with the activation of signalling pathways implicated in cancers, it was our objective to explore their tissue-specific expression further.

In agreement with previously reported results [12][74], HK2 appears to be derepressed and transcriptionally active starting from the 8-cell and 16-cell

stage (morula) of human embryonic development, and varying levels of transcription in the blastocyst and in its outgrowths.

Breast cancer and testicular germ cell tumours are all cancers that derive from organs of the endocrine system. It is tempting to speculate that this link between HERV-K expression and some these cancer types is due to the expression of steroid hormones in glandular tissue. For example, within the U3 region of the HERV-K LTR, there are several predicted oestrogen, progesterone, and androgen response elements. *env* transcription has been shown to be upregulated up to 10-fold in breast cancer cell lines that had been treated with estradiol followed by progesterone. Moreover, HERV-K reverse transcriptase is overexpressed in breast cancer cell lines that have been treated with the estradiol/progesterone combination, suggesting that these steroid hormones indeed have a regulatory effect on HK2 transcription [21][75].

Because of the novel method we developed in processing and analysing the expression of highly similar repetitive elements in RNA-seq data, the strict parameters we used in interpreting the significance of the results may have led to false negatives. The Wilcoxon signed rank-sum test was chosen to compare the relative expression of the HK2 genes from each of the unmatched samples within each cancer type compared to non-diseased tissue, and the values were corrected for multiple comparisons using the Bonferroni method. p-values < 0.0033 were considered significant. Although, according to our parameters of significance, only a few cancers were determined as having *np9*, and *rec* transcripts upregulated in comparison to non-diseased tissue (**Fig. 2.3**, **Table 2.3**), there is a general trend of HK2 upregulation in the majority of the cancers surveyed, in comparison to the median expression of non-diseased tissue, including the cancer types that did not met our threshold of significance. Thus, the lack of significance in correlations and

comparisons is likely to reflect lack of power by adopting a conservative p-value rather than true absence of correlation.

We observed a bias in the abundance of transcription from the 5' end of HERV-K113 LTR, where the U3 region is (**Fig. 2.1a**), compared to the 3' end. In particular, transcription is most abundant at the GA-rich motif found at the 5' end. The uneven distribution of mapped reads could be due to biological biases in provirus expression. For example, in teratocarcinoma cell lines, both LTRs of full proviruses are transcribed, although the 3' LTR appears to be transcribed far less abundantly than the 5' LTR [79]. However, the significance or functional importance of truncated expression from the 3' LTR is unknown.

HK2 expression is very well-documented in testicular germ cell tumours – germ cell tumour cell lines have been shown to express HK2 non-infectious particles [17][76][77]. HERV-K has also been shown to be overexpressed at specific stages of early embryonic development [12][56][57], which we also observed using our novel method. Germ cell tumours also share with early embryonic cells similar traits of embryonic stem cell identity, such as the upregulation of the pluripotency factors Oct3/4, Nanog, and Sox2 [78].

During the early stages of normal development, primordial germ cells are also totipotent stem cells: the expression of pluripotency factors Oct3/4, Nanog, Sox2 gives primordial germ cells high self-renewal capacity by the suppression of differentiation [78]. A characteristic chromosomal abnormality seen in testicular germ cell tumours is a gain of the short arm of chromosome 12 (chr12p), resulting in the amplification of genes that map to that location, which includes stem-cell associated factors like Nanog, cell cycle regulators like cyclin D2 (*CCND2*), the oncogene KRAS, the growth factor receptor tumour necrosis factor receptor 1 (*TNFRSF1A*), and genes

involved in metabolism such as GAPDH and GLUT3 [78]. Oct4 and Nanog are known to bind to HERV LTRs, mediating the transcription of both host genes and the provirus [12]. It has been suggested that the increased levels of Nanog due to the chr12p gain may be enough to overexpress the other core pluripotency factors Oct3 and Sox2, encoded by genes located on different chromosomes [78].

Surprisingly, we found that *rec* is highly upregulated in uveal melanoma by more than a thousandfold compared to the non-diseased tissues (**Fig. 2.5b** & **Table 2.4b**). The high presence of *rec* in uveal melanoma suggests a yet undescribed role of the HK2 gene in the molecular biology of this rare cancer. Uveal melanoma is the most common intraocular malignant tumour in adults, and is biologically distinct from cutaneous melanoma, the most common melanoma subtype. Whereas cutaneous melanoma arises from the melanocytes in the basal layer of the epidermis, uveal melanoma arises from melanocytes in the uveal tract, the layer of tissue between the outer layer (cornea and sclera) and the inner layer (retina) of the eye. The metastatic and clinical behaviour of uveal melanoma also differs from cutaneous melanoma because of its tendency to disseminate in the bloodstream and metastasise to the liver (reviewed in [504][505][506]). Consequently, due to the high incidence of metastases, the global overall survival at five years is as low as 60% [507]. Uveal melanoma is characterised by mutations in the GNAQ or GNA11 genes, which activate signalling pathways, including the mitogen-activated protein kinase (MAPK) pathway, similar to the effect of the BRAF mutation in cutaneous melanoma [508][509]. Np9 protein has been shown to interact with MAPK1 [26], and we show in this study that *np9* is upregulated by ~13x compared to undiseased tissue in the uveal melanoma dataset. The immunnosuppressive environment of the eye helps protect eyesight, and uveal melanoma cells also uses similar processes to facilitate escape from immune surveillance [505].

The formation of cancers often does not have a single aetiological agent that directly forms the disease. Tumourigenesis and the formation of malignancies is a complex, multifactorial process that requires the presence of both genetic predisposition of the individual to the cancer as well as life history factors such as environmental exposure to carcinogens. It is possible that the polymorphic nature of HK2 as well as its potential to express oncogenes like *np9* and *rec* are key components in increasing the genetic susceptibility of an individual to that cancer.

It would also be interesting to further explore the scope of *np9* and *rec* upregulation in even more cancers. Although we used RNA-seq files from 15 different types of cancer, there are more cancer types reported in the literature to overexpress HERVs: colorectal cancer [80], chronic lymphocytic leukaemia (CLL) [29], and Burkitt's lymphoma [81]. Further research could also be performed to characterise the broad network that indicates how Np9 and Rec proteins interact with effectors in the MAPK/Akt/NOTCH signalling cascade that ultimately leads to cell survival and proliferation, specific to the cancers they are upregulated in.

5. Acknowledgements

Thank you to Fabricia Nascimento, Robert Belshaw, Tim Coulson, and the Palaeovirology group for helpful discussions and comments. Thank you to Paul Klenerman for his help in mediating communication between the co-authors.

6. Funding Statement

GM and TK have been supported by an MRC fellowship to GM (MR/K010565/1). CS is funded by Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile, and AK is funded by The Royal Society. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

7. Ethics Approval and Consent to Participate

All human data used were obtained through The Cancer Genome Atlas and used according to the Data Use Certification as stipulated by the TCGA Data Access Committee and dbGaP authorisation (https://dbgap.ncbi.nlm.nih.gov/aa/). Permission to use controlled-access data was approved by dbGaP under projects #57108 and #7621.

8. Supplement

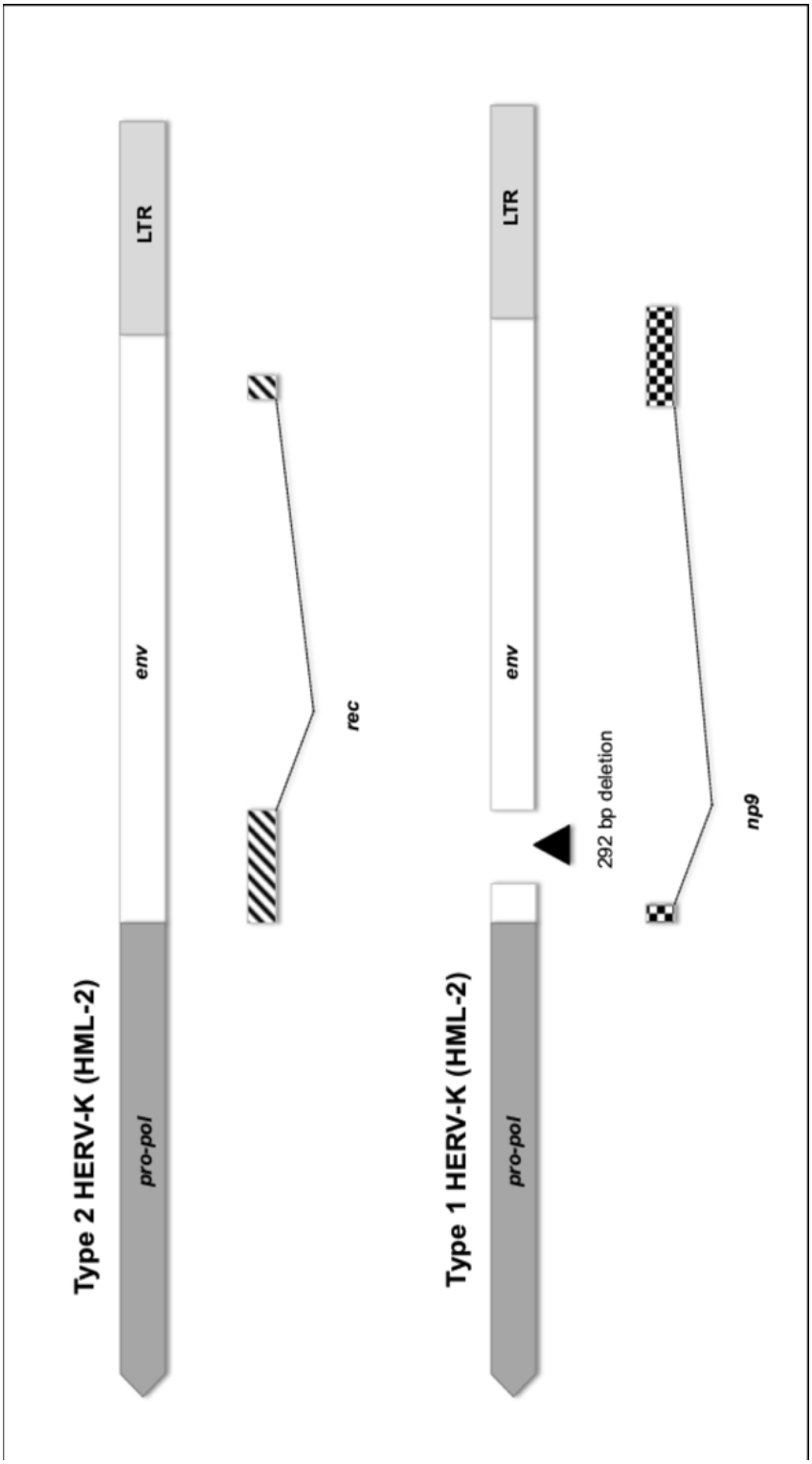


Figure S2.1. Partial schematics of HERV-K (HML-2) proviruses encoded gene products

Type 2 HERV-K (HML-2) endogenous retroviruses (top) encode for *rec*; Type I HERV-K (HML-2) HERVs have a 292 base pair deletion in the junction of *pol* and *env*, resulting in a different splice donor that produces an *np9* transcript instead of *rec*.

Table S2.1. HERV-K (HML-2) reference sequences used

Gene	Sequence
np9	GAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCC ATCGGAGATGCAAAGAAAAGGGCCTCCACGGAGATGTCTGCAGGTGTACCCAACA GCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGACGATGGCGGTTTTG TCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTG TGTCTGTGTAGAAAGAAGTAGACATGGGAGACTCCATTTTGTTATGTGTTAA
rec	ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCA ATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATG AAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCT GACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGA GTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGTCTGCAGGTGTACCCAAC AGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGACGATGGCGGTTTT GTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACT GTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTACTAAGAA AAATTCTTCTGCCTTGAGATTCTGTT

HERV-K (HML-2) ***pol***

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATT
AAAATTCTTTTAAAAAGAGGGGGAGTTAAAGTATCTACAAAAAATCTAATCAAGCT
ATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATCT
AAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAAT
ATCATTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCA
TTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAGCTGTATAAT
AGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATT
GCGAATATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAAT
CAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGA
ATTAGTGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTC
AGGTGCCTGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCG
CCAGTAGCCTATCAATACTGGCCTCCGGCTGAACTTCAGTATCGGCCACCCCCAGA
AAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGGGCGCCATAC
CCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGTAG
TAAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGC
AATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAGGAGCCCAAGAGGGAGA
GCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAAAGCTAAAAGATA
TGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCCTTATATGAGGACATTATTA
GATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTGGGAGATTCTGGCAAA
ATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGT
ACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAG
ATCAACTATTAGGAATAGGTCAAAATTGGAGTACTATTAGTCAACAAGCATTAATG
CAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT
CCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAG
AGCCCTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCT
GATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACGCCAATCC
TGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATG
TAATCTCAGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAAGCT
ATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAGGACAAGTTAGAACATT
TGGAAGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAATTGCCCAG
TCTTAAATAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCT
GACTTATGTCCAAGATGTAAAAAAGGAAAACATTGGGCTAGTCAATGTCGTTCTAA
ATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAG
GCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCA
GGGACAACAACCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAAT
ACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAG

S2.2 Supplementary spreadsheet

TCGA_tumor samples

Sample ID	Cancer type	Sample Type
TCGA-A7-A13G	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A7-A26F	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-A8-A06Q	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A8-A07C	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-AC-A2QH	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-AC-A3QQ	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-AN-A0FZ	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-AR-A24L	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-AR-A251	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BH-A1EU	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BH-A1EX	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BH-A204	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-D8-A1JH	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-D8-A1JL	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-E2-A1B1	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-E9-A1N3	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-E9-A1RC	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-E9-A22D	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-PE-A5DE	Breast Invasive Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-C5-A1BF	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-C5-A3HL	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DS-A0VM	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DS-A0VN	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DS-A1OD	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EA-A1QT	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EA-A411	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EK-A2IR	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EK-A2RB	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EK-A2RJ	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EK-A2RK	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EK-A3GJ	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EX-A3L1	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FU-A40J	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HM-A3JJ	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IR-A3LC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IR-A3LH	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-Q1-A5R3	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VS-A953	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-ZJ-AAXA	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-3X-AAV9	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-3X-AAVB	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2O	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2Q	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2R	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2T	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2U	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2X	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA2Z	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA30	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA31	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W5-AA33	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W6-AA0S	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WD-A7RX	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-YR-A95A	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ZH-A8Y2	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ZH-A8Y4	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-ZH-A8Y5	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ZH-A8Y6	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ZH-A8Y8	Cholangiocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-5656	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-5657	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-5659	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-6648	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-6780	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-A6-6781	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-CK-4950	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-D5-6923	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A1D0	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A1D8	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A1D9	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A1DA	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A1HB	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DM-A285	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-QG-A5YW	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-QG-A5Z2	Colon Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FA-8693	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FA-A4XK	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FA-A6HN	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FA-A6HO	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-G8-6326	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-GR-A4D5	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-GR-A4D9	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-RQ-A68N	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-RQ-A6JB	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BC-4072	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BC-A10T	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BC-A10X	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BC-A216	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BC-A3KF	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BD-A2L6	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-CC-A1HT	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-CC-A3MB	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-CC-A7IL	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-A113	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-A11C	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-A1ED	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-A4NI	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-A4NP	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-AACJ	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-AACK	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-AADK	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-AADV	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DD-AADW	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-UB-A7MC	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4244	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4382	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4395	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4398	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4418	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4420	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-05-4422	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4426	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4430	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4432	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-05-4434	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-35-3615	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-38-4625	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-38-4626	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-38-4628	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-38-4629	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-38-4630	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-44-2655	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-44-2656	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-44-2659	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-55-8089	Lung Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-18-5592	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-18-5595	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-21-1081	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-21-5786	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-21-5787	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-0940	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-0944	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-4591	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-4604	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5474	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5477	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5478	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5480	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5483	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5489	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-22-5492	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-33-4587	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-33-6737	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-34-5231	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-34-5236	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-85-6175	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-04-1348	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-04-1365	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-09-0369	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-09-1659	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-13-1498	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-24-1434	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-24-1558	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-24-2020	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-24-2038	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-25-1631	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-25-2399	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-29-1696	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-57-1994	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-59-2354	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-59-A5PD	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-61-2002	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-61-2095	Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-3A-A9IB	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-F2-6879	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-F2-7273	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-F2-7276	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-7923	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-7924	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-7925	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-8315	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-8519	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-8636	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-8637	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HZ-8638	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-7647	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-7654	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-7885	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-7887	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-7893	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-8127	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IB-A5SS	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-IB-AAUV	Pancreatic Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-CH-5753	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-5509	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-5511	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-5527	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-5530	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-7781	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-7791	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EJ-8470	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-G9-6366	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-G9-6384	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-G9-6498	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HC-7211	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HC-7212	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HC-7736	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HC-8261	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HI-7171	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-J4-8198	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-J4-A83L	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-XK-AAJR	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-YL-A9WX	Prostate Adenocarcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-A1PU	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-A1PV	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-A1PX	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-A1PZ	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-A3DN	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BF-AAP1	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A24C	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A24D	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A3XB	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A430	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A431	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A4IS	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EB-A6R0	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ER-A2NF	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FR-A2OS	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-FR-A726	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-FS-A1ZA	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-GF-A2C7	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-GF-A769	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-GN-A26C	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-HR-A5NC	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-IH-A3EA	Skin Cutaneous Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-W4-A7U4	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAEW	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAL5	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAGY	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XE-AANJ	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XY-A89B	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XE-AAO3	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-SN-A84W	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAHG	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-SN-A6IS	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAF6	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor

TCGA-S6-A8JY	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XE-A8H5	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XE-A8H1	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-W4-A7U2	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-YU-A90P	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-SB-A6J6	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2X-A9D6	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-2G-AAGA	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-XY-A9T9	Testicular Germ Cell Cancer (TCGA, Provisional)	Primary Solid Tumor
TCGA-BJ-A0ZB	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-BJ-A28V	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DJ-A1QH	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DJ-A2PX	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-DJ-A2Q4	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-E8-A414	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EL-A3MZ	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EL-A3N2	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EL-A3T1	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-EL-A3T6	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EM-A3O8	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EM-A3O9	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EM-A4FQ	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-EM-A4FR	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ET-A25G	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-ET-A25I	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-H2-A421	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-J8-A3YG	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EE	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EI	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EJ	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9ES	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EW	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EX	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9F0	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9F1	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9F7	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-VD-A8K7	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8K8	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8K9	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KA	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KG	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KH	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KM	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-AA8R	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-AA8T	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-A87T	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-AA9A	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-AA9E	Thyroid Carcinoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EE	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EI	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EJ	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9ES	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EW	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9EX	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor

TCGA-V4-A9F0	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9F1	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-V4-A9F7	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8K7	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8K8	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8K9	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KA	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KG	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KH	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-A8KM	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-AA8R	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-VD-AA8T	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-A87T	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-AA9A	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor
TCGA-WC-AA9E	Uveal Melanoma (TCGA, Provisional)	Primary Solid Tumor

TCGA_normal samples

Green highlight indicates samples in Normal Tissue Reference group

Case Submitter ID	Project Name	Primary Site	Sample Type
TCGA-BH-A18N	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0H9	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0BW	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A204	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-E9-A1RC	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A1F0	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-A7-A13G	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0DZ	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-A7-A0CH	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A1ET	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-A7-A0DC	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-E9-A1N6	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-E2-A15K	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-E9-A1R7	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0H7	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal

TCGA-E9-A1ND	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A18V	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0B5	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-BH-A0HK	Breast Invasive Carcinoma (TCGA, Provisional)	Breast	Solid Tissue Normal
TCGA-FU-A3EO	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-MY-A5BF	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-HM-A3JJ	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-MY-A5BF	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-FU-A3EO	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-HM-A3JJ	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)	Cervix	Solid Tissue Normal
TCGA-ZU-A8S4	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-W5-AA34	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-W5-AA30	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-W5-AA31	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-W5-AA2R	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-ZU-A8S4	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Blood Derived Normal
TCGA-W5-AA34	Cholangiocarcinoma (TCGA, Provisional)	Bile Duct	Solid Tissue Normal
TCGA-AA-3522	Colon Adenocarcinoma (TCGA, Provisional)	Colorectal	Solid Tissue Normal

TCGA-AA-3517	Colon Adenocarcinoma (TCGA, Provisional)	Colorectal	Solid Tissue Normal
TCGA-AA-3527	Colon Adenocarcinoma (TCGA, Provisional)	Colorectal	Solid Tissue Normal
TCGA-AA-3520	Colon Adenocarcinoma (TCGA, Provisional)	Colorectal	Solid Tissue Normal
TCGA-A6-2682	Colon Adenocarcinoma (TCGA, Provisional)	Colorectal	Solid Tissue Normal
TCGA-DD-A1EL	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A39X	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A1EJ	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-BC-A10Q	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A3A2	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-BC-A10U	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-ES-A2HT	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A39Z	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A11A	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-EP-A26S	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A1EJ	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A1EC	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-DD-A1EI	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal
TCGA-BC-A216	Liver Hepatocellular Carcinoma (TCGA, Provisional)	Liver	Solid Tissue Normal

TCGA-55-6985	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-55-6986	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-6147	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-2661	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-3398	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-50-5933	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-5645	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-2657	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-50-6595	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-44-2665	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-38-4626	Lung Adenocarcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-56-7222	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-90-6837	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-22-5489	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-43-6771	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-22-5491	Lung Squamous Cell Carcinoma (TCGA, Provisional)	Lung	Solid Tissue Normal
TCGA-H6-A45N	Pancreatic Adenocarcinoma (TCGA, Provisional)	Pancreas	Solid Tissue Normal
TCGA-H6-8124	Pancreatic Adenocarcinoma (TCGA, Provisional)	Pancreas	Solid Tissue Normal

TCGA-YB-A89D	Pancreatic Adenocarcinoma (TCGA, Provisional)	Pancreas	Solid Tissue Normal
TCGA-HV-A5A3	Pancreatic Adenocarcinoma (TCGA, Provisional)	Pancreas	Solid Tissue Normal
TCGA-G9-6496	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-EJ-7782	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-G9-6356	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-EJ-7786	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-J4-A83J	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-EJ-7793	Prostate Adenocarcinoma (TCGA, Provisional)	Prostate	Solid Tissue Normal
TCGA-GN-A4U8	Skin Cutaneous Melanoma (TCGA, Provisional)	Skin	Solid Tissue Normal
TCGA-BJ-A2N8	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-EL-A3T3	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-BJ-A290	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-KS-A41L	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-EL-A3ZQ	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-H2-A2K9	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal
TCGA-EL-A3TB	Thyroid Carcinoma (TCGA, Provisional)	Thymus	Solid Tissue Normal

Sequence read archive (SRA) samples

Green highlight indicates samples in Normal Tissue Reference group

Sequence read archive (SRA) ID	Cell type
SRR445718	oocyte
SRR445719	oocyte
SRR445720	oocyte
SRR445721	zygote
SRR445722	zygote
SRR445723	zygote
SRR445724	2 cell embryo
SRR445726	2 cell embryo
SRR445727	2 cell embryo
SRR445728	2 cell embryo
SRR490961	4 cell embryo
SRR490962	4 cell embryo
SRR490963	4 cell embryo
SRR490964	4 cell embryo
SRR490965	4 cell embryo

SRR490966	4 cell embryo
SRR490967	4 cell embryo
SRR490968	4 cell embryo
SRR490969	4 cell embryo
SRR490970	4 cell embryo
SRR490971	4 cell embryo
SRR490972	4 cell embryo
SRR490973	8 cell embryo
SRR490974	8 cell embryo
SRR490975	8 cell embryo
SRR490976	8 cell embryo
SRR490977	8 cell embryo
SRR490978	8 cell embryo
SRR490979	8 cell embryo
SRR490980	8 cell embryo
SRR490981	8 cell embryo
SRR490982	8 cell embryo
SRR490983	8 cell embryo

SRR490984	8 cell embryo
SRR490985	8 cell embryo
SRR490986	8 cell embryo
SRR490987	8 cell embryo
SRR490988	8 cell embryo
SRR490989	8 cell embryo
SRR490990	morulae
SRR490991	morulae
SRR490992	morulae
SRR491039	morulae
SRR490993	morulae
SRR490994	morulae
SRR490995	morulae
SRR490996	morulae
SRR490997	morulae
SRR491001	morulae
SRR491002	late blastocyst
SRR491003	late blastocyst

SRR491004	late blastocyst
SRR491005	late blastocyst
SRR491009	late blastocyst
SRR491010	late blastocyst
SRR491011	late blastocyst
SRR491012	late blastocyst
SRR491013	late blastocyst
SRR491021	late blastocyst
SRR491022	hESC
SRR491023	hESC
SRR491024	hESC
SRR491025	hESC
SRR445730	hESC
SRR445731	hESC
SRR491039	hESC
SRR491040	hESC
SRR491041	hESC
SRR491042	hESC

SRR491043	hESC
SRR491045	hESC
ERR030873	healthy testis
ERR030878	healthy lymph node
ERR030874	healthy ovary

tissue_type	sample number	np9_raw_count	rec_raw_count
BRCA	TCGA-A7-A13G-01	18	0
BRCA	TCGA-A7-A26F-01	10	1.25
BRCA	TCGA-A8-A06Q-01	3	0
BRCA	TCGA-A8-A07C-01	7	0
BRCA	TCGA-AC-A2QH-01	3	0
BRCA	TCGA-AC-A3QQ-01	0	0
BRCA	TCGA-AN-A0FZ-01	3	0
BRCA	TCGA-AR-A24L-01	3	0
BRCA	TCGA-AR-A251-01	5	0
BRCA	TCGA-BH-A1EU-01	0	0
BRCA	TCGA-BH-A1EX-01	2	0
BRCA	TCGA-BH-A204-01	0	0.625
BRCA	TCGA-D8-A1JH-01	0	0
BRCA	TCGA-D8-A1JL-01	2	3.75
BRCA	TCGA-E2-A1B1-01	3	0
BRCA	TCGA-E9-A1N3-01	0	0
BRCA	TCGA-E9-A1RC-01	2	0
BRCA	TCGA-E9-A22D-01	0	0.625
BRCA	TCGA-PE-A5DE-01	0	0
CESC	TCGA-DS-A0VM	11	1.2
CESC	TCGA-EK-A3GJ	16	0
CESC	TCGA-FU-A40J	9	0
CESC	TCGA-EK-A2RK	11	0
CESC	TCGA-EA-A411	3.833333333	0.6
CESC	TCGA-Q1-A5R3	0	3
CESC	TCGA-VS-A953	0	0
CESC	TCGA-DS-A0VN	0	0
CESC	TCGA-IR-A3LH	7	
CESC	TCGA-C5-A3HL	0	0
CESC	TCGA-ZJ-AAXA	2	0
CESC	TCGA-EA-A1QT	2	0
CESC	TCGA-EK-A2IR	0	0.6
CESC	TCGA-IR-A3LC	3	0
CESC	TCGA-DS-A1OD	0	0
CESC	TCGA-EK-A2RJ	39	0
CESC	TCGA-HM-A3JJ	1	0
CESC	TCGA-EK-A2RB	13	1.2
CESC	TCGA-C5-A1BF	0	0
CESC	TCGA-EX-A3L1	0	0
CHOL	TCGA-ZH-A8Y4	1	0
CHOL	TCGA-WD-A7RX	3	0
CHOL	TCGA-W5-AA2O	2	0
CHOL	TCGA-W5-AA2X	3	0
CHOL	TCGA-W5-AA2U	46	0
CHOL	TCGA-3X-AAV9	1	0
CHOL	TCGA-W5-AA2Q	1	0
CHOL	TCGA-ZH-A8Y2	4	0.6
CHOL	TCGA-ZH-A8Y8	3	0
CHOL	TCGA-YR-A95A	0	0
CHOL	TCGA-3X-AAVB	1	0.6
CHOL	TCGA-W5-AA30	4	0.6
CHOL	TCGA-W6-AA0S	0	0
CHOL	TCGA-W5-AA2Z	1	1.2
CHOL	TCGA-ZH-A8Y5	4	0
CHOL	TCGA-W5-AA2T	2	0
CHOL	TCGA-W5-AA2R	3	0.6
CHOL	TCGA-W5-AA31	0	0
CHOL	TCGA-ZH-A8Y6	0	0
CHOL	TCGA-W5-AA33	1	0.6
COAD	TCGA-DM-A1D9	1	0
COAD	TCGA-A6-5657	3	0
COAD	TCGA-CK-4950	1	0
COAD	TCGA-D5-6923	0	0
COAD	TCGA-A6-5656	0	0

COAD	TCGA-A6-6780	6	0
COAD	TCGA-QG-A5YW	0	0
COAD	TCGA-A6-5659	0	0
COAD	TCGA-QG-A5Z2	2	0
COAD	TCGA-DM-A285	2	0
COAD	TCGA-A6-6648	0	0
COAD	TCGA-DM-A1HB	2	0
COAD	TCGA-A6-3810	0	0
COAD	TCGA-DM-A1DA	1	0
COAD	TCGA-A6-3810	0	2.4
COAD	TCGA-A6-5656	0	0
COAD	TCGA-DM-A1D8	3	0
COAD	TCGA-DM-A1D0	2	0.6
COAD	TCGA-A6-6781	0	0
COAD	TCGA-A6-6781	0	0
DBCL	TCGA-GR-A4D5	2	0
DBCL	TCGA-GS-A9TX	0	0.6
DBCL	TCGA-GR-A4D9	3	0.6
DBCL	TCGA-RQ-A6JB	1	0
DBCL	TCGA-FA-A6HO	0	1.2
DBCL	TCGA-FA-8693	1	0.6
DBCL	TCGA-FF-A7CQ	1	0
DBCL	TCGA-GS-A9TQ	3	0.6
DBCL	TCGA-FA-A6HN	1	0
DBCL	TCGA-GS-A9TT	9	0
DBCL	TCGA-RQ-A68N	6	0
DBCL	TCGA-FA-A7Q1	5	0.6
DBCL	TCGA-GS-A9TZ	4	0
DBCL	TCGA-FA-A4XK	6	0
DBCL	TCGA-FA-A86F	3	0
DBCL	TCGA-FF-A7CR	2	0
DBCL	TCGA-G8-6326	15	0
DBCL	TCGA-GS-A9U3	7	0
DBCL	TCGA-GS-A9TV	0	0.6
DBCL	TCGA-GR-7353	0	0
LIHC	TCGA-BC-4072	2	0
LIHC	TCGA-DD-AACK	2	0
LIHC	TCGA-DD-A1ED	0	0.6
LIHC	TCGA-BC-A216	3	0
LIHC	TCGA-DD-A4NP	4	0
LIHC	TCGA-DD-AADV	3	0
LIHC	TCGA-DD-A113	1	0
LIHC	TCGA-DD-A4NI	0	0
LIHC	TCGA-CC-A1HT	33	35.5625
LIHC	TCGA-DD-AADW	0	0.6
LIHC	TCGA-BD-A2L6	0	0
LIHC	TCGA-BC-A216	17	0
LIHC	TCGA-DD-A11C	2	0
LIHC	TCGA-CC-A7IL	2	0
LIHC	TCGA-BC-A10T	7	0.625
LIHC	TCGA-DD-AADK	3	0
LIHC	TCGA-BC-A10X	1	0
LIHC	TCGA-UB-A7MC	2	0
LIHC	TCGA-CC-A3MB	0	0
LIHC	TCGA-DD-AACJ	2	0
LUAD	TCGA-38-4630	1	0
LUAD	TCGA-05-4422	18	0.6
LUAD	TCGA-05-4418	2	0
LUAD	TCGA-38-4628	0	4.2
LUAD	TCGA-05-4432	4	0
LUAD	TCGA-05-4395	7	0
LUAD	TCGA-55-8089	9	0
LUAD	TCGA-05-4398	4	0.6
LUAD	TCGA-44-2659	13	0
LUAD	TCGA-44-2656	6	1.2
LUAD	TCGA-05-4398	1.0001	0.6
LUAD	TCGA-05-4420	18	0.6
LUAD	TCGA-05-4382	16	12
LUAD	TCGA-38-4626	4	258.4125

LUAD	TCGA-05-4426	5	0
LUAD	TCGA-38-4625	9.833333333	1.2
LUAD	TCGA-35-3615	7	0
LUAD	TCGA-05-4430	0	1.2
LUAD	TCGA-05-4244	6	0.6
LUSC	TCGA-22-5489	1	0
LUSC	TCGA-22-4604	2	0
LUSC	TCGA-22-5474	1	0
LUSC	TCGA-22-5480	1	0
LUSC	TCGA-33-6737	0	3.75
LUSC	TCGA-22-4591	2	0.625
LUSC	TCGA-33-4587	3.833333333	0
LUSC	TCGA-22-5492	0	0
LUSC	TCGA-21-1081	1.833333333	0
LUSC	TCGA-34-5231	2	1.875
LUSC	TCGA-18-5592	0	0
LUSC	TCGA-22-5477	1	0
LUSC	TCGA-18-5595	0	0
LUSC	TCGA-21-5787	4	0.625
LUSC	TCGA-22-5483	1	0
LUSC	TCGA-22-5478	0	0.625
LUSC	TCGA-21-5786	0	0
LUSC	TCGA-22-0940	5	0.6
LUSC	TCGA-22-0944	6	1.25
LUSC	TCGA-34-5236	2	0.625
NT	TCGA-44-5645	1.001	0
NT	TCGA-44-2657	3	0
NT	TCGA-50-6595	2	0
NT	ERR030873	0.001	0.001
NT	TCGA-IH-A3EA	1.05	1.05
NT	TCGA-43-3394	2	1.25
NT	TCGA-CC-A3MB	1	1
NT	ERR030878	0	0
NT	TCGA-44-2665	3	0
NT	TCGA-AA-3517	0	0
NT	TCGA-H6-8124	1	1
NT	TCGA-G9-6363	4	0
NT	TCGA-EL-A3TB	5	0.6
NT	TCGA-KS-A41L	4	0
NT	TCGA-FU-A3EO	3	0
NT	TCGA-MY-A5BF	0.001	0.6
NT	TCGA-HM-A3JJ	0.001	0.6
NT	TCGA-W5-AA34	0	0
NT	TCGA-ZU-A8S4	0	0
NT	ERR030874	0.001	0.001
NT	TCGA-BH-A0HK	0	0.625
OV	TCGA-04-1348	32	0.9375
OV	TCGA-10-0926	2	4.4875
OV	TCGA-61-2002	6	0
OV	TCGA-61-2095	1.001	0
OV	TCGA-13-1498	1.001	0
OV	TCGA-25-2399	4	0
OV	TCGA-25-1631	50	2.3625
OV	TCGA-29-1696	9.611111111	0
OV	TCGA-24-2020	0.001	0.7875
OV	TCGA-13-1482	9	1.5875
OV	TCGA-24-1558	2	0
OV	TCGA-13-1499	0.001	0
OV	TCGA-57-1994	0.001	0
OV	TCGA-24-2038	0.001	0
OV	TCGA-59-2354	6	0
OV	TCGA-59-A5PD	1.001	0
OV	TCGA-25-2409	5	0
OV	TCGA-24-1418	3	5.75
OV	TCGA-24-1549	10	0
OV	TCGA-61-2000	8	1.475
PAAD	TCGA-IB-7893	0	0
PAAD	TCGA-HZ-7925	1	0
PAAD	TCGA-IB-A5SS	0	0.6

PAAD	TCGA-IB-8127	1	0
PAAD	TCGA-IB-7647	3	0.6
PAAD	TCGA-HZ-8638	2	0.6
PAAD	TCGA-3A-A9IB	0	0
PAAD	TCGA-F2-7276	0	0
PAAD	TCGA-HZ-7924	0	0
PAAD	TCGA-F2-7273	2	0.6
PAAD	TCGA-F2-6879	1	0
PAAD	TCGA-IB-7887	7	0.6
PAAD	TCGA-HZ-7923	5	0
PAAD	TCGA-IB-AAUV	2	0
PAAD	TCGA-HZ-8315	0	0
PAAD	TCGA-HZ-8636	1	0
PAAD	TCGA-IB-7885	0	0
PAAD	TCGA-HZ-8637	2	0
PAAD	TCGA-HZ-8519	2	0.6
PAAD	TCGA-IB-7654	1	0
PRAD	TCGA-EJ-7791	1	0
PRAD	TCGA-XK-AAJR	0	0
PRAD	TCGA-HC-7736	2	0
PRAD	TCGA-J4-A83L	2	0
PRAD	TCGA-HC-7212	0	1.2
PRAD	TCGA-EJ-5509	0	0.6
PRAD	TCGA-EJ-7781	2	0
PRAD	TCGA-G9-6366	0	0
PRAD	TCGA-G9-6498	0	0
PRAD	TCGA-CH-5753	0	0
PRAD	TCGA-J4-8198	2	0
PRAD	TCGA-G9-6384	0	0
PRAD	TCGA-EJ-5530	0	0
PRAD	TCGA-YL-A9WX	1	1.2
PRAD	TCGA-EJ-5527	0	0
PRAD	TCGA-EJ-8470	1	0
PRAD	TCGA-HI-7171	0	0
PRAD	TCGA-HC-8261	0	0
PRAD	TCGA-HC-7211	0	0
PRAD	TCGA-EJ-5511	0	0
SKCM	TCGA-BF-A1PZ	3	22.8
SKCM	TCGA-FS-A1ZA	5	0
SKCM	TCGA-FR-A726	4	107.625
SKCM	TCGA-EB-A3XB	3	0
SKCM	TCGA-BF-A1PX	27	40.15
SKCM	TCGA-BF-A1PV	1	27.5
SKCM	TCGA-GF-A2C7	0	0
SKCM	TCGA-FR-A2OS	2	1.2
SKCM	TCGA-EB-A430	3	0
SKCM	TCGA-EB-A24C	0	0.6
SKCM	TCGA-BF-AAP1	6.611111111	63.5
SKCM	TCGA-EB-A24D	1	1.2
SKCM	TCGA-ER-A2NB	0	6
SKCM	TCGA-BF-A1PU	3	3.6
SKCM	TCGA-EB-A431	4	10.2
SKCM	TCGA-GN-A26C	1	0
SKCM	TCGA-BF-A3DN	2	0
SKCM	TCGA-EB-A6R0	5	54.725
SKCM	TCGA-ER-A2NF	2743.777778	1351.3875
SKCM	TCGA-GF-A769	2	0
SKCM	TCGA-EB-A4IS	1	0.6
SKCM	TCGA-IH-A3EA	0	0
SKCM	TCGA-HR-A5NC	0	0
TGCT	TCGA-W4-A7U4	98.11111111	24.6
TGCT	TCGA-2G-AAEW	9.166666667	31.2
TGCT	TCGA-2G-AAL5	1.001	0.6
TGCT	TCGA-2G-AAGY	4	1.2
TGCT	TCGA-XY-A89B	46	0.6
TGCT	TCGA-XE-AANJ	46	0.6
TGCT	TCGA-XE-AAO3	154	26.925
TGCT	TCGA-SN-A84W	2	1.8
TGCT	TCGA-2G-AAHG	27	40.15

TGCT	TCGA-SN-A6IS	15	18.6
TGCT	TCGA-2G-AAF6	32.5	4.1375
TGCT	TCGA-S6-A8JY	238.444444	11.4
TGCT	TCGA-XE-A8H5	133.5	158.175
TGCT	TCGA-XE-A8H1	53.4444444	39.5875
TGCT	TCGA-W4-A7U2	161.833333	35.9375
TGCT	TCGA-YU-A90P	39	36.5375
TGCT	TCGA-SB-A6J6	160.722222	18
TGCT	TCGA-2X-A9D6	23.1677777	47.4
TGCT	TCGA-2G-AAGA	53.8333333	18
TGCT	TCGA-XY-A9T9	27.6777777	57
THCA	TCGA-ET-A25I	2	0.6
THCA	TCGA-EM-A3O8	2	0
THCA	TCGA-BJ-A0ZB	2	0.6
THCA	TCGA-ET-A25G	2	1.8
THCA	TCGA-EM-A2CM	3	0
THCA	TCGA-H2-A421	0	1.8
THCA	TCGA-DJ-A2PX	1	0
THCA	TCGA-EL-A3N2	0	0
THCA	TCGA-BJ-A28V	1	0
THCA	TCGA-EL-A3MZ	2	0
THCA	TCGA-EM-A4FR	3	0.6
THCA	TCGA-EM-A4FQ	3	1.2
THCA	TCGA-DJ-A3UV	0	0
THCA	TCGA-EL-A3T1	4	0.6
THCA	TCGA-E8-A414	3	3
THCA	TCGA-DJ-A1QH	2	0.6
THCA	TCGA-J8-A3YG	0	0
THCA	TCGA-DJ-A2Q4	5	0.6
THCA	TCGA-EM-A3O9	0	3.6
THCA	TCGA-EL-A3T6	0	4.2
UVM	TCGA-V4-A9EE	1.05	158.2625
UVM	TCGA-VD-A8K9	2	7.2
UVM	TCGA-VD-A8KG	51.77777778	1365.025
UVM	TCGA-VD-A8KA	2	99.5
UVM	TCGA-V4-A9EI	19	3.6
UVM	TCGA-VD-A8KA	3	3.6
UVM	TCGA-WC-AA9A	12	233.9135
UVM	TCGA-VD-A8KM	1	66.55
UVM	TCGA-V4-A9EX	1	78.25
UVM	TCGA-VD-AA8R	0	0.001
UVM	TCGA-VD-AA8T	2	1.2
UVM	TCGA-VD-A8K8	0	5.3375
UVM	TCGA-V4-A9EJ	9.666666667	1988.6625
UVM	TCGA-V4-A9EW	22	36.5375
UVM	TCGA-V4-A9F0	0	1.8
UVM	TCGA-WC-A87T	4	1.15
UVM	TCGA-V4-A9F1	0	3
UVM	TCGA-WC-AA9E	18	11.4
UVM	TCGA-V4-A9F7	0	8.4
UVM	TCGA-VD-A8K7	4	1.2
oocyte	SRR445718	0	0
oocyte	SRR445719	0	0
oocyte	SRR445720	0	0
zygote	SRR445721	0	0
zygote	SRR445722	0	0
zygote	SRR445723	0	0
2-cell	SRR445724	0	0
2-cell	SRR445726	0	0

2-cell	SRR445727	0	0
2-cell	SRR445728	0	0
4-cell	SRR490961	0	0
4-cell	SRR490962	0	0
4-cell	SRR490963	0	0
4-cell	SRR490964	0	0
4-cell	SRR490965	0	0
4-cell	SRR490966	0	0
4-cell	SRR490967	0	0
4-cell	SRR490968	0	0
4-cell	SRR490969	0	0
4-cell	SRR490970	0	0
4-cell	SRR490971	0	0
4-cell	SRR490972	0	0
8-cell	SRR490973	1	3.675
8-cell	SRR490974	3	0
8-cell	SRR490975	1	0.7
8-cell	SRR490976	3	6.15
8-cell	SRR490977	0	0
8-cell	SRR490978	1	2.6375
8-cell	SRR490979	1	2.5125
8-cell	SRR490980	0	5.675
8-cell	SRR490981	0	10.4625
8-cell	SRR490982	0	0.8
8-cell	SRR490983	1	0
8-cell	SRR490984	0	0
8-cell	SRR490985	0	8.0125
8-cell	SRR490986	18	17.2625
8-cell	SRR490987	0	6.0125
8-cell	SRR490988	6	0
8-cell	SRR490989	3	4.5625
morula	SRR490990	0	0
morula	SRR490991	11	0
morula	SRR490992	12	0
morula	SRR491039	0	0
morula	SRR490993	0	0
morula	SRR490994	1	0
morula	SRR490995	0	0
morula	SRR490996	0	1.6375
morula	SRR490997	0	0
morula	SRR491001	0	0
late_blastocyst	SRR491002	0	1.6
late_blastocyst	SRR491003	0	0
late_blastocyst	SRR491004	0	0
late_blastocyst	SRR491005	0	0
late_blastocyst	SRR491009	1	19.5375
late_blastocyst	SRR491010	0	0
late_blastocyst	SRR491011	0	0
late_blastocyst	SRR491012	0	0
late_blastocyst	SRR491013	0	0
late_blastocyst	SRR491021	0	0
hESC	SRR491022	0	9.9
hESC	SRR491023	0	0
hESC	SRR491024	0	0
hESC	SRR491025	0	0
hESC	SRR445730	0	0
hESC	SRR445731	9	6.5375
hESC	SRR491039	0	0
hESC	SRR491040	21	0
hESC	SRR491041	3	0
hESC	SRR491042	13	0
hESC	SRR491043	6	0
hESC	SRR491045	5	0

Chapter 3, "HERV-K HML-2 transcription in diverse cancers is related with cancer stem cell and epithelial-mesenchymal transition markers."

This work has been submitted to bioRxiv in 2018. **doi:** https://doi.org/10.1101/451997

Statement of Contribution

Research design

I developed the overall project with GM.

Data

I curated and categorised all the CSC/EMT genes to be analysed from the CSC/EMT literature and obtained the TCGA data from all the CSC/EMT genes from www.cbioportal.org. The controlled-access Cancer Genome Atlas Project data was used according to the Data Use Certification as stipulated by the TCGA Data Access Committee and dbGaP authorisation (https://dbgap.ncbi.nlm.nih.gov/aa/). I acquired permission to use controlled-access dbGAP data under project #7621.

Method development and analysis

I developed the formula to calculate the transcriptional distance with GM. I performed all of the data analyses and the statistical analyses under the supervision of GM.

Manuscript

I wrote the manuscript with additional input from CGS, FFN, TK, AK, and GM. I generated all the tables and figures. I curated and edited the information presented in the Supplement.

Author list and affiliations

Authors

Audrey T. Lin[1], Cindy G. Santander[1], Fabricia F. Nascimento[1,2], Emanuele Marchi[3], Timokratis Karamitros[1,4], Aris Katzourakis[1], Gkikas Magiorkinis[5]**

* Corresponding Authors: audrey.lin@gtc.ox.ac.uk and gmagi@med.uoa.gr

Affiliations

1. Department of Zoology, University of Oxford, Oxford, UK
2. Department of Infectious Disease Epidemiology, Imperial College London, London, UK
3. Nuffield Department of Medicine, University of Oxford, Oxford, UK
4. Public Health Laboratories, Department of Microbiology, Hellenic Pasteur Institute, Athens, Greece
5. Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Athens, Greece

Abstract

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections that make up 8% of the human genome. Although these elements are mostly fragmented and inactive, many proviruses belonging to the HERV-K (HML-2) family, the youngest lineage in the human genome, have intact open reading frames, some encoding for accessory genes called *np9* and *rec* that interact with oncogenic pathways. Many studies have established that ERVs are transiently expressed in both stem cells and cancer, resulting in aberrant self-renewal and uncontrolled proliferation. *np9* and *rec* expression are significantly correlated with a range of cancer stem cell (CSC) and epithelial to mesenchymal transition (EMT) biomarkers, including cellular receptors, transcription factors, and histone modifiers. Surprisingly, these ERV genes are negatively correlated with genes known to promote pluripotency in embryonic stem cell lines, such as Oct4. These results suggest that HERV-K (HML-2) is part of the transcriptional landscape responsible for cancer cells undergoing the phenotypic switch that characterises EMT. The discovery of *np9* and *rec*'s correlation with CSC and EMT biomarkers suggest a yet undescribed role affecting the transitional CSC-like state in EMT and the shift towards cancer malignancy.

1. Introduction

Human endogenous retroviruses (HERVs) are remnants of ancient retroviral infections that propagated in the germlines of our distant mammalian ancestors and are present in multiple copies and lineages in their host genomes [82]. HERVs stopped proliferating in our ancestors' genomes millions of years ago, with one exception, HERV-K (HML-2) (referred here as HK2), which continued to expand after the human-chimpanzee split [83]. The majority of HERVs are defective, due to accumulation of random mutations, selection pressure from host-derived restriction factors, or recombination, resulting in frameshifts, premature stop codons, and large deletions that remove the coding capacity of the provirus [82]. Many HERVs are also tightly regulated by epigenetic modification [84]. HK2 integrations, on the other hand, have frequently more intact open reading frames compared to older HERVs. Some HK2 proviruses have the capacity to produce the accessory proteins Np9 and Rec depending on the presence (type I proviruses encoding for *np9*) or absence (type 2 proviruses encoding for *rec*) of a deletion in the *env* gene [53][54]. These proteins are known to be highly expressed in a number of cancers, germ cell tumour cell lines, and human embryonic stem cells [12]. HK2-encoded proteins have been shown to have potential oncogenic function: Np9 and Rec have been shown to interact with cellular pathways linked to cancer [19][26][27], including the activation of the oncogene *myc* [26], and the Notch signalling pathway [26]. Rec protein functions in regulating viral gene expression by transporting viral mRNA from the nucleus into the cytosol [55][12], and its expression has been found to interfere with the development of germ cells and also cause carcinoma in mice [55]. In addition, Np9 protein is reported to bind and regulate ubiquitin ligase MDM2, resulting in the repression of p53 tumour suppressor [28]. Transcripts of *np9* and *rec* are found in human embryonic stem cells and human induced pluripotent stem cell lines and are associated with

maintenance of pluripotency [56][57], and recently it has been demonstrated that the activation and silencing of a specific HERV-K LTR LTR5HS is associated with reciprocal up- and downregulation of hundreds of human genes, and can occur over long genomic distances [85].

Within a tumour, there is a small subset of biologically distinct and rare "tumour-initiating" cells with attributes that parallel the functional properties of normal stem cells, referred to as cancer stem cells (CSCs). CSCs have been proposed to be the source of all malignant cells in a primary tumour, and their presence also gives rise to metastatic growth. It has also been suggested that the failure to eliminate CSCs results in relapse and a drug-resistant phenotype of the tumour regrowth, following chemotherapy-induced remission [86]. Because of the tumour-initiating ability of CSCs and the association of stem cell behaviours by normal and neoplastic cells during the activation of epithelial to mesenchymal transition (EMT) [86], there has been a significant effort to develop novel treatments targeting CSC and EMT-specific signalling pathways and cell surface markers.

Since HK2 is transcribed in many different cancers, during specific stages of early development, and is involved in the maintenance of pluripotency in embryonic stem cells [74][12], HK2 overexpression in seemingly diverse cancer types might reflect the stem cell-like identity of the tumour. Whether the upregulation of HK2 contributes to tumorigenesis and cancer progression or is simply an unrelated side-effect of increased cellular dysfunction is not very well understood. To better address this problem, we characterise in this study a list of transcriptional interactions between HK2 and specific components of stem cell/EMT-associated pathways.

2. Materials and Methods

2.1 Data Analysed

The method described in [87] was used in capturing *np9* and *rec*-specific transcripts in RNA-seq data. Raw FASTQ files of Illumina HiSeq RNA-seq reads, originally sequenced from 305 samples derived from primary solid tumour tissues and blood-derived normal tissues from different individuals, were downloaded from Cancer Genomics Hub (data now available at Genomics Data Commons, https://gdc.cancer.gov). A total of 15 different cancer types were analysed. Cancer types reported to be associated with HK2 expression were selected for study: breast cancer, lymphoma, lung cancer (lung adenocarcinoma and lung squamous cell carcinoma), ovarian cancer, prostate cancer, skin melanoma, liver hepatocellular carcinoma, and testicular germ cell cancer. The other 6 cancer types were selected by tumour type: melanoma (uveal melanoma), adenocarcinomas (colon, pancreatic, and thyroid adenocarcinomas), carcinomas (cervical carcinoma, cholangiocarcinoma) **(Table 3.1)**. Normal matched controls are not available for the majority of the tumour samples in TCGA, so a selection of random normal non-tumour tissues (cervix, bile duct, colorectum, liver, lung, ovary, pancreas, prostate, testis, and thymus) from TCGA and Illumina Human Body Map 2.0 was used. The non-diseased samples used for the comparison group (referred in the text as "NT") were taken from.

Table 3.1. Summary of The Cancer Genome Atlas (TCGA) RNA-seq data

Breast invasive carcinoma (BRCA)
Cervical squamous cell carcinoma and Endocervical adenocarcinoma (CESC)
Cholangiocarcinoma (CHOL)
Colorectal adenocarcinoma (COAD)
Lymphoid neoplasm diffuse large B-cell lymphoma (LBCL)
Liver hepatocellular carcinoma (LIHC)
Lung adenocarcinoma (LUAD)
Lung squamous cell carcinoma (LUSC)
Ovarian serous cystadenocarcinoma (OV)
Pancreatic adenocarcinoma (PAAD)
Prostate adenocarcinoma (PRAD)
Skin cutaneous melanoma (SKCM)
Testicular germ cell cancer (TGCT)
Thyroid carcinoma (THCA)
Uveal melanoma (UVM)

Note: To analyse mRNA expression of EMT and CSC genes, the data was obtained from TCGA, provisional by using www.cbioportal.org.

2.2 Testicular germ cell cancer as cancer stem cell reference

Testicular germ cell cancer is used as a CSC reference in this study because of the distinctly similar molecular and cellular phenotypes observed in germ cell tumours and human embryonic stem cells. These shared traits include lack of contact inhibition in culture, a rapid proliferation rate, high activity of telomerase, and similarities in gene expression and epigenetic patterns, like the upregulation of stem cell markers *POU5F1* (Oct4), *NANOG*, and *SOX2*, and oncogenes *MYC* and *KLF4*, and the downregulation of the tumour suppressor *TP53* (reviewed in [55]).

2.3 Analysis of EMT and CSC genes from TCGA datasets

For expression analysis of EMT and CSC genes, the data was obtained from TCGA by using www.cbioportal.org. EMT and CSC genes analysed were classified depending on either being reportedly overexpressed (or upregulated) and transcriptionally repressed (downregulated) in stem cells and EMT **(Table 3.2)**. The gene groups are classified as follows: EMT_up genes (77 genes); EMT_down genes (13 genes); CSC_up genes (47 genes); CSC_down genes (8 genes).

Table 3.2. Lists of gene groups analysed

Group	Genes
EMT – upregulated genes (EMT_up)	AKT1, ALDH1A1, AXIN2, BMI1, BMP4, CD44, CSN2, CTNNB1, EHMT2, EZH2, FGF2, FOXC2, GLI1, GRHL2, GSC, GSK3B, HDAC1, HDAC2, HDAC3, HIF1A, IL6, JAK1, JAK2, KAT5, KDM1A, KDM5B, KLF4, KLF8, LATS2, LBX1, LIN28A, LOXL2, MAPK1, MYC, NANOG, NODAL, NOG, NOTCH1, NOTCH2, NOTCH3, NOTCH4, OVOL1, OVOL2, PAK1, PDGFRB, POU5F1, PRDM14, PRRX1, QKI, SALL4, SETDB1, SHH, SIN3A, SIRT1, SIX1, SMAD2, SMAD3, SMAD4, SMO, SNAI1, SNAI2, SOX2, SRSF2, STAT3, SUV39H1, SUZ12, T, TCF3, TCF4, TGFB1, TGFB2, TWIST1, WNT2, YBX1, ZEB1, ZEB2, ZFP42

CSC – upregulated genes (CSC_up)	ALDH1A1, APC, AXIN2, CD34, CD44, CTNNB1, DLL4, EP300, FGF2, KDM1A, KDM5B, KLF2, KLF4, KLF8, LATS2, LIN28A, MAPK1, MYC, NANOG, NODAL, NOG, NOTCH1, NOTCH2, NOTCH3, NOTCH4, NRB01, POU5F1, PROM1, SALL4, SHH, SIN3A, SMAD2, SMAD3. SMAD4, SMO, SOX18, SOX2, STAT3, SUV39H1, TCF3, TCF4. TGFB1, TGFB2, THY1, WNT2, YBX1, ZFP42
EMT – downregulated genes (EMT_down)	APC, CD24, CD38, CDH1, CTNND1, ENAH, ESRP1, ESRP2, FBXL14, HSP90AA1, MDM2, PRKD1, TP53
CSC – downregulated genes (CSC_down)	BMP4, CD24, CD38, CTNND1, GSK3B, IL6, SETDB1, SUZ12

Note: EMT – epithelial to mesenchymal transition; CSC – cancer stem cells. Upregulated genes – genes in this group are reported by the literature to be upregulated in the context of EMT/CSC. Downregulated genes – genes in this group are reported by the literature to be downregulated in the contexts of EMT/CSC.

2.4 Calculation of transcriptional distance of HK2 genes from transcriptional profile of all EMT/CSC genes

Testicular germ cell cancer (TGCT) is used as a CSC reference because of distinctly similar molecular and cellular phenotypes observed in germ cell tumours and human embryonic stem cells [88]. For the TCGA-TGCT RNA-seq dataset, the mean expression value was computed for each gene, and the similarity of the expression profile of each tumour sample to the expression profile of the CSC reference was calculated using Formulas 1 and 2. In Formula 1, within each tumour sample per cancer, φ is the geometric distance from the cancer stem cell (CSC) phenotype (either EMT or stem cell

genes) in EMT/CSC_up genes in a given tumour sample, μg_{tc} is the average expression of the gene in all the TCGA-TGCT samples, and g_z is the expression of the gene in a given tumour sample. Therefore, the more negative φ is, the more similar the expression profile of the tumour sample is to the CSC reference. In either EMT_down or CSC_down genes, Formula 2 is used, where ψ is the geometric distance from the CSC reference (either EMT or stem cell genes), and the more positive ψ is, the more similar the expression profile of the given tumour sample to the CSC reference.

1. $$\varphi = \frac{\mu g_{tc} - g_z}{\mu g_{tc}}$$

2. $$\psi = \frac{g_z - \mu g_{tc}}{\mu g_{tc}}$$

Because some of the transcript counts were zero or negligible, transformation of these values were necessary in order to conduct further statistical analyses to find the strength of association between HK2 expression in each tumour sample to the transcriptional profiles of the "pooled" EMT and CSC genes (represented by φ and ψ). These values were transformed using formulas 3 and 4, respectively. The integers 10,000+1 used in these formulas is to account for the non-negative values that would otherwise result with the log transformation. The smaller the calculated value of the transcriptional distance, the stronger the correlation between the EMT/CSC transcriptional profile and the expression of the HK2 genes. In other words, values that are closer to 1 indicates more similarity to the "pooled" EMT/CSC transcriptional profile.

3. $$\varphi = log_{10}[(1 - \varphi) * 10,000 + 1]$$

4. $$\psi = log_{10}[(25 - \psi) * 10,000 + 1]$$

2.5 Statistical Analysis

The Wilcoxon signed rank-sum test was used to compare the relative expression of the HK2 genes from each of the unmatched samples within each cancer type compared to NT. Multiple linear regression with interaction terms was used to estimate the association between the expression of the HK2 targets and up (φ) and down (ψ) EMT and CSC genes in the different cancers. The chi-square test of independence was performed to determine the statistical significance of the correlations between the expression of *np9* or *rec* and activity of functional categories of CSC/EMT genes. The gene categories consist of cell surface receptors (10 genes), transcription factors (29 genes), and genes categorised as "Other": negative regulators (3 genes), histone modifiers (11 genes), kinases (5 genes), metabolic enzyme (1 gene), ubiquitin ligase (1 gene), splicing factors (2 genes), receptor ligands (4 genes), and cell adhesion molecules (3 genes). Multiple linear regression with interaction terms was used to test the correlation of HK2 genes with the individual EMT/CSC upregulated and downregulated genes. For all statistical methods, associations with p-value < 0.05 were considered statistically significant. All statistical analyses were performed in R version 3.3.1 [89].

3. Results

3.1 HK2 expression is high in cells with CSC/EMT-like transcriptional profiles

HK2 is generally overexpressed across different cancer types [87], thus we wanted to determine whether there is a correlation between *np9* and *rec* activity and the expression of genes associated with CSCs and EMT. We could not clearly determine whether high transcription of *np9* and *rec* is only observed in cancers with highly similar EMT/CSC transcriptional profiles (positive correlation between expression of the HK2 gene and the EMT/CSC transcriptional profile would be represented by x-axis values closer to 1) (**Fig. 3.1**). These results suggest that the CSC-like transcriptional profile is not sufficient by itself as an explanation for high HK2 transcription. The specific genes that could be independently associated with the observed transcriptional pattern between CSC/EMT genes and HK2 were next investigated.

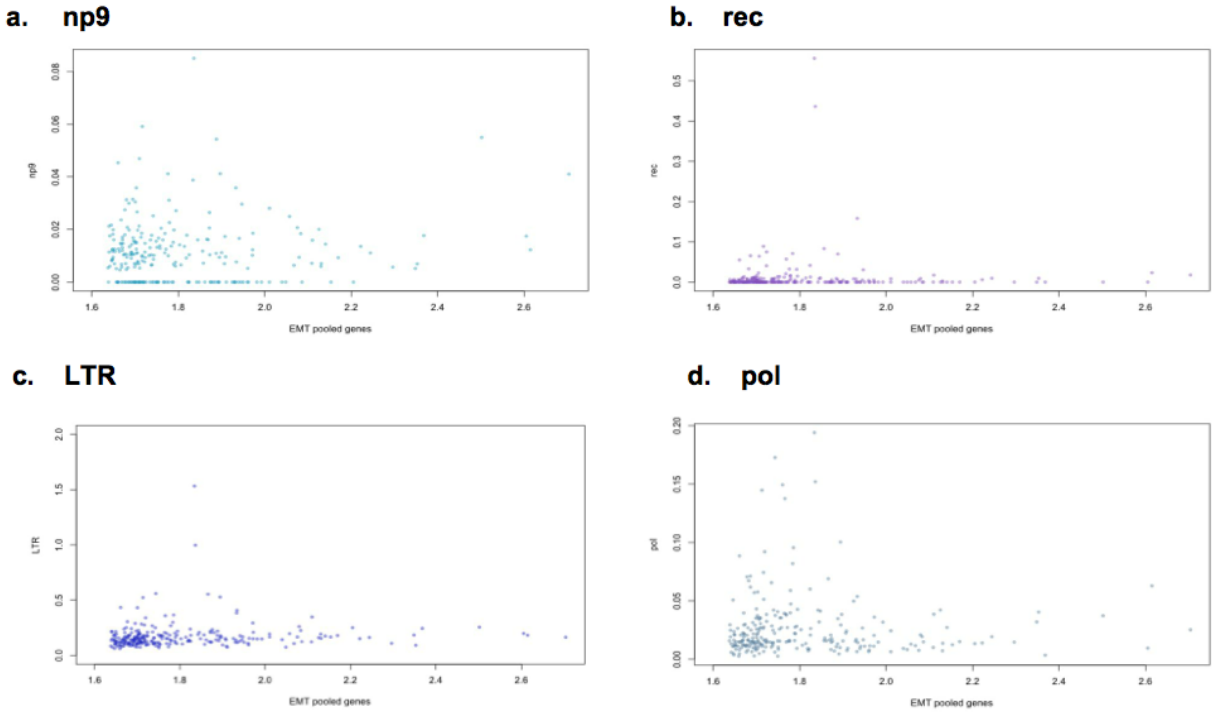


Figure 3.1. HK2 upregulation in cancers in comparison to epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) gene expression signature

x-axis indicates the pooled gene expression signature of all the EMT/CSC genes within all the tumour samples. A value closer to 1 on the x-axis indicates more similarity to the EMT/CSC phenotype. y-axis indicates the relative expression of **(a.)** *np9*; **(b.)** *rec*; **(c.)** HERV-K113 LTR; **(d.)** *pol* in all tumour samples. These values have been transformed by square root.

3.2 Independent correlates of HK2 expression and CSC/EMT gene expression

The overall transcriptional landscape was described for the CSCs and the CSC-like transitory state in EMT. The results indicate that HK2 transcription is significantly correlated with a number of specific CSC/EMT genes: there are 22 significant positive and negative correlations between the expression of *np9* and EMT/CSC genes, and 29 significant correlations between the expression of *rec* and the EMT/CSC genes (**Table 3.4**).

Table 3.3. Chi-square contingency tables showing significant correlations between different functional gene categories and transcription of *rec* and *np9*

(a) *rec* is significantly correlated with the expression of cell surface receptor and transcription factor genes

Contingency table of significant correlation between *rec* and gene categories			
Gene Category	**Positive**	**Negative**	**Not significant**
Cell surface receptor	4 (1.14) [7.14]	1 (2.71) [1.08]	5 (6.14) [0.21]
Transcription factor	1 (3.31) [1.62]	8 (7.87) [0.0]	20 (17.81) [0.27]
Other	3 (3.54) [0.08]	10 (8.41) [0.30]	18 (19.04) [0.06]
Column Total	8	19	43
$X^2 = 10.76$** **p = 0.029			

Note: The gene categories "cell surface receptor" contains 10 genes and "transcription factor" contains 29 genes. The gene category "Other" consists of the following gene categories: negative regulator (3 genes), histone modifier (11 genes), kinase (5 genes), metabolic enzyme (1 gene), ubiquitin ligase (1 gene), splicing factor (2 genes), receptor ligand (4 genes), cell adhesion (3 genes). Each cell contains the following information: the observed cell totals, the expected cell totals in parentheses ( ), and the chi-square (X^2) statistic in brackets [ ].

(b) *np9* is independently correlated with the expression of cell surface receptor and transcription factor genes

Contingency table of significant correlation between *np9* and gene categories			
Gene Category	**Positive**	**Negative**	**Not significant**

Cell surface receptor	1 (1.03) [0.0]	4 (1.41) [4.73]	4 (6.56) [1.0]
Transcription factor	2 (3.43) [0.60]	2 (4.71) [1.56]	26 (21.86) [0.79]
Other	5 (3.54) [0.60]	5 (4.87) [0.0]	21 (22.59) [0.11]
Column Totals	8	11	51
$X^2 = 9.38$** **p = 0.052			

Note: Chi-square contingency tables showing significant correlations between different gene categories and transcription of *np9*. The gene categories "cell surface receptor" contains 10 genes and "transcription factor" contains 30 genes. The gene category "Other" consists of the following gene categories: negative regulator (3 genes), histone modifier (11 genes), kinase (5 genes), metabolic enzyme (1 gene), ubiquitin ligase (1 gene), splicing factor (2 genes), receptor ligand (4 genes), cell adhesion (3 genes). Each cell contains the following information: the observed cell totals, the expected cell totals in parentheses ( ), and the chi-square (X^2) statistic in brackets [ ]

---

np9 transcription is significantly correlated with the transcription of 5 transcription factor (NODAL, OVOL2, TBXT (Brachyury), CTNNB1 (ß-catenin), and TCF3), 6 cell surface receptors (CD38, CD133, NOTCH1, NOTCH2, CD44, NOTCH4), 3 receptor ligands (DLL4, BMP4, WNT2), 2 histone modifiers (EHMT2 (G9a), SUV39H1), 2 kinases (EP300, LATS2), 1 metabolic enzyme (ALDH1A1), and 1 ubiquitin ligase (FBXL14) (**Table 3.4a, Table 3.4c, Fig. 3.2, Fig. 3.5**). *rec* transcription is significantly correlated with 10 transcription factors (GRHL2, GSC (Goosecoid), MYC, POU5F1 (Oct4), SMAD2, STAT3, TCF4, TWIST1, CTNNB1 (ß-catenin), and TCF3), 5 cell surface receptors (CD24, PDGFRB, CD44, NOTCH4, and SMO), 3 receptor ligands (FGF2, BMP4, and WNT2), 2 histone modifiers (SUZ12 and SUV39H1), 2 kinases (PRKD1 and LATS2), 2 negative regulators (APC and CSN2), 2 splicing factors (ESPR1/2), and 1 cell adhesion molecule

(CTNND1) (**Table 3.4b, Table 3.4d, Fig. 3.2, Fig. 3.5**). HK2 transcription is most prominently correlated with the expression of genes encoding for transcription factors and cell surface receptors (**Fig. 3.2**). However, only *rec* transcription is independently and significantly correlated with expression of gene types that fall into the functional categories of cell surface receptors, transcription factors, and "Other" gene types (**Table 3.3a**), while *np9* is not independently correlated with any of these functional categories (**Table 3.3b**).

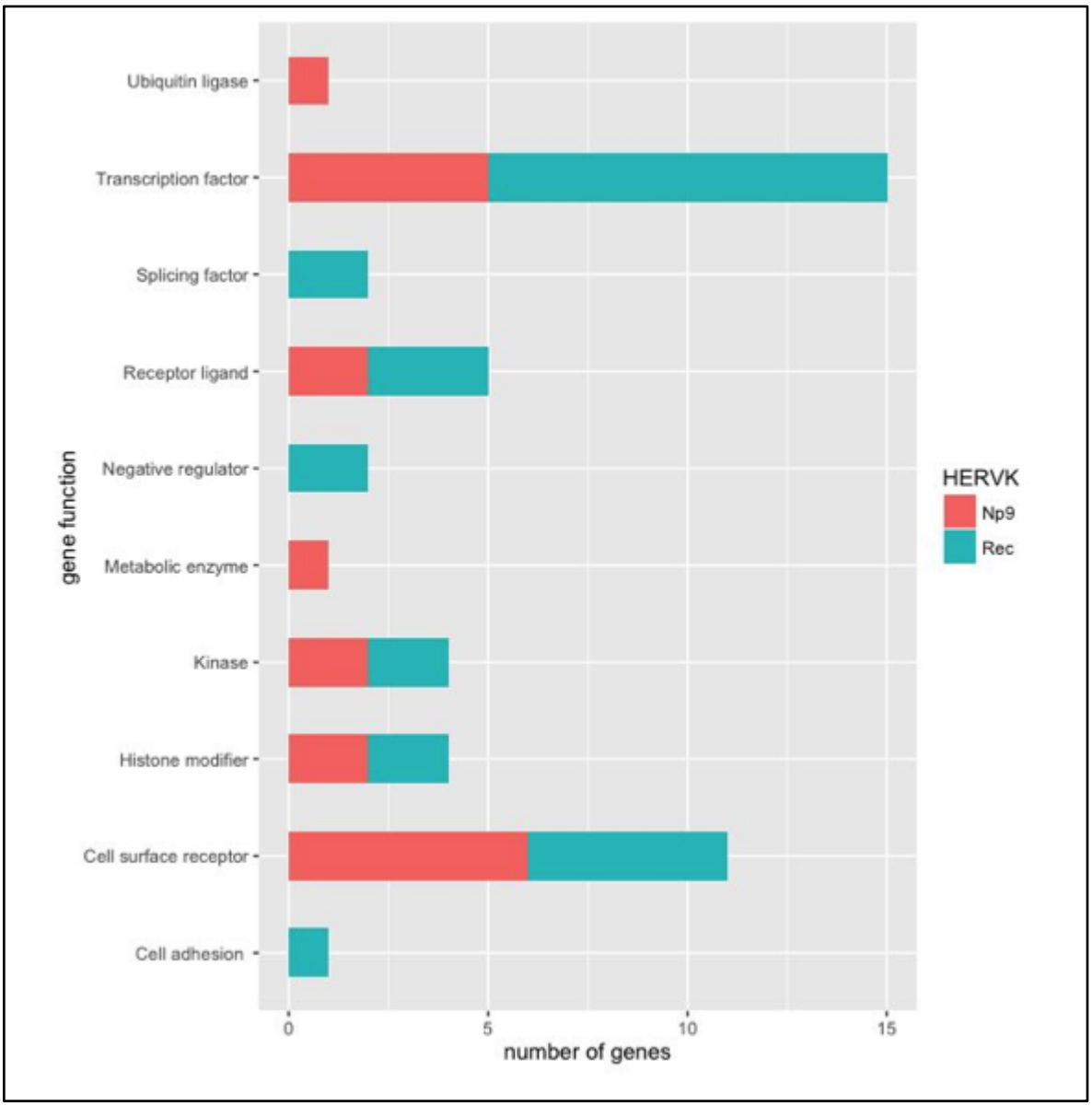


Figure 3.2. Breakdown of different gene categories that *np9* and *rec* are significantly correlated with

x-axis indicates the number of genes *np9* and *rec* are significantly correlated with; y-axis indicates the generalised function of the host genes that are significantly correlated with *np9* and *rec* expression.

The significant correlation between *rec* activity with overexpression of the cell surface molecules CD44 (EMT group: $p = 1.18 \cdot 10^{-7}$; CSC group: $p = 4.50 \cdot 10^{-5}$) and CD24 underexpression (CSC group: $p = 6.90 \cdot 10^{-3}$) indicates an association with the $CD44^{high}/CD24^{low}$ phenotype, known to be enriched in breast cancer CSCs [90][91]. Strikingly, there is conflicting expression states between *np9* and *rec* and CD44 expression, where *rec* is positively correlated (CSC group: $p = 4.50 \cdot 10^{-5}$), and *np9* is negatively correlated (CSC group: $p = 2.49 \cdot 10^{-2}$). Summaries of all the significant correlations between upregulated and downregulated genes within the CSC/EMT groups with *np9* and *rec* expression are in **Table 3.4**.

Table 3.4. *np9* and *rec* correlates with CSC/EMT genes

(a) *np9* is significantly correlated with specific CSC/EMT upregulated genes

Gene function category	Gene	Group	Coefficient	p-value
Transcription factor	NODAL	EMT	0.13	$2.25 \cdot 10^{-3}$
	OVOL2	EMT	-0.05	$1.74 \cdot 10^{-3}$

	T (Brachyury)	EMT	0.07	$9.23 \cdot 10^{-3}$
	CTNNB1 (ß-catenin)	CSC	0.26	$1.35 \cdot 10^{-2}$
	TCF3 (E47)	EMT	-0.52	$6.89 \cdot 10^{-7}$
	TCF3 (E47)	CSC	0.26	$1.35 \cdot 10^{-2}$
Cell surface receptor	NOTCH4	CSC	0.42	$1.50 \cdot 10^{-4}$
	CD133 (PROM1)	CSC	0.06	$5.00 \cdot 10^{-3}$
	NOTCH1	CSC	-0.2	$2.85 \cdot 10^{-2}$
	NOTCH2	CSC	-0.24	$2.00 \cdot 10^{-2}$
	CD44	CSC	-0.11	$2.49 \cdot 10^{-2}$
Receptor ligand	DLL4	CSC	-0.31	$4.20 \cdot 10^{-4}$
	WNT2	CSC	-0.08	$3.67 \cdot 10^{-4}$
Histone modifier	EHMT2 (G9a)	EMT	0.4	$9.47 \cdot 10^{-4}$
	SUV39H1	CSC	0.47	$5.81 \cdot 10^{-5}$

Kinase	EP300 (p300)	CSC	-0.3	$3.52 \cdot 10^{-2}$
	LATS2	CSC	0.21	$2.83 \cdot 10^{-2}$
Metabolic enzyme	ALDH1A1	EMT	-0.08	$2.54 \cdot 10^{-2}$

(b) *rec* is significantly correlated with specific CSC/EMT upregulated genes

Gene function category	Gene	Group	Coefficient	p-value
Transcription factor	GRHL2	EMT	-0.04	$2.30 \cdot 10^{-2}$
	GSC (Goosecoid)	EMT	-0.08	$1.61 \cdot 10^{-4}$
	MYC	CSC	0.15	$2.49 \cdot 10^{-2}$
	POU5F1 (Oct4)	CSC	-0.12	$1.46 \cdot 10^{-2}$
	POU5F1 (Oct4)	EMT	-0.12	$1.69 \cdot 10^{-2}$
	SMAD2	EMT	-0.37	$1.07 \cdot 10^{-2}$
	STAT3	EMT	-0.35	$2.40 \cdot 10^{-2}$
	TCF4 (E2-2)	CSC	-0.2	$9.65 \cdot 10^{-3}$

	TWIST1	EMT	-0.12	$5.98 \cdot 10^{-3}$
	CTNNB1 (ß-catenin)	EMT	0.66	$2.53 \cdot 10^{-5}$
	TCF3	EMT	-0.31	$1.17 \cdot 10^{-2}$
Cell surface receptor	PDGFRB	EMT	0.2	$1.36 \cdot 10^{-2}$
	CD44	EMT	0.28	$1.18 \cdot 10^{-7}$
	CD44	CSC	0.21	$4.50 \cdot 10^{-5}$
	NOTCH4	CSC	0.27	$3.08 \cdot 10^{-2}$
	SMO	EMT	0.15	$1.18 \cdot 10^{-2}$
Receptor ligand	WNT2	CSC	-0.07	$5.09 \cdot 10^{-3}$
	FGF2	CSC	-0.11	$3.46 \cdot 10^{-2}$
	BMP4	EMT	-0.1	$4.32 \cdot 10^{-2}$
Histone modifier	SUV39H1	EMT	-0.41	$5.85 \cdot 10^{-4}$
Kinase	LATS2	CSC	0.27	$9.97 \cdot 10^{-3}$

Negative regulator	CSN2	EMT	0.03	$1.93 \cdot 10^{-2}$

(c) *np9* is significantly correlated with specific CSC/EMT downregulated genes

Gene function category	Gene	Group	Coefficient	p-value
Cell surface receptor	CD38	CSC	12.8	$7.97 \cdot 10^{-4}$
	CD38	EMT	5.14	$2.51 \cdot 10^{-3}$
Receptor ligand	BMP4	CSC	2	$4.96 \cdot 10^{-2}$
Ubiquitin ligase	FBXL14	EMT	11.3	$1.45 \cdot 10^{-2}$

(d) *rec* is significantly correlated with specific CSC/EMT downregulated genes

Gene function category	Gene	Group	Coefficient	p-value
Cell surface receptor	CD24	CSC	2.4	$6.90 \cdot 10^{-3}$
Histone modifier	SUZ12	CSC	43.2	$1.92 \cdot 10^{-4}$
Kinase	PRKD1	EMT	-0.33	$4.40 \cdot 10^{-2}$
Negative regulator	APC	EMT	7.37	$4.21 \cdot 10^{-2}$

Splicing factor	ESRP2	EMT	4.2	$5.82 \cdot 10^{-6}$
	ESRP1	EMT	14.7	$3.01 \cdot 10^{-3}$
Cell adhesion	CTNND1	EMT	-8.67	$6.26 \cdot 10^{-4}$

Summary of significant correlations between upregulated genes within the CSC or EMT groups with **(a.)** *np9* and **(b.)** *rec*. Summary of significant correlation between downregulated genes within the CSC or EMT groups with **(c.)** *np9* and **(d.)** *rec*. Parentheses ( ) indicates other names for the same gene.

4. Discussion

Previous studies have also shown that HK2 expression is high in specific stages of early embryonic development [12][56][57][87]. HK2 is upregulated in tumours with greater transcriptional similarity with testicular germ cell cancer, a type of malignancy known for a strong stem-cell identity. This suggests that overexpression of HK2 can be found in diverse cancers that have a stem cell-like transcriptome. Together these observations indicate that HK2 transcription is a strong marker of CSC identity within a tumour. This is the first study, to our knowledge, that associates *np9* and *rec* activity with the expression of many transcriptional regulatory genes and signalling pathway genes implicated in the maintenance of cellular self-renewal and cancer malignancy through the cell state changes of EMT.

The transient expression of HK2 may be part of the transcriptional repertoire that shapes stem cell behaviour in CSCs. For example, *rec* is associated with the $CD44^{high}CD24^{low}$ antigenic profile in the cancer types surveyed. This phenotype is linked with increased malignancy and the presence of CSCs in

breast tissues [90][91][92]. $CD44^{high}CD24^{low}$ cells also share many features with normal stem cells and are shown to be highly tumourigenic when injected into immunocompromised mice [92].

In addition, *rec* correlates with EMT biomarkers controlled by the ESPR (epithelial splicing regulatory protein) proteins, which regulates the splicing of cell-specific mRNAs in epithelial cells [93] and when highly expressed, promotes an epithelial phenotype. ESPR1 and its paralog ESPR2 differentially splice the transcripts of CTNND1 (catenin delta-1, also known as p120 catenin), CD44, and ENAH into epithelium-specific shorter isoforms. The absence of ESPR1/2 results in the longer isoforms being transcribed, resulting in a mesenchymal cell phenotype. CTNND1 functions as a cell adhesion and signal transduction molecule and the longer, epithelial-specific isoform of CTNND1 stabilises E-cadherin and promotes cell-cell adhesion [93]. Negative correlations were observed between *rec* expression and both ESPR1 and ESPR2, although out of the genes that ESPR1/2 regulates, there is a negative association between transcription of *rec* and CTNND1, and a positive association with CD44. The results of this study suggest that the cells where *rec* is transcriptionally active also repress the transcription of ESPR1/2, perhaps due to the tumour cells favouring the mesenchymal cell state and stem cell markers. It may be useful in future studies to shed more light on HK2 transcriptional presence within the more epithelial or mesenchymal phenotype by determining the isoform type of CTNND1 and CD44 that is transcribed in concert with high *rec* activity.

There was also a significant association between HK2 transcription and the deregulation of members of the Notch family of receptor proteins. In mammals, one of the five known ligands that bind to Notch includes DLL4, which interacts specifically with NOTCH1 or NOTCH4, activating signal transduction and promoting vascular development and self-renewal in stem

cells [94]. Interestingly, there is an inverse relationship between expression of *np9* and DLL4 in CSC genes, yet there is a positive correlation between NOTCH4 and *np9.* In addition, the *np9*-NOTCH4 upregulation is also positively correlated with the expression of growth differentiation factor NODAL, which has been associated with a more aggressive melanoma phenotype [94]. There is also some evidence that breast CSCs are more resistant to chemotherapy compared to more differentiated cancer cells [95], and this could be partly explained by preferential NOTCH4 activation over NOTCH1 signalling in breast CSC populations compared to differentiated cells [96]. It would be interesting to explore in future studies whether Np9 protein physically interacts with Notch4 protein or its effectors, or impacts their transcription in *trans*. Although there was no significant correlation observed between *np9* and the other Notch receptors, it has been reported that in leukaemia progenitor/stem cell lines, Np9 protein interacts with NOTCH1 and other important signalling genes implicated in cancers, namely AKT1, MYC, MAPK1, and CTNNB1 [26].

A surprising result of this study is the negative correlation between *rec*, *pol*, and POU5F1 (Oct4) (**Table 3.4b**, **Table S3.1b**), a transcription factor necessary for self-renewal of undifferentiated stem cells and cellular reprogramming of differentiated cells to a pluripotent state [97][98]. It has been previously reported that hypomethylation of HK2 LTRs and the transactivation of Oct4 synergistically promotes expression of HK2 [12], however this does not seem to be the case for cancer tumours. *np9* is positively correlated with EHMT2 (G9a), a histone-lysine N-methyltransferase that regulates cellular self-renewal in early development by silencing the expression of Oct4, NANOG, and DNMT3L [98][99]. In addition, G9a also has the ability to silence the transcription of tumour suppressor genes that would consequently induce EMT. The deactivation of Oct4 through heterochromatinisation appears as a complex, multistep

process involving the activity of numerous transcriptional repressors, including G9a which functions as both a methylase and also a recruiter of histone deacetylases and other histone methylases to the Oct4 promoter [98]. It is likely that HK2 expression in the maintenance of self-renewal is more complex than previously thought.

The results of this study indicate that there is not a direct relationship between known "core" pluripotency factors and HK2. Instead, this relationship is likely to be a far more complex interactive process, which might depend on the sensitivity of CSCs and tumour cells to extrinsic factors within the dynamic physical tissue microenvironment, as well as with stochastic variation in gene expression levels that promote cell state changes [100]. Within a heterogeneous tumour population, cells with stem cell-like properties and metastatic potential may exist in a spectrum. For example, it has been suggested that CSCs are not distinct entities but rather malignant cells that transiently acquire stem cell-like characteristics as a result of EMT [100]. In EMT, epithelial cells gain and disseminate new invasive behaviours such as motility, environmental adaptability, and self-renewal that are not only associated with metastatic growth but also with the stem cell phenotype. EMT appears to be a dynamic state, with "metastable" hybrid cells expressing both mesenchymal and epithelial traits, with the ability to induce or reverse the process [101][102]. This intermediate, plastic, "partial EMT" state observed in fibrosis, development, and cancer is initiated and maintained by transcriptional regulation, which is possibly influenced by aberrant expression of previously silenced HK2 insertions and other transposable elements that promote the misregulation of host genes [103]. The activation of HK2 activity could be an advantageous trait selected for during the tumour's evolution, because the derepression of *np9* and *rec*'s oncogenic potential would favour the tumour's growth.

A fundamental property in early development and normal cell and tissue differentiation is reversible cell state switching, and the transient microenvironments generated are ordered interactions between thousands of genes in each cell, all within an intricate gene regulatory network [100] that HK2 is a part of (**Fig. 3.4, 3.5**) [85]. HK2, and specifically its genes *np9* and *rec*, would potentially be good candidate biomarkers in identifying the transitional cancer stem cell-like state in EMT and the shift towards cancer malignancy.

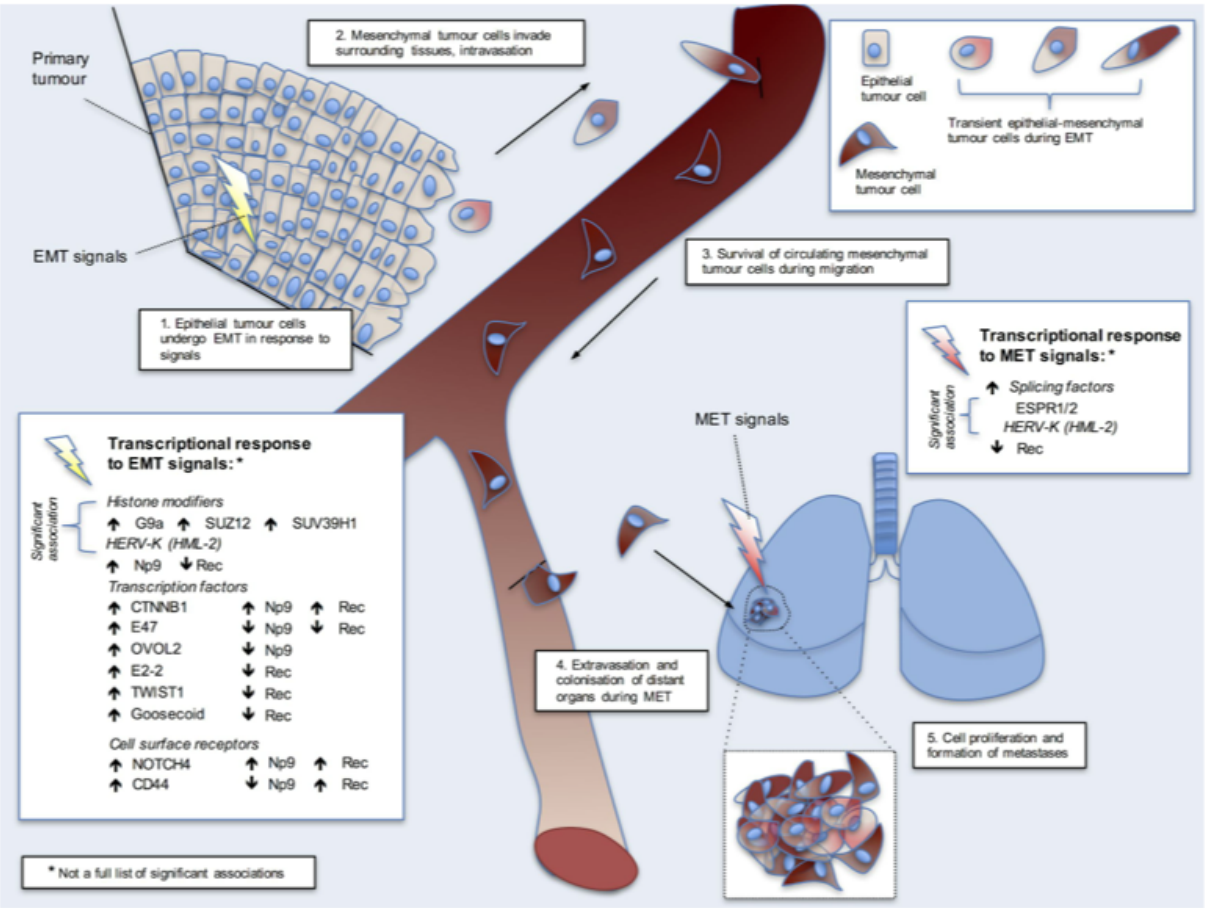


Figure 3.4. Model of *np9* and *rec* transcriptional activity during epithelial-mesenchymal transition (EMT)

1. Epithelial tumour cells undergo EMT in response to signals that include the activity of transcriptional response of histone modifiers, transcription factors, and cell surface receptors. When the histone modifiers G9a, SUZ12, and SUV39H1 are upregulated, there is an upregulation of *np9*, and a

downregulation of *rec*. The inverse is seen with the upregulation of cell surface receptor CD44 with a downregulation of *np9* and an upregulation of *rec*. When the cell surface receptor gene NOTCH4 is upregulated, there is also an upregulation of *np9* and *rec*. The upregulation of transcription factor CTNNB1 occurs with the co-expression of *np9* and *rec*; the upregulation of E47 occurs with the co-repression of *np9* and *rec*; the upregulation of OVOL2 occurs with the downregulation of *np9*; and the upregulation of E2-2, TWIST1, and Goosecoid all result with the downregulation of *rec*. **2.** Mesenchymal tumour cells then invade the surrounding tissues, and intravasation occurs in vasculature. **3.** The circulating mesenchymal tumour cells survive as they migrate within vasculature. **4.** Extravasation of the mesenchymal tumour cells and colonisation of a new site. These tumour cells undergo mesenchymal-epithelial transition (MET) in response to MET-specific signals that include the transcriptional response of splicing factors, specifically the upregulation of splicing factors ESPR1/2 and the downregulation of *rec*. **5.** Cellular proliferation and the formation of metastases at the new site.

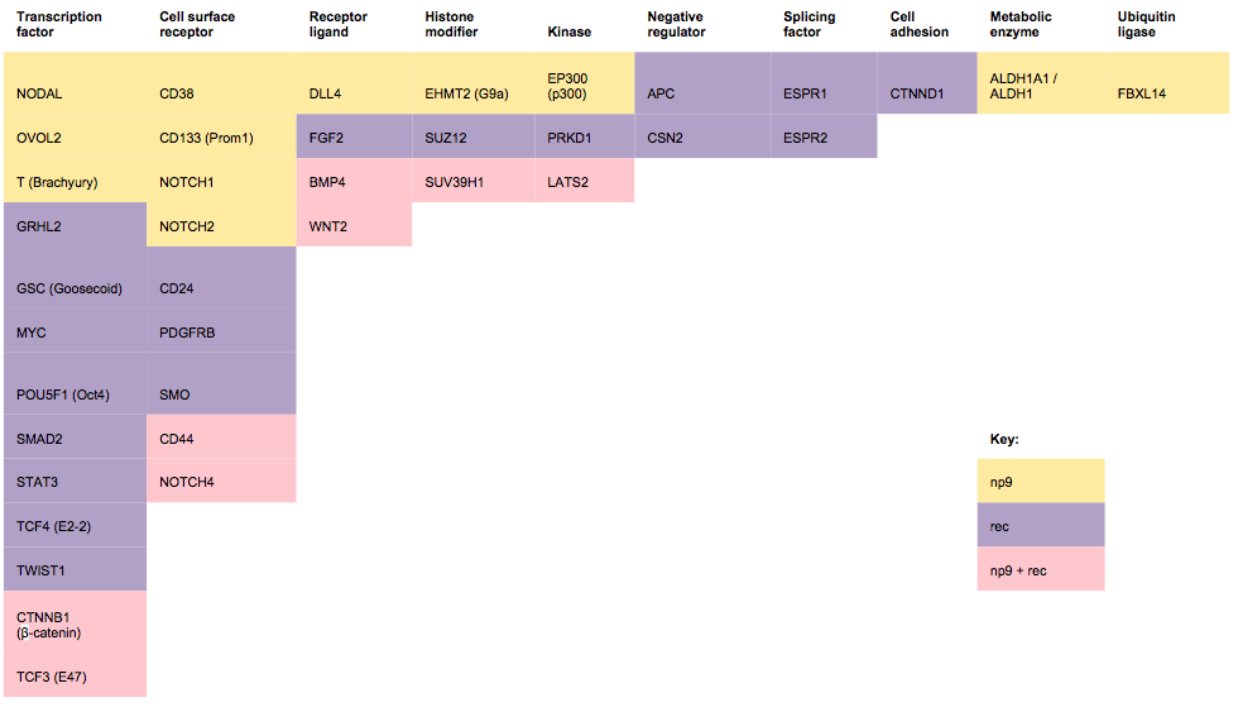


Figure 3.5. Host gene function categories significantly correlated with HK2 gene expression

Genes that are significantly correlated with the expression of *np9* and/or *rec* are categorised according to different host functions. The host genes that only *np9* is significantly correlated with are in the yellow cells, the host genes that only *rec* is significantly correlated with are in the purple cells, and the genes that both *np9* and *rec* are significantly correlated with are in the pink cells.

---

5. Acknowledgements

Thank you to Robert Belshaw, Tim Coulson, Samuel Campbell, and the Palaeovirology group for helpful discussions and comments. Thank you to Paul Klenerman for his help in mediating communication between the co-authors.

6. Funding Statement

GM and TK have been supported by an MRC fellowship to GM (MR/K010565/1). CS is funded by Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile, and AK is funded by The Royal Society. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

7. Ethics Approval and Consent to Participate

All human data used were obtained through The Cancer Genome Atlas and used according to the Data Use Certification as stipulated by the TCGA Data Access Committee and dbGaP authorisation (https://dbgap.ncbi.nlm.nih.gov/aa/). Permission to use controlled-access data was approved by dbGaP under projects #57108 and #7621.

8. Supplement

Table S3. EMT and CSC upregulated and downregulated genes

(a) LTR is significantly associated with specific CSC/EMT gene expression

Gene	Group	Coefficient	p-value
TCF3	CSC	-0.41	$3.66 \cdot 10^{-5}$
WNT2	CSC	-0.07	$3.33 \cdot 10^{-4}$
SNAI1	EMT	-0.17	$3.80 \cdot 10^{-4}$
TCF3	EMT	-0.31	$1.31 \cdot 10^{-3}$
KDM5B	CSC	0.31	$2.11 \cdot 10^{-3}$
HDAC3	EMT	-0.52	$5.14 \cdot 10^{-3}$
SIN3A	CSC	0.51	$5.52 \cdot 10^{-3}$
MAPK1	CSC	-0.40	$6.30 \cdot 10^{-3}$
MAPK1	EMT	-0.35	$1.19 \cdot 10^{-2}$
ALDH1A1	EMT	-0.07	$1.75 \cdot 10^{-2}$
NOTCH2	CSC	-0.23	$1.76 \cdot 10^{-2}$

NRB01 (Dax1)	CSC	-0.03	$2.12 \cdot 10^{-2}$

(b) *pol* is significantly associated with specific CSC/EMT gene expression

Gene	Group	Coefficient	p-value
SOX2	EMT	-0.10	$3.54 \cdot 10^{-5}$
KDM5B	CSC	0.41	$8.79 \cdot 10^{-4}$
SOX2	CSC	-0.08	$1.0 \cdot 10^{-3}$
MYC	EMT	0.22	$1.03 \cdot 10^{-3}$
SMAD3	CSC	-0.23	$2.84 \cdot 10^{-3}$
TCF4	CSC	-0.24	$3.83 \cdot 10^{-3}$
CD44	CSC	0.16	$5.29 \cdot 10^{-3}$
POU5F1 (Oct4)	EMT	-0.15	$6.15 \cdot 10^{-3}$
KDM5B	EMT	0.004	$1.43 \cdot 10^{-2}$
NANOG	CSC	0.31	$2.67 \cdot 10^{-2}$
FGF2	CSC	0.13	$2.69 \cdot 10^{-2}$

KLF8	CSC	0.78	$3.31 \cdot 10^{-2}$
DLL4	CSC	0.20	$3.76 \cdot 10^{-2}$
ZEB2	EMT	0.15	$3.84 \cdot 10^{-2}$

(c) LTR is significantly associated with specific CSC/EMT gene expression

Gene	Group	Coefficient	p-value
CD38	CSC	10.4	$8.21 \cdot 10^{-3}$
CD38	EMT	4.3	$1.60 \cdot 10^{-2}$
BMP4	CSC	2.6	$1.94 \cdot 10^{-2}$
SUZ12	CSC	18.90	$4.31 \cdot 10^{-2}$

(d) *pol* is significantly associated with specific CSC/EMT gene expression

Gene	Group	Coefficient	p-value
CD24	EMT	-1.3	$1.69 \cdot 10^{-4}$
HSP90AA1	EMT	-23.3	$2.67 \cdot 10^{-3}$
APC	EMT	-10.1	$4.94 \cdot 10^{-3}$

ESRP1	EMT	13.0	$1.05 \cdot 10^{-2}$
CDH1	EMT	1.6	$2.15 \cdot 10^{-2}$

Summary of significant correlations between upregulated genes within the CSC (CSC_up) or EMT (EMT_up) groups with LTR **(a.)** and *pol* **(b.)** Summary of significant correlation between downregulated genes within the CSC or EMT groups with LTR **(c.)** and *pol* **(d.)** Parentheses ( ) indicates other names for the same gene.

Chapter 4, "Discussion and Conclusions concerning Part I"

1. Concerning Chapter 1: Endogenous retroviruses and human genetic diseases

Chapter 1 is a literature review discussing the recent developments on specific genomic elements of viral origin that contribute to the human host's genetic and transcriptional identity. I discuss the growing body of evidence that specific "domesticated" retroelements such as endogenous retroviruses (ERVs) are not genomic "junk," but have biological significance in humans. Much of the recent research indicate that nucleic acids, rather than proteins of retroelements, modify human gene expression and regulatory networks in a cell and tissue-specific manner. These retroelements act as promoters, enhancers, noncoding RNAs, and proteins in human biological processes such as hematopoiesis, human innate immunity pathways, anticancer mechanisms, and central nervous system function.

In this discussion, I reiterate the main results of the previous research chapters as case studies that contribute to the field of ERV research described in Chapter 1. In Chapters 2 and 3, I develop methods to try to elucidate the role of ERVs in host immunity and cancer, and I focus on the evolutionarily young family HERV-K (HML-2). Chapter 8 (Appendix) focus on the mechanism by which a specific HERV-K (HML-2) insertion in the RASGRF2 gene can regulate the host cell with potential consequences for pathogenicity. I discuss new relevant research in the field that had been published since my work on these chapters. I also describe potential limitations and implications of my research, and offer suggestions for future directions for this field.

2. Concerning Chapter 2: Targeting endogenous retrovirus gene transcription in human cancers

Although there are clear associations between ERV expression and many human genetic diseases, it is unclear if ERVs directly cause or exacerbate disease pathogenesis. For example, whether the presence of HERV-K (HML-2) nucleic acids and its gene products are merely byproducts of inflammation or immunological dysfunction, or if they indeed have a part in the aetiology of disease development. Because copies of ERVs are ubiquitous in the genome, particularly those of fixed insertions, their presence is not considered unique in disease. However, many loci of evolutionarily young ERVs like HERV-K (HML-2) are not fixed, and their presence and aberrant activity may likely contribute and be a part of various complex human diseases with multifactorial causes, such as cancers. Therefore, the investigation of the role ERVs in disease pathogenesis requires careful detection of viral expression in the disease of study.

This is demonstrated in Chapter 2, where I present a novel and a highly specific pipeline that is able to capture and measure transcription of HERV-K (HML-2)-encoded putative oncogenes called *np9* and *rec*. Using this method, I observed high transcription of *np9* and *rec* in testicular germ cell cancer, and cancers of the eye, skin, liver, lung, and thymus. Previous studies have also shown that HERV-K (HML-2) expression is high in specific stages of early embryonic development [12][56][57][87], and I also demonstrate that *np9* and *rec* are specifically upregulated during the 8- and 16-stage cell stages of human embryogenesis.

Previous studies ([104][105][106]), as well as my analysis of RNA-seq data in Chapter 2 suggest that breast cancer patients and cell lines are capable of producing HERV-K (HML-2) protein. In a recent genome-wide transcriptome

analysis of triple-negative breast cancer, it was found that HERV long non-coding RNA named TROJAN is not only highly expressed, but promotes cellular proliferation, invasion, and indicates poor patient outcomes. TROJAN specifically upregulates metastasis-related genes, and also binds to a metastasis-repressing factor termed ZMYND8, and promoting its degradation through the ubiquitin-proteosome pathway [107].

In another recent study using breast cancer cell lines [108], it was found that 15 full-length HERV-K (HML-2) proviruses were significantly expressed, primarily in the antisense orientation and from proviruses integrated within introns. The LTR-driven transcription was found only in tumourigenic cells, which suggests that HERV-K (HML-2) transcriptional activity is dependent upon the cellular environment of cancer malignancy. Surprisingly, five proviruses (1q21.3, 1q22, 3q12.3, 4p16.1b, and 4p16.3a) that were transcribed in the sense orientation do not contain any open open reading frames for any retroviral genes, including alternative splice products of *env* [108]. Stage-specific transcription factors contribute partly to LTR promoter activity during malignant transformation, and in some cases, transcription factor binding site polymorphisms at the 5' LTR may be responsible for the differential HERV-K (HML-2) expression seen between individuals [109].

The results illustrate the significant heterogeneity of HERV-K (HML-2) provirus expression amongst different breast cancer cell lines as well as patients, and how poorly understood the mechanisms HERV-K (HML-2)'s putative role in cancer development and progression is.

Limitations and future directions

A limitation to my study in Chapter 2 is that additional isoforms or variants of *np9* and *rec* would not be detected by my novel method. Specifically, there

is a 414 bp isoform of *rec* that has a 58 bp deletion containing the splice donor. Because my method isolates *rec* transcripts specifically by the splice junction, it would not be able to detect transcripts of the shorter isoform, but only of the longer 471 bp "wild type" *rec*. Because I could not perform protein assay experiments for this case study, my rationale for choosing the 471 bp "wild type" *rec* for my method is that the full *rec* transcript ss more likely to be translated into protein rather than the truncated isoform.

However, it is also possible that there can be no proxy for protein assay experiments using RNA transcripts, because many mRNAs are degraded by RNAses in the cytoplasm before translation [110]. Conversely, *np9*/*rec* "expressed" from intronic HERV-K (HML-2) could be artefactual, and only present as a result of incomplete splicing, where the intron has not been yet removed from the transcript, or from spliced introns that were captured and sequenced before cellular RNAses were able to degrade them. It is also possible that *np9*/*rec* were never translated into protein, but remain in the cell as tissue-specific non-coding RNA [111].

Some HERV-K (HML-2) reads in the RNA-seq data may derive from within gene introns, hence it would be useful to be able to measure the amount of splicing that goes on in the sample. Specifically, when HERV-K (HML-2) proviruses are upregulated, does the amount of subsequent splicing change? Many HERV-K (HML-2) transcripts will have varying amounts of transcriptional processing: the main *gag-pro-pol* polypeptide is from an unspliced transcript, *env* is from a single spliced transcript, and *np9*/*rec* is from a doubly spliced transcript.

In RNA-seq data, reads that map to gene introns are often observed, particularly from nascent RNAs, or genes that are transcribed but where the RNA polymerase has not yet reached the end of the gene [112]. As a result, some intermediate splicing products end up being detected, and also explains

why longer introns have higher RNA-seq coverage compared to shorter introns. It is also possible that some HERV-K (HML-2) mapped reads may be incidentally present for alternative reasons. For example, although rare, if a full-length ERV is in a gene intron and that gene is highly expressed and processed, some gene splicing intermediates could mimic ERV expression.

One strategy to measure the spliced and unspliced transcripts could be by first identifying chimeric reads that span the spliced transcript and then to compare that number to reads that span the splice donor and acceptor sites. If we can discriminate and isolate reads that span spliced transcripts by detecting intermediate and post-splicing-generated reads, we can then determine that the expression of HERV-K (HML-2) (and subsequently *np9*/*rec*) is genuine and not artefactual.

We also don't know which HERV-K (HML-2) loci the *np9*/*rec* transcripts are expressed from, whether they are from insertionally polymorphic proviruses are not. A possible method in determining the chromosomal loci where the *np9*/*rec* transcripts originate would be to combine the DNA-seq, RNA-seq, and ChIP-seq data for each individual sample, if all three datasets are available. If we know where the chromatin is accessible for the start of transcription, we would have a restricted area to investigate candidate expressed HERV-K (HML-2) loci. Additional evidence that the particular chromosomal locus is transcriptionally active would be to see if nearby genes are actively expressed as well.

3. Concerning Chapter 3: HERV-K HML-2 transcription in diverse cancers is related with cancer stem cell and epithelial-mesenchymal transition markers

In Chapter 3, I take the results gained from the novel method presented in

Chapter 2 and I ask whether *np9* and *rec* are associated with biomarkers of cancer stem cells and metastatic cancer progression.

Specific ERV families drive pluripotency and early embryonic development in human embryonic stem cells and early human embryos, and chromatin remodelling and methylation dynamics allow for temporally regulated transcription of these retroelements in early and later stages of embryogenesis. The results of Chapter 3 indicate that there is not a direct relationship between known "core" pluripotency factors and HERV-K (HML-2). Instead, this relationship is likely to be a far more complex interactive process, which might depend on the sensitivity of cancer stem cells and tumour cells to extrinsic factors within the dynamic physical tissue microenvironment, as well as with stochastic variation in gene expression levels that promote cell state changes during epithelial to mesenchymal transition (EMT) [100].

I also find that the upregulation of *np9* is significantly correlated with the expression of CD133, a cell surface receptor and a recognised marker of cancer stem cells. This observation is supported by recent evidence in malignant melanoma cell lines, where HERV-K (HML-2) is required to expand and maintain a CD133+ melanoma cell subpopulation with a stem cell phenotype in response to modifications in the microenvironment [113].

HERV-K (HML-2) is also overexpressed in tumours with greater transcriptional similarity with testicular germ cell cancer, a type of malignancy known for a strong stem-cell identity by upregulation of pluripotency genes such as Oct3/4, Nanog, and Sox2 [78]. However, surprisingly I find a negative correlation between HERV-K (HML-2) gene targets and Oct4. Both Oct4 and Nanog are known to bind to HERV LTRs, mediating the transcription of both host genes and the provirus [12], and

these results indicate this may not be the case for certain cancer tumours.

Limitations and future directions

The major limitation in Chapter 3 is that the experiments are designed based on the results of Chapter 2, using a method that has its own limitations (as I discuss in the previous section). For example, we do not know which chromosomal loci *np9*/*rec* are expressed from, whether they are inserted within gene introns or not, and whether the proviruses are fixed or insertionally polymorphic. The identification of the proviral loci may allow us to better understand epigenetic changes and cancer or tissue-specific differences [114].

A larger issue, however, is that the biology of epithelial to mesenchymal transition (EMT) in the context of cancer is so poorly understood. Many EMT genes are frequently expressed in non-epithelial tumours, from glioblastomas, melanomas, sarcomas, and leukaemia [115][116]. Many functions of EMT-related genes (specifically of transcription factors) are tissue-specific, non-redundant, and have pleiotropic roles in cancer initiation and metastasis [116]. For example, SNAIL triggers metastasis in breast cancer [117], but has no effect on metastasis in pancreatic cancer [118], which instead favours activation by ZEB1. In addition, members of the same EMT-associated transcription factor family can have antagonistic functions: in melanomas, ZEB1 is tumour-promoting, whereas ZEB2 reduces tumour aggressiveness. However, it is critical to continue to elucidate the pleiotropy of EMT genes and their effects in different tumour types and stages of cancer progression. It is important to also consider the role of signalling genes such as Notch in intracellular and extracellular networks that induce, regulate, and modulate EMT [116]. The presence of HERV-K (HML-2) in these

important signalling pathways and cancer stem cell biomarkers indicate that they are factors within the specific tumour environment that can help our understanding of cancer stem cells during tissue-specific colonisation and metastatic stages of EMT.

Taken altogether, the results in Chapter 3 indicate that HERV-K (HML-2) transcription is a marker of cancer stem cell identity within a tumour, and also appears to be part of the tumour microenvironment during EMT. Interestingly, recent research has revealed that an ancestral ERV envelope protein termed HEMO may also serve as another "stemness" marker in normal cells and as an immunotherapeutic target in cancer. HEMO is found actively shed in the blood of pregnant women and expressed in pluripotent stem cells and some tumours, particularly of the germ cell, breast, and ovary [119]. *np9* and *rec* activity is associated with the expression of many human-specific transcriptional regulatory genes and signalling pathway genes implicated in the maintenance of cellular self-renewal and cancer malignancy through the cell state changes of EMT. Such research will support the development of novel therapeutic strategies in targeting cancer progression from initiation to metastasis [116].

4. Concerning Chapter 9 (Appendix): Human Endogenous Retrovirus-K HML-2 integration within *RASGRF2* is associated with intravenous drug abuse and modulates transcription in a cell-line model

HERV-K (HML-2) may still be proliferating in the human genome, and is also insertionally polymorphic in the human population. Chapter 8 (Appendix) focuses primarily on one specific HERV-K (HML-2) provirus LTR found within an intron of a gene called RASGRF2. This gene is involved in

modulating dopaminergic activity in the brain, is more frequently found in drug addicts, and is significantly associated with addiction.

In the paper, we experimentally show that HERV-K (HML-2) can manipulate nearby genes, even in the antisense orientation. We find in two different human populations that the HERV-K (HML-2) insertion in RASGRF2 is significantly more present in the genomes of people who inject drugs (PWIDs) compared to the non-PWID controls. The results of this paper suggests that polymorphic HERV-K (HML-2) can contribute to the burden of pathogenesis of genetic disease in humans.

Limitations and future directions

Within the LTRs of ERVs are core promoter motifs that can have bidirectional promoter activity, promoting transcription both upstream and downstream of their insertion site [120]. ERV LTRs also contain *cis*-regulatory elements such as enhancers, with the potential to further influence the expression of a host gene it is inserted nearby [121]. ERVs also impact host gene expression by driving the transcription of non-coding RNAs. The majority of ERV LTRs in the mammalian genome drive the transcription of non-coding RNAs (ncRNAs) from ERV insertions in introns, intergenic regions, and the 3' region of protein-coding genes [122][123].

In this our study, we proposed a model where the intronic ERV LTR regulates the host transcriptome by transcriptional collision or RNA interference through cis-natural antisense transcription (cis-NATs) [124], resulting in the disruption (but still sufficiently functional) transcription of RASGRF2 .

The major limitation to this study is that the association of the pathogenic burden of HERV-K (HML-2) comes from the small epidemiological studies

performed on PWID and controls from two European (Greek and British) populations. It is unknown the degree of pathogenicity that the disrupted RASGRF2 transcription due to the presence of the intronic ERV actually has on the individual's health. All individuals who screened positive for the intronic ERV were heterozygous for the insertion; there were no homozygotes identified. This could be because of the small sample size used – both the Greek and British populations are only represented by ~100 PWIDs and ~100 controls. It is possible that the homozygous genotype of the ERV insertion is deleterious for the host and its absence is a result of negative selection.The heterozygous genotype could therefore be a result of balancing selection with yet-unidentified beneficial effects, however, the true evolutionary reasons for these observations are unknown.

As for all reported studies linking ERV activity in human disease, there are many associations but causality has yet to be proven. Larger studies would be needed to clarify the possible existence of homozygotes for the intronic ERV insertion, determine their potential phenotype, and to test for the hypothesis of balancing selection.

5. Conclusions

More than 100 years have passed since retroviruses were first discovered [125], and the last ten years were marked by significant progress in our understanding of how ERVs and other retroelements shape genome evolution and contribute to host gene regulation [103]. Novel tools for direct genetic manipulations such as the CRISPR-Cas9 system used in Chapter 8 (Appendix), show that retroelements like HERV-K (HML-2) have been co-opted by the host for the regulation of cellular functions, in processes such as immunity [12][1][11][10], early embryonic development [12][56][57][87][119][78][126], and dosage compensation [103][127].

Recent studies beyond the scope of my research chapters have further elaborated on the role of *np9* and *rec* in disease pathogenesis, specifically within the context of ERV transactivation alongside co-infection with other exogenous viruses. Kaposi's sarcoma-associated herpesvirus (KSHV)-induced Kaposi's sarcoma is one of the most common tumours associated with AIDS [128], and there is evidence showing the upregulation of HERV-K in HIV-1-infected patients [129]. Dai *et al.* [130] have recently demonstrated that transactivation of HERV-K (HML-2) by KSHV promotes development of Kaposi's sarcoma. *np9* rather than *rec* was found to be more prominently expressed in KSHV and AIDS-associated Kaposi's sarcoma tumour tissues, and *np9* is more closely associated to KSHV-induced invasion and anchorage-independent growth of primary endothelial cells, resulting in enhanced viral pathogenesis in the infected cells. RNAi-mediated silencing of *np9* in a KSHV-infected cell line dramatically reduced in vitro cell growth and also suppressed KSHV-induced tumours in nude mice, suggesting a functional role for *np9* in KSHV-induced tumourigenesis [130][131].

An additional complication to our understanding of the role of ERVs in pathogenic burden is a paradigm shift: the possibility that HERV-K (HML-2) may have a role in preventing rather than contributing to tumour development. Because aberrant ERV expression in healthy cells is commonly prevented by epigenetic control mechanisms like DNA methylation, a potential therapeutic option is the use of innate immune sensing of ERV nucleic acids to promote anti-tumour activity. ERV expression from the provirus LTRs would be induced by epigenetic drugs such as methyltransferase inhibitors or KRAB zinc-finger proteins [132], resulting in the detection of ERV nucleic acids in the cytoplasm and other cellular components by TLRs and other innate immune sensors. The activated pattern recognition receptors would then initiate antiviral signal transduction pathways resulting in the production of type I and III IFNs, pro-

inflammatory cytokines, apoptosis, and the promotion of an adaptive immune response through anti-ERV antibodies [114]. Such an approach has been described as "viral mimicry," analogous to a cellular response to exogenous viral infection [114]. Elevated expression of ERV nucleic acids and protein would enhance tumour immunogenicity in cancer cells, functioning as an encoded "warning system" that alerts the host cellular and immune defences to any genetic or epigenetic damage [133].

To summarise, I show in the research performed in these previous chapters that the human host sustains a balancing act in modulating the positive and deleterious forces of retroelements such as HERV-K (HML-2), with implications for healthy development, immune function, and susceptibility to disease. The adverse effect of ERV nucleic acids on tumour cells offers intriguing possibilities for novel therapeutic approaches. The presence of HERV-K (HML-2) in the acquisition of the cellular "stemness" phenotype within the context of the tumour microenvironment and metastatic aggression necessitates further research in further clarifying their functional mechanisms in cancer plasticity. Such studies would provide us with new insights and strategies for developing anticancer therapeutics.

Part II: Using large-scale ancient animal DNA datasets to address fundamental evolutionary questions

Chapter 5, Introduction

"Ancient DNA for evolution-focused enquiry"

I joined Palaeo-BARN in April 2017 to start a new project with Laurent Frantz and Greger Larson. My goal was to use bioinformatic methods to take advantage of the wealth of ancient DNA that was being generated in Palaeo-BARN in order to explore fundamental evolutionary biology questions as well as early animal domestication processes. Following this introduction, the rest of this thesis represents work that I have completed as a member of Palaeo-BARN. This also includes the Appendix chapters "Dire wolves were the last of an ancient New World canid lineage" (Chapter 10) and "Specialised sledge dogs aided Inuit dispersal across the North American Arctic" (Chapter 11).

Chapter 6, "Time-dependent estimates of evolutionary rates in ancient animal mitogenomes", was co-supervised by Professor Simon Ho, University of Sydney. This chapter addresses the hypothesis of time-dependent rates by measuring the substitution rates of modern and ancient mitochondrial genomes of seven animal taxa, using both rapid-dating and Bayesian methods.

Chapter 7, "Ancient pigs reveal a near complete genomic turnover following their introduction to Europe" was published in PNAS in 2019 (doi:10.1073/pnas.1901169116). This paper was the combined effort of more than 100 co-authors, and my contribution focused on phylogeographic analyses from pig mitochondrial haplotype markers in pigs We show the near-complete genetic replacement of Near Eastern ancestry in domestic pigs in Europe over the timespan of several thousand years.

1. Rates of evolution

1a. The molecular clock

One of the primary goals of biological research is to understand the evolutionary process. Efforts using direct analysis of nucleotide sequences have had a tremendous impact in achieving this goal, providing a wide range of valuable insights into evolutionary relationships and timescales. The inference of phylogenies and evolutionary divergence times are dependent upon reliable models of the evolutionary process [134][135], one of the simplest and most powerful being the molecular clock model. The term "molecular clock" was introduced in the 1960s by Zuckerkandl and Pauling [136][137] in their studies of haemoglobin proteins. They suggested that amino acid differences accumulate at a constant rate, so that there is a linear relationship between the degree of sequence divergence between species and the time since their last common ancestor [136]. Consequently, this meant that if the rate of amino acid change is known, the times of evolutionary divergence between species could be estimated.

The molecular clock concept was soon supported by additional amino acid studies in 1971 by Dickerson, focusing on rates of amino acid change in fibrinopeptides, hemoglobin, and cytochrome *c* [138]. The discussion of protein molecular clocks led to the investigation whether the molecular clock concept could be applied to DNA. In the last few decades, the molecular clock has been used to estimate evolutionary rates and divergence times of some of the deepest nodes in the Tree of Life including eukaryotes [139][140] and bacteria [141][142][143].

1b. Evolutionary rate variation

Several years after Zuckerkandl and Pauling proposed the molecular clock hypothesis, Kimura suggested that the "neutral theory of molecular evolution" could explain the observations of constant evolutionary rate and that a large proportion of mutations seem to not alter the overall fitness of an organism [144]. The accumulation of spontaneous neutral mutations is governed by the random process of genetic drift [144][145], and is independent of population size. The rate of evolutionary change will therefore be constant among species, assuming that their generation times are identical [146]. The reasoning is that most inherited mutations are thought to occur as a result of copying errors during germline DNA replication, and species with shorter generation times tend to accumulate more deleterious mutations per year [147].

The "strict" molecular clock model provides a useful null hypothesis of evolutionary rate, however it does not account for rate variation across different taxa and timescales, which can often be attributed to differences in the biological context of the studied species [148]. When measuring rates within the same species, there have been significant differences in the reported evolutionary rates. It has been shown that mammalian [149][150] and bird [151] rates measured directly (by counting mutations between generations) and by using fossil calibrations can differ by up to 1.5 orders of magnitude. In addition, there are considerable disparities between the low rates estimated from fossil data compared to pedigree studies that infer remarkably high rates [152].

The growing recognition of rate variation that contradicts the assumption of rate constancy of the "strict" molecular clock motivated the development of more nuanced models that incorporate rate heterogeneity across lineages

[153][154][155][147]. These "relaxed clock" models allow the rate to vary across the branches of the phylogeny [154][156].

1c. Time-dependent molecular evolution

Another explanation for the inconstancy of the "tick" of the molecular clock is the hypothesis of time-dependent molecular evolution, or time-dependent rates [152][157][158][159][151]. This is the observation that there is a measurable transition from an increased, short-term mutation (the instantaneous rate at which nucleotide changes occur in the genome) rate to a slow, long-term substitution rate (the rate at which mutations are fixed in the population) [152]. This "rate curve" decays exponentially from the present to the past even when the underlying rate of evolution is constant (**Fig. 5.1**).

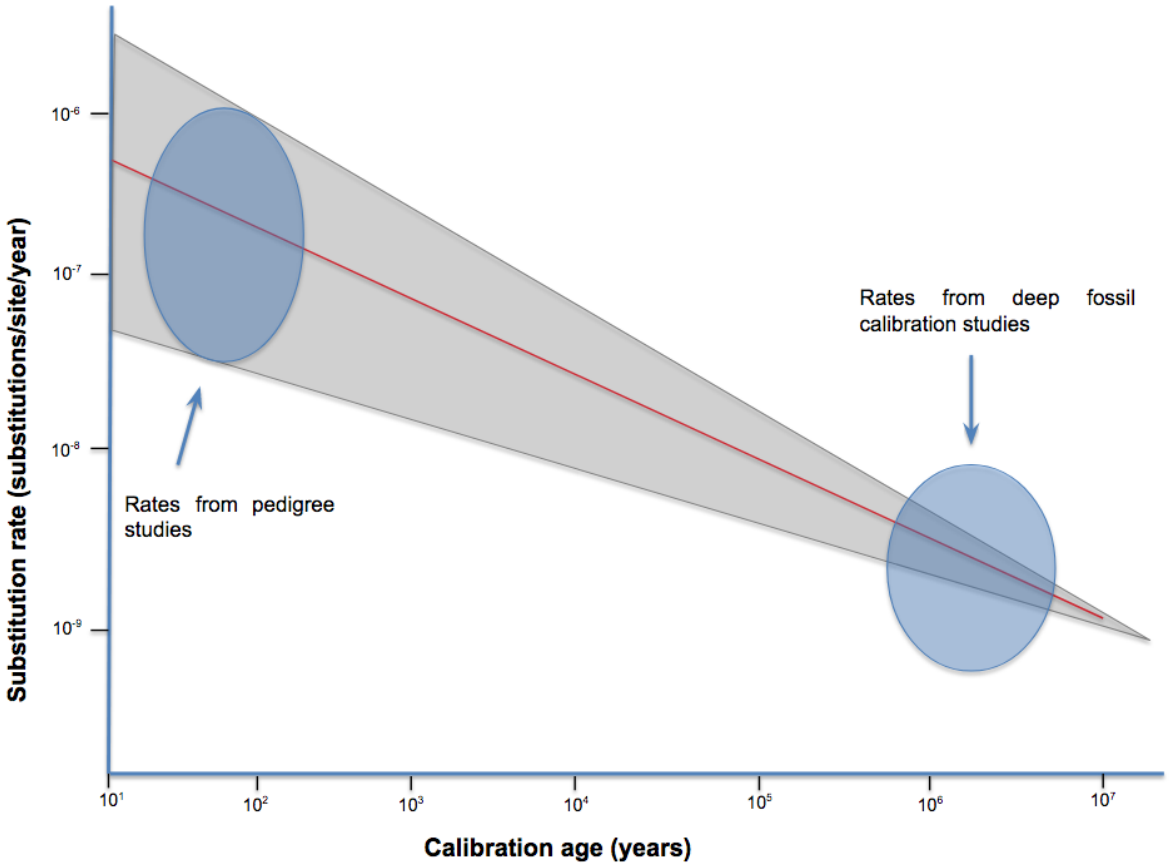


Figure 5.1. A schematic of the declining rate curve of time-dependent molecular evolution on a log-log chart

The x-axis depicts calibration times of the molecular clock in years, and the y-axis depicts the molecular rate of evolution in terms of nucleotide substitutions/site/year. The red line is the regression of the time-dependent substitution rates derived from previous studies [151], and the grey shaded areas represent the confidence intervals. The blue shaded oval on the left represents the high rates generated from pedigree studies that take place over a shallow timescale, and the blue shaded oval on the right represents the low rates generated from deep fossil calibration studies that take place over large timescales.

Time-dependent biases in molecular rate estimates, as seen by the negative relationship between calibration time and molecular rate, has been observed in multiple taxa (**Fig. 5.2**) including amniotes [160], primates and birds [157][161], woolly mammoth [162][163], insects [164][165], fish [166][167], as well as bacteria [168] and viruses [169][170][171][172][173][174][163] .

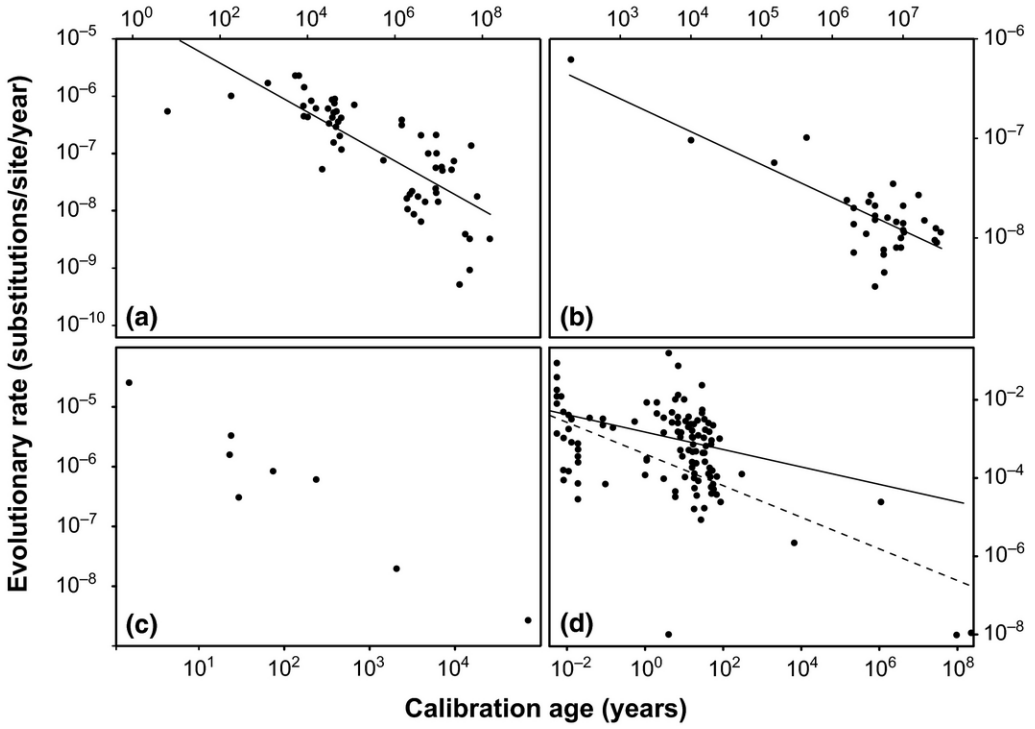


Figure 5.2. Time-dependent patterns in evolutionary rate estimates have been observed in a variety of taxa

(a) Non-coding mitochondrial DNA from amniotes (data from Fig. 1b in [160]); **(b)** Mitochondrial DNA from insects (data from Fig. 2 in [165]); **(c)** Genomic DNA from bacteria (data from Fig. 4 in [168]); **(d)** Genomic RNA and DNA from viruses (data from Fig. 1a in [173]). Trend lines are based on those estimated in the original analyses of these data sets. In panel **D**, separate trend lines are given for RNA viruses (solid) and DNA viruses (dashed). This figure is taken from Fig. 1 in [158].

1d. Calibrating the molecular clock

The strict and relaxed clock models may offer different descriptions of rate variation across the branches of the phylogeny, but they provide only relative timescales of evolutionary divergences. This is due to genetic divergence being a product of mutation rate and time. To transform these relative rates and dates to absolute ones, the molecular clock has to be calibrated, and independent information is needed on the timing of one or more evolutionary divergence events. This information necessary to calibrate the molecular clock usually comes from geological or fossil evidence, which is then used to constrain the ages of the internal nodes [147][175]. For example, well-dated geological events such as island formation and the rifting of continents can be used to constrain the ages of divergences between species affected by these events [176].

An alternative to this approach in calibrating the molecular clock is to take advantage of information from the sequenced samples themselves – assigning dates to the tips (or terminal nodes) of the tree, also known as tip-dating. This is possible in viruses with rapidly evolving genomes that were serially sampled through time, or with ancient genomes [177][178]. The increasing availability of genomic data from ancient DNA allows for tip-dating calibration to a variety of taxa that previously had a scarce fossil record,

including extinct megafauna such as woolly mammoths, ancient humans, and other organisms including fish and amphibians [179][180][181][182][183].

Bayesian methods are a powerful framework for estimating phylogenetic trees, evolutionary rates, and evolutionary timescales. Complex parameter-rich models can be easily incorporated in a Bayesian framework [184][185][186][187], such as detailed descriptions of molecular evolution [188][156], explicit model assumptions [189], and large parameter spaces can be explored efficiently [135]. The estimation of substitution rates, divergence times, and demography can be accounted for simultaneously while also accounting for phylogenetic uncertainty by using Markov chain Monte Carlo (MCMC) integration [153][190]. By reconstructing phylogenetic trees in a Bayesian framework, molecular clocks that are calibrated with ancient DNA have also been used to estimate the age of biological material that date beyond the range of radiocarbon dating. More calibration points from tip dates used in the tree generally leads to an improvement of molecular clock estimates **(Fig. 5.3)** [191][192], and in Bayesian phylogenetic inference, a specified informative lognormal or exponential prior distribution allows for calibration uncertainty in the node ages [193][194][195].

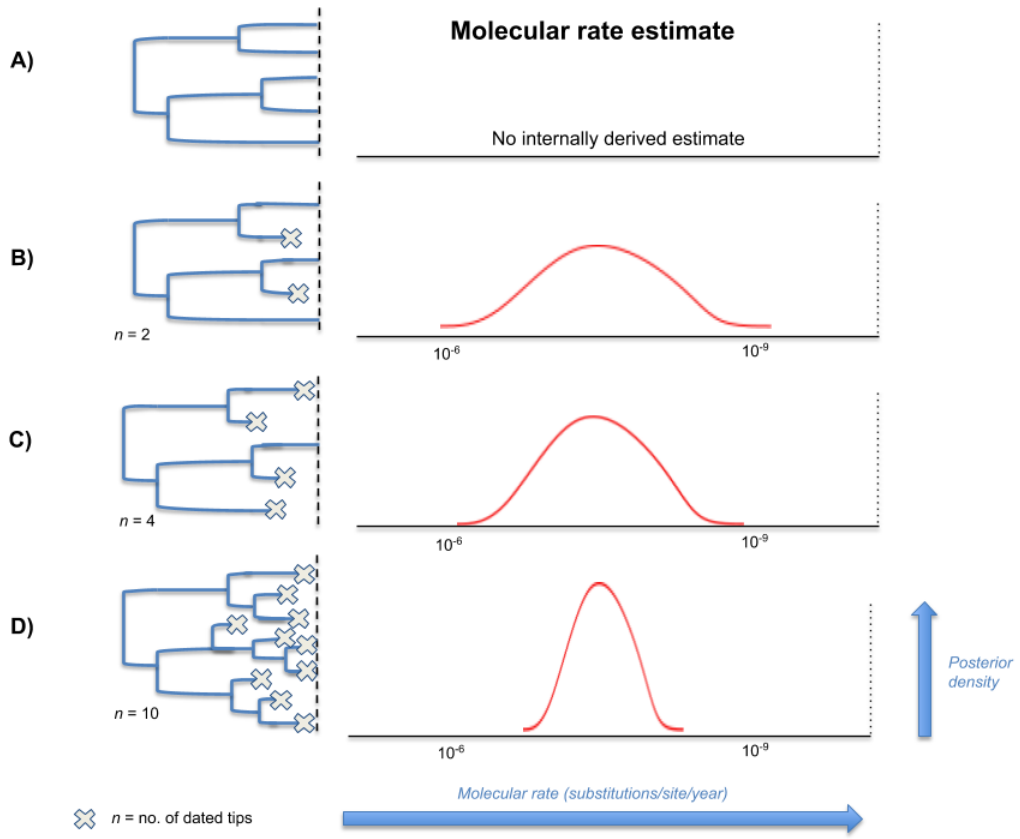


Figure 5.3. A schematic demonstrating the value of increasing the sample size of dated samples in a molecular phylogeny

The left panel depicts phylogenetic trees with varying numbers of radiocarbon dated tips marked as Xs. The right panel depicts an idealised posterior density distribution of a molecular rate parameter (obtained from a Bayesian molecular clock analysis). The narrower the distribution, the more precise the inference. Panel **A** shows a tree that only includes modern samples. Without internally dated tips, it is not possible to derive an internally derived molecular rate estimate. Panels **B-D** depict trees with an increasing number of dated samples which allows for greater certainty in the estimates of evolutionary rates.

There are now a number of statistical models implemented with tip-dating methods that can account for heterogeneity in various aspects of the evolutionary process while at the same time accommodating for very large datasets, such as whole genomes. This allows for inferences to be made on

divergence times, substitution rates, the dates of specific mutations, and past demographics [182].

2. Animal domestication

2a. The concept of domestication

Charles Darwin famously first wrote about artificial selection in *The Variation of Animals and Plants Under Domestication*, where he described how the results of artificial selection are similar to that of natural selection. He observed that the dramatic phenotypic changes can be made through the "gradual and accumulative force of selection," but emphasised that evolution by selection can only work where genetic variation is also present: "The power of selection… absolutely depends on the variability of organic beings" [196].

In the decades since Darwin had penned his observations, geneticists, zoologists, and zooarchaeologists have concentrated a great deal of effort on understanding the genetic underpinnings of developmental, anatomical, physiological, and behavioural changes associated with the domestication process [197][198][199][200].

The domestication-associated phenotypic changes in mammalian species is referred to as the "domestication syndrome," and the most crucial trait that all domesticates share is a tolerance or lack of fear of humans. Other features include reproductive cycle changes that result in year-round breeding [201], adaptations to an anthropogenic diet [202], and other physical and physiological alterations such as variation in coat colour and tail morphology, brain size reduction, changes in craniofacial morphology, floppy and small

ears and other manifestations of neonety (i.e. retention of juvenile characteristics into sexual maturity) [203][198][200].

The term “domestication” has been proposed to refer only to the original process that produced an early domestic population from their wild progenitors. If the original domestication process is followed by admixture with wild populations not involved with the original population, this should not be regarded as domestication but rather part of the evolutionary history of the domestic species [204].

2b. Early domestication

The domestication of animals and plants was one of the most important cultural and evolutionary transitions in the history of the human species, and is valued as a model of evolutionary and demographic change [205][204][206]. The products of domestication by humans over the last 11,500 years has transformed the planet, with consequences for the environment, population size, and evolution of both humans and the domesticated organisms.

Early perceptions of domestication was dominated by the assumption of near-universal adoption of agriculture and that animals were domesticated in or nearby a limited number of “core zones” [207][208]. However, domestication studies in recent years has facilitated a shift in this perspective – for example, indigenous plants were domesticated in up to 20 different geographically distinct regions [209][205]. In the Fertile Crescent, the domestication of cereals is also a result of multiple independent domestication of multiple grain species in multiple centres of domestication from Southern Anatolia to Syria and the southern Levant [210][198]. As for animals, there appears to be multiple independent domestications of geographically and genetically divergent populations of pigs [211][212][213],

cattle [214][215], goats [216][217], sheep [218][219] and horses [220][221][222].

The first domestic animals however most likely did not result from deliberate, goal-directed selection by humans, but rather through a protracted co-evolutionary process with multiple stages along different pathways, such as the commensal pathway [223][224][225][213][226] (e.g. the wild animals were initially attracted to human settlements were tamer, less aggressive individuals), the prey pathway [227][228][229][230][231][232] (e.g. animals that were targeted by humans who then altered their hunting strategies to maximise access to these animals), or the directed pathway (e.g. where there is a deliberate objective to domesticate a species) [222][233][234][235][236][237].

Early domestication was a co-evolutionary process with multiple stages along multiple pathways; the domestication events occurring as many animal species became more involved in the lives of humans and their environments, to the point where humans had a direct role in the reproduction and survival of these animals [222]. Although the exact mechanisms leading to the genetic changes of early domestication remain elusive, the early domesticates frequently bred with other domestic species and wild populations, with gene flow significantly affecting the genotypes and phenotypes of the domestic animal populations. Random mutations resulting in novel traits would emerge after the establishment of the domestic animal populations, and the selection by humans for these particular traits are most likely economically motivated (e.g. shorter reproductive cycles, milk production, wool quality, egg-laying, and coat colour) [222].

2c. Using ancient DNA to study domestication

The modern field of palaeogenomics emerged with HTS (high-throughput sequencing), which although was originally developed for other uses, significantly reduced the amount of time, costs, and sample amounts required for sequencing ancient DNA. HTS allows for millions to billions of DNA templates to be sequenced in parallel, dramatically increasing the amount of genetic material of interest that can be obtained from each extract that often contains extremely low amounts of endogenous DNA [238][195][239]. Improved molecular techniques to access ancient DNA molecules have also been developed. There are now methods that can recover ultra-short (~25-50 bp) DNA fragments [240][241][242]; large genomic regions within samples with low endogenous DNA content can be enriched using hybridisation enrichment methods [243][244][245][246][240][228]; and new protocols are available for extracting DNA from hair [247], petrous bone [248], teeth [249][250][251], and soil [252][253]. Improved DNA library construction methods allow for access to DNA with significant postmortem damage [254][255][256], and computational tools were developed that optimised read alignment parameters for ancient DNA [257][258][259][195][260].

Ancient DNA has the ability to shed light on the timing, location, and evolutionary mechanisms and dynamics of domestication, such as admixture between wild and domestic animals, and human-mediated selection for domestication genes [222].

Modern animal genomics studies that use selective sweep detection, candidate genes, and other tools to evaluate the evolutionary forces that shape genetic variability has resulted with increased understanding of the genes and genomic regions that are most influenced through the selection processes during domestication [261]. For example, such studies have identified the regions responsible for muscle growth in pigs [262] and a

putative “domestication gene” that influenced the origins of chicken domestication [263]. The reduction of genetic variability in commercial chicken populations relative to noncommercial breeds have also been quantified using high-density genetic markers [264][204].

However, inferring evidence from modern domestic animals from the earliest stages of domestication from modern data has been unsuccessful, mainly due to confounding genetic signatures from long-term gene flow within and between wild and domestic species through intensive breeding practices over the last 200 years. The impact of artificial selection is so strong that it is difficult to infer domestication history from modern populations that bear little genetic resemblance to their early domesticated ancestors [223].

As demonstrated in palaeogenomic studies of domestic pigs and wild boar [212][265], the genetic signatures of their early domestication heritage had disappeared nearly completely following their geographic dispersal and subsequent admixture with the native wild boar population, and had been replaced by genetic signatures of the latter. Therefore, the breeding and domestication history that underlies the characteristics of the animal’s domestic phenotype cannot be reconstructed from patterns of genetic variation found in modern livestock and companion animals. As a result, the ubiquity of a particular phenotypic trait in modern breeds does not necessarily indicate that this trait was an early target of domestication.

Another example can be seen in the domestic chicken. For the vast majority of commercial chicken breeds, yellow skin is very common and it was suggested that this trait was selected for prior to the arrival of domestic chickens in Europe between 2,700-2,900 years ago [266]. However, when 80 ancient European chickens dating from 280 BC to the 18th century AD were genotyped for the *BCDO2* locus, it was found that none of these animals could express the yellow skin phenotype. This result indicates that the yellow

skin trait only became prominent during the late domestication history of chickens, most likely during the development of commercial breeds following the Industrial Revolution [266][200]. The results of this study illustrate the limitation of using modern genomes to infer the history of early domestication, and emphasises the importance of using time series ancient DNA datasets when investigating how domestication processes shaped the modern genome.

Ancient DNA studies of European domestic chickens have further demonstrated that selection for other traits such as reduced fear of humans and increased egg production only occurred over the past 1,100-500 years [267][266]. These results in combination with other archaeological and historical evidence supports the hypothesis that chicken husbandry had intensified in Medieval Europe as a consequence of popular religious fasting practices that limited the consumption of meat from four-legged animals [267][268].

Long-held perceptions about the domestication has also been contradicted by recent evidence from ancient genomic data. For example, the reduced level of diversity in domesticated crops is commonly attributed to the "domestication bottleneck" – a dramatic reduction in the population size associated with the subsampling of the wild progenitor species and selection pressure on the subsampled species based on traits associated with the domestication syndrome. A sharp decline in genetic diversity is then presumed to be a result of the domestication bottleneck. However, analysis of ancient genomes of sorghum, maize, and barley show instead that there was no drop in genetic diversity in the early stages of domestication [269]. This observed loss of genetic diversity is unlikely to be directly due to domestication, and this study clearly demonstrates how ancient genomes can better inform our understanding of the evolutionary processes of domestication.

2d. Mitochondrial genomes in ancient DNA studies

Extracting and sequencing DNA from archaeological remains can allow for the characterisation of ancient genomes from animals that have not been subject to as much introgression as is seen in their modern counterparts [204]. However, technological limitations had restricted the focus of earlier ancient DNA studies on the mitochondrial genome, which is a maternally inherited non-recombining single locus with a high mutation rate [270]. Although mitochondrial DNA was sufficient for differentiating between broadly defined population (e.g. haplotypes) and tracking the movement of these populations, they lacked the power to fully reconstruct evolutionary histories [271]. For example, with mitochondrial DNA it is difficult to specifically to quantify with precision genetic signatures of early domestication such as degrees of admixture between populations, human-mediated migration, and intensive sex-specific breeding practices [272][204].

However, mitochondrial genomes are still a valuable asset to domestication studies because of the vast amount of high-quality ancient data from multiple that is available – particularly sequences of the control region and cytochrome *b* [211][273][274][220][275][216][276][277][278][279][280][281] – and specific hypothesis-testing models for non-recombining genetic markers [282][204]. Finally, there is an increasing amount of complete mitochondrial genomes from many different species and diverse genetic and geographic populations that are available and will provide an excellent resource for future evolutionary studies [283][284][150][285][217][221][286][287].

2e. Nuclear genomes in ancient DNA studies

Recent developments in HTS technology has allowed for the nuclear genome to be analysed in a population genetics framework, thereby overcoming the

limitation of ancient mitochondrial DNA studies. The application of models [288] increases the resolution in studying the nuclear genome, leading to observations that gene flow is not only common between geographically diverse domestic populations of the same species but also between domestic and wild populations that never independently gave rise to a domestic population [205], as demonstrated in European domestic pigs [211][212][213] in Chapter 7.

3. Origins, expansion, and genetics of animal domestication

The next 4 sub-sections are case studies describing recent ancient DNA literature of four different domesticated species: pigs, goats, horses, and dogs.

3a. Case study: Pigs

According to zooarchaeological evidence, the protracted long-term processes of pig domestication only took place independently twice – first in the Near East, as seen by demographic changes in *Sus* remains excavated from archaeological sites in Eastern Anatolia approximately 8,300 years ago [226], and in China, where by 8,000 BP pig assemblages were prevalent in archaeological sites in both northern and southern China [289][290][291]. Ancient DNA studies using mitochondrial genomes [211][292][293], nuclear markers [294][295][296], and whole genomes [297][298][299] confirmed the unique genetic identities of European and East Asian wild boar, and also genetic differences between their domesticated descendents [291].

Once pigs were domesticated in Eastern Anatolia and China, they accompanied farmers during the migrations beyond the domestication

regions - from the Near East into Europe, and from China into Polynesia. Following the introduction and establishment of the domestic pigs in the new environment, the pigs hybridised with the local wild boar population and acquired mitochondrial and nuclear signatures dominated by the local wild boar, erasing almost completely their original biogeographic domestication heritage. As a result, domestic pigs in Europe no longer possess Near Eastern ancestry and instead have genetic markers of European wild boar [212][265][300], and Polynesian native pigs do not have a mitochondrial signature seen in modern wild boar of central and northern China [301][291].

Livestock and companion animals have not evolved in isolation. In domestic pigs, the patterns of genetic diversity demonstrates that there was significant admixture with their wild counterparts, and it is clear that domestic phenotypes can be maintained despite strong homogenising gene flow from wild boar. To explain this observation, Frantz *et al.* [285] proposed the hypothesis of "domestication islands" in the pig genome which resist the effects of gene flow and therefore prevents the back-introgression of maladaptive genes of the wild boar [285].

Echoing the results reported in a study on ancient domesticated crops [269], a study that had analysed over 100 pig genome sequences did not demonstrate any evidence of domestication bottlenecks. There was also no evidence of reproductive isolation resulting in a drop of genetic diversity during pig domestication. Instead, there was substantial gene flow between the domestic and wild populations [285].

3b. Case study: Goats

According to the archaeological record, the domestic phenotype of the goat emerged approximately 9-10,000 years ago in Anatolia [233][302]. Genomic

evidence for early selection for stature, pigmentation, milking, reproduction, and other responses to dietary change was found in 8,000 year old goat genomes [217].

During the Neolithic, multiple divergent wild populations contributed to the origin of modern goats and over time, one mitochondrial haplotype spread and established dominance worldwide. However, the ancient whole genome data revealed that modern goat populations are a mix of goats from different sources in the Near East, and contributed differently to modern goats in Africa, Asia, and Europe. The results support the idea that there were multiple dispersal routes out of the Fertile Crescent region [217], as opposed to dispersal from a central core [303], and also aligns with archaeozoological evidence for disparate early management strategies in Neolithic sites in Anatolia, Iran, and the Levant [304][305]. Multiple lines of evidence suggests that during the period of early domestication, there was not large-scale migrations of population across Southwest Asia, but instead there was a culture of shared domestication and agriculture techniques. There is also some evidence of parallel domestication of crop plants in the Near East [306], corresponding to a divergence in the genomes of Northwest Anatolian and Iranian Neolithic humans [307][308][309].

3c. Case study: Horses

According to archaeological evidence, horses were domesticated by the Eneolithic Botai culture of the Central Asian steppes approximately 5,500 years ago [310]. Studying the patterns of variation in the mitochondrial genomes of ancient and modern horses suggest extensive gene flow from wild into domestic horses [311][312][313][314].

Time and time again, evidence from ancient DNA contradicts long-held notions about wild and domestic animals. For example, Przewalski's horses were once thought to be the world's last wild horse that had never been domesticated [315]. However, a recent study demonstrates that surprisingly, Przewalski's horses are feral descendent of Botai horses, and not truly wild horses. All domestic horses dated to ~4,000 years ago to the present day only show ~2.7% of Botai-related ancestry, indicating a massive genomic turnover during the expansion of domestic horses, coinciding with large-scale human population spread during the Early Bronze Age [316].

Strong phylogeographic structure is not always seen in modern domestic animal populations. For example, evidence at the mitochondrial and microsatellite level in horses show that human-mediated population movements and large-scale dispersal along transcontinental trade routes have resulted in a homogenised geographic population structure [311][220][200].

In cases where the wild progenitors of domestic animals have gone extinct, ancient DNA provides the only method to discriminate the genetic presence of the extinct animal. The sequencing of two complete genomes from pre-domestic Late Pleistocene horses in Siberia revealed the existence of previously unknown and now extinct population of wild horses that had significantly contributed to the genetic makeup of modern domesticated horses [317]. Another recent study also revealed another two extinct horse lineages that existed during early domestication, one located in Iberia and the other in Siberia. These extinct horse lineages however did not contribute significantly to modern diversity. In addition, the authors showed the increased influence of Persian-related horse lineages following the Islamic conquests in Europe and Asia [221].

In a different study focusing on 13 ancient genomes of domestic horses dating from 2,700 to 2,300 years BP in the Kazakh Steppe in Asia [318], enrichment for genes expressed in mammary and pituitary tissues was found within genomic regions that appeared to be under positive selection. This pattern may be due to human-directed selection for phenotypes associated with horse-milking [318], a practice that, according to archaeological pottery lipid analysis and ethnographic records, was conducted for millennia in the geographic region [310][268].

Additionally, multiple alleles associated with elite racing, including MSTN, also known as the "speed gene", were also positively selected for by humans, and only rose in popularity within the last 1,000 years of horse domestication. Modern breeding practices also impacted genetic diversity in horses far more than the previous millennia of human management [221].

3d. Case study: Dogs

Recent palaeogenomic studies of ancient dogs have only revealed how complex the processes of domestication really are. For example, most phylogeographic analyses of dogs indicate that the domestication process occurred only once, however it is disputed whether the lone original domestication centre is in Europe [283], Central Asia [319], or East Asia [320]. Drawn from the genome-wide mutation rate calibrated from the nuclear genome sequence of a ~4,800 year old domestic dog excavated at Newgrange, a Neolithic tomb in eastern Ireland, Frantz *et al.* [150] suggest that there are two domestication centres. Specifically, Western European and East Asian dogs had formed two distinct groups that had diverged between 14,000 to 6,400 years ago [150]. The proposed scenario is that dogs were independently domesticated in Europe and East Asia but migrated with humans from Asia to Europe, eventually replacing the indigenous European

dogs. This claim is supported from mitochondrial DNA evidence of 59 ancient dogs, which revealed that haplogroups C and D, the most abundant in Europe prior to ~3,000 years ago, have been replaced by haplogroups A and B [150].

This scenario of dog domestication is further complicated by another ancient genome obtained from a ~35,000 year old gray wolf that lived in the Taymir Peninsula in Central North Siberia. This ancient wolf was found to belong to a population whose descendents had only contributed to the genetic makeup of Greenland sledge dogs and other Arctic breeds [321].

Mutation rates calibrated from the genomes of the Taymir wolf [321] and the Newgrange dog [150] suggest that modern wolf and dog populations diverged between 20,000 to 60,000 years ago. However, the oldest zooarchaeological remains of domestic dogs indicate that domestic dogs most likely only arrived in the high latitudes of the Arctic within the last ~15,000 years.

Ancient DNA studies are also shedding light on previously unknown extinct wolf populations that seem to bear some morphological resemblance to early domestic dogs. Researchers have argued that canid skull remains found in Goyet, Belgium that date to ~32,000 years ago can be interpreted as "proto-dogs", and that dogs had been domesticated much earlier than previously thought [322]. At the same time, it has also been argued that because of the paucity of Pleistocene wolf material, the range of variation in Pleistocene wolf populations is unknown, so the identification of domesticated dogs prior to the Late Upper Palaeolithic cannot be conclusively accepted or rejected [323]. Recent geometric morphometric analyses demonstrate that the Goyet specimen as well as a 13,000 year old canid cranium from the Upper Palaeolithic Eliseevichi site in Russia, which had first been thought of as a dog [324], are actually wolves [325]. Furthermore, there are also no traces of

the Goyet canid in the mitochondrial DNA of modern dogs, which suggests that this population may not have played a role in dog domestication [200].

Dogs were present in North America before the arrival of the European colonists, and appeared in the archaeological record ~4,500 years after the earliest evidence of human activity in North America [326][327]. A study of 71 ancient mitochondrial genomes and 7 nuclear genomes from ancient North America and Siberian dogs [284] demonstrates that American dogs were not descendants of North American wolves, but rather form a monophyletic clade that most likely had its origins in Siberia and then dispersed throughout the Americas alongside people. These dogs most likely arrived in the Americas not during the first human migration but during a later arrival [328] before the flooding of the Bering land bridge approximately 11,000 years ago [327]. This pre-contact dog population was isolated in North America for at least 9,000 years, but within the past 1,000 years, at least three independent reintroductions of dogs have occurred. The first reintroduction was likely the Arctic dogs that arrived with the Thule (or proto-Inuit) culture approximately 1,000 years ago [329].

The second reintroduction came with the arrival of the Europeans, whose dogs almost completely replaced the native dogs. The final wave came from Siberian huskies that were introduced to the American Arctic during the Alaskan gold rush in the 19th century [330]. As a result of these three dog introductions, native American dogs had almost completely disappeared, leaving only minimal traces of their presence in modern dog populations, which are primarily derived from Eurasian breeds. The only evidence that detects this extant lineage to pre-contact American dogs is the canine transmissible venereal tumour, derived from an individual dog that lived up to 8,000 years ago [284].

4. Conclusions

The revolution in ancient DNA and palaeogenomics research over the past decade is inextricably intertwined with the power of next generation sequencing platforms and the ability to target and sequence short and degraded DNA molecules. Archaeological specimens that were previously unsuitable for DNA analyses due to extensive postmortem sequence damage can now be used successfully as source material [238]. In addition, the analytical power obtained by analysing billions of sequence reads allows for the quantification of contamination issues that had previously hindered the field [238], [182]. As a result, whole genome sequences up to hundreds of thousands of years old have now been retrieved [331].

In addition, the increasing availability of ancient DNA allows for greater precision and accuracy in the molecular dating of phylogenetic trees, where sequence data is used to estimate the timing of evolutionary events as well as rates of molecular evolution. A primary requirement of all molecular dating methods is to convert the genetic divergence between sequences into units of absolute time. In the past, this step could only be achieved by associating dates of fossil or biogeographical events to internal nodes of the phylogenetic tree. This would then allow the tree to be calibrated by associating sampling dates directly to the sequences that represent the tips (e.g. terminal nodes) of the phylogeny [182].

Ancient DNA has allowed for tip-based calibration to encompass a variety of taxa that typically has a scarce fossil record, including Denisovans [254] and Neanderthals [332][333], extinct megafauna such as Beringian steppe bison [334][335], woolly mammoth [336][337], and ancient and historically important pathogens including *Phytophora infestans* (the cause of the Irish potato blight in the 1840s) [338], 1918 Spanish influenza [339], hepatitis B virus [340][341], *Yersinia pestis* bacterium [342][343][344][345], *Salmonella*

enterica [346], and *Mycobacterium tuberculosis* [347] and *leprae* [348][349]. Statistical models have been developed to account for heterogeneity in many aspects of the evolutionary process such as time dependency of molecular rates [158] and variation in substitution rates between phylogenetic branches [350]. These methods are accommodating for whole genomes, allowing for tip-dating methods to make inferences on past demography, times of divergence, substitution rates, or even the age of specific mutations on a variety of spatiotemporal scales [182]. The wealth of data provided for by ancient genomes can allow more accurate phylogenetic reconstructions, where crucially the sampling dates have enough sufficient temporal spread to capture measurable amounts of evolutionary change. As such, the systematic investigation of tip-dating methodologies can be implemented [351][352] to avoid rate and divergence time estimates biased by prior information [182].

Finally, ancient DNA has allowed us to better understand the timing, location, mechanisms, and evolutionary dynamics of domestication. Analyses of ancient genomes of domesticated animals have allowed for the characterisation of their evolutionary origins and their phylogeographic dispersal during different human cultural periods and over long stretches of time [317]. Such studies can be a valuable glimpse into the human decisions that favoured the selection of specific traits in the animals [268], as shown in functional gene changes associated with domestication [353][266]. The massive amount of effort and data has been generated in domestication-related studies, significantly increasing our understanding of domestication patterns and processes [205]. The importance of domestication cannot be understated: over the last 11,500 years, the domestication of plants and animals by humans has transformed the planet, irrevocably changing the environment, population size, and evolution of both the organisms we have domesticated and ourselves.

Chapter 6, "Time-dependent estimates of evolutionary rates in ancient animal mitogenomes."

Statement of Contribution

Research design

I developed the overall project with GL, LAFF, and SYWH.

Data

Sequences were obtained from LAFF, DB, OL, IM, LO, BS, and GL. I curated the data analysed in this study.

Method development

The methodology used in this study was developed with SYWH and GL. I wrote the scripts for partitioning the mitochondrial genomes and creating the BEAST2 XML files with assistance from LdP.

Analysis

I performed all of the analyses, with input by SYWH, LdP, LAFF, and GL. I created alignments of the curated sequences and partitioned the alignments. I generated the phylogenetic trees of the datasets for each sampling window of all the animal taxa. I analysed all the data using date-randomisation tests and using the four statistical dating methods chosen for this study. I evaluated and interpreted the results of all of the analyses.

Manuscript

I wrote the manuscript with additional input from SYWH, LdP, LAFF, and GL. I generated all the tables and figures. I curated and edited the information presented in the Supplement.

Author list and affiliations

Authors

Audrey T. Lin[1,2], Laurent A.F. Frantz[1,3], Louis du Plessis[2], Dan Bradley[4], Ophélie Lebrasseur[5], Ivica Međugorac[6], Ludovic Orlando[7,8], Beth Shapiro[9], Simon Y.W. Ho[10], and Greger Larson[1]

Affiliations

1. The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
2. Department of Zoology, University of Oxford, Oxford, UK
3. School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
4. Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland
5. Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK
6. Veterinärwissenschaftiches Department, AG Populationsgenomik, Ludwig-Maximilians-Universität München, Munich, Germany
7. Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Toulouse, France
8. Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark
9. Department of Ecology and Evolutionary Biology and Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, USA
10. School of Life and Environmental Sciences, University of Sydney, Sydney, Australia

Abstract

Estimating rates of molecular evolution forms the foundation for inferring demographic dynamics and evolutionary timescales from genetic data. However, estimates of evolutionary rates appear to vary with the timeframe over which the rate is measured—this is known as the time-dependence of evolutionary rate estimates. Time-series ancient DNA datasets provide an excellent opportunity to clarify the discrepancy between short- and long-term rate estimates of pedigree and fossil-calibrated phylogenetic studies. In this study, we analyse mitochondrial genomes from modern and ancient and historic archaeological specimens of seven taxa: bison, chickens, dogs, goats, horses, pigs, and sheep. We use three different rapid dating methods and Bayesian inference to estimate mitochondrial substitution rates in these animal taxa across different time scales. We find some evidence of time-dependence in the estimates of evolutionary rates, but the degree of this effect varies among parts of the mitogenome and with the estimation method being used. Only the bison dataset shows a consistent time-dependent pattern in all the rate estimates across all of the different methods and mitogenome partitions. Most of the other taxa show discrepancies between rate estimates from the narrower and wider sampling time windows, but a consistent time-dependent effect is seen only in the faster-evolving D-loop and 3rd codon sites. Our study shows the importance of taking into consideration that the different parts of the mitogenome evolve at different rates, especially when measuring the effect of time-dependence in rates using datasets from slower-evolving species on more recent timescales.

1. Introduction

One of the fundamental goals of biological research is to calculate rates of molecular evolution. Direct analyses of nucleotide sequences have contributed greatly to this understanding [177]. One of the simplest but most powerful models of the evolutionary process is the molecular clock, which posits that evolutionary change occurs at an approximately constant rate over time [136][137]. The theory of neutral evolution provides the foundation for molecular dating, which makes use of the molecular clock to estimate the timing of evolutionary divergences across the tree of life.

The assumption that the molecular clock "ticks" regularly over time is a powerful one, but does not appear to be a general property of evolution. In fact, the application of molecular dating has frequently shown that molecular rates can vary across lineages [153]. This variation is often attributed to differences among the species being studied. For example, generation time has been found to a positive predictor of male mutation bias, and a negative predictor of substitution rates [354]. Even when measuring rates within the same species, there can be significant differences in the reported evolutionary rates estimated across different timescales being measured. For example, Ho and colleagues [157] analysed a range of avian and primate mitochondrial datasets and found that rate estimates based on recent calibration points were markedly higher than those calibrated by older nodes.

The hypothesis of time-dependent molecular evolution, or time-dependent rates [152][157][158][159] is an explanation for the inconstancy of the molecular clock. This is the hypothesis that there is a measurable transition from a high short-term mutation (the instantaneous rate at which nucleotide changes occur in the genome) rate to a low long-term substitution rate (the rate at which mutations are fixed in the population) [152]. Time-dependent

biases in molecular rate estimates can be seen in the negative relationship between calibration time and molecular rate. For example, rates of spontaneous mutation measured in human [355] and birds [151] (by counting mutations between generations) differ by an order of magnitude from rates inferred in phylogenetic studies of the same taxa [356][357].

A possible explanation for the time-dependence of rate estimates is incomplete purifying selection [157]. Specifically, slightly deleterious mutations are still observed on shorter timescales, inflating the estimated rate of change. In other words, these weakly deleterious mutations can persist for some time in the population despite their lower probability of fixation, leading to an apparent elevation of short-term evolutionary rates [358]. Over a longer timescales, these transient mutations are purged by purifying selection [159]. It has also been suggested that the time-dependent effect can be a result of samples that were drawn from population that was contracting in size [359][360].

These explanations for the time-dependent effect in molecular rates, however, have been contested, with some authors suggesting that the pattern can be caused by factors other than purifying selection (e.g. [361][362][363][364][365]). For example, time-dependent biases in rate estimation can arise from analyses of short sequences [366], the methods used to estimate the rates [364], rate variation among lineages [367], tree imbalance [368], phylo-temporal clustering (i.e. closely related samples having the same age) [351][369], and the presence of strong population structure [365].

Recent studies of time-dependent patterns in rate estimates have focused on bacteria [168][178] and rapidly evolving viruses [173][170][370][163], which can accumulate sequence divergence over years or even months. Most

recently, Membrebe *et al.* [163] developed a formal approach in a Bayesian statistical framework that draws direct inference from primarily virus data by using a flexible epoch model that allows rate variation through time. The authors' method was also tested on ancient DNA data from woolly mammoths, where there was a much weaker, but still present, time-dependent rate effect on estimates of node ages [163].

In recent years, there has been a wealth of ancient DNA from animal remains collected from archaeological sites of human settlements (the ages of which generally fall within the radiocarbon age limit of 50,000 years). This is especially true of remains of domesticated animals, of which the dog is the first domesticate (at least 15,000 years ago [371]).

Analysis of ancient DNA sequences for rate estimation has been mostly dominated by Bayesian phylogenetic methods [182]. BEAST is one of the most sophisticated tools for estimating time trees, with the ability to sample many possible histories to evaluate posterior distributions of evolutionary rates, divergence times, and many other parameters. BEAST also can implement many different complex phylogenetic and phylogeographic models [372][373]. However, such methods are limited by the length of time it takes for the program to sample trees: analyses can take days to weeks for moderately large datasets of only a few hundred sequences [374]. Bayesian phylogenetic analysis of even larger datasets may be computationally prohibitive [375].

There are much more rapid tools that have been released in recent years that can infer time-scaled phylogenies from sequence data [374][376][377][378][379]. These methods were primarily designed to handle large heterochronous viral sequence alignments [374][376][377] but can be utilised to study "measurably evolving" [380] animal populations with

ancient DNA. Ancient DNA data can be analysed using different rapid dating methods in addition to Bayesian inference to produce substitution rate estimates that can be compared with each other [360].

For molecular dating, the amount of evolutionary change is compared between homologous DNA sequences (i.e. mitochondrial genomes) in order to infer the time since their most recent common ancestor. Calibration of the molecular clock can be achieved by assigning a date to an internal node ("node-dating") or to the tips ("tip dating"). With node-dating, the date assigned to the internal node represents the time to the most recent common ancestor between lineages using information from the fossil record or a dated biogeographical event [182]. If the sequences were sampled at sufficiently different timepoints, then the sampling dates are the only temporal information required to obtain a rate estimate. However, "tip-dating" methods of calibrating the molecular clock for a phylogeny are only valid if there is sufficient temporal signal in the data being analysed. Temporal signal will be absent if the sampling period was too short for measurable evolutionary change to occur, or if evolutionary rates vary excessively between lineages [351][369][381][382]. In other words, the temporal signal in the data will only be present if there is sufficient genetic change between sampling times to reconstruct a statistical relationship between time and genetic divergence [383]. A consequence of not taking into account this time-dependent effect and using higher short-term rates is overestimation of the timeframe of recent evolutionary events [151].

There is evidence of a time-dependent pattern in evolutionary rates in many other taxa, including amniotes [160], primates and birds [157][161], woolly mammoth [162][163], insects [164][165], fish [166][167], and bacteria [168], but most notably in fast-evolving viruses [163][169][170][171][172][173][174]. For ancient animals, this effect is mainly seen in taxa whose samples come

from remains that are well preserved because they have been excavated from the cold and dry deposition conditions of permafrost. As a result, the archaeological material tends to be very old and come from extinct species and populations. With recent technological advances in extraction and sequencing methods, there has been a growing amount of ancient DNA retrieved from younger remains from warmer and wetter climates. For example, full mitogenomes and even nuclear genomes have been sequenced from ancient and historic pig, goat, and cattle remains from important Neolithic sites in Western Asia and Europe [384][217][385].

Is a time-dependent effect detectable in mitochondrial rate estimates from these different animal taxa? In this study, we take advantage of recent ancient DNA datasets recently made available, primarily from domesticated animals. We also test the data for temporal signal in different parts of the mitochondrial genome. We then use three different rapid dating methods and Bayesian inference to estimate the rate of change in seven different animal taxa across different time scales.

2. Materials and methods

2.1 Data collection and organisation

Seven taxa were chosen for this study based on recent availability of ancient mitochondrial sequence data: some datasets were only recently published and some are unpublished. As a result, these samples had not been used in studies of evolutionary rate estimation.

Modern and ancient and historic mitogenome sequences of 47 bison, 98 chicken, 466 dogs, 155 goats, 344 horses, 160 pigs, and 599 sheep were used

in this study (**Table 6.1**). The mitogenome sequences of 38 ancient and historic chickens, 152 modern and ancient canids, and 295 modern and ancient sheep are unpublished (**Supplement**). We included the bison data from [386], canid data from [284] and [283], goat data from [217], horse data from [221], and pig data from [384] because they are the most complete published time-series datasets of ancient full mitochondrial genomes of these respective species. Modern domestic pig samples for the pig dataset were randomly selected from [285], where all sequences were reported to have an average 10× coverage. All modern sheep and a number of modern horse and domestic dog full mitogenomes used were downloaded from Genbank, and the accession numbers are available in the **Supplement**.

The general criteria for inclusion in the datasets were as follows: all ancient samples with direct radiocarbon dates, ancient samples dated by stratigraphy, and samples with a strong association with an archaeological site. The mean age of each sample was used in our dating analyses, and any sample with a wide age uncertainty (e.g. >5,000 years) was excluded. All samples at least 3× coverage were included following visual inspection for blocks of unresolved nucleotide sequences. In order to cover a broader temporal range, for the pig and dog data, those samples with less than 1× average coverage were also included after stretches of unresolved sequences were manually removed. A scatterplot showing the ages of all the samples in the datasets is provided in **Fig. 6.1**.

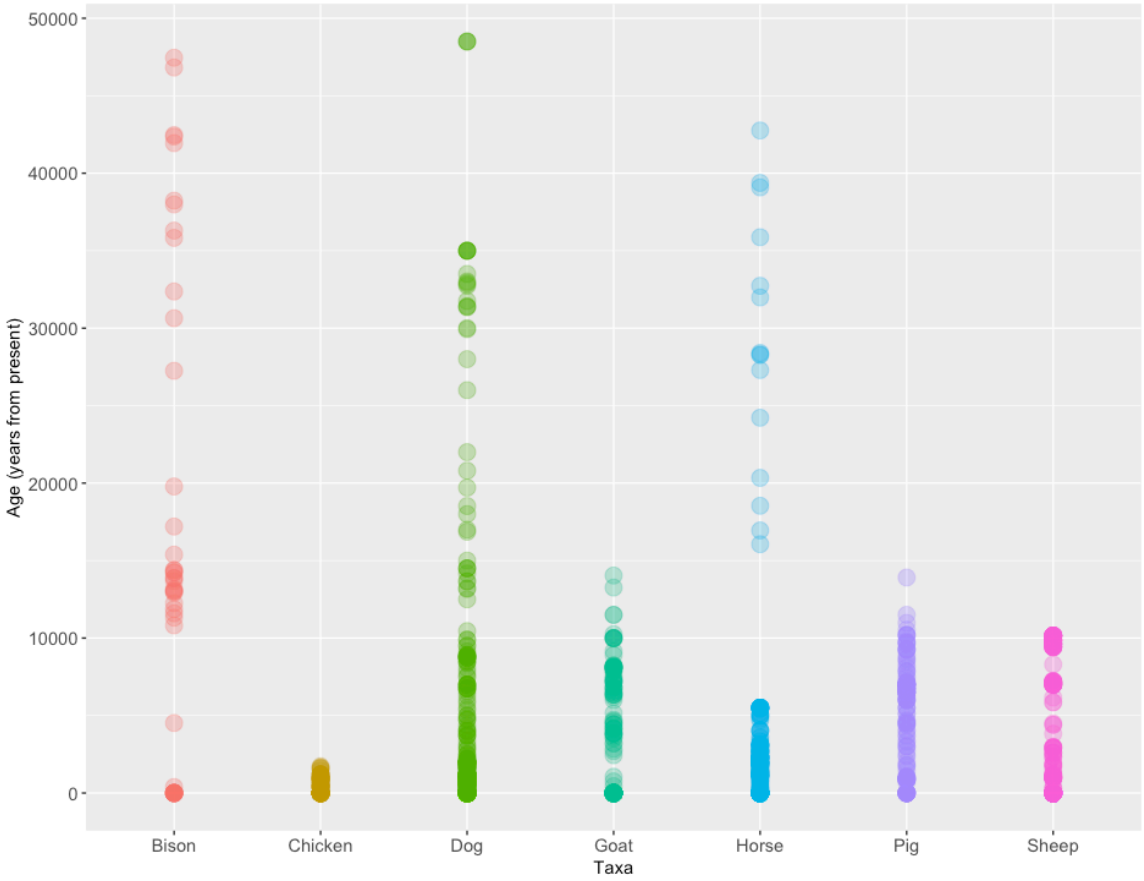


Figure 6.1 Plot showing the temporal distribution of all ancient and modern samples used for analysis

x-axis shows the seven different taxa analysed in this study, and the y-axis shows the mean ages (years from present) of the samples used. The transparency of the points reflect the abundance of samples at each timepoint.

For testing the hypothesis of time-dependent rates, the sequence data within each taxon was organised into two temporal sampling windows by subsampling. In previous studies, temporal subsamples were produced by the removal of progressively older sequences from the complete dataset, with an associated decrease in the number of sequences [387]. Additionally, a study that used simulated datasets suggest that rate overestimation could be due to narrow calibration windows being uninformative, resulting in a large

uncertainty and upward bias in rate estimates [351]. We therefore decided to use only two different temporal sampling windows per taxon dataset to increase the power of our analyses. For example, in a given dataset, the sampling windows of 0–5,000 years and 0–7,000 years might only have two ancient samples representing the difference of 2,000 years between the two timeframes. We decided to instead make each dataset within the sampling windows dependent on sample number: the wider sampling window t2 has 100% of all ancient and modern samples, whereas the narrower sampling window t1 has all of the modern sequences and approximately 1/3rd of the youngest ancient and historic sequences (**Table 6.1**).

Table 6.1. Number of samples in each taxon per each sampling window (in years)

Species	Sampling window (years)	*n*	*n**	Number of ancient & modern samples	Full mitogenome sequence length (bp)	Generation time (years)
Bison priscus* & *Bison bison	0-14,500 (t1)	30	30	37 / 10	16,328	8 [388]
	0-125,000 (t2)	47	47			
Gallus gallus domesticus	0-900 (t1)	72	72	51 / 47	16,775	1 [389]
	0-1,800 (t2)	98	98			
Canis lupus* & *Canis lupus familiaris	0-2,100 (t1)	361	361	302 / 164	16,727	4.4 [390]
	0-33,000 (t2)	466	466			
Capra hircus	0-7,000 (t1)	115	115	82 / 77	16,643	2.5 [217]
	0-15,000 (t2)	155	155			
Equus caballus	0-2,300 (t1)	233	233	221 / 123	16,451	8 [331]
	0-43,000 (t2)	344	344			
Sus scrofa	0-6,500 (t1)	96	96	127 / 33	16,613	5 [298]
	0-14,000 (t2)	160	160			
Ovis aries	0-7,100 (t1)	524	125	150 (50) / 449 (150)	16,315	3.6 [391]
	0-10,150 (t2)	599	200			

Note: *n* indicates datasets used in rapid-dating analyses; *n** indicates datasets used in BEAST2 analyses; parentheses indicate subsampled datasets for BEAST2 analyses

Mitogenome sequence alignments were performed using MUSCLE integrated into Aliview [392]. The aligned sequences were then partitioned using in-house scripts. The partitions were defined using the NCBI annotation (chicken: accession # X52392; dog: accession # NC_002008.4; goat: accession # NC_005044; pig: accession # NC_000845.1) for tRNAs, rRNAs, the 1st codon sites, 2nd codon sites, and 3rd codon sites, and D-loop. The partitioned bison sequences (RNAs, 1st+2nd codon sites, 3rd codon position, and D-loop) were obtained from [386], and the partitioned horse sequences from Prof. Orlando (Universite de Toulouse). Both published horse sequences from [221] and unpublished data were obtained using the same methodology described in [221]. Many dog and pig mitogenome sequences had up to 10% unresolved nucleotides due to DNA degradation, so the sequences of these taxa were only partitioned into all RNAs, 1st+2nd codon sites, 3rd codon sites, and D-loop. After partitioning, large gaps in the alignments were removed with Gblocks 0.91b [393], using default parameters, prior to further analysis.

2.2 Date-randomisation tests

In order to validate the estimates of rates by assessing the extent of temporal signal in the datasets, we chose to implement the date-randomisation test [382][394]. In this test, the ages of the samples are randomly reassigned to the full mitogenome sequences and the analysis of the data is repeated to generate a set of rate estimates from the date-randomised datasets. The random assignment of sampling times to the sequences effectively breaks the association between the nucleotide substitution and time [351]. The expectation is that the random assignment of dates to the sequences will lead to rate estimates that are considerably different from those of the original and correct sampling times if the original data are temporally structured

[351]. This method of randomisation was performed on the bison, dog, goat, horse, and pig datasets.

The chicken and sheep datasets only had a few distinct sample ages (that is, many of the aDNA sequences are the same age), a clustered version of date-randomisation [351][369] was performed instead. The cluster-randomisation approach involves shuffling only the distinct ages. For example, if a dataset contains 10 modern sequences (age 0 years) and 10 sequences of age 50,000 years, then there is only a single alternative configuration of ages: swapping the ages of the two sets of sequences so that the modern sequences are assigned an age of 50,000 years and the ancient sequences are assigned an age of 0 years. This is because unless there are many distinct ages, it is not possible to generate 20 distinct permutations of the dates.

We used two criteria that have been proposed for the date-randomisation test, referred to as CR1 and CR2 [351][360]. For CR1, the rate estimate from the original dataset should not fall within the 95% confidence (or in the Bayesian context, credibility) interval of the rate estimates from the date-randomised replicates [382]. The conditions of CR2 are more stringent: the 95% confidence/credibility interval of the original rate estimate should not overlap with those of any of the date-randomised datasets [394][395]. Because of time constraints, our results are based on only five date-randomised replicates of the original mitogenome datasets for each taxon and each method. The date-randomisation tests using both CR1 and CR2 criteria were performed using Treedater, Treetime, and BEAST2.

2.3 Rapid dating methods

Individual maximum-likelihood trees for the full mitogenome and the different partitions were inferred using RAxML v8.2.12 [396], using the GTR+Γ substitution model. In each case, a rapid bootstrapping analysis with 100 replicates was followed by a search for the best-scoring tree, which was used for further analysis. Outgroup taxa (yak as outgroup to bison; red jungle fowl as outgroup to chickens; coyote as outgroup to dogs; Nubian ibex as outgroup to goats; Somali wild ass as outgroup to horses; warthog as an outgroup to pigs; goat as outgroup to sheep) were used to root each tree. The outgroup was then pruned from the tree using ape [397] prior to rate estimation.

Sequences were then analysed using a likelihood-based Gamma-Poisson mixture model in Treedater [377], an approximate maximum-likelihood approach in Treetime [374], and a root-to-tip dating method in TempEST [383]. Evolutionary rates for all methods were estimated using the strict molecular clock. Default parameters were applied for all the analyses using Treedater and Treetime (with the flag --keep-root).

Treedater, an R-based package, is a fast dating method that can use a Gamma-Poisson mixture model of substitution rates to estimate the substitution rate for each branch while being scalable to large phylogenies. Treedater can estimate unknown sample dates, the root of the tree, confidence intervals for rates, dates, and tip dates by using both parametric and non-parametric bootstrap approaches. The program offers flexibility in performing analyses – the user can specify strict or relaxed molecular clock models, with or without temporal constraints on nodes, or rooted trees. There are also options for detecting outlier lineages, root-to-tip regression, and to perform a statistical test for selecting molecular clock models [377].

Treetime is another fast dating method originally developed to analyse large heterochronous sequence alignments in the context of real-time phylogenetics pipelines of viral epidemics [398]. Its algorithm is based on iterative optimisation of simpler subproblems, similar to sequential quadratic programming or expectation maximisation, to arrive at a consistent approximation of the global optimum. The iteration proceeds at multiple levels, such as iterating optimisation of branch lengths, ancestral sequences, relaxed clock parameters, and coalescent models. The run time of the algorithm is linear to the size of the dataset. The Treetime strategy is approximate and assumes short branch lengths, but is tractable for trees with thousands of tips [374].

TempEST is a simple method that can be used to visualise temporally sampled sequence data. It works as a simple diagnostic for molecular clock models by performing a regression of root-to-tip genetic distance against sampling time. For example, a linear trend with small residuals from the regression line suggests that a strict molecular clock will adequately represent the molecular evolution of that data set. The absence of a linear relationship suggests that the dataset contains little temporal signal, whereas a strong non-linear trend indicates that the evolutionary rate has changed systematically through time. TempEST can also be used to provide quality control of the dataset [383].

Least squares dating (LSD) is another rapid maximum-likelihood-based method [376] and has been employed to analyse ancient data [360]. However, we have chosen not to use LSD in this study because of its reported poorer sensitivity when handling larger datasets, compared with Treetime and BEAST [374].

Correlation analyses comparing the depth of the time windows (the upper boundaries of t1 and t2) and the estimated rates were performed using the Spearman correlation formula. This non-parametric test was selected after the Shapiro-Wilk test for normality was performed and we found that the data was not normally distributed. All statistical analyses were performed with R (version 3.3.1) in the integrated development environment (IDE) "Rstudio" (version 1.1.383).

2.4 Bayesian inference

The number of sheep mitogenomes was too large to allow efficient analysis, so these datasets were subsampled to 200 mitogenomes total. Outgroup sequences were excluded in order to minimise violation of the assumptions of the coalescent tree prior (which assumes a panmictic population). Evolutionary rates were estimated for both full and subsampled datasets by using the strict molecular clock in BEAST v2.4.7 [399]. For the prior on clock rates, we chose a lognormal distribution with a mean in real space of $1.0 \cdot 10^{-8}$, upper bound of $1.0 \cdot 10^{-4}$ substitutions/site/year, lower bound of $1.0 \cdot 10^{-10}$ substitutions/site/year (these bounds are actually part of a separate uniform prior and are not part of the lognormal distribution itself), and standard deviation of 1.25. These bounds were informed by the rate estimates from the rapid dating methods. The HKY+Γ substitution model was used, with four rate categories for gamma-distributed rates across sites. We used an exponential prior for kappa and a lognormal prior for the gamma shape prior, with default parameters for both.

Although Bayesian inference allows the incorporation of age or date uncertainty in the prior distribution, we did not do this even though not all of

the samples were radiocarbon dated. We used the mean date estimate so that the results would be comparable across all of the different methods, as the likelihood-based methods do not allow for age uncertainty. Moreover, it has been shown [400] that accounting for age uncertainty had negligible or minimal impacts on the resulting estimates in BEAST. A constant-size coalescent model was used as the tree prior for the chicken, dog, goat, and sheep datasets, and a Bayesian skyline population model was used for the bison, horse, and pig datasets, as in previous studies [386][221]. Posterior distributions of parameters were estimated by Markov chain Monte Carlo (MCMC) sampling. Two independent runs of chains were performed, and samples were drawn every 5,000 steps over a total of at least 50 million steps, with the first 20% of samples discarded as burn-in. For all of the datasets, we considered that sampling was sufficient when the effective sample size of each parameter exceeded 100. When required, we ran additional MCMC analyses to achieve sufficient sampling. The trace files were assessed using Tracer [401] and the samples from the independent runs were merged using LogCombiner [372].

3. Results

3.1 Date-randomisation test results

For the Treedater estimates, the datasets from both sampling time windows passed the date-randomisation test 11 out of 14 times according to CR1, and 8 out of 14 times according to CR2. For the Treetime estimates, the datasets from both sampling time windows passed the date-randomisation test 12 out of 14 times according to both CR1 and CR2 (**Table 6.2b**). For both Treedater and Treetime, both sampling time windows of the chicken datasets failed the date-randomisation tests according to CR1 and CR2, and for the BEAST2

analysis, only the full sampling time window (t2) failed the date-randomisation test, whereas the smaller sampling time window (t1) passed. As for the rest of the BEAST2 analyses, only the bison, goat, and horse t1 and t2 datasets passed both CR1 and CR2 (**Table 6.2**).

Date-randomisation tests were not performed using TempEST because the original rate estimates had been reported as zero for at least one of the sampling time windows in 4 out of 7 taxa, indicating a lack of temporal signal in the data (**Fig 6.1d**).

Table 6.2 Date-randomisation test results for full mitogenomes of 7 taxa

Taxa	Sampling time window	Treedater CR1	Treedater CR2	Treetime CR1	Treetime CR2	BEAST2 CR1	BEAST2 CR2
Bison	t1	Pass	Pass	Pass	Pass	Pass	Pass
	t2	Pass	Fail	Pass	Pass	Pass	Pass
Chickens	t1	Fail	Fail	Fail§	Fail§	Pass	Pass
	t2	Fail	Fail	Fail	Fail	Fail	Fail
Dogs	t1	Pass	Pass	Pass§	Pass§	NA	NA
	t2	Pass	Pass	Pass§	Pass§	NA	NA
Goats	t1	Pass	Pass	Pass	Pass	Pass	Pass
	t2	Pass	Fail	Pass	Pass	Pass	Pass
Horses	t1	Pass	Pass	Pass	Pass	Pass	Pass
	t2	Pass	Pass	Pass	Pass	Pass	Pass

Pigs	t1	Fail	Fail	Pass§	Pass§	Pass	Pass
	t2	Pass	Fail	Pass§	Pass§	NA	NA
Sheep	t1	Pass	Pass	Pass	Pass	Fail	Fail
	t2	Pass	Pass	Pass§	Pass§	NA	NA

Date-randomisation test results of 5 replicates performed using Treedater, Treetime, and BEAST2. The CR1 criterion requires that the rate estimate from the original dataset should not fall within the 95% confidence (or credibility) interval of the rate estimates from the date-randomised replicates [382]. CR2 is a more stringent criterion requiring that the 95% confidence/credibility interval of the original rate estimate should not overlap with those of the date-randomised datasets [394][395]. Cells marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.
§ At least one of the date-randomised replicates produced a negative rate estimate

---

3.2 Difference in rate estimates between the sampling windows

The difference in rate estimates between datasets from the two sampling windows (t2 and t1) is an indication of time-dependent estimates of evolutionary rates. The narrower sampling window t1 is expected to produce rate estimates that are higher than t2, the wider sampling window. **Fig. 6.9a-b** shows a summary of rate estimates of the D-loop and full mitogenome, respectively, generated by Treedater for all seven taxa within seven different sampling windows, scaled by the generation time for each taxon (**Table 6.1**). Although the bison dataset has the largest time window of 125,000 years (scaled by generation time in **Fig. 6.9**), the difference in rate estimate between t2 ($7.42 \cdot 10^{-7}$ s/s/y [$5.91 \cdot 10^{-7}$—$7.42 \cdot 10^{-7}$]; 0-125,000 years) to t1 ($2.29 \cdot 10^{-6}$ s/s/y [$1.54 \cdot 10^{-6}$ —$3.41 \cdot 10^{-6}$]; 0-14,500 years) is less dramatic than the rate estimates of the horse sampling time windows t1 ($1.35 \cdot 10^{-5}$ s/s/y [$9.56 \cdot 10^{-6}$—$1.90 \cdot 10^{-5}$]; 0-2,300 years) and t2 ($1.59 \cdot 10^{-7}$s/s/y

[$2.85 \cdot 10^{-8}$—$8.91 \cdot 10^{-7}$]; 0-43,000 years).

When comparing the results of **Fig 6.9a** to **Fig. 6.9b** (a summary of rate estimates from the full mitogenome), striking differences also can be observed. There is a considerable difference in the scale of the rate estimates, where the rate of the D-loop is two orders of magnitude greater than that of the full mitogenome. Another view of the pattern can be seen when comparing **Fig. 6.3b** to **Fig. 6.2b**.

3.3 Difference in rate estimates among partitions and methods

The rate estimates differed depending on the part of the mitogenome being analysed and on the choice of dating method (**Fig 6.10**). The bison is the only species for which the time-dependent effect can be seen in all of the mitochondrial partitions, for all four methods used. The results of the other taxa are more variable, but patterns can be seen in their results. These discrepancies can be seen in **Figure 6.10** for all of the taxa except for the bison. The strength of time-dependency does not appear to be correlated with the length of the time period over which samples were collected (**Fig. 6.9a-b**, **Fig. 6.11**). This observation is most striking between the bison and chicken, where the difference between the t1 and t2 sampling windows is only 900 years for the chicken dataset, and 110,500 years for the bison (**Table 6.1**). Yet there is only 0.04% difference between the rho statistics of both taxa (rho = -0.38 for the bison; rho = -0.34 for the chicken) **(Fig. 6.11)**. Indeed, the strongest regression of the sampling windows against rate is seen in the horse dataset (rho = -0.58), where the difference between the t1 and t2 sampling windows is 40,700 years (**Table 6.1** and **Fig. 6.11**).

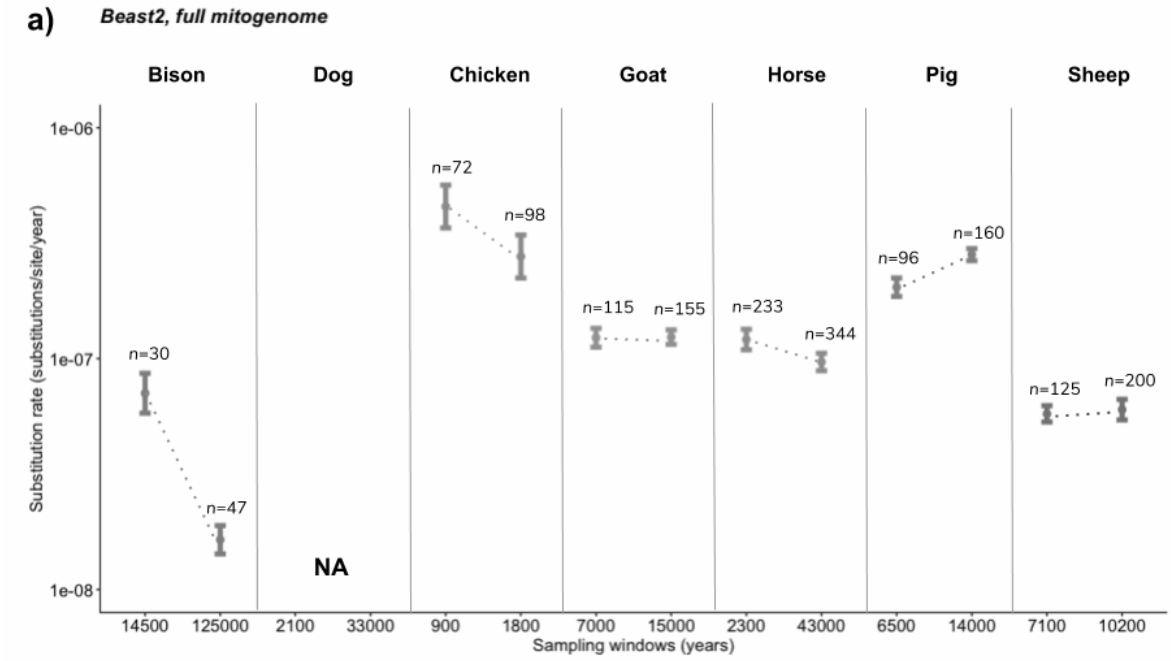
Although the time-dependent effect seems to vary in the full mitogenome of all the taxa surveyed, it can be seen in other partitions where the mutation rate is expected to be higher. The D-loop and 3rd codon sites are generally the only partitions that show a time-dependent rate effect, particularly in datasets where the temporal signal seems to be poor, such as those from the chicken, goat, and sheep. Notably, although the chicken datasets mostly failed the date-randomisation test (with the exception of t1 rates estimated by BEAST2; **Table 6.2**), the t2 estimates < t1 estimates for the full mitogenome, D-loop, 3rd codon sites, and the rRNA alignments for all methods except for TempEST (**Fig 6.2-6.8a,b,d** and **Fig 6.10**). The goat dataset is likely to have a temporal signal, since it passed most of the date-randomisation tests (**Table 6.2**), and it does show the time-dependent effect in the D-loop (using Treetime) and 3rd codon sites (Treetime and Treedater) alignments, but no time-dependency can be seen in any of the other partitions (**Fig. 6.2b,d** and **Fig 6.10**). A similar pattern is also seen in the sheep dataset, where only the D-loop and 3rd codon sites (as well as the rRNA) have a time-dependent effect in the rate estimates generated by Treedater and Treetime (**Fig. 6.3b,d, Fig. 6.4b,d; Fig 6.7b,d; Fig 6.10**). The sheep full mitogenome dataset does seem to have a stronger temporal signal than that of the goat, however (**Fig 6.2b,d** and **Fig 6.10**).

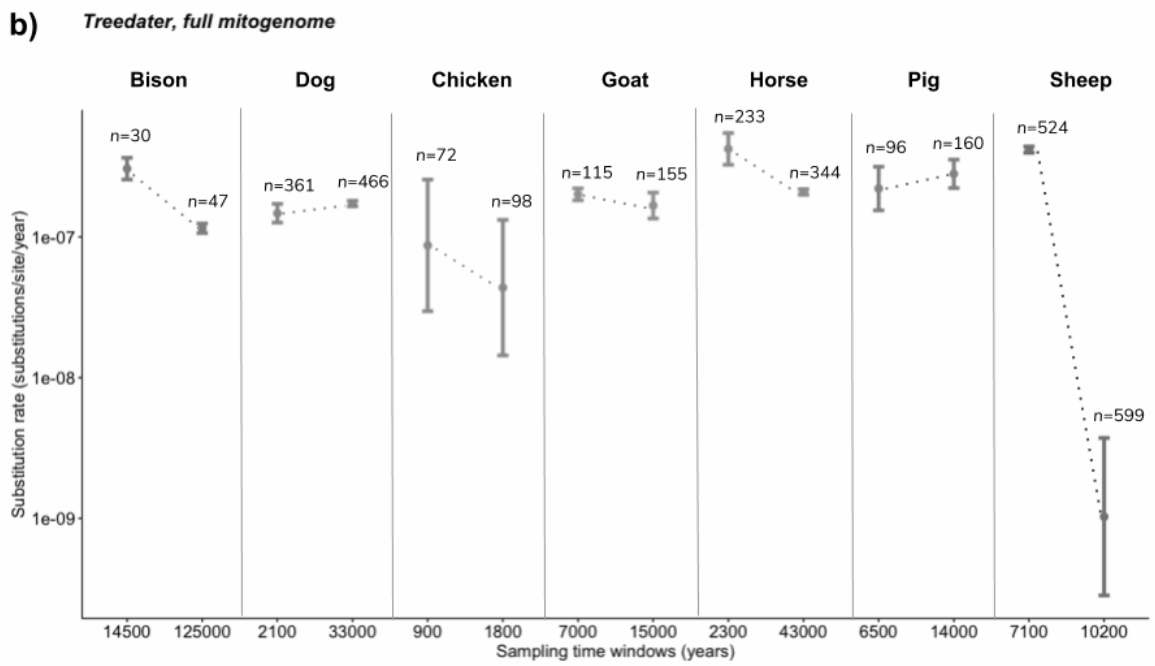
Many of the rate results from Treetime and TempEST, in contrast with those from BEAST2 and Treedater, were not informative because the rate estimates were negative, indicating a lack of temporal signal. Negative rate estimates are reported in the figures as asterisks (*) **(Fig. 6.2c,d-6.8c,d)**.

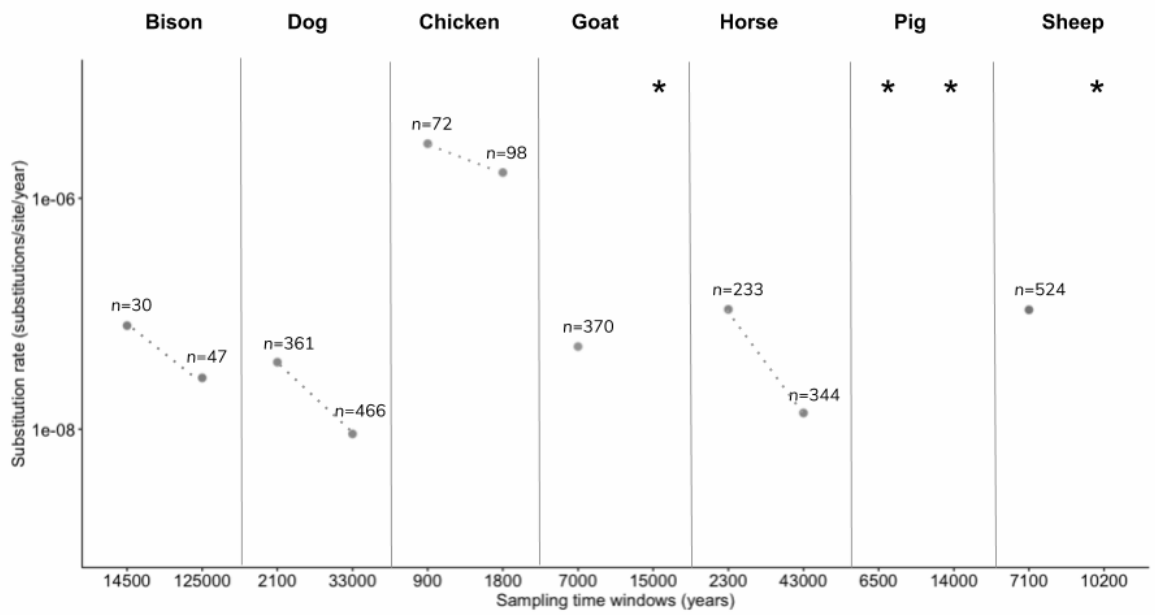
The difference in rates estimated by different methods is also seen by the example of the rate estimate of the full mitogenome of the sheep t2 in Treedater **(Fig 6.2b)**, seen to be surprisingly low at $1.03 \cdot 10^{-9}$ s/s/y ($1.10 \cdot 10^{-10}$ s/s/y—$9.57 \cdot 10^{-9}$ s/s/y). This contrasts sharply with the rate estimates for the

same partition generated by BEAST2 at $5.80 \cdot 10^{-8}$ s/s/y ($5.00 \cdot 10^{-8}$ s/s/y — $6.62 \cdot 10^{-8}$ s/s/y) **(Fig 6.2a)** and Treetime at $3.59 \cdot 10^{-7}$ s/s/y ($3.13 \cdot 10^{-7}$ s/s/y — $4.06 \cdot 10^{-7}$ s/s/y) **(Fig 6.2d)**, with Treedater giving a median rate estimate differing by up to two orders of magnitude **(Fig 6.2b)**.

Also interesting is that the horse full mitogenome and rRNA has a time-dependent effect for rates estimated by all four of the different statistical methods, but the effect is not seen in any of the other partitions, including the faster-evolving D-loop and 3rd codon sites (**Fig. 6.10**).




b)
Treedater, full mitogenome
Bison
Dog
Chicken
Goat
Horse
Pig
Sheep
n=30
n=47
n=361
n=466
n=72
n=98
n=115
n=155
n=233
n=344
n=96
n=160
n=524
n=599
Substitution rate (substitutions/site/year)
1e-07
1e-08
1e-09
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
Sampling time windows (years)


c)
Tempest, full mitogenome
Bison
Dog
Chicken
Goat
Horse
Pig
Sheep
*
*
*
*
n=72
n=98
n=30
n=47
n=361
n=466
n=370
n=233
n=344
n=524
Substitution rate (substitutions/site/year)
1e-06
1e-08
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
Sampling time windows (years)

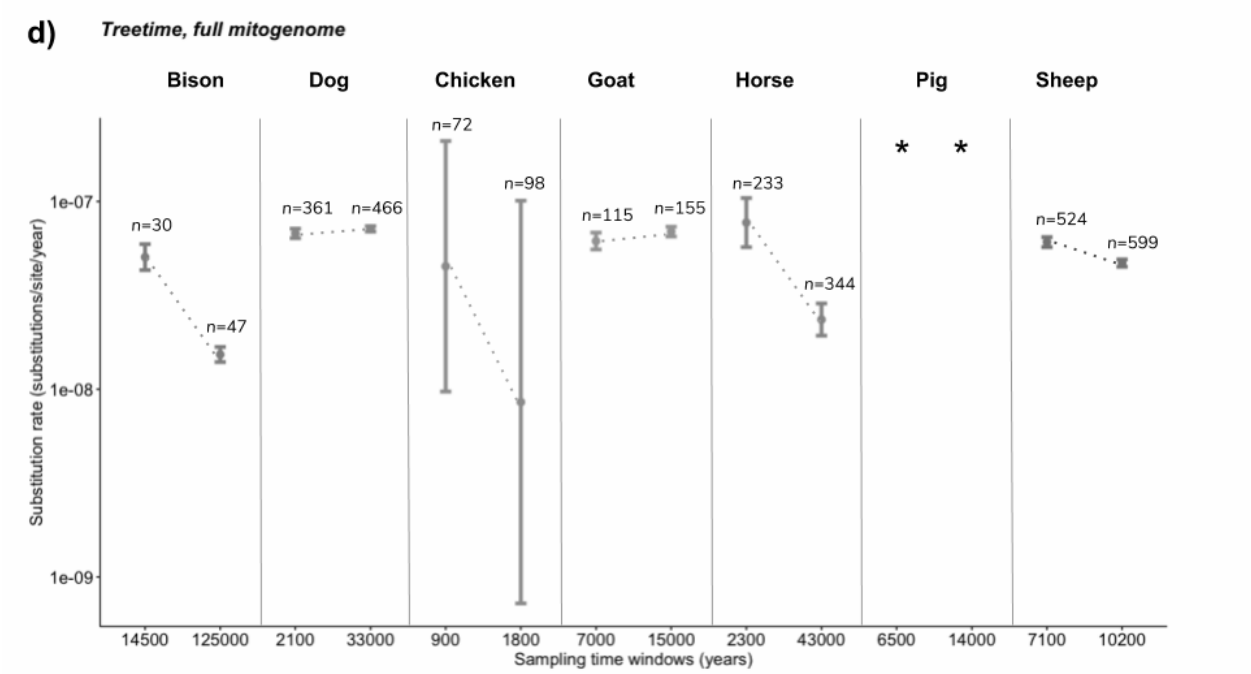
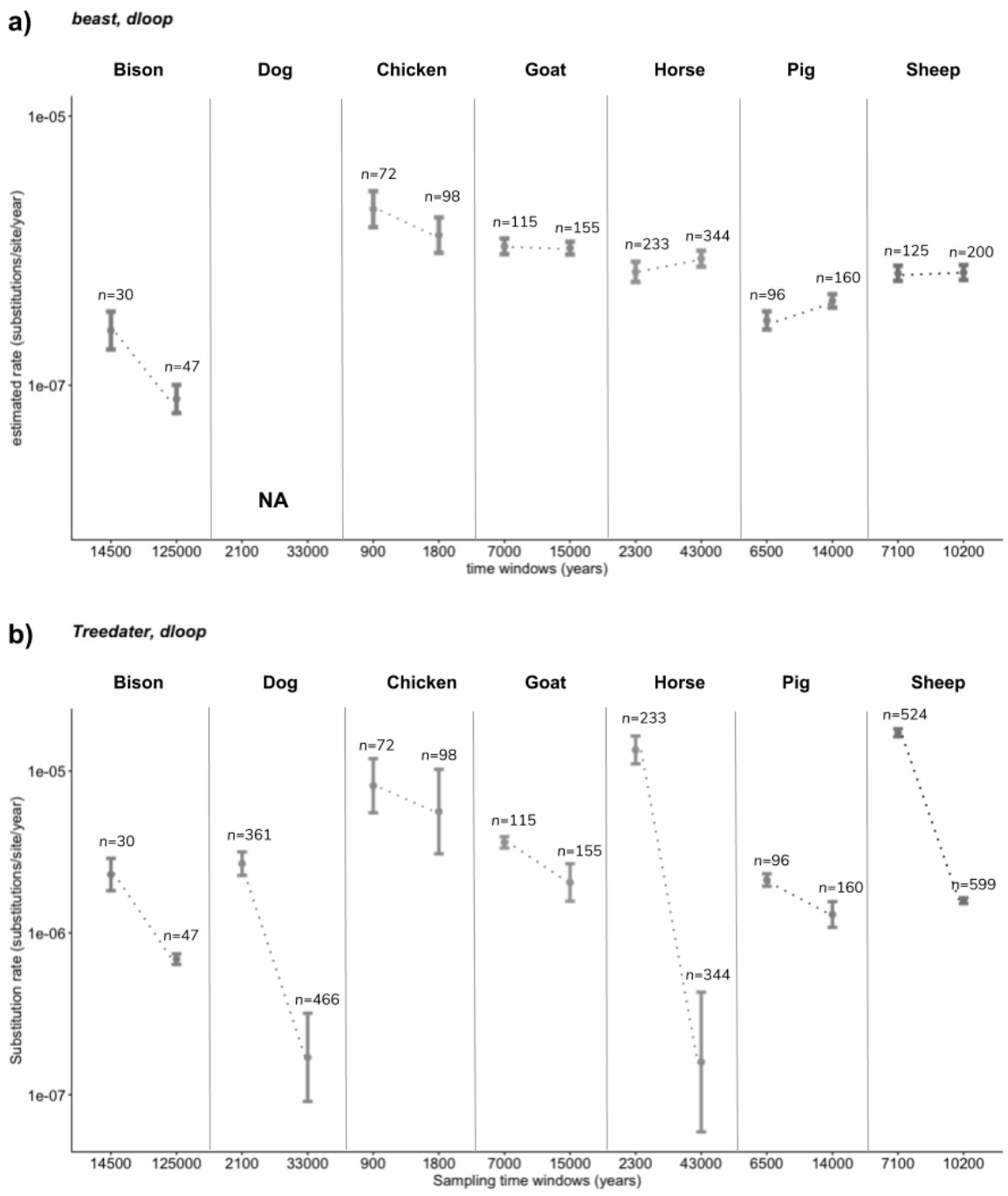


Figure 6.2 Substitution rate estimates for full mitogenomes from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicate the number of samples within a sampling time window. Dotted lines join the mean rate estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.


a)
beast, dloop
Bison
Dog
Chicken
Goat
Horse
Pig
Sheep
1e-05
1e-07
estimated rate (substitutions/site/year)
n=30
n=47
NA
n=72
n=98
n=115
n=155
n=233
n=344
n=96
n=160
n=125
n=200
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
time windows (years)
b)
Treedater, dloop
Bison
Dog
Chicken
Goat
Horse
Pig
Sheep
1e-05
1e-06
1e-07
Substitution rate (substitutions/site/year)
n=30
n=47
n=361
n=466
n=72
n=98
n=115
n=155
n=233
n=344
n=96
n=160
n=524
n=599
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
Sampling time windows (years)

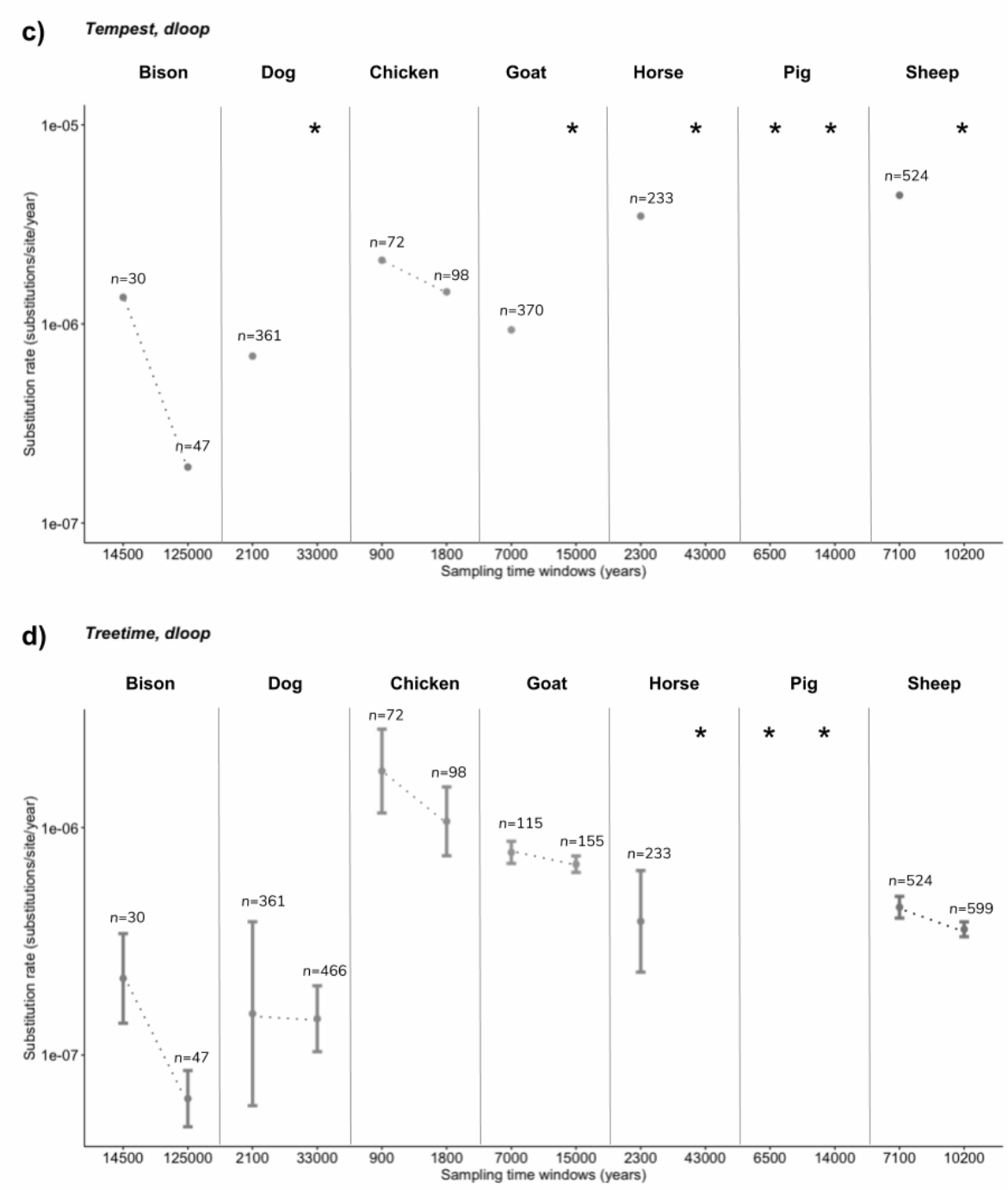
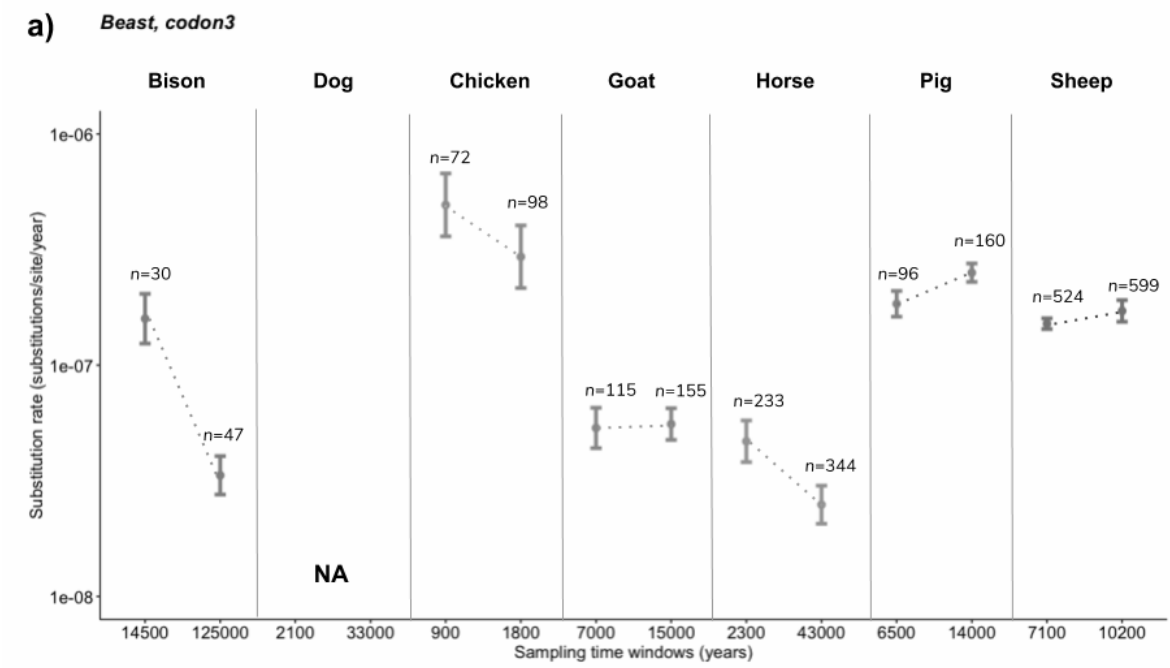


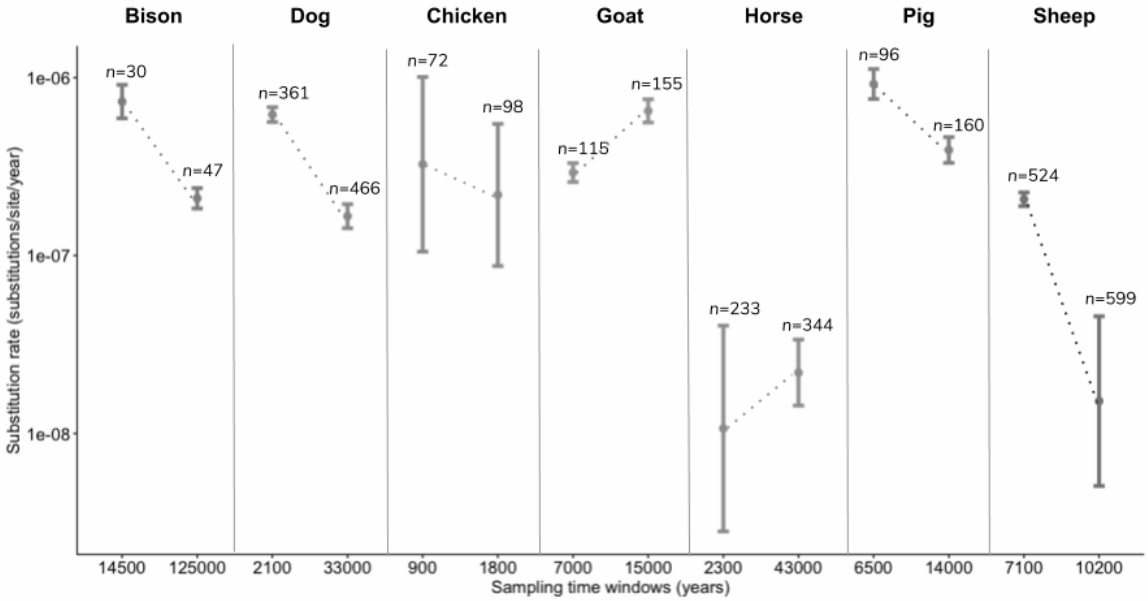
Figure 6.3 Substitution rate estimates for D-loop generated from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicate the number of

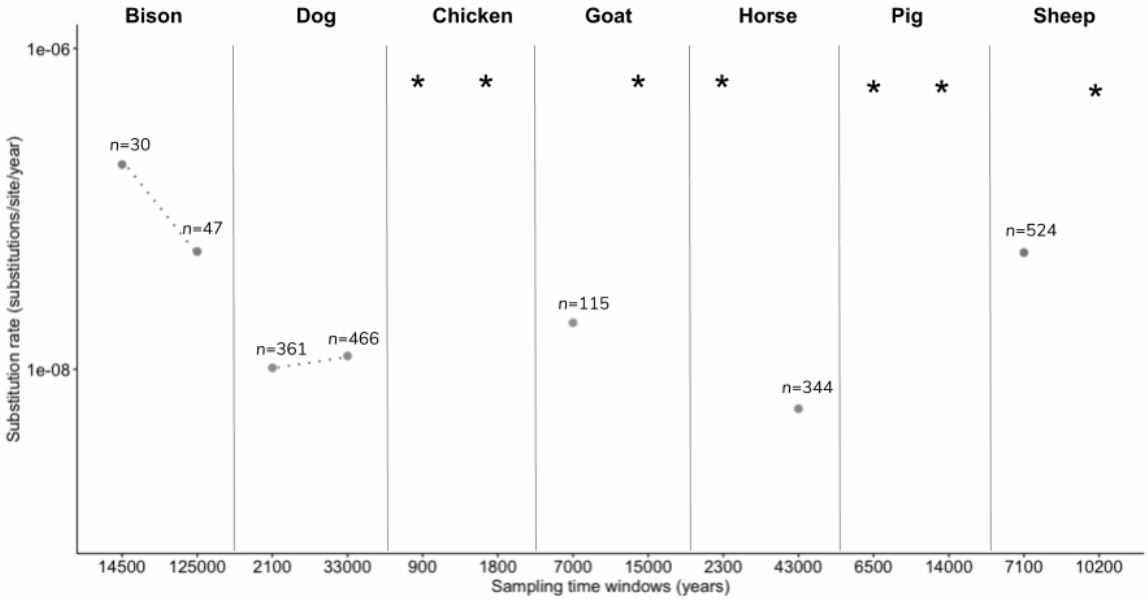
samples within a sampling time window. Dotted lines join the mean estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.



b) *Treedater, codon3*



c) *Tempest, codon3*



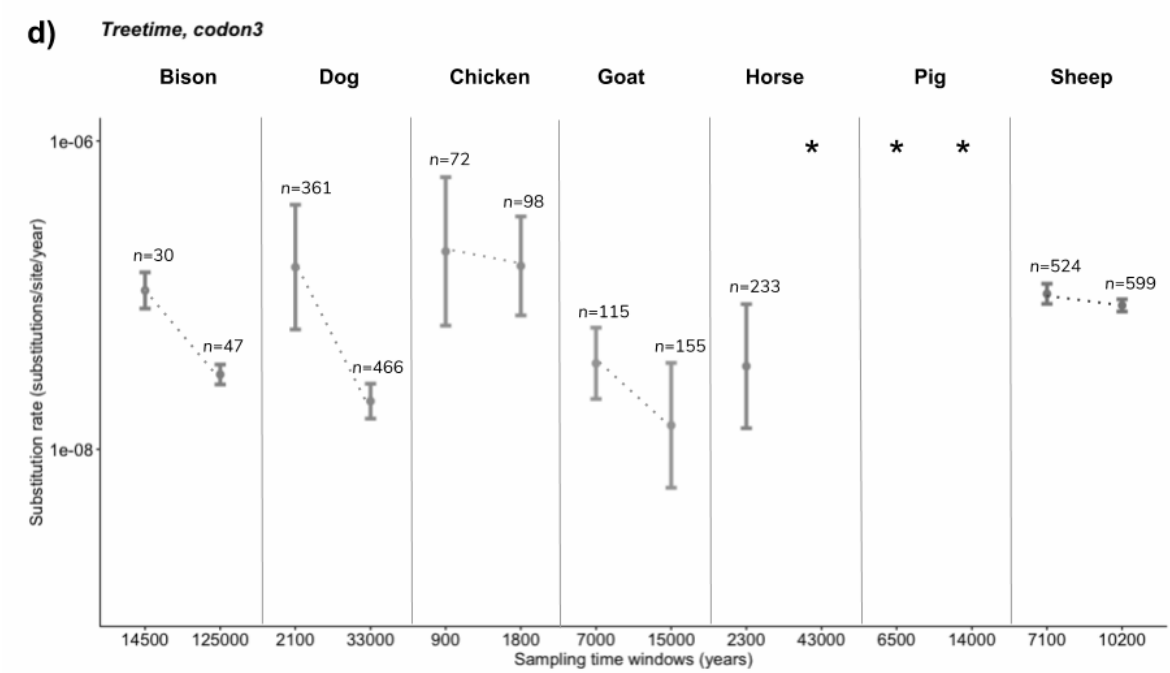
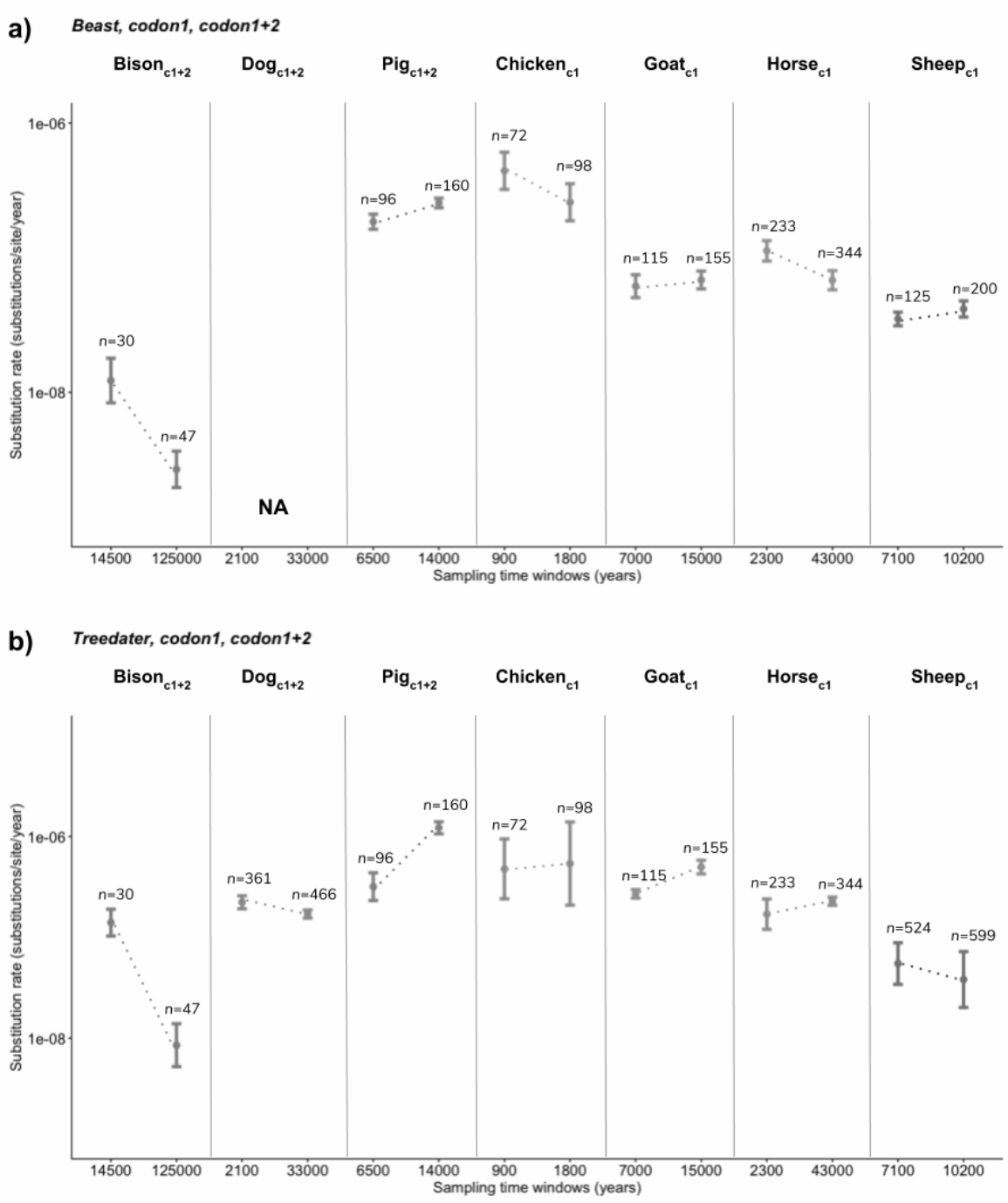


Figure 6.4 Substitution rate estimates for 1st codon sites (chicken, goat, horse) or 1st+2nd codon sites (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicates the number of samples within a sampling time window. Dotted lines join the mean estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.


a)
Beast, codon1, codon1+2
Bison c1+2
Dog c1+2
Pig c1+2
Chicken c1
Goat c1
Horse c1
Sheep c1
Substitution rate (substitutions/site/year)
1e-06
1e-08
n=30
n=47
NA
n=96
n=160
n=72
n=98
n=115
n=155
n=233
n=344
n=125
n=200
14500
125000
2100
33000
6500
14000
900
1800
7000
15000
2300
43000
7100
10200
Sampling time windows (years)
b)
Treedater, codon1, codon1+2
Bison c1+2
Dog c1+2
Pig c1+2
Chicken c1
Goat c1
Horse c1
Sheep c1
Substitution rate (substitutions/site/year)
1e-06
1e-08
n=30
n=47
n=361
n=466
n=96
n=160
n=72
n=98
n=115
n=155
n=233
n=344
n=524
n=599
14500
125000
2100
33000
6500
14000
900
1800
7000
15000
2300
43000
7100
10200
Sampling time windows (years)

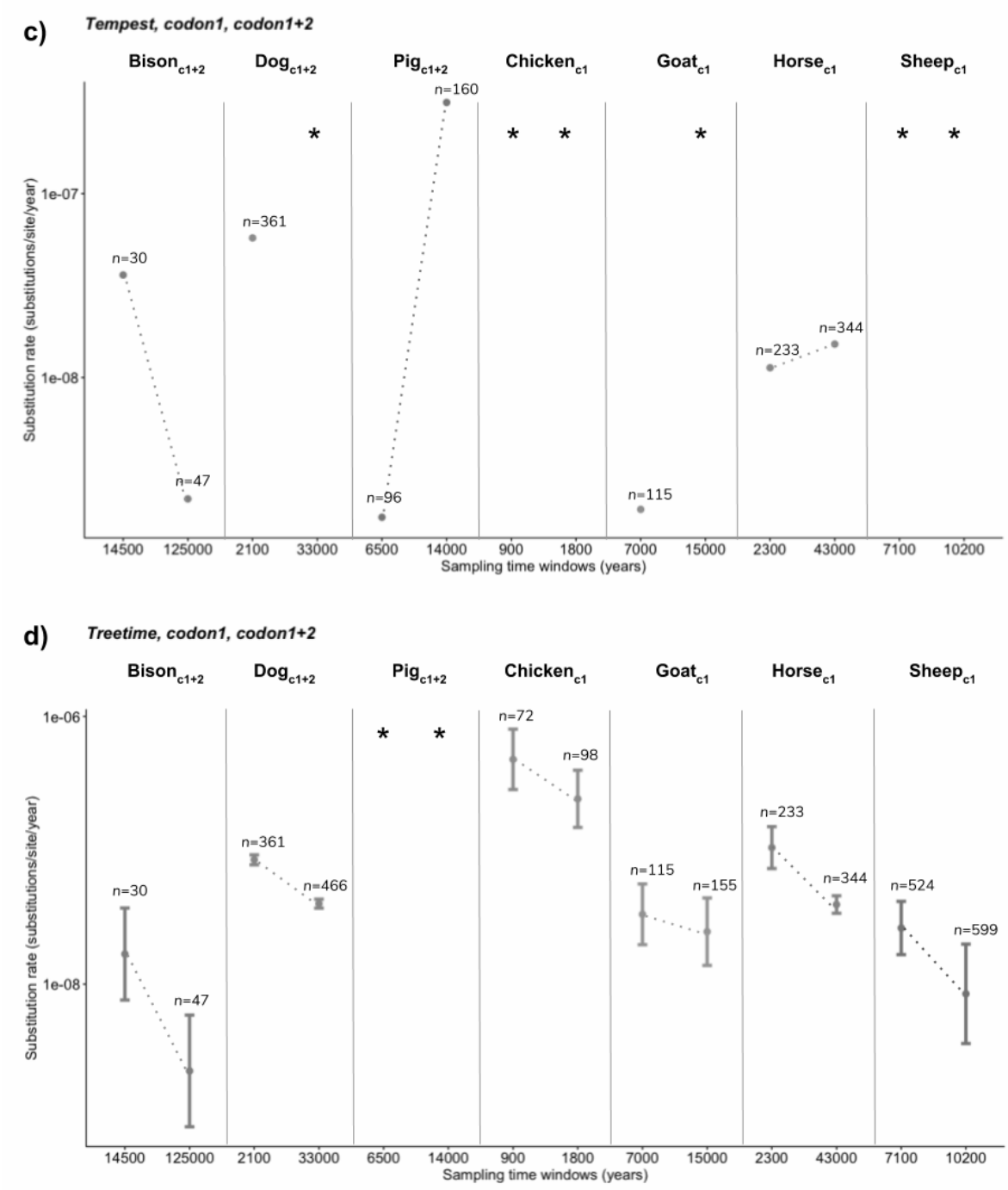
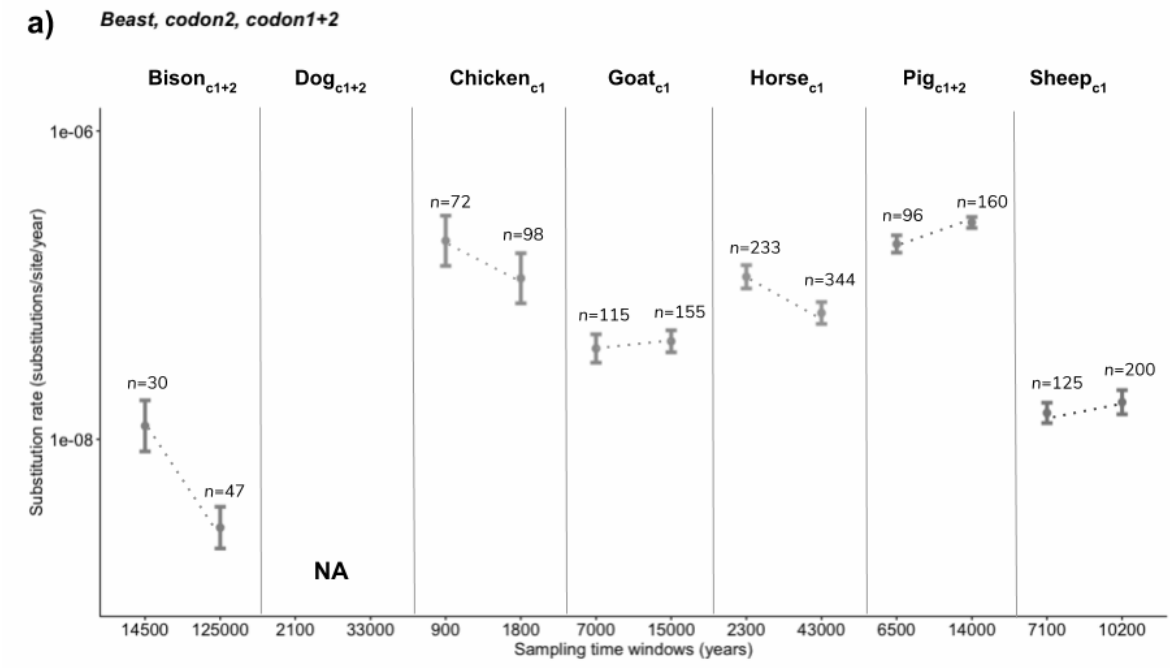
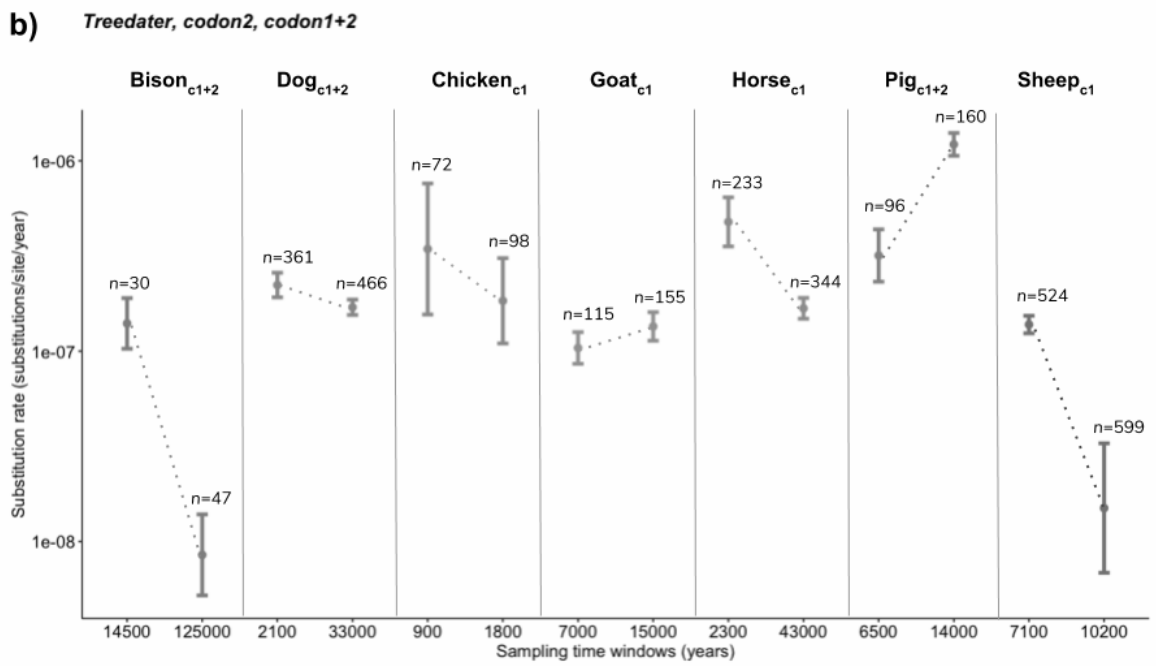


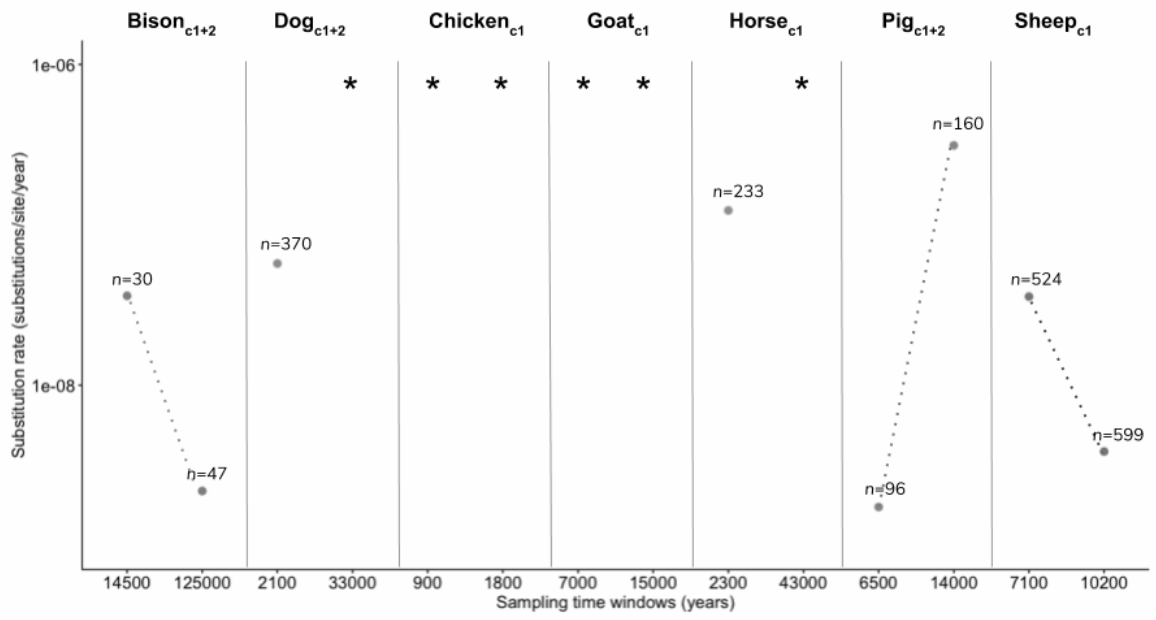
Figure 6.5 Substitution rate estimates from 2nd codon sites (chicken, goat, horse) or 1st+2nd codon sites (bison, dog, pig, sheep) from seven taxa with seven different samping time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each

taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicates the number of samples within a sampling time window. Dotted lines join the mean estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.




b)
Treedater, codon2, codon1+2
Bison c1+2
Dog c1+2
Chicken c1
Goat c1
Horse c1
Pig c1+2
Sheep c1
Substitution rate (substitutions/site/year)
1e-06
1e-07
1e-08
n=30
n=47
n=361
n=466
n=72
n=98
n=115
n=155
n=233
n=344
n=96
n=160
n=524
n=599
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
Sampling time windows (years)


c)
Tempest, codon2, codon1+2
Bison c1+2
Dog c1+2
Chicken c1
Goat c1
Horse c1
Pig c1+2
Sheep c1
Substitution rate (substitutions/site/year)
1e-06
1e-08
*
n=30
n=47
n=370
n=233
n=160
n=96
n=524
n=599
14500
125000
2100
33000
900
1800
7000
15000
2300
43000
6500
14000
7100
10200
Sampling time windows (years)

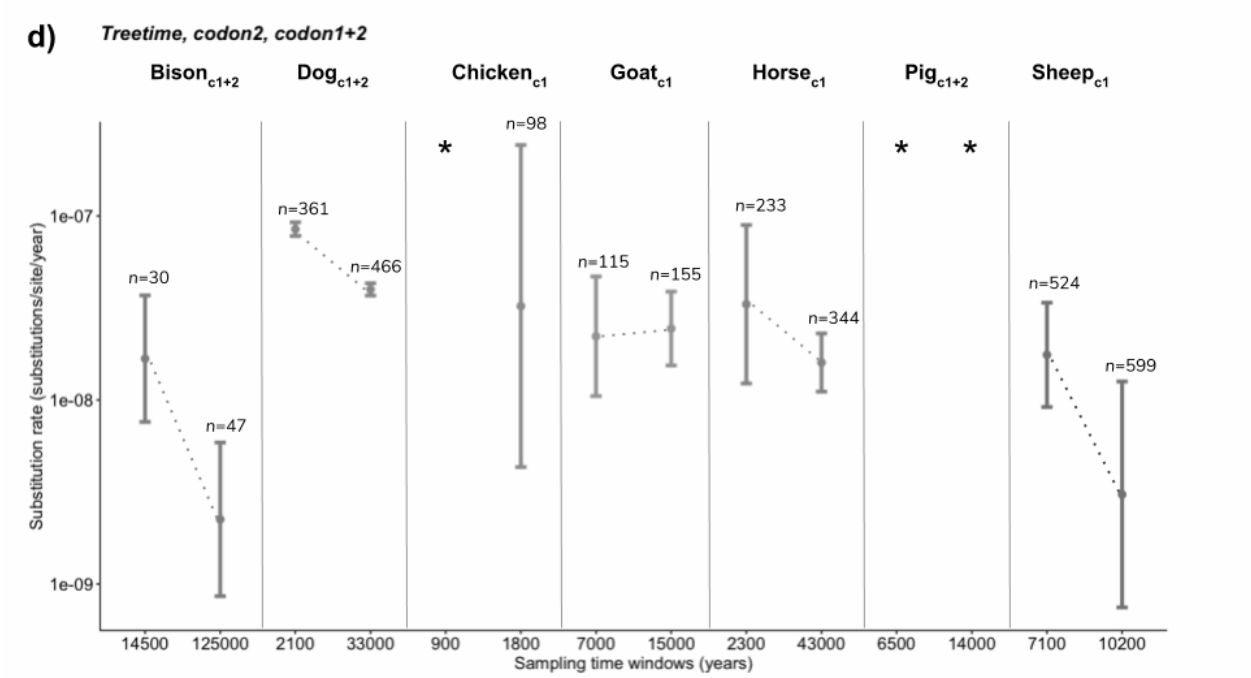
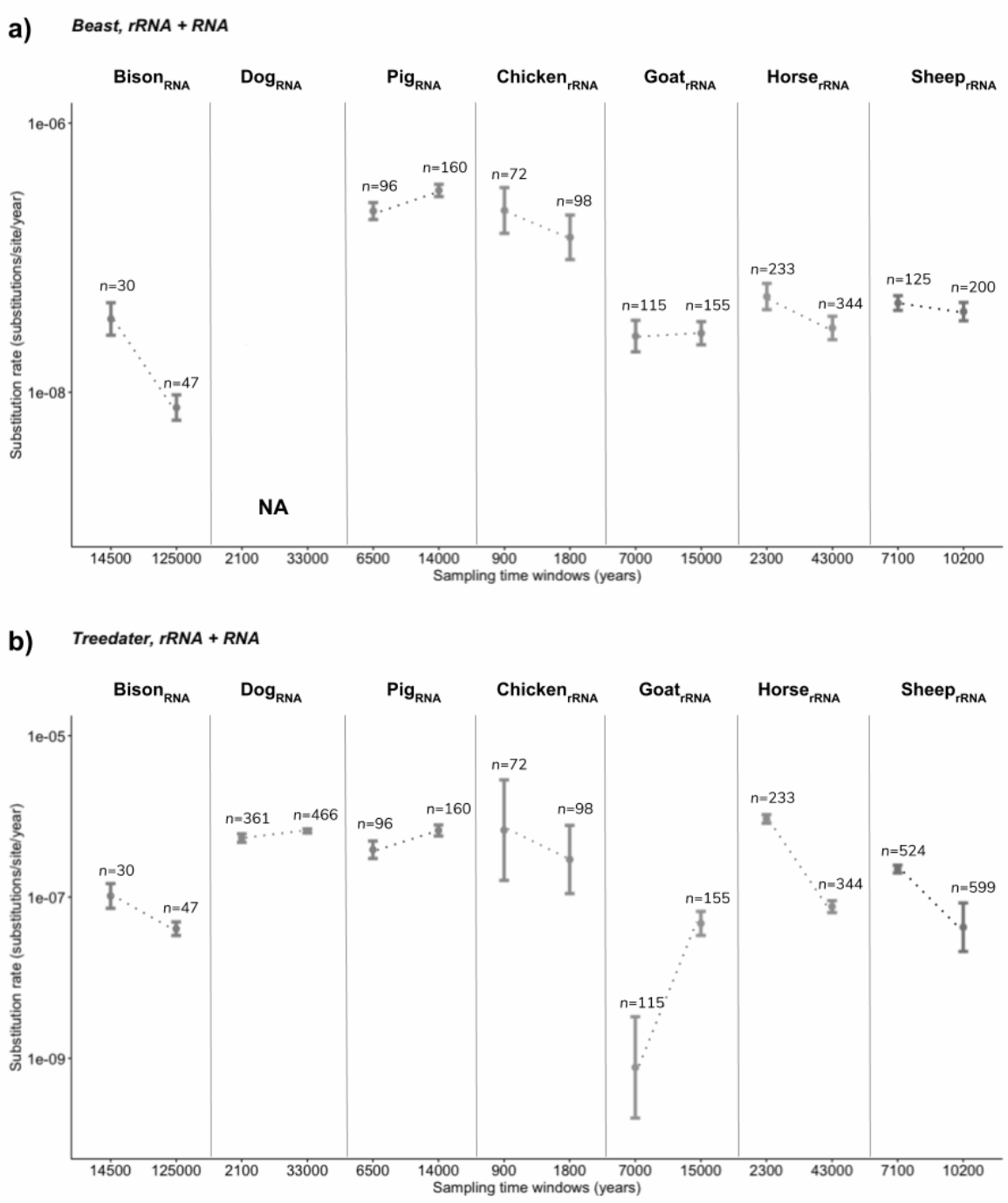


Figure 6.6 Substitution rate estimates from 3rd codon sites from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicates the number of samples within a sampling time window. Dotted lines join the mean estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.


a)
Beast, rRNA + RNA
Bison_RNA
Dog_RNA
Pig_RNA
Chicken_rRNA
Goat_rRNA
Horse_rRNA
Sheep_rRNA
Substitution rate (substitutions/site/year)
1e-06
1e-08
n=30
n=47
NA
n=96
n=160
n=72
n=98
n=115
n=155
n=233
n=344
n=125
n=200
14500
125000
2100
33000
6500
14000
900
1800
7000
15000
2300
43000
7100
10200
Sampling time windows (years)
b)
Treedater, rRNA + RNA
1e-05
1e-07
1e-09
n=361
n=466
n=524
n=599

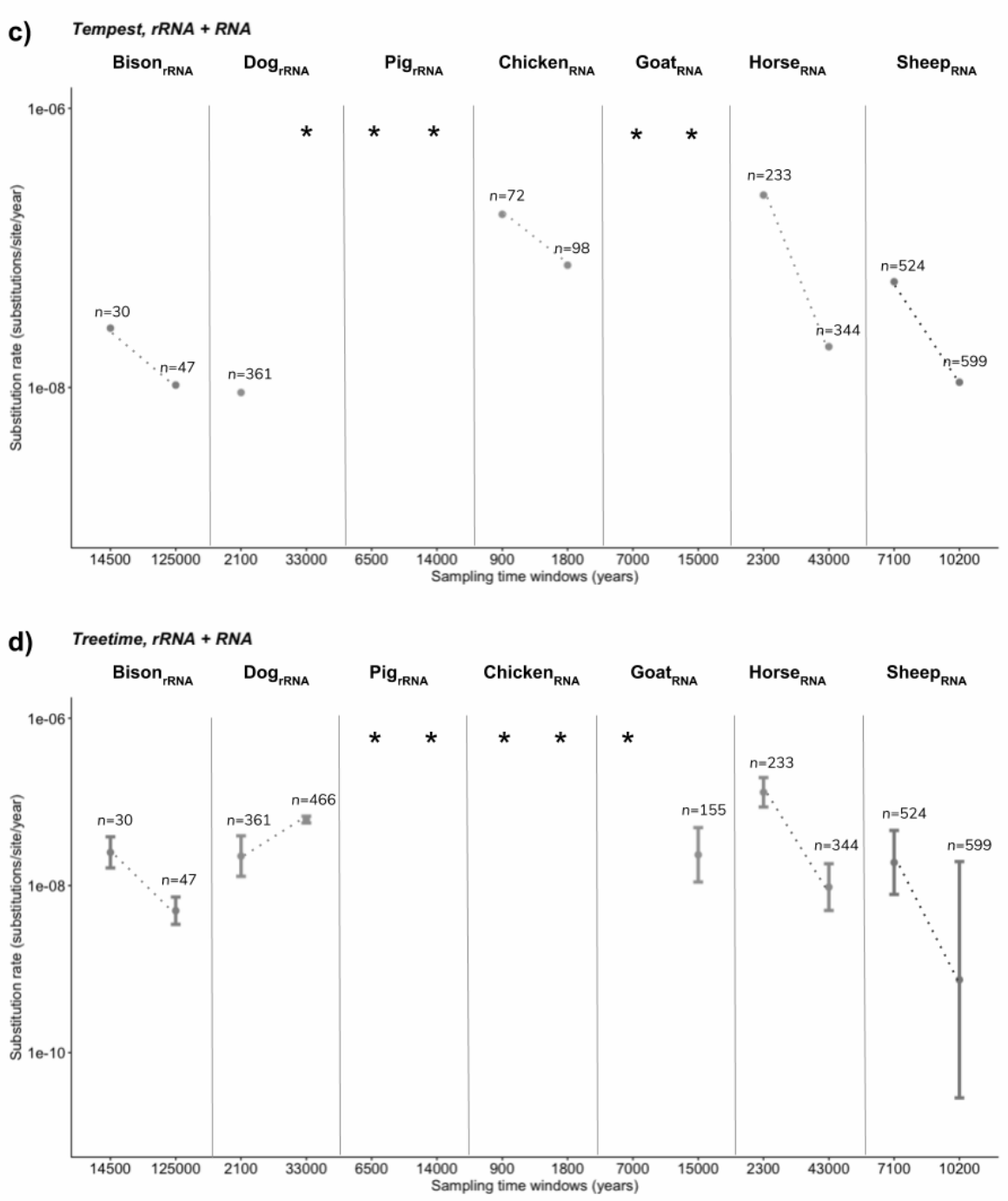
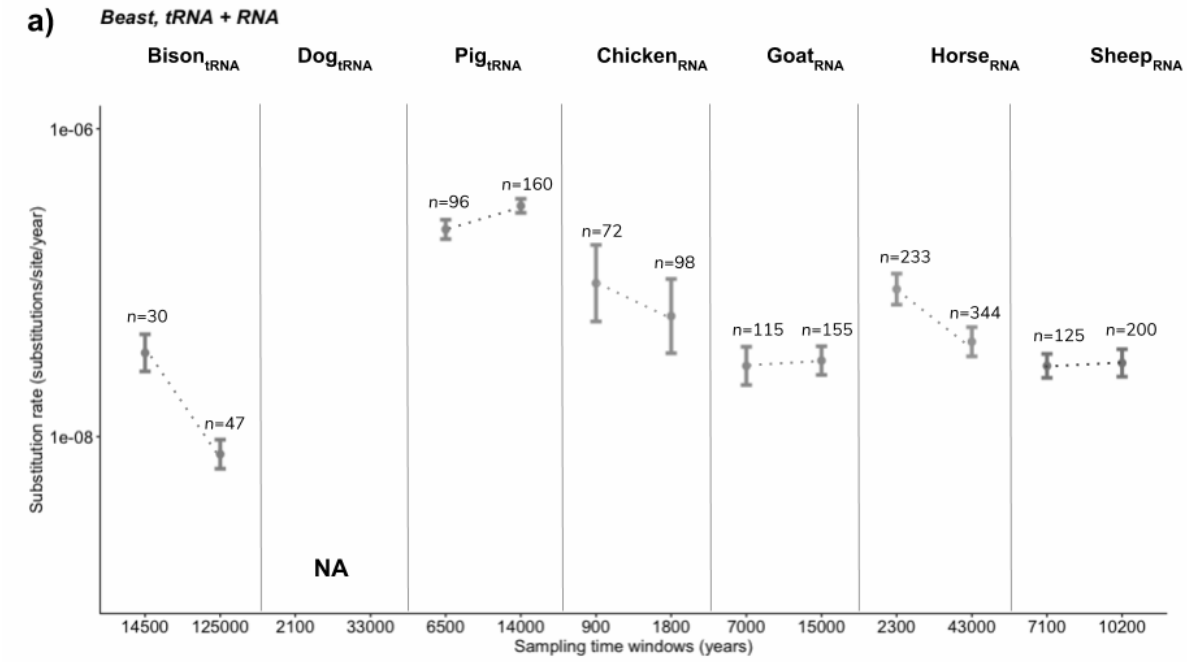
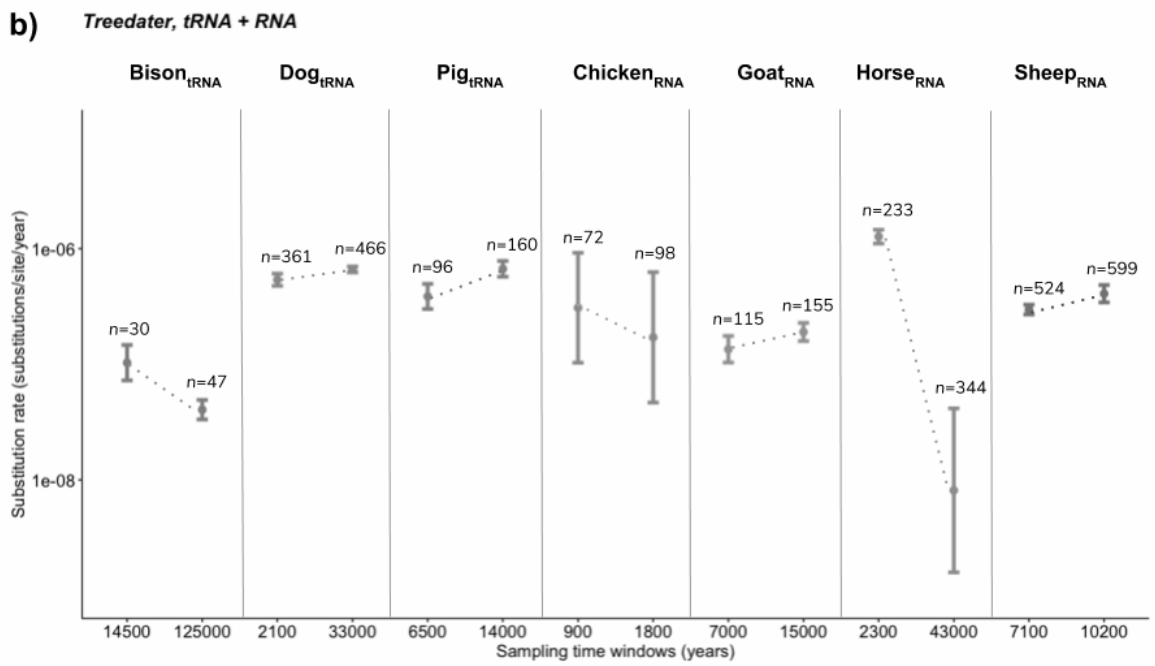


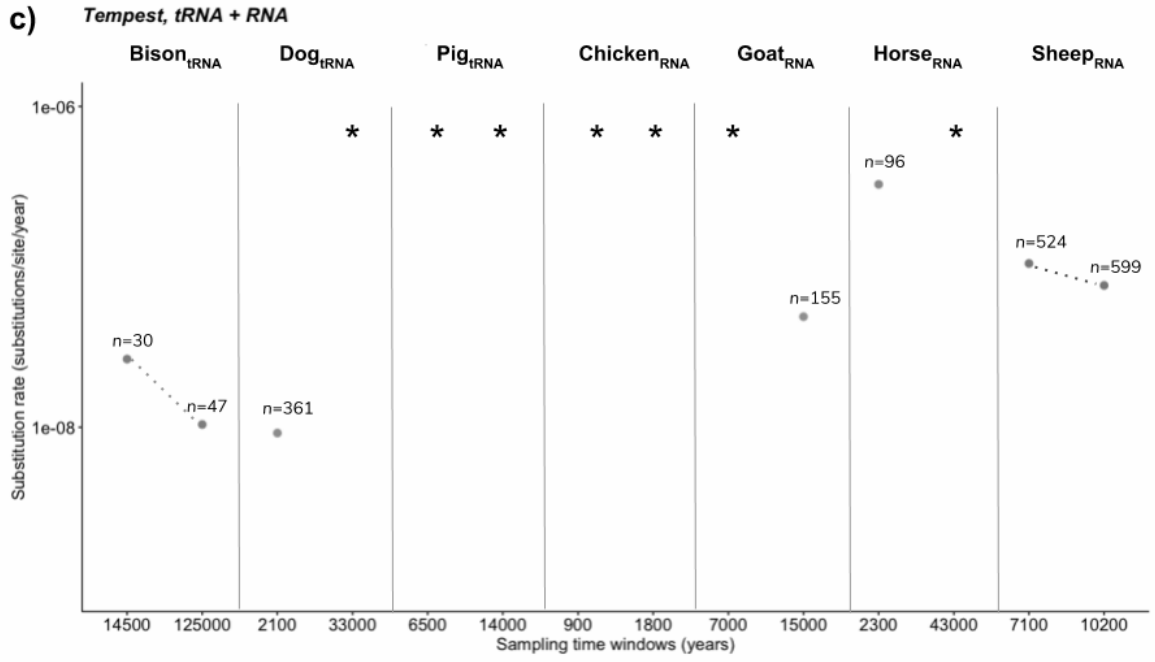
Figure 6.7 Substitution rate estimates for rRNA (chicken, goat, horse) or all RNA (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a

log scale. Numbers along the top of each error bar indicates the number of samples within a sampling time window. Dotted lines join the mean estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.




b)
Treedater, tRNA + RNA
Bison tRNA
Dog tRNA
Pig tRNA
Chicken RNA
Goat RNA
Horse RNA
Sheep RNA
Substitution rate (substitutions/site/year)
1e-06
1e-08
n=30
n=47
n=361
n=466
n=96
n=160
n=72
n=98
n=115
n=155
n=233
n=344
n=524
n=599
14500
125000
2100
33000
6500
14000
900
1800
7000
15000
2300
43000
7100
10200
Sampling time windows (years)


c)
Tempest, tRNA + RNA
Bison tRNA
Dog tRNA
Pig tRNA
Chicken RNA
Goat RNA
Horse RNA
Sheep RNA
Substitution rate (substitutions/site/year)
1e-06
1e-08
*
n=30
n=47
n=361
n=155
n=96
n=524
n=599
14500
125000
2100
33000
6500
14000
900
1800
7000
15000
2300
43000
7100
10200
Sampling time windows (years)

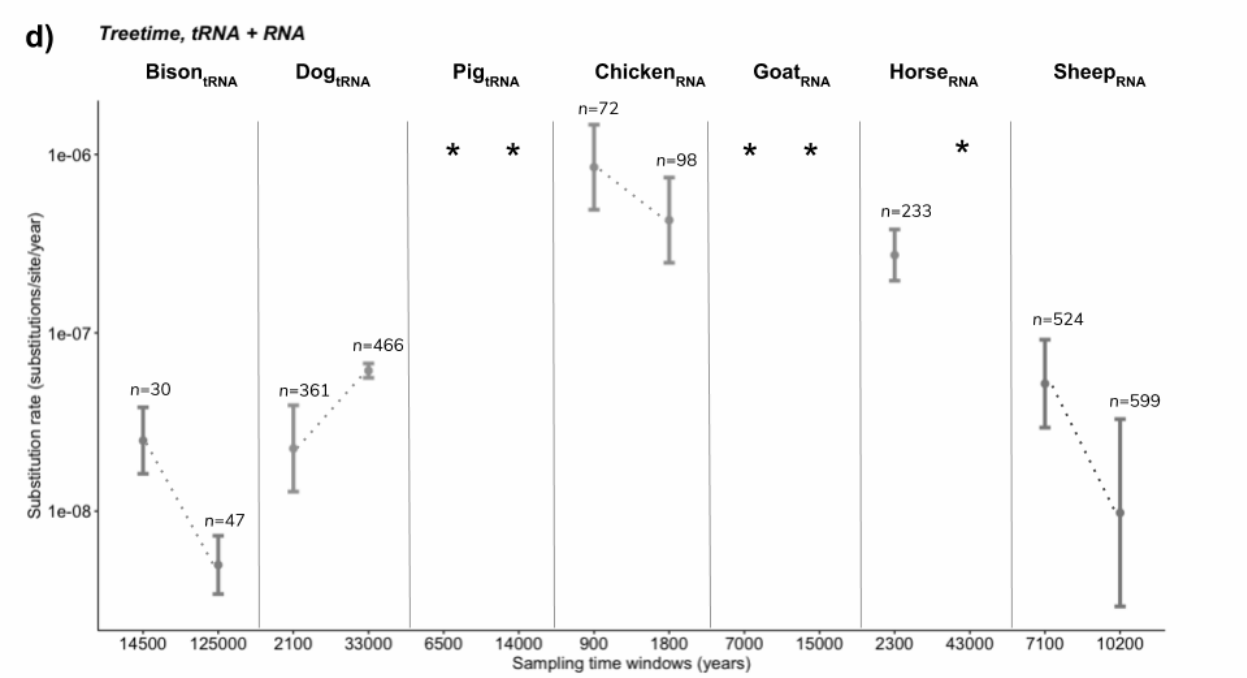


Figure 6.8 Substitution rate estimates for tRNA (chicken, goat, horse) or all RNA (bison, dog, pig, sheep) from seven taxa with seven different sampling time windows

Data were analysed using **a)** BEAST2; **b)** Treedater; **c)** TempEST; **d)** Treetime. X-axis indicates the sampling time windows (in years) for each taxon. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. Numbers along the top of each error bar indicates the number of samples within a sampling time window. Dotted lines join the estimates of t1 to t2 for each taxon. Error bars indicate 95% credibility intervals (BEAST2), 95% confidence intervals (Treedater), or 1 standard deviation from the mean (Treetime). Asterisks (*) indicate a negative rate estimate and lack of temporal signal in the dataset. Windows marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.

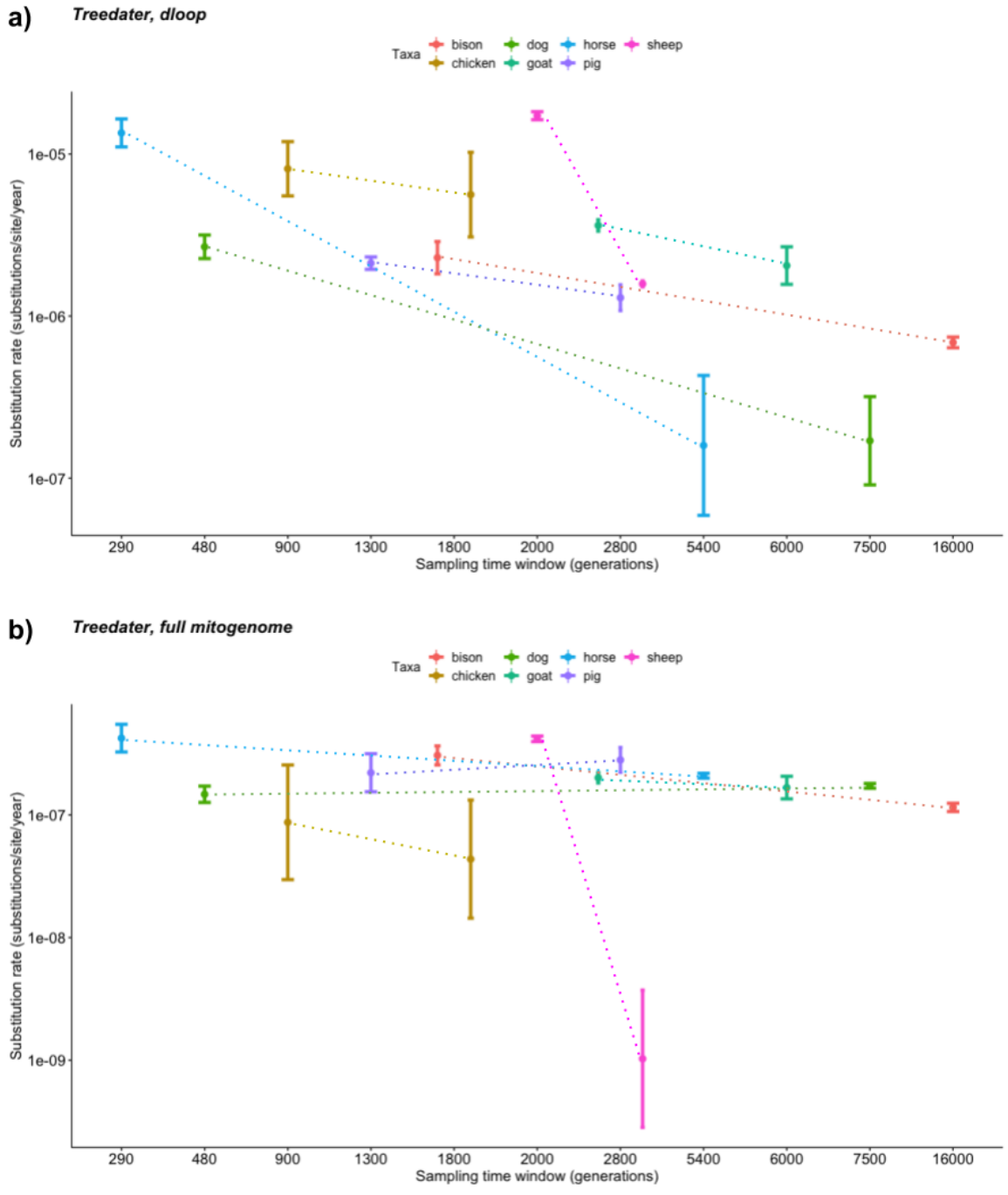
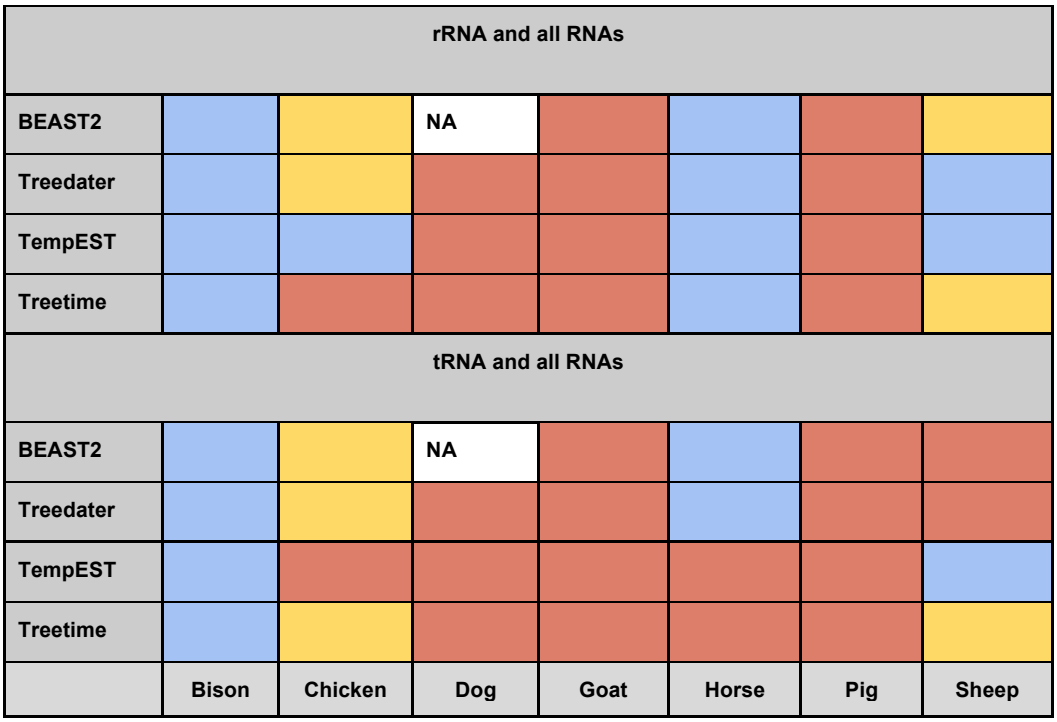


Figure 6.9 Summary of substitution rate estimates from the D-loop (a) and full mitogenome (b) generated by Treedater for seven taxa with seven different sampling time windows, scaled by the generation time for each taxon.

Dotted lines join the mean rate estimates from the datasets with sampling windows t2 and t1. Error bars indicate 95% confidence intervals.

Full mitogenome							
	Bison	Chicken	Dog	Goat	Horse	Pig	Sheep
BEAST2			NA				
Treedater							
TempEST							
Treetime							
D-loop							
BEAST2			NA				
Treedater							
TempEST							
Treetime							
Codon3							
BEAST2			NA				
Treedater							
TempEST							
Treetime							
Codon1 and Codon1+2							
BEAST2			NA				
Treedater							
TempEST							
Treetime							
Codon2 and Codon1+2							
BEAST2			NA				
Treedater							
TempEST							
Treetime							



Key:

	t2 rate estimate < t1 rate estimate
	t2 rate estimate > t1 rate estimate
	t2 mean values < t1 values (CIs and/or the 95% HPD intervals overlap).

Figure 6.10 Summary of t2 rate estimates that are lower than t1 rate estimates

Blue cells indicate that the t2 rate estimate is lower than the t1 rate estimate. Red cells indicate that the t2 rate estimate is not lower than the t1 rate estimate. Yellow cells indicate that the t2 mean values are lower than the t1 mean values, but that the 95% credibility interval (BEAST2) or the 95% confidence intervals (Treedater and Treetime) overlap. White cells marked with NA indicate incomplete results from BEAST2 at the time of the thesis submission.

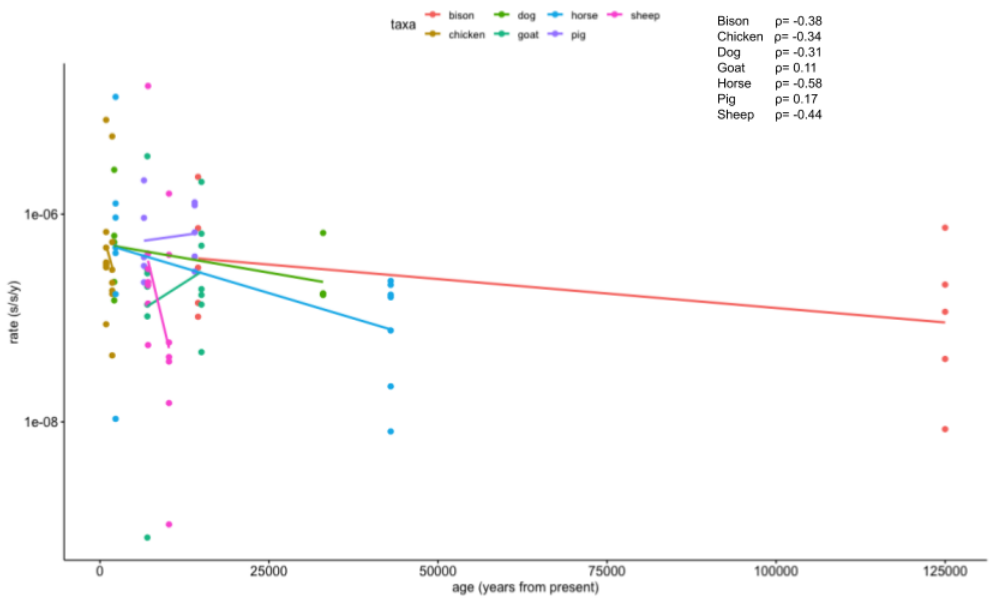


Figure 6.11 Regression of mean rate estimates from all mitogenome partitions generated from Treedater

X-axis indicates age of sampling windows as years from present, spanning the upper time boundaries of the t1 and t2 sampling windows. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. The points indicate the mean rate estimates of all sampling time windows and the solid lines indicate the regression. The Spearman rho statistic from correlation analyses of all the taxa are reported in the top right of the figure.

Because the D-loop and 3rd codon sites are under the least selective constraint, it is expected that we find in the bison data set higher rate estimates in 3rd codon sites in comparison to not only the 1st+2nd codon sites, but for other partitions including the full mitogenome. **Fig 6.12** shows that the D-loop and 3rd codon sites have higher rates compared to the other mitogenome partitions, and the pattern is seen in rate estimates generated by BEAST2, Treedater, and Treetime. However, this pattern of higher rate estimates seen in 3rd codon sites is more variable in the other taxa (**Fig 6.2-6.8**).

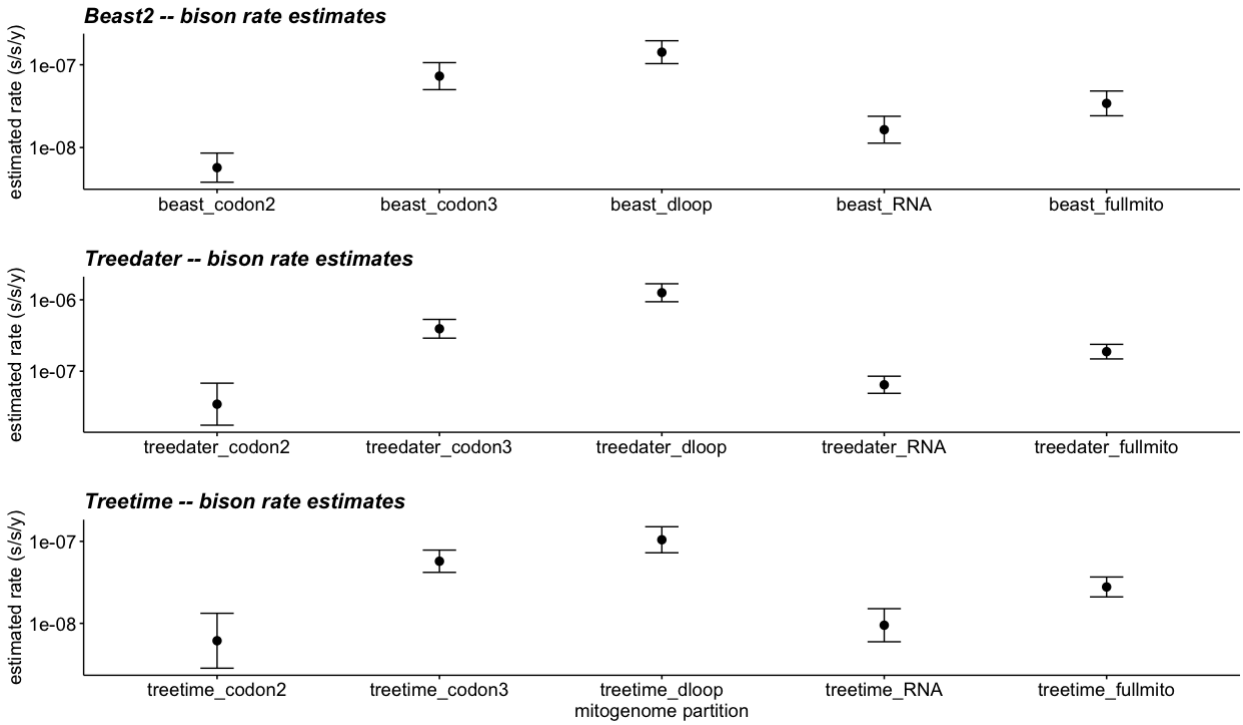


Figure 6.12 Summary of bison t2 rate estimates of all mitogenome partitions from BEAST2, Treedater, and Treetime

X-axis indicates the separate mitochondrial partitions analysed. Y-axis is the estimated substitution rate (substitution/site/year) on a log scale. The points indicate the mean rate estimates of the t2 sampling time window and error bars indicate 95% credibility intervals (BEAST2 panel), 95% confidence intervals (Treedater panel), and 1 standard deviation from the mean (Treetime panel).

To explore why the chicken datasets consistently failed the date-randomisation tests, we went back to the maximum-likelihood trees (**Figure 6.13**) that show that the ancient/historic samples form a clade in a highly nested position. This is due to their E1 haplotype lineage which the vast majority of the modern samples are not members of.



Figure 6.13 RAxML tree of chicken dataset with red jungle fowl as outgroup

Best RAxML tree from 100 bootstrap replicates of the full chicken dataset

rooted with red jungle fowl (highlighted in yellow) as the outgroup. The ancient/historic chicken samples cluster in a highly nested position.

---

4. Discussion

Using Bayesian and maximum-likelihood methods to analyse ancient and historic animal mitogenomes, we have found some evidence of time-dependent estimates of evolutionary rates. Rate estimates differed between sampling windows, depending on the mitogenome partition and on the method of analysis. Interestingly, only the bison dataset showed consistent time-dependent rate estimates across all of the different methods and mitogenome partitions. All of the other taxa showed discrepancies between the rate estimates from datasets with narrower and wider sampling windows.

We show in this study the importance of taking into consideration the different evolutionary rates across parts of the mitogenome. We demonstrate that if the dataset is of a slow-evolving species (i.e. animals instead of viruses) from shallower timescales, and if the full mitochondrial genome is being analysed, there will not be much disparity between the rate estimates based on different sampling windows.

Surprisingly, the strength of the time-dependency does not appear to be directly correlated with the length of the time period over which samples were collected (**Fig. 6.9a-b**, **Fig. 6.11**). We expected that there would be a minimum sampling period needed in order to detect time-dependency in the substitution rate is correlated with the mean (or long-term) substitution rate. As different genes evolve at different rates, it stands to reason that time-dependence will become visible over shorter time periods in fast-evolving than slow-evolving genes. In models that do not account for rate heterogeneity between genes (as for the maximum-likelihood analyses), it is

possible that averaging the substitution rate across all genes can obscure the time-dependence, if slow-evolving genes make up a significant portion of the mitogenome.

Therefore, for datasets that are on deeper timescales (e.g. the bison dataset that includes modern and 125,000 year old samples), the time-dependent effect is seen in all of the mitochondrial partitions (**Fig. 6.10)**. If time-dependency is to be detected in datasets from shallower timescales, the parts of the mitogenome that evolve more quickly than others (e.g. D-loop and 3rd codon sites) should be targeted. This is because those partitions that are slower-evolving (e.g. the RNAs and 1st+2nd codon sites of protein-coding genes) will not have accumulated enough mutations to see a relative time-dependent effect. This likely explains the discrepancy between the pig and dog t1 and t2 rates of the D-loop (**Fig 6.9a**) compared with those from the full mitogenome (**Fig 6.9b**).

4.1 Purifying selection

By looking at the rates of the individual partitions, it may be possible to see which part of the genome is driving the signal of time-dependency. As mentioned earlier, it has been suggested that purifying selection is one of the main driving mechanisms for the time-dependency in rate estimates, since some portions of the genome are more likely to be subject to selective constraints than others [402][403]. If there is strong purifying selection on a specific region, then the majority of genetic differences over longer time periods should represent substitutions rather than transient polymorphisms. It has also been suggested that over long time frames, purifying selection gives way to mutational saturation as the main driver in the time-dependent

effect in rate estimates [160]. Molak and Ho [160] warn that if nucleotide substitution models do not account for the multiple saturation hits sufficiently, the rates over longer timescales will be more severely underestimated [160]. This can result in time-dependent patterns in other evolutionary parameters, in particular the ratio of transitions to transversions [174] and the ratio of nonsynonymous to synonymous substitutions [404]. Some studies of human mitochondrial DNA have proposed corrections based on models of purifying selection (e.g. [358][405]).

Moreover, it has been suggested that a large effective population size are required to explain the rate curve, if the main driver of time-dependency is indeed purifying selection and the purging of deleterious mutations [501]. Large populations would be required to maintain these mutations over shallow timescales with higher observed rate estimates [501]. In the future, time-series ancient DNA datasets could be used to test whether purifying selection is responsible for the observed time-dependency of molecular rates. For example, the rates estimated from coding positions 1 and 2 can be compared to that of coding position 3 – the expectation being that codons 1/2 should show a higher gradient of change if purifying selection is involved.

4.2 Ancient DNA and tip-calibration

Time-series ancient DNA datasets provide a useful opportunity to bridge the gap between the clear discrepancy between short- and long-term rate estimates of pedigree and fossil-calibrated phylogenetic studies [406]. This is because as ages can be assigned directly to the sequences in the analysis (e.g. tip-dating or tip-calibration), no assumptions need to be made about the timing of genetic divergence to population splits [407]. However, tip

calibrations should only be implemented if there is a detectable amount of evolutionary change in the sampling window, that is, the population is "measurably evolving" [177]). It has been found that even a small number of ancient sequences can be sufficient for calibration, given that they are old enough [158][408].

In an ideal tip-calibrated phylogenetic study, the sequences would be sampled randomly from a panmictic population. In practice, however, datasets (even those that do not use ancient DNA) are sampled in a nonrandom manner, resulting in a nonuniform and biased distribution of sampling times [365][351]. This can have a substantial impact on evolutionary timescales and substitution rate estimates. For example, an erroneously low rate estimate can lead to overestimation of node times, and vice versa [409].

Specific to ancient DNA studies, sampling bias can be caused by the availability of the sampling material recovered from archaeological sites, the abundance of the remains at the archaeological site, and the success of DNA preservation from postmortem damage. The utility of ancient DNA is dependent on its preservation conditions: cold, dry, permafrost conditions preserves biological material better than hot, humid climates. In addition, ancient DNA analyses are also constrained by the limited reach of radiocarbon dating.

Emerson *et al.* [363][364] are specifically critical of rate estimates derived from tip-dated sequences within a Bayesian framework, asserting that these estimates exaggerate the importance of time-dependency over larger timescales. Using simulated data, the authors found that BEAST consistently obtains incorrect tree topologies that are consistent with substantial overestimation of rates and underestimation of population size [363].

Population structure within the datasets or species with less than simple demographic histories has been shown previously to dramatically affect population parameter estimates, and has been demonstrated to inflate short-term rate estimates [364][365][502]. Moreover, Emerson *et al.* [363] found that data generated from lower effective population sizes were less likely to fail the date-randomisation test, yet still resulted in higher rate estimates. In future studies, to address the issue of biased inflation of rates due to population structure, known substructure within the datasets could be partitioned out, to see whether or not the rate estimates decrease as a consequence. Unfortunately, these tests would likely come at the cost of reduced power.

4.3 Postmortem damage

Although the error rate likely varies among the samples in the different datasets, it is possible that some of the rate estimates could have been impacted by postmortem damage (PMD) not accounted for in sequences generated using high-throughput methods. This DNA damage manifests itself as extraneous mutations along lineages leading to "ancient" specimens, artificially increasing the length of terminal branches in a phylogeny, and can potentially confound evolutionary and demographic analyses [410]. For Bayesian inference methods, there are two models of postmortem damage that can estimate the amount of damage in DNA data and accounts for its effects.

These models of postmortem damage also differ according to whether they treat sequence damage as age-dependent [411] or age-independent [410]. The age-dependent model assumes that as DNA damage will accumulate through time, it is more likely for sequences derived from older specimens to have

more miscoding lesions. The model is the probability that any given nucleotide remains undamaged is assumed to decay exponentially with the age of the sample [411]. The age-independent model of postmortem damage adds a single parameter, δ, which represents the average amount of damage per nucleotide. The addition of δ additional substitutions per site to each external branch represents substitutions due to DNA damage. The δ parameter is then sampled in the Bayesian MCMC kernel during the run [410]. These models have been implemented in BEAST, but not BEAST2.

In analyses from [410], the authors estimated the damage rates in a range of aDNA data sets that had various levels of quality control and processing. The authors have found that final (consensus) aDNA sequences typically have a very low error rate due to damage, so using a model of postmortem damage is usually unnecessary. Furthermore, when the two models of postmortem damage were directly compared, very low rates of estimated damage were found [410].

It must be emphasised that these postmortem decay models were developed for aDNA data generated by molecular cloning and PCR focusing on sequence replication. These models pre-date the explosion of data generation in the field of palaeogenomics due to the use of high-throughput sequencing technologies. Moreover, there are no studies yet testing the models of postmortem damage on datasets carrying large amount of damage, such as single-coverage sequences. However, in order to recover as much data as possible from all ancient and archaeological samples, it may be prudent to implement models of postmortem damage on low-coverage consensus sequences as part of the general aDNA bioinformatic pipeline, following damage pattern analyses in MapDamage. In this way, rapid dating methods can be applied more reliably to aDNA data.

The alignments for the dog and pig datasets had varying amounts of unresolved sequences. The pig dataset in particular contains many older samples with majority consensus sequences that have very low average coverage (~1–2× average coverage) (**Supplement**). After sequencing and read alignment, software such as mapDamage [412][413] and pmdtools [414] are used to test the presence of nucleotide misincorporation patterns from overrepresentation of cytosine deamination, and these patterns can be fit to statistical models of post-mortem DNA damage [415][416][417]. The signature patterns of DNA damage caused by cytosine deamination are used to "authenticate" the ancient DNA molecules from modern DNA [418][419][420], and these damage patterns have been seen in all ancient sample types to date.

For the pig mtDNA data specifically [384], the majority consensus sequences were generated (using reads with BQ>=20 and MAPQ>=30) for all samples that had at least 2× average coverage, excluding bases within 5 bp of the start and end of a read. Many of these pig sequences did not pass the stringent criterion of 10× depth of coverage, but because only a few reads are necessary to provide enough resolution for a robust haplogroup assignment, the threshold for 10 reads with MAQP>=30 and BQ>=30 was lowered to 1 read with these requirements, and then a consensus sequence was created [384]. In other words, the pig mitogenomes included were considered complete enough to include in the analysis if there was enough genetic information in the sequence to determine a sample's haplotype. Haplotype diversity is typically found in the D-loop or control region of the mitochondrial genome. Although large gaps of unresolved nucleotides throughout the full mitogenome alignments were removed prior to analysis, it is possible that these sequences simply were not informative enough to for

molecular rate estimation. The pertinent information necessary for estimating rate was lost in the missing data: they may be in those specific nucleotide positions, in specific mitochondrial partitions.

4.4 Bayesian vs. rapid dating methods

The root-to-tip regression method offered by TempEST yielded the poorest results in our study, but has been previously shown to produce rate estimates comparable to least-squares fitting and Bayesian inference methods in analyses of mitogenome datasets [360]. TempEST should not be used as a primary method of estimating rates, but should be used in conjunction with other more robust methods, particularly if the sequence data do not show an obvious relationship between sampling time and root-to-tip distance [360]. TempEST should be used instead as a qualitative tool to rapidly assess the presence of a temporal signal in the data [383].

Although we used the strict clock model for analysing the datasets using Treedater and Treetime, these methods can implement relaxed clock models, provided that accurate data on sampling times are provided. Moreover, as with BEAST/BEAST2, both programs allow for uncertainty in the dates of the samples if the user specifies an interval within which the date is assumed to be uniformly distributed [374][377]. Treetime in particular determines the probability distribution of the node date based on the distribution of the ancestor and the number of substitutions that have occurred since the ancestor [374].

All of the rate estimates were included regardless of whether or not the datasets passed the date-randomisation test. For the chicken data, all failed the date-randomisation tests except for the t1 dataset that was performed by

BEAST2 (**Table 6.2**). An outgroup sequence was not included in the BEAST2 analyses, because the program samples the root position along with the rest of the nodes in the tree to generate a posterior probability for the root position. The results of the date-randomisation test, or the degree of temporal structure depends on many factors: the method of analysis, the depth and breadth of the sampling window, the shape of the tree, and the clock model.

Although it is beyond the scope of this study, potential biases arising from sampling effects based on what is seen in pedigree analyses should be considered. Although there are reported disparities between the low rates estimated from fossil data compared to pedigree studies that infer remarkably high rates [152], the methodologies generating these pedigree rate estimates have been criticised. For example, the high human pedigree rates that are often cited come from analyses of the hyper-variable region of the mitochondrial D-loop (reviewed in [364]), which we have also demonstrated in this study has a faster rate than other parts of the mitogenome. In contrast, several human pedigree studies have yielded rate estimates comparable to phylogenetically-derived substitution rates [503]. If gender and the probability of intra-individual fixation are not corrected for in the dataset, the rate estimates from the pedigree data will be overestimated and in excess to rates derived in phylogenetic analyses [364]. Analyses on whole genome pedigrees and population genetics also show a human generational rate estimate that is less than the phylogenetic estimate (reviewed in [364]). Given that pedigree estimates are based on a very small number of mutations [364][503], it would be useful to explore how many mtDNA transmissions would be necessary for a good statistical measurement of novel mutation rates.

Finally, it would be interesting to explore whether there is a difference in the time-dependent effect in rate estimates from wild species compared to

domestic species, as studies comparing evolutionary rates in wild and domestic species are scarce. It has been suggested by Haldane in 1949 that the rate in domestic species would be much higher due to higher selection pressure as a result of the domestication processes [421]. This expectation is intuitive given the vast morphological and behavioural changes that appear in domestic species by human intervention in a relatively short amount of time. On the other hand, in plants, evolutionary rates during the domestication are lower, or at least similar in speed, as those of wild populations [422][423]. Some evidence suggests that GxE (gene-by-environment) and epistasis (gene-by-gene) interactions may have been important in slowing down evolutionary rates during domestication [424]. A recent study analysing the rate of evolution in the morphological data of pigs and dogs found that domestication rates are not any faster than those found in the wild [425]. It is unclear how any of these selection processes, whether in wild or domesticated animal species, would impact the mitochondrial genome.

5. Conclusions

We have presented the most comprehensive study yet known testing the hypothesis of time-dependent molecular rates using ancient animal mitogenome data, particularly that of domestic animals. We have found that the difference in rate estimates between the sampling windows and the rate estimates themselves depend on what part of the mitogenome is analysed and the statistical method used. Only the bison dataset shows consistent time-dependent rate estimates across all of the different methods and mitogenome partitions. The majority of the other taxa (chicken, dog, goat, horse, and sheep) show discrepancies between the rate estimates of shallower and the deeper sampling time windows depending on the method used and the mitogenome partition analysed. The variability in time-dependency and

temporal signal in these different datasets may be affected in part by postmortem damage. Time-dependent molecular evolution has important implications for studies of recent evolutionary timescales. For example, the dates of coalescent events within populations can be overestimated if they are based on fossil-calibrated rates. We show in this study that the different parts of the mitogenome evolves at different rates, an important distinction especially when examining time-dependent rates in datasets from of slower-evolving species across shallow evolutionary timescales.

6. Supplement

Spreadsheets of all the datasets can be accessed in an attached PDF copy.

ID	Reference/Accession	Species	Age (mean years BP)	Directly dated	Coverage (mean)
Yak1_BosGrunniens_0	AY684273	*Bos grunniens*	0	-	-
Bison1_BisonBison_0	EU177871	*Bison bison*	0	-	-
B790_BisonBison_0	GU946976	*Bison bison*	0	-	-
B853_BisonBison_0	GU946977	*Bison bison*	0	-	-
B854_BisonBison_0	GU946978	*Bison bison*	0	-	-
B855_BisonBison_0	GU946979	*Bison bison*	0	-	-
B877_BisonBison_0	GU946980	*Bison bison*	0	-	-
B880_BisonBison_0	GU946981	*Bison bison*	0	-	-
B897_BisonBison_0	GU946982	*Bison bison*	0	-	-
B925_BisonBison_0	GU946983	*Bison bison*	0	-	-
B929_BisonBison_0	GU946984	*Bison bison*	0	-	-
Yak4_BosGrunniens_0	KM233416	*Bos grunniens*	0	-	-
Yak2_BosGrunniens_0	KJ463418	*Bos grunniens*	0	-	-
Yak3_BosGrunniens_0	KJ704989	*Bos grunniens*	0	-	-
AF027_Edmonton_385	KX269120	*Bison priscus*	385	Y	101.03
PH025_Edmonton_4518	KX269143	*Bison priscus*	4518	Y	22.2
AE061_RauRiv_10817	KR350472	*Bison priscus*	10817	Y	163.97
BL227_Inuvik_11314	KX269125	*Bison priscus*	11314	Y	161.3
AE035_CharLakeCave_11598	KX269118	*Bison priscus*	11598	Y	56.57
AE033_CharLakeCave_11871	KX269117	*Bison priscus*	11871	Y	21.76
BS337-AE032_CharLakeCave_12253	KX269126	*Bison priscus*	12253	Y	57.61
AF019_Edmonton_12927	KX269119	*Bison priscus*	12927	Y	15.08
PH027_Edmonton_13008	KX269144	*Bison priscus*	13008	Y	61.04
PH034_Edmonton_13023	KX269145	*Bison priscus*	13023	Y	16.81
BL225_Tuktoyaktuk_13094	KX269124	*Bison priscus*	13094	Y	18.09
AF028_Edmonton_13165	KX269121	*Bison priscus*	13165	Y	98.41
BL008_Tsiigehtchic_13686	KX269123	*Bison priscus*	13686	Y	33.47
MS213_OldCrow_13861	KX269136	*Bison priscus*	13861	Y	22.85
MS242_OldCrow_13873	KX269142	*Bison priscus*	13873	Y	18.61
MS219_OldCrow_14219	KX269137	*Bison priscus*	14219	Y	258.93
MS223_OldCrow_14254	KX269140	*Bison priscus*	14254	Y	36.96
MS220_OldCrow_14360	KX269138	*Bison priscus*	14360	Y	871.7
MS221_OldCrow_14409	KX269139	*Bison priscus*	14409	Y	72.28
MS014_Klondike_15385	KX269128	*Bison priscus*	15385	Y	56.7
MS056_Klondike_17197	KX269133	*Bison priscus*	17197	Y	33.75
MS047_Klondike_19783	KX269131	*Bison priscus*	19783	Y	54.9
AE006_Chukotka_27247	KX269109	*Bison priscus*	27247	Y	898.42
MS018_Klondike_30641	KX269129	*Bison priscus*	30641	Y	33.13
MS002_Klondike_32368	KX269127	*Bison priscus*	32368	Y	22.53
AE012_Taimyr_35827	KX269113	*Bison priscus*	35827	Y	647.18
AE014_Taimyr_36296	KX269114	*Bison priscus*	36296	Y	19.52
MS058_Klondike_37986	KX269135	*Bison priscus*	37986	Y	27.27
AE017_Taimyr_38227	KX269116	*Bison priscus*	38227	Y	51.4
MS057_Klondike_41941	KX269134	*Bison priscus*	41941	Y	16.23
MS022_Klondike_42364	KX269130	*Bison priscus*	42364	Y	19.82
MS055_Klondike_42462	KX269132	*Bison priscus*	42462	Y	59.81
AE015_Taimyr_46823	KX269115	*Bison priscus*	46823	Y	21
AE010_Taimyr_47452	KX269112	*Bison priscus*	47452	Y	788.85
B5493_BisonLatifrons_Colorado_124000	KX269122	*Bison latifrons*	124000 (124±8.5 ka)	N	6.6
MS226-DC009_ChijeesBluff_OldCrow_125000	KX269141	*Bison priscus*	125000 (125 ± 4.5 kyBP)	N	159.34
AP003323_gallus_gallus_bankiva_0	AP003323	*Gallus gallus*	0	-	-
reference_X52392_0	X52392	*Gallus gallus domesticus*	0	-	-
GU261674_0	GU261674	*Gallus gallus domesticus*	0	-	-
GU261675_0	GU261675	*Gallus gallus domesticus*	0	-	-
GU261676_0	GU261676	*Gallus gallus domesticus*	0	-	-
GU261677_0	GU261677	*Gallus gallus domesticus*	0	-	-
GU261678_0	GU261678	*Gallus gallus domesticus*	0	-	-
GU261679_0	GU261679	*Gallus gallus domesticus*	0	-	-
GU261680_0	GU261680	*Gallus gallus domesticus*	0	-	-
GU261681_0	GU261681	*Gallus gallus domesticus*	0	-	-
GU261682_0	GU261682	*Gallus gallus domesticus*	0	-	-
GU261683_0	GU261683	*Gallus gallus domesticus*	0	-	-

GU261684_0	GU261684	*Gallus gallus domesticus*	0	-	-
GU261685_0	GU261685	*Gallus gallus domesticus*	0	-	-
GU261686_0	GU261686	*Gallus gallus domesticus*	0	-	-
GU261687_0	GU261687	*Gallus gallus domesticus*	0	-	-
GU261688_0	GU261688	*Gallus gallus domesticus*	0	-	-
GU261689_0	GU261689	*Gallus gallus domesticus*	0	-	-
GU261690_0	GU261690	*Gallus gallus domesticus*	0	-	-
GU261691_0	GU261691	*Gallus gallus domesticus*	0	-	-
GU261692_0	GU261692	*Gallus gallus domesticus*	0	-	-
GU261693_0	GU261693	*Gallus gallus domesticus*	0	-	-
GU261694_0	GU261694	*Gallus gallus domesticus*	0	-	-
GU261695_0	GU261695	*Gallus gallus domesticus*	0	-	-
GU261696_0	GU261696	*Gallus gallus domesticus*	0	-	-
GU261697_0	GU261697	*Gallus gallus domesticus*	0	-	-
GU261698_0	GU261698	*Gallus gallus domesticus*	0	-	-
GU261699_0	GU261699	*Gallus gallus domesticus*	0	-	-
GU261700_0	GU261700	*Gallus gallus domesticus*	0	-	-
GU261701_0	GU261701	*Gallus gallus domesticus*	0	-	-
GU261702_0	GU261702	*Gallus gallus domesticus*	0	-	-
GU261703_0	GU261703	*Gallus gallus domesticus*	0	-	-
GU261704_0	GU261704	*Gallus gallus domesticus*	0	-	-
GU261705_0	GU261705	*Gallus gallus domesticus*	0	-	-
GU261706_0	GU261706	*Gallus gallus domesticus*	0	-	-
GU261707_0	GU261707	*Gallus gallus domesticus*	0	-	-
GU261708_0	GU261708	*Gallus gallus domesticus*	0	-	-
GU261709_0	GU261709	*Gallus gallus domesticus*	0	-	-
GU261710_0	GU261710	*Gallus gallus domesticus*	0	-	-
GU261711_0	GU261711	*Gallus gallus domesticus*	0	-	-
GU261712_0	GU261712	*Gallus gallus domesticus*	0	-	-
GU261713_0	GU261713	*Gallus gallus domesticus*	0	-	-
GU261714_0	GU261714	*Gallus gallus domesticus*	0	-	-
GU261715_0	GU261715	*Gallus gallus domesticus*	0	-	-
GU261716_0	GU261716	*Gallus gallus domesticus*	0	-	-
GU261717_0	GU261717	*Gallus gallus domesticus*	0	-	-
GU261718_0	GU261718	*Gallus gallus domesticus*	0	-	-
GU261719_0	GU261719	*Gallus gallus domesticus*	0	-	-
OL1009_E1_1650	Unpublished	*Gallus gallus domesticus*	1650	No	20.2
OL1014_E1_1100	Unpublished	*Gallus gallus domesticus*	1100	No	22.2
OL1041_E1_625	Unpublished	*Gallus gallus domesticus*	625	No	25.8
OL1145_E1_600	Unpublished	*Gallus gallus domesticus*	600	No	13
OL1206_E1_1450	Unpublished	*Gallus gallus domesticus*	1450	No	18
OL1212_E1_1582	Unpublished	*Gallus gallus domesticus*	1582	No	86.5
OL1215_E1_624	Unpublished	*Gallus gallus domesticus*	624	No	73.1
OL1224_E1_1233	Unpublished	*Gallus gallus domesticus*	1233	No	23.2
OL1312_E1_1212	Unpublished	*Gallus gallus domesticus*	1212	No	30.2
OL1331_E1_1150	Unpublished	*Gallus gallus domesticus*	1150	No	17.5
OL1383_E1_1150	Unpublished	*Gallus gallus domesticus*	1150	No	46.9
OL1385_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	76.5
OL1389_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	42
OL1395_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	97.2
OL1396_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	24.9
OL1397_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	63.8
OL1398_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	14.1
OL1409_E1_900	Unpublished	*Gallus gallus domesticus*	900	No	13.5
OL1412_E1_1609	Unpublished	*Gallus gallus domesticus*	1609	No	16.9
OL1557_E1_1725	Unpublished	*Gallus gallus domesticus*	1725	No	52
OL1576_E1_949	Unpublished	*Gallus gallus domesticus*	949	No	59.2
OL1581_E1_550	Unpublished	*Gallus gallus domesticus*	550	No	23.1
OL1584_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	59.5
OL1585_E1_350	Unpublished	*Gallus gallus domesticus*	350	No	21.7
OL1590_E1_700	Unpublished	*Gallus gallus domesticus*	700	No	70.4
OL1591_E1_700	Unpublished	*Gallus gallus domesticus*	700	No	24.8
OL1594_E1_800	Unpublished	*Gallus gallus domesticus*	800	No	28.1
OL1653_E1_612	Unpublished	*Gallus gallus domesticus*	612	No	26.2

OL1656_E1_900	Unpublished	*Gallus gallus domesticus*	900	No	19.5
OL1673_E1_800	Unpublished	*Gallus gallus domesticus*	800	No	16.5
OL1677_E1_912	Unpublished	*Gallus gallus domesticus*	912	No	23.9
OL1682_E1_862	Unpublished	*Gallus gallus domesticus*	862	No	95.3
OL1685_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	52.7
OL1687_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	19.6
OL1688_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	17.4
OL1697_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	33.3
OL1698_E1_400	Unpublished	*Gallus gallus domesticus*	400	No	75.3
OL1765_E1_1200	Unpublished	*Gallus gallus domesticus*	1200	No	25.8
OL1771_E1_1200	Unpublished	*Gallus gallus domesticus*	1200	No	121.1
OL1780_E1_1200	Unpublished	*Gallus gallus domesticus*	1200	No	18.1
OL1782_E1_975	Unpublished	*Gallus gallus domesticus*	975	No	16.4
OL1783_E1_975	Unpublished	*Gallus gallus domesticus*	975	No	24.9
OL1800_E1_975	Unpublished	*Gallus gallus domesticus*	975	No	60.6
OL1801_E1_975	Unpublished	*Gallus gallus domesticus*	975	No	116.1
OL1802_E1_924	Unpublished	*Gallus gallus domesticus*	924	No	95.2
OL1809_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	47.8
OL1810_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	95.5
OL1812_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	13.7
OL1813_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	20.7
OL1814_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	27.4
OL1815_E1_950	Unpublished	*Gallus gallus domesticus*	950	No	20.4
OL1817_E1_750	Unpublished	*Gallus gallus domesticus*	750	No	69.7
OL1820_E1_450	Unpublished	*Gallus gallus domesticus*	450	No	127.8
A1_01_0	Daly et al 2018	*Capra hircus*	0	-	-
A1a_02_0	Daly et al 2018	*Capra hircus*	0	-	-
A1a_03_0	Daly et al 2018	*Capra hircus*	0	-	-
A1a_04_0	Daly et al 2018	*Capra hircus*	0	-	-
A1a_05_0	Daly et al 2018	*Capra hircus*	0	-	-
A1a_06_0	Daly et al 2018	*Capra hircus*	0	-	-
A2_07_0	Daly et al 2018	*Capra hircus*	0	-	-
A2_08_0	Daly et al 2018	*Capra hircus*	0	-	-
A2_09_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a_10_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a_11_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_12_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_13_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_14_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_15_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_16_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_17_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_18_0	Daly et al 2018	*Capra hircus*	0	-	-
A2a1_19_0	Daly et al 2018	*Capra hircus*	0	-	-
A3_20_0	Daly et al 2018	*Capra hircus*	0	-	-
A3_21_0	Daly et al 2018	*Capra hircus*	0	-	-
A3_22_0	Daly et al 2018	*Capra hircus*	0	-	-
A3_23_0	Daly et al 2018	*Capra hircus*	0	-	-
A3_24_0	Daly et al 2018	*Capra hircus*	0	-	-
A4_25_0	Daly et al 2018	*Capra hircus*	0	-	-
A4_26_0	Daly et al 2018	*Capra hircus*	0	-	-
A4_27_0	Daly et al 2018	*Capra hircus*	0	-	-
A4_28_0	Daly et al 2018	*Capra hircus*	0	-	-
A5_29_0	Daly et al 2018	*Capra hircus*	0	-	-
A5_30_0	Daly et al 2018	*Capra hircus*	0	-	-
A5_31_0	Daly et al 2018	*Capra hircus*	0	-	-
A5_32_0	Daly et al 2018	*Capra hircus*	0	-	-
A6_33_0	Daly et al 2018	*Capra hircus*	0	-	-
A6_34_0	Daly et al 2018	*Capra hircus*	0	-	-
A7_36_0	Daly et al 2018	*Capra hircus*	0	-	-
A7_37_0	Daly et al 2018	*Capra hircus*	0	-	-
A7_38_0	Daly et al 2018	*Capra hircus*	0	-	-
A_39_0	Daly et al 2018	*Capra hircus*	0	-	-

A_40_0	Daly et al 2018	*Capra hircus*	0	-	-
A_41_0	Daly et al 2018	*Capra hircus*	0	-	-
A_42_0	Daly et al 2018	*Capra hircus*	0	-	-
A_43_0	Daly et al 2018	*Capra hircus*	0	-	-
A_44_0	Daly et al 2018	*Capra hircus*	0	-	-
A_45_0	Daly et al 2018	*Capra hircus*	0	-	-
A_46_0	Daly et al 2018	*Capra hircus*	0	-	-
A_47_0	Daly et al 2018	*Capra hircus*	0	-	-
A_48_0	Daly et al 2018	*Capra hircus*	0	-	-
A_49_0	Daly et al 2018	*Capra hircus*	0	-	-
A_50_0	Daly et al 2018	*Capra hircus*	0	-	-
A_51_0	Daly et al 2018	*Capra hircus*	0	-	-
A_52_0	Daly et al 2018	*Capra hircus*	0	-	-
A_53_0	Daly et al 2018	*Capra hircus*	0	-	-
A_54_0	Daly et al 2018	*Capra hircus*	0	-	-
A_55_0	Daly et al 2018	*Capra hircus*	0	-	-
A_56_0	Daly et al 2018	*Capra hircus*	0	-	-
A_57_0	Daly et al 2018	*Capra hircus*	0	-	-
A_58_0	Daly et al 2018	*Capra hircus*	0	-	-
A_59_0	Daly et al 2018	*Capra hircus*	0	-	-
A_60_0	Daly et al 2018	*Capra hircus*	0	-	-
A_61_0	Daly et al 2018	*Capra hircus*	0	-	-
A_62_0	Daly et al 2018	*Capra hircus*	0	-	-
A_63_0	Daly et al 2018	*Capra hircus*	0	-	-
A_64_0	Daly et al 2018	*Capra hircus*	0	-	-
A_65_0	Daly et al 2018	*Capra hircus*	0	-	-
D1_67_0	Daly et al 2018	*Capra hircus*	0	-	-
D1_68_0	Daly et al 2018	*Capra hircus*	0	-	-
G_69_0	Daly et al 2018	*Capra hircus*	0	-	-
G_70_0	Daly et al 2018	*Capra hircus*	0	-	-
G_71_0	Daly et al 2018	*Capra hircus*	0	-	-
G_72_0	Daly et al 2018	*Capra hircus*	0	-	-
G_73_0	Daly et al 2018	*Capra hircus*	0	-	-
G_74_0	Daly et al 2018	*Capra hircus*	0	-	-
B1_78_0	Daly et al 2018	*Capra hircus*	0	-	-
C1a_81_0	Daly et al 2018	*Capra hircus*	0	-	-
C1a_82_0	Daly et al 2018	*Capra hircus*	0	-	-
C1a_83_0	Daly et al 2018	*Capra hircus*	0	-	-
Capra_hircus_NC_005044.2_0	NC_005044.2	*Capra hircus*	0	-	-
Nubian_ibex_NC_020624.1_0	NC_020624.1	*Capra nubiana*	0	-	-
Darre2_443	Daly et al 2018	*Capra hircus*	443	Y	-
Geor2_750	Daly et al 2018	*Capra hircus*	750	N	-
Kazbeg1_1050	Daly et al 2018	*Capra hircus*	1050	N	-
Azer4_2440	Daly et al 2018	*Capra hircus*	2440	N	-
Miqne5_2700	Daly et al 2018	*Capra hircus*	2700	N	-
Tac2_2850	Daly et al 2018	*Capra hircus*	2850	N	-
Tac1_3200	Daly et al 2018	*Capra hircus*	3200	N	-
Tac3_3200	Daly et al 2018	*Capra hircus*	3200	N	-
Potterne1_3515	Daly et al 2018	*Capra hircus*	3515	N	-
Yoqneam2_3595	Daly et al 2018	*Capra hircus*	3595	N	-
Acem2_3700	Daly et al 2018	*Capra hircus*	3700	N	-
Bulak1_3850	Daly et al 2018	*Capra hircus*	3850	N	-
Bulak2_3850	Daly et al 2018	*Capra hircus*	3850	N	-
Bulak3_3850	Daly et al 2018	*Capra hircus*	3850	N	-
Bulak4_3850	Daly et al 2018	*Capra hircus*	3850	N	-
Bulak5_3850	Daly et al 2018	*Capra hircus*	3850	N	-
Azer3-5_4150	Daly et al 2018	*Capra hircus*	4150	N	-
Qazvin1_4150	Daly et al 2018	*Capra hircus*	4150	N	-
Chalow1_4150	Daly et al 2018	*Capra hircus*	4150	N	-
Acem1_4193	Daly et al 2018	*Capra hircus*	4193	Y	-
Yarmut7_4425	Daly et al 2018	*Capra hircus*	4425	N	-
Kan19_4450	Daly et al 2018	*Capra hircus*	4450	N	-
Kan25_4450	Daly et al 2018	*Capra hircus*	4450	N	-
Yarmut1_4600	Daly et al 2018	*Capra hircus*	4600	N	-

Kan23_4649	Daly et al 2018	*Capra hircus*	4649	Y	-
Safi2_4735	Daly et al 2018	*Capra hircus*	4735	N	-
Kohneh2_5150	Daly et al 2018	*Capra hircus*	5150	N	-
Shiqmim1_6000	Daly et al 2018	*Capra hircus*	6000	N	-
Azer6_6200	Daly et al 2018	*Capra hircus*	6200	N	-
Pie17_6350	Daly et al 2018	*Capra hircus*	6350	N	-
Gilat2_6350	Daly et al 2018	*Capra hircus*	6350	N	-
Gilat8_6350	Daly et al 2018	*Capra hircus*	6350	N	-
Gilat10_6350	Daly et al 2018	*Capra hircus*	6350	N	-
Dra34_6467	Daly et al 2018	*Capra hircus*	6467	Y	-
Darre1_6500	Daly et al 2018	*Capra hircus*	6500	N	-
Fars1_6600	Daly et al 2018	*Capra hircus*	6600	N	-
Monjukli1_6800	Daly et al 2018	*Capra hircus*	6800	N	-
Monjukli2_6800	Daly et al 2018	*Capra hircus*	6800	N	-
Monjukli3_6800	Daly et al 2018	*Capra hircus*	6800	N	-
Monjukli4_6800	Daly et al 2018	*Capra hircus*	6800	N	-
Uiv17_7150	Daly et al 2018	*Capra hircus*	7150	N	-
AP45_7150	Daly et al 2018	*Capra hircus*	7150	N	-
AP50_7150	Daly et al 2018	*Capra hircus*	7150	N	-
AP46_7175	Daly et al 2018	*Capra hircus*	7175	Y	-
AP44_7250	Daly et al 2018	*Capra hircus*	7250	N	-
Fars4_7336	Daly et al 2018	*Capra hircus*	7336	Y	-
AP49_7350	Daly et al 2018	*Capra hircus*	7350	N	-
AP38_7392	Daly et al 2018	*Capra hircus*	7392	Y	-
Ovc11_7600	Daly et al 2018	*Capra hircus*	7600	N	-
Kov27_7900	Daly et al 2018	*Capra hircus*	7900	N	-
Kov57_7900	Daly et al 2018	*Capra hircus*	7900	N	-
Kov60_7900	Daly et al 2018	*Capra hircus*	7900	N	-
Blagotin3_7994	Daly et al 2018	*Capra hircus*	7994	Y	-
Blagotin16_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan7_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan8_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan9_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan10_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan13_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan17_8100	Daly et al 2018	*Capra hircus*	8100	N	-
Semnan3_8109	Daly et al 2018	*Capra hircus*	8109	Y	-
Monjukli5_8150	Daly et al 2018	*Capra hircus*	8150	N	-
Monjukli6_8150	Daly et al 2018	*Capra hircus*	8150	N	-
Monjukli7_8150	Daly et al 2018	*Capra hircus*	8150	N	-
Monjukli8_8150	Daly et al 2018	*Capra hircus*	8150	N	-
Monjukli9_8150	Daly et al 2018	*Capra hircus*	8150	N	-
Blagotin2_8229	Daly et al 2018	*Capra hircus*	8229	Y	-
Blagotin1_8248	Daly et al 2018	*Capra hircus*	8248	Y	-
Fars2-5_8910	Daly et al 2018	*Capra hircus*	8910	Y	-
Semnan1-2_9152	Daly et al 2018	*Capra hircus*	9152	Y	-
Lur12_9958	Daly et al 2018	*Capra hircus*	9958	Y	-
Ghosh5_10000	Daly et al 2018	*Capra hircus*	10000	N	-
Ainghazal1_10000	Daly et al 2018	*Capra hircus*	10000	N	-
Ainghazal2_10000	Daly et al 2018	*Capra hircus*	10000	N	-
Ainghazal3_10000	Daly et al 2018	*Capra hircus*	10000	N	-
Ainghazal4_10000	Daly et al 2018	*Capra hircus*	10000	N	-
Lur9_10250	Daly et al 2018	*Capra hircus*	10250	N	-
Direkli5_11500	Daly et al 2018	*Capra hircus*	11500	N	-
Direkli6_11500	Daly et al 2018	*Capra hircus*	11500	N	-
Direkli1-2_13259	Daly et al 2018	*Capra hircus*	13259	Y	-
Direkli4_14037	Daly et al 2018	*Capra hircus*	14037	Y	-
Przewalski_Belina_AP012267_0	AP012267	*Equus przewalskii*	0	-	-
Przewalski_Anushka_AP012268_0	AP012268	*Equus przewalskii*	0	-	-
Przewalski_NOUMA2_AP013095_0	AP013095	*Equus przewalskii*	0	-	-
Zhongdian_EF597512_0	EF597512	*Equus caballus*	0	-	-
Naqu_EF597513_0	EF597513	*Equus caballus*	0	-	-
Deqin_EF597514_0	EF597514	*Equus caballus*	0	-	-

Debao_EU939445_0	EU939445	*Equus caballus*	0	-	-
AkhalTeke_HQ439441_0	HQ439441	*Equus caballus*	0	-	-
AkhalTeke_HQ439442_0	HQ439442	*Equus caballus*	0	-	-
Altai_HQ439443_0	HQ439443	*Equus caballus*	0	-	-
Altai_HQ439444_0	HQ439444	*Equus caballus*	0	-	-
Kladruber_HQ439445_0	HQ439445	*Equus caballus*	0	-	-
Appaloosa_HQ439446_0	HQ439446	*Equus caballus*	0	-	-
Arabian_rus_HQ439447_0	HQ439447	*Equus caballus*	0	-	-
Arabian_syr1_HQ439448_0	HQ439448	*Equus caballus*	0	-	-
Arabian_syr2_HQ439449_0	HQ439449	*Equus caballus*	0	-	-
Ardennais_HQ439450_0	HQ439450	*Equus caballus*	0	-	-
Bashkir_Curly_HQ439451_0	HQ439451	*Equus caballus*	0	-	-
Bashkir_Curly_HQ439452_0	HQ439452	*Equus caballus*	0	-	-
Barb_HQ439453_0	HQ439453	*Equus caballus*	0	-	-
Camargue_HQ439454_0	HQ439454	*Equus caballus*	0	-	-
Clydesdale_HQ439455_0	HQ439455	*Equus caballus*	0	-	-
German_Sport_Horse_HQ439456_0	HQ439456	*Equus caballus*	0	-	-
Hanoverian_HQ439457_0	HQ439457	*Equus caballus*	0	-	-
Holstein_HQ439458_0	HQ439458	*Equus caballus*	0	-	-
Oldenburg_HQ439459_0	HQ439459	*Equus caballus*	0	-	-
Westphalian_HQ439460_0	HQ439460	*Equus caballus*	0	-	-
German_Riding_Pony_HQ439461_0	HQ439461	*Equus caballus*	0	-	-
Thoroughbred_HQ439462_0	HQ439462	*Equus caballus*	0	-	-
Norwegian_Fjord_HQ439463_0	HQ439463	*Equus caballus*	0	-	-
Haflinger_HQ439464_0	HQ439464	*Equus caballus*	0	-	-
Icelandic_HQ439465_0	HQ439465	*Equus caballus*	0	-	-
Icelandic_HQ439466_0	HQ439466	*Equus caballus*	0	-	-
Yakut_HQ439467_Lippold_0	HQ439467	*Equus caballus*	0	-	-
Kabardin_HQ439468_0	HQ439468	*Equus caballus*	0	-	-
Kinsky_HQ439469_0	HQ439469	*Equus caballus*	0	-	-
Kinsky_horse_HQ439470_0	HQ439470	*Equus caballus*	0	-	-
Kladruber_HQ439471_0	HQ439471	*Equus caballus*	0	-	-
Kladruber_HQ439472_0	HQ439472	*Equus caballus*	0	-	-
Konik_HQ439473_0	HQ439473	*Equus caballus*	0	-	-
Kuznet_HQ439474_0	HQ439474	*Equus caballus*	0	-	-
Kuznet_HQ439475_0	HQ439475	*Equus caballus*	0	-	-
Kustanai_HQ439476_0	HQ439476	*Equus caballus*	0	-	-
Lewitzer_HQ439477_0	HQ439477	*Equus caballus*	0	-	-
Liebenthaler_HQ439478_0	HQ439478	*Equus caballus*	0	-	-
Liebenthaler_HQ439479_0	HQ439479	*Equus caballus*	0	-	-
Noriker_HQ439480_0	HQ439480	*Equus caballus*	0	-	-
Orlov_Trotter_HQ439481_0	HQ439481	*Equus caballus*	0	-	-
Paint_Horse_HQ439482_0	HQ439482	*Equus caballus*	0	-	-
Percheron_HQ439483_0	HQ439483	*Equus caballus*	0	-	-
Przewalski_HQ439484_0	HQ439484	*Equus przewalskii*	0	-	-
Rhineland_Heavy_Draft_HQ439485_0	HQ439485	*Equus caballus*	0	-	-
Russian_Riding_Horse_HQ439486_0	HQ439486	*Equus caballus*	0	-	-
Black_Forest_HQ439487_0	HQ439487	*Equus caballus*	0	-	-
Shagya_Arabian_HQ439488_0	HQ439488	*Equus caballus*	0	-	-
Shetland_HQ439489_0	HQ439488	*Equus caballus*	0	-	-
Shire_HQ439490_0	HQ439490	*Equus caballus*	0	-	-
Rhineland_HeavyDraft_HQ439491_0	HQ439491	*Equus caballus*	0	-	-
Spotted_horse_HQ439492_0	HQ439492	*Equus caballus*	0	-	-
Trakehner_HQ439493_0	HQ439493	*Equus caballus*	0	-	-
Hungarian_Coldblood_HQ439494_0	HQ439494	*Equus caballus*	0	-	-
Viatka_HQ439495_0	HQ439495	*Equus caballus*	0	-	-
Welsh_Pony_D_HQ439496_0	HQ439496	*Equus caballus*	0	-	-
Welsh_Pony_A_HQ439497_0	HQ439497	*Equus caballus*	0	-	-
Welsh_Pony_B1_HQ439498_0	HQ439498	*Equus caballus*	0	-	-
Welsh_Pony_B2_HQ439499_0	HQ439499	*Equus caballus*	0	-	-
Vladimir_Heavy_Draught_HQ439500_0	HQ439500	*Equus caballus*	0	-	-
Przewalski_JN398402_0	JN398402	*Equus przewalskii*	0	-	-
Przewalski_JN398403_0	JN398403	*Equus przewalskii*	0	-	-
Yakutia_Horse1_KT368727_0	KT368727	*Equus caballus*	0	-	-

Yakutia_Horse2_KT368728_0	KT368728	*Equus caballus*	0	-	-
Yakutia_Horse3_KT368729_0	KT368729	*Equus caballus*	0	-	-
Yakutian_Yak1_KT368730_0	KT368730	*Equus caballus*	0	-	-
Yakutian_Yak2_KT368731_0	KT368731	*Equus caballus*	0	-	-
Yakutian_Yak3_KT368732_0	KT368732	*Equus caballus*	0	-	-
Yakutian_Yak4_KT368733_0	KT368733	*Equus caballus*	0	-	-
Yakutian_Yak5_KT368734_0	KT368734	*Equus caballus*	0	-	-
Yakutian_Yak6_KT368735_0	KT368735	*Equus caballus*	0	-	-
Yakutian_Yak7_KT368736_0	KT368736	*Equus caballus*	0	-	-
Yakutian_Yak8_KT368737_0	KT368737	*Equus caballus*	0	-	-
Yakutian_Yak9_KT368738_0	KT368738	*Equus caballus*	0	-	-
Mongolian_KT368741_0	KT368741	*Equus przewalskii*	0	-	-
Przewalski_Viola_KT368742_0	KT368742	*Equus przewalskii*	0	-	-
Przewalski_Vjuga_KT368743_0	KT368743	*Equus przewalskii*	0	-	-
Przewalski_Bars_KT368744_0	KT368744	*Equus przewalskii*	0	-	-
Przewalski_Vika_KT368745_0	KT368745	*Equus przewalskii*	0	-	-
Przewalski_Pacifique_KT368746_0	KT368746	*Equus przewalskii*	0	-	-
Przewalski_Kuporovitch_KT368747_0	KT368747	*Equus przewalskii*	0	-	-
Przewalski_Basil_KT368748_0	KT368748	*Equus przewalskii*	0	-	-
Przewalski_Bonnette_KT368749_0	KT368749	*Equus przewalskii*	0	-	-
Przewalski_Sigor_KT368750_0	KT368750	*Equus przewalskii*	0	-	-
Przewalski_Rolf_KT368751_0	KT368751	*Equus przewalskii*	0	-	-
Przewalski_Kolette_KT368752_0	KT368752	*Equus przewalskii*	0	-	-
Przewalski_hybrid_KB7903_KT368757_0	KT368757	*Equus przewalskii*	0	-	-
Arabian_KT757739_0	KT757739	*Equus caballus*	0	-	-
Icelandic_KT757745_0	KT757745	*Equus caballus*	0	-	-
Norwegian_Fjord_KT757760_0	KT757760	*Equus caballus*	0	-	-
Przewalski_KT757761_0	KT757761	*Equus przewalskii*	0	-	-
Standardbred_KT757762_0	KT757762	*Equus caballus*	0	-	-
Thoroughbred_Twilight_KT757764_0	KT757764	*Equus caballus*	0	-	-
Heavy_Warmblood_0269A_0	Orlando	*Equus caballus*	0	-	-
Icelandic_0247A_0	Orlando	*Equus caballus*	0	-	-
Przewalski_0217_0	Orlando	*Equus przewalskii*	0	-	-
Przewalski_0218_0	Orlando	*Equus przewalskii*	0	-	-
Quarter_0	Orlando	*Equus caballus*	0	-	-
Shetland_0250A_0	Orlando	*Equus caballus*	0	-	-
Thoroughbred_0290A_0	Orlando	*Equus caballus*	0	-	-
Somali_Wild_Ass_MG885769_0	MG885769.1	*Equus africanus somaliensis*	0	-	-
Przewalski_NOUMA1_NC_024030_0	NC_024030	*Equus przewalskii*	0	-	-
Connemara_0004A_0	PRJEB19970	*Equus caballus*	0	-	5.38
Shetland_0249A_0	SAMEA3367610	*Equus caballus*	0	-	365.12
FMontagnes_0065A_0	SAMEA3498888	*Equus caballus*	0	-	133.22
Standardbred_0081A_0	SAMEA3499832	*Equus caballus*	0	-	1057.12
Quarter_0073A_0	SAMEA3499837	*Equus caballus*	0	-	3742.11
Morgan_0096A_0	SAMEA3499838	*Equus caballus*	0	-	534.52
Friesian_0296A_0	SAMEA3951218	*Equus caballus*	0	-	1029.98
Icelandic_0144A_0	SAMN00857894	*Equus caballus*	0	-	606.36
Jeju_0275A_0	SAMN01057172	*Equus caballus*	0	-	351.06
Duelmener_0238A_0	SAMN02422919	*Equus caballus*	0	-	476.23
Arabian_0237A_0	SAMN02439777	*Equus caballus*	0	-	308.33
Sorraia_0236A_0	SAMN02439778	*Equus caballus*	0	-	392.78
Hanoverian_0235A_0	SAMN02439779	*Equus caballus*	0	-	481.1
Mongolian_0215A_0	SAMN04002338	*Equus przewalskii*	0	-	532.23
Marwari_0239A_0	SRR1275408	*Equus caballus*	0	-	227.96
Prze_0212A_Holotype_mtDNA_BQ_25	Orlando	*Equus przewalskii*	25	-	-
Przewalski_Theodore_KT368758_90	KT368758	*Equus przewalskii*	90	-	-
Mongolian_Emgl1_KT368739_104	KT368739	*Equus przewalskii*	104	-	-
Mongolian_Emgl3_KT368740_105	KT368740	*Equus przewalskii*	105	-	-
Przewalski_Bijsk1_KT368753_109	KT368753	*Equus przewalskii*	109	-	-
Przewalski_Bijsk2_KT368754_116	KT368754	*Equus przewalskii*	116	-	-
Przewalski_Paratype_KT368756_118	KT368756	*Equus przewalskii*	118	-	-
Tumeski_CGG101397_KT368726_192	KT368726	*Equus caballus*	192	N	819.63
WitterPlace_UK16_217	Fages et al 2019	*Equus caballus*	217	N	61.4
WitterPlace_UK17_267	Fages et al 2019	*Equus caballus*	267	N	75.7

WitterPlace_UK18_267	Fages et al 2019	*Equus caballus*	267	N	53.5
Odjuluun_ODJ6_KT368724_267	KT368724	*Equus caballus*	267	-	-
Diriali_Georgia2_317	Fages et al 2019	*Equus caballus*	317	N	1.53
CortaDeLaCartuja_Jal5886_368	Orlando	*Equus caballus*	368	-	-
Beauvais_GVA122_417	Fages et al 2019	*Equus caballus*	417	Y	46.5
Belgheis_TrBWBX116_485	Fages et al 2019	*Equus caballus*	485	Y	113
Sittard_SS54_550	Orlando	*Equus caballus*	550	-	-
Beauvais_GVA375_567	Fages et al 2019	*Equus caballus*	567	N	159
TavanTolgoi_GEP13_730	Fages et al 2019	*Equus caballus*	730	N	549
TavanTolgoi_GEP14_730	Fages et al 2019	*Equus caballus*	730	N	455
TavanTolgoi_GEP21_730	Fages et al 2019	*Equus caballus*	730	N	185
Oktyabrsky_Rus37_830	Fages et al 2019	*Equus caballus*	830	Y	199
Demuin_GVA401_867	Orlando	*Equus caballus*	867	-	-
Demuin_GVA402_867	Orlando	*Equus caballus*	867	-	-
Musumae_EREsto20_967	Orlando	*Equus caballus*	967	-	-
Hungate_VEM107_1063	Orlando	*Equus caballus*	1063	-	-
Berufjordur_VHR102_1067	Fages et al 2019	*Equus caballus*	1067	N	22.1
BroughOfDeerness_VHR010_1067	Fages et al 2019	*Equus caballus*	1067	N	26.1
BroughOfDeerness_VHR011_1067	Fages et al 2019	*Equus caballus*	1067	N	35.3
BroughOfDeerness_VHR062_1067	Fages et al 2019	*Equus caballus*	1067	N	32
Quoygrew_VHR017_1067	Fages et al 2019	*Equus caballus*	1067	N	41.1
Marvele_21_1087	Fages et al 2019	*Equus caballus*	1087	Y	79.1
Marvele_01_1117	Fages et al 2019	*Equus caballus*	1117	N	66.9
Saadjarve_Saa1_1117	Fages et al 2019	*Equus caballus*	1117	N	95.1
Pada_EREsto5_1117	Orlando	*Equus caballus*	1117	-	-
Zhanaturmus_Issyk1_1143	Fages et al 2019	*Equus caballus*	1143	Y	77.7
KashkhatolVtaroyMagini_VP2004N97_1143	Orlando	*Equus caballus*	1143	-	-
Marvele_32_1144	Fages et al 2019	*Equus caballus*	1144	Y	95.5
Yenikapi_Tur145_1156	Fages et al 2019	*Equus caballus*	1156	N	321
Otepaa_Ote2_1184	Fages et al 2019	*Equus caballus*	1184	Y	114
Nustar_4_1187	Fages et al 2019	*Equus caballus*	1187	Y	4.43
Nustar_5_1187	Fages et al 2019	*Equus caballus*	1187	Y	119
Marvele_18_1189	Fages et al 2019	*Equus caballus*	1189	Y	83
Grigorevka4_PAVH2_1192	Fages et al 2019	*Equus caballus*	1192	N	112
KashkhatolVtaroyMagini_VP2004B95_1213	Orlando	*Equus caballus*	1213	-	-
Loebnitz_SS28_1250	Orlando	*Equus caballus*	1250	-	-
BozAdyr_KYRH10_1267	Fages et al 2019	*Equus caballus*	1267	N	102
BozAdyr_KYRH8_1267	Fages et al 2019	*Equus caballus*	1267	N	112
Yenikapi_Tur140_1289	Fages et al 2019	*Equus caballus*	1289	N	147
Khotont_UCIE2012x85_1291	Fages et al 2019	*Equus caballus*	1291	N	95.9
Burgast_GVA9046_1304	Orlando	*Equus caballus*	1304	-	-
BapskaGradac_BAPSKA_1305	Fages et al 2019	*Equus caballus*	1305	Y	1.48
Otok_OTOK16_1307	Fages et al 2019	*Equus caballus*	1307	Y	6.1
Yenikapi_Tur194_1360	Fages et al 2019	*Equus caballus*	1360	Y	320
Mainz_Mzr35_1373	Fages et al 2019	*Equus caballus*	1373	Y	90.8
Yenikapi_Tur142_1396	Fages et al 2019	*Equus caballus*	1396	N	197
Zayukovo_OSCAE2016xSP7x1_1421	Orlando	*Equus caballus*	1421	-	-
Yenikapi_Tur141_1430	Fages et al 2019	*Equus caballus*	1430	N	208
Yenikapi_Tur150_1443	Fages et al 2019	*Equus caballus*	1443	N	178
Yenikapi_Tur170_1443	Fages et al 2019	*Equus caballus*	1443	N	212
Yenikapi_Tur193_1443	Fages et al 2019	*Equus caballus*	1443	N	91
Yenikapi_Tur229_1443	Fages et al 2019	*Equus caballus*	1443	N	159
Zayukovo_OSCAE2016SP8_1444	Orlando	*Equus caballus*	1444	-	-
CanRoqueta_TR279_1523	Orlando	*Equus caballus*	1523	-	-
ShahrlQumis_AM115_1557	Fages et al 2019	*Equus caballus*	1557	Y	201
WhitehallRomanVilla_UK08_1667	Fages et al 2019	*Equus caballus*	1667	N	3.35
Yenikapi_Tur171_1689	Fages et al 2019	*Equus caballus*	1689	Y	330
ShahrlQumis_AM181_1694	Fages et al 2019	*Equus caballus*	1694	N	104
KoulianCave_MV178_1694	Orlando	*Equus caballus*	1694	-	-
Yenikapi_Tur172_1695	Fages et al 2019	*Equus caballus*	1695	N	223
Boves_GVA191_1717	Fages et al 2019	*Equus caballus*	1717	N	79.3
Yenikapi_Tur146_1730	Fages et al 2019	*Equus caballus*	1730	N	222
Macon_GVA201_1767	Fages et al 2019	*Equus caballus*	1767	N	81.6
Sittard_SS124_1815	Orlando	*Equus caballus*	1815	-	-

Sittard_SS127_1815	Orlando	*Equus caballus*	1815	-	-
FrankfurtHeddenheim_Fr1_1863	Fages et al 2019	*Equus caballus*	1863	Y	109
Evreux_GVA133_1867	Fages et al 2019	*Equus caballus*	1867	N	66.8
FMontauban_GVA126_1872	Fages et al 2019	*Equus caballus*	1872	N	62.5
Chartres_GVA111_1917	Fages et al 2019	*Equus caballus*	1917	N	3.42
Chartres_GVA112_1917	Fages et al 2019	*Equus caballus*	1917	N	4.04
Chartres_GVA115_1917	Fages et al 2019	*Equus caballus*	1917	N	8.61
Chartres_GVA26_1917	Fages et al 2019	*Equus caballus*	1917	Y	123
Chartres_GVA28_1917	Fages et al 2019	*Equus caballus*	1917	N	8.96
Chartres_GVA4_1917	Fages et al 2019	*Equus caballus*	1917	N	74
Chartres_GVA43_1917	Fages et al 2019	*Equus caballus*	1917	N	95.2
Chartres_GVA53_1917	Fages et al 2019	*Equus caballus*	1917	N	3.44
Chartres_GVA56_1917	Fages et al 2019	*Equus caballus*	1917	N	74.6
Chartres_GVA60_1917	Fages et al 2019	*Equus caballus*	1917	N	3.83
Chartres_GVA75_1917	Fages et al 2019	*Equus caballus*	1917	N	3.36
Chartres_GVA81_1917	Fages et al 2019	*Equus caballus*	1917	N	110
Chartres_GVA9_1917	Fages et al 2019	*Equus caballus*	1917	N	5.25
AugustaRaurica_JG160_1927	Fages et al 2019	*Equus caballus*	1927	N	109
Haunstetten_1979	Fages et al 2019	*Equus caballus*	1979	N	96.5
GolModII_Mon28_1988	Orlando	*Equus caballus*	1988	N	143
GolModII_Mon24_1993	Fages et al 2019	*Equus caballus*	1993	Y	215
GolModII_Mon26_2002	Orlando	*Equus caballus*	2002	N	170
GolModII_Mon23_2007	Fages et al 2019	*Equus caballus*	2007	Y	383
GolModII_Mon25_2011	Fages et al 2019	*Equus caballus*	2011	Y	150
GolModII_Mon27_2011	Fages et al 2019	*Equus caballus*	2011	Y	150
KashkhatolVtaroyMagini_VP2004M16N3x2_2099	Orlando	*Equus caballus*	2099	-	-
Biluut_Mon2017x152x1_2103	Orlando	*Equus caballus*	2103	-	-
Actiparc_GVA307_2127	Fages et al 2019	*Equus caballus*	2127	Y	45.9
Actiparc_GVA124_2143	Fages et al 2019	*Equus caballus*	2143	Y	94.6
Sittard_SS118_2150	Orlando	*Equus caballus*	2150	-	-
Sittard_SS120_2150	Orlando	*Equus caballus*	2150	-	-
Sittard_SS121_2150	Orlando	*Equus caballus*	2150	-	-
Sittard_SS122_2150	Orlando	*Equus caballus*	2150	-	-
SaintJuste_GVA242_2250	Fages et al 2019	*Equus caballus*	2250	Y	102
Actiparc_GVA311_2253	Fages et al 2019	*Equus caballus*	2253	Y	45.7
Vermand_GVA199_2267	Fages et al 2019	*Equus caballus*	2267	N	34.8
Berel_BER01_A_2300	Fages et al 2019	*Equus caballus*	2300	N	384
Berel_BER02_B_2300	Fages et al 2019	*Equus caballus*	2300	N	346
Berel_BER04_D_2300	Fages et al 2019	*Equus caballus*	2300	N	316
Berel_BER05_E_2300	Fages et al 2019	*Equus caballus*	2300	N	482
Berel_BER06_F_2300	Fages et al 2019	*Equus caballus*	2300	N	650
Berel_BER07_G_2300	Fages et al 2019	*Equus caballus*	2300	N	301
Berel_BER08_H_2300	Fages et al 2019	*Equus caballus*	2300	N	261
Berel_BER09_I_2300	Fages et al 2019	*Equus caballus*	2300	N	306
Berel_BER10_K_2300	Fages et al 2019	*Equus caballus*	2300	N	181
Berel_BER11_L_2300	Fages et al 2019	*Equus caballus*	2300	N	212
Berel_BER12_M_2300	Fages et al 2019	*Equus caballus*	2300	N	151
NewSiberianIslands_154_KT757746_2301	KT757746	*Equus caballus*	2301	-	-
AlordaPark_Horse9020_2307	Orlando	*Equus caballus*	2307	-	-
PuigdeSantAndreu_UE14029_2310	Orlando	*Equus caballus*	2310	-	-
Actiparc_GVA308_2312	Fages et al 2019	*Equus caballus*	2312	Y	123
Khatuu_Kha2_t1_2312	Fages et al 2019	*Equus caballus*	2312	Y	57.1
Syrgal_Syr1t1c3_2317	Fages et al 2019	*Equus caballus*	2317	N	64.6
Actiparc_GVA310_2319	Fages et al 2019	*Equus caballus*	2319	N	119
Hereuet_2012x137_2329	Orlando	*Equus caballus*	2329	-	-
OlonKurinGol_OKG2_2367	Fages et al 2019	*Equus caballus*	2367	N	122
Sikarra_UE172_2398	Orlando	*Equus caballus*	2398	-	-
Biluut_Mon2017x156_2467	Orlando	*Equus caballus*	2467	-	-
ArzhanII_Rus9_r_2500	Fages et al 2019	*Equus caballus*	2500	N	322
ElsVilars_Fetusx9m_2517	Orlando	*Equus caballus*	2517	-	-
Halvai_KSH4_2542	Fages et al 2019	*Equus caballus*	2542	N	79.6
Monediere_UE10287_2640	Orlando	*Equus caballus*	2640	-	-
TepeHasanlu_V31E_2667	Orlando	*Equus caballus*	2667	-	-
Vilars_UE4618_2672	Orlando	*Equus caballus*	2672	-	-

Asva_EREsto3_2675	Orlando	*Equus caballus*	2675	-	-
Sayangorsk_Rus41_2677	Fages et al 2019	*Equus caballus*	2677	Y	152
Asva_EREsto2_2681	Orlando	*Equus caballus*	2681	-	-
TepeHasanlu_1140_2682	Fages et al 2019	*Equus caballus*	2682	Y	27
Vilars_UE11080x11082_2686	Orlando	*Equus caballus*	2686	-	-
BorShorooniiAm_Mon2017x127_2691	Orlando	*Equus caballus*	2691	-	-
Ridala_Rid1_2717	Fages et al 2019	*Equus caballus*	2717	N	251
Ridala_Rid2_2717	Fages et al 2019	*Equus caballus*	2717	N	200
Arzhan_I-K2_Arz1_2727	PRJEB19970	*Equus caballus*	2727	N	150
Arzhan_I-K3_Arz2_2727	PRJEB19970	*Equus caballus*	2727	N	823
TepeHasanlu_3394_2808	Fages et al 2019	*Equus caballus*	2808	Y	8.51
TepeHasanlu_2405_2886	Fages et al 2019	*Equus caballus*	2886	N	162
Yerqorqan_YER28_2895	Fages et al 2019	*Equus caballus*	2893	N	53.9
Fengtai_Fen4_2917	Fages et al 2019	*Equus caballus*	2917	Y	142
TepeHasanlu_3461_2930	Fages et al 2019	*Equus caballus*	2930	Y	4.15
TastsiinEreg_GVA9036_3018	Orlando	*Equus caballus*	3018	-	-
ZuunKhangai257_Mon2017x188_3025	Orlando	*Equus caballus*	3025	-	-
UushgiinUvur_Mon86_3039	Fages et al 2019	*Equus caballus*	3039	N	186
ZuniiGol_Mon2017x172_3068	Orlando	*Equus caballus*	3068	-	-
UushgiinUvur_Mon45_3080	Fages et al 2019	*Equus caballus*	3080	Y	205
Burgast_GVA9042_3101	Orlando	*Equus caballus*	3101	-	-
UushgiinUvur_Mon37_3105	Fages et al 2019	*Equus caballus*	3105	N	90.3
UushgiinUvur_Mon39_3105	Fages et al 2019	*Equus caballus*	3105	N	130
Sagzabad_SAGS27_3117	Fages et al 2019	*Equus caballus*	3117	Y	48
UushgiinUvur_Mon87_3117	Fages et al 2019	*Equus caballus*	3117	Y	175
UushgiinUvur_Mon43_3120	Fages et al 2019	*Equus caballus*	3120	Y	180
UushgiinUvur_Mon84_3123	Fages et al 2019	*Equus caballus*	3123	N	308
UushgiinUvur_Mon42_3130	Fages et al 2019	*Equus caballus*	3130	Y	323.4
ZuunKhangai1_Mon2017x138x1_3133	Orlando	*Equus caballus*	3133	-	-
ZeerdegchingiinKhoshuu_Mon2017x159_3140	Orlando	*Equus caballus*	3140	-	-
Bateni_Rus14_3318	Fages et al 2019	*Equus caballus*	3318	N	129
KarasukRriver_Rus16_3350	Orlando	*Equus caballus*	3350	-	-
Garbovat_Gar3_3574	Fages et al 2019	*Equus caballus*	3574	N	55.4
TachtiPerda_TP4_3604	Fages et al 2019	*Equus caballus*	3604	N	100
ElAcequion_Spain39_3993	Fages et al 2019	*Equus caballus*	3993	Y	173
Halvai_KSH5_4017	Fages et al 2019	*Equus caballus*	4017	N	0.509
Sintashta_838-64_NB46_4023	PRJEB19970	*Equus caballus*	4023	N	431
ElAcequion_Spain38_4055	Fages et al 2019	*Equus caballus*	4055	Y	386
Dunaujvaros_Duk2_4077	Fages et al 2019	*Equus caballus*	4077	N	41.5
CaminoDeLasYeseras_CDY2_4678	Fages et al 2019	*Equus caballus*	4678	N	19.5
Cantorella_UE2275x2_4791	Fages et al 2019	*Equus caballus*	4791	Y	149
Borly4_PAVH4_4974	Fages et al 2019	*Equus caballus*	4974	N	197
Borly4_PAVH9_4977	Fages et al 2019	*Equus caballus*	4977	N	269
Borly4_PAVH8_4978	Fages et al 2019	*Equus caballus*	4978	N	198
Borly4_PAVH6_5012	Fages et al 2019	*Equus caballus*	5012	N	465
Borly4_PAVH11_5015	Fages et al 2019	*Equus caballus*	5015	N	231
Batagai_KT368725_5155	KT368725	*Equus caballus*	5155	N	500
Yukagir_KT368723_5451	KT368723	*Equus caballus*	5451	-	-
Botai_1_5500	Fages et al 2019	*Equus caballus*	5500	N	280
Botai_2_5500	Fages et al 2019	*Equus caballus*	5500	N	599
Botai_3_5500	Fages et al 2019	*Equus caballus*	5500	N	72.5
Botai_4_5500	Fages et al 2019	*Equus caballus*	5500	N	239
Botai_5_5500	Fages et al 2019	*Equus caballus*	5500	N	248
Botai_6_5500	Fages et al 2019	*Equus caballus*	5500	N	479
Botai_8_5500	Fages et al 2019	*Equus caballus*	5500	N	155
Botai_C_5500	Fages et al 2019	*Equus caballus*	5500	N	127
Botai_D1_5500	Fages et al 2019	*Equus caballus*	5500	N	176
Botai_D4_5500	Fages et al 2019	*Equus caballus*	5500	N	237
Botai_D5_5500	Fages et al 2019	*Equus caballus*	5500	N	239
Botai_D6_5500	Fages et al 2019	*Equus caballus*	5500	N	250
Botai_F_5500	Fages et al 2019	*Equus caballus*	5500	N	136
Botai_G_5500	Fages et al 2019	*Equus caballus*	5500	N	221
Botai_I_5500	Fages et al 2019	*Equus caballus*	5500	N	130
Botai_K_5500	Fages et al 2019	*Equus caballus*	5500	N	192

Botai_L_5500	Fages et al 2019	*Equus caballus*	5500	N	169
Botai_P_5500	Fages et al 2019	*Equus caballus*	5500	N	113
Botai_Petrous_5500	Fages et al 2019	*Equus caballus*	5500	N	56
Botai_R_5500	Fages et al 2019	*Equus caballus*	5500	N	74.4
Taymyr_CGG10023_KT757741_16056	SAMEA2821681	*Equus caballus*	16056	N	245
UralMountains_158_KT757759_16946	Orlando	*Equus caballus*	16946	-	-
UralMountains_149_KT757758_18538	Orlando	*Equus caballus*	18538	-	-
NewSiberianIslands_155_KT757747_20340	Orlando	*Equus caballus*	20340	-	-
NewSiberianIslands_156_KT757748_24220	Orlando	*Equus caballus*	24220	-	-
Taymyr_CGG10026_KT757742_27298	Orlando	*Equus caballus*	27298	-	-
Taymyr_CGG10032_KT757744_28264	Orlando	*Equus caballus*	28264	-	-
UralMountains_160_KT757755_28308	Orlando	*Equus caballus*	28308	-	-
Taymyr_CGG10027_KT757743_28403	Orlando	*Equus caballus*	28403	-	-
UralMountains_148_KT757754_31984	Orlando	*Equus caballus*	31984	-	-
UralMountains_159_KT757756_32734	Orlando	*Equus caballus*	32734	-	-
Goyet_Vert311_35870	Fages et al 2019	*Equus caballus*	35870	Y	22.6
UralMountains_151_KT757757_39088	Orlando	*Equus caballus*	39088	-	-
NewSiberianIslands_152_KT757750_39377	Orlando	*Equus caballus*	39377	-	-
Taymyr_CGG10022_KT757740_42758	SAMEA2821680	*Equus caballus*	42758	N	127
AA348_0	Frantz et al 2019	*Sus scrofa*	0	-	0.30265455
GL946_0	DQ872971	*Sus scrofa*	0	-	2.33058448
GL951_0	DQ872975	*Sus scrofa*	0	-	2.57870343
GL943_0	Frantz et al 2019	*Sus scrofa*	0	-	4.14440498
GL950_0	DQ872974	*Sus scrofa*	0	-	9.40552579
AA133_0	Frantz et al 2019	*Sus scrofa*	0	-	129.061699
AA349_0	Frantz et al 2019	*Sus scrofa*	0	-	274.535484
AS01F01_0	Frantz et al 2015	*Sus scrofa*	0	-	-
AS01F09_0	Frantz et al 2015	*Sus scrofa*	0	-	-
CA01F14_0	Frantz et al 2015	*Sus scrofa*	0	-	-
CM01F17_0	Frantz et al 2015	*Sus scrofa*	0	-	-
CM01F18_0	Frantz et al 2015	*Sus scrofa*	0	-	-
CS01F02_0	Frantz et al 2015	*Sus scrofa*	0	-	-
CT01F13_0	Frantz et al 2015	*Sus scrofa*	0	-	-
GO01F04_0	Frantz et al 2015	*Sus scrofa*	0	-	-
KF888634.1_reference	KF888634.1	*Sus scrofa*	0	-	-
LW37F01_0	Frantz et al 2015	*Sus scrofa*	0	-	-
warthog_DQ409327.1_0	DQ409327	*Phacochoerus africanus*	0	-	-
AA280_700	Frantz et al 2019	*Sus scrofa*	700	N	2.32715343
KD024_850	Frantz et al 2019	*Sus scrofa*	850	N	36.563655
AA511_865	Frantz et al 2019	*Sus scrofa*	865	Y	3.34822127
AA333_875	Frantz et al 2019	*Sus scrofa*	875	N	3.84620478
AA411_885	Frantz et al 2019	*Sus scrofa*	885	N	1.09709264
AA415_885	Frantz et al 2019	*Sus scrofa*	885	N	1.161801
AA444_885	Frantz et al 2019	*Sus scrofa*	885	N	1.17564558
AA442_885	Frantz et al 2019	*Sus scrofa*	885	N	1.19388431
AA439_885	Frantz et al 2019	*Sus scrofa*	885	N	11.3019322
AA443_885	Frantz et al 2019	*Sus scrofa*	885	N	32.6945765
AA114_900	Frantz et al 2019	*Sus scrofa*	900	N	3.64876904
AA449_950	Frantz et al 2019	*Sus scrofa*	950	N	1.1539758
AA448_950	Frantz et al 2019	*Sus scrofa*	950	N	2.4493469
AA451_950	Frantz et al 2019	*Sus scrofa*	950	N	224.875158
AA016_1000	Frantz et al 2019	*Sus scrofa*	1000	N	2.78667309
AL1062_6850	Frantz et al 2019	*Sus scrofa*	1062	N	2.08926744
AA401_1065	Frantz et al 2019	*Sus scrofa*	1065	N	2.33594173
AL1068_6850	Frantz et al 2019	*Sus scrofa*	1068	N	1.54511527
AA231_1450	Frantz et al 2019	*Sus scrofa*	1450	N	1.08427135
AA284_1700	Frantz et al 2019	*Sus scrofa*	1700	N	2.6237886
AA409_1800	Frantz et al 2019	*Sus scrofa*	1800	N	17.331066
AA301_1800	DQ873163	*Sus scrofa*	1800	N	36.2903148
AA266_2000	DQ873163	*Sus scrofa*	2000	N	35.4696322
KD025_2375	Frantz et al 2019	*Sus scrofa*	2375	N	305.079998
AA090_2435	Frantz et al 2019	*Sus scrofa*	2435	N	18.899055
AL613_2718	Frantz et al 2019	*Sus scrofa*	2718	N	24.9794137

AA091_2999	Frantz et al 2019	*Sus scrofa*	2999	N	3.18557756
AL1107_3000	Frantz et al 2019	*Sus scrofa*	3000	N	32.8135195
AA119_3050	Frantz et al 2019	*Sus scrofa*	3050	N	492.275989
AA089_3300	Frantz et al 2019	*Sus scrofa*	3300	N	1.09516644
AA079_3350	Frantz et al 2019	*Sus scrofa*	3350	N	6.23993258
AA324_3700	Frantz et al 2019	*Sus scrofa*	3700	Stratigraphically	18.0204057
AA325_3700	Frantz et al 2019	*Sus scrofa*	3700	Contextual	27.2520315
AA525_4000	Frantz et al 2019	*Sus scrofa*	4000	Stratigraphically	2.30096912
AA361_4050	Frantz et al 2019	*Sus scrofa*	4050	N	7.73346175
AA021_4400	Frantz et al 2019	*Sus scrofa*	4400	N	2.63281767
AA097_4400	Frantz et al 2019	*Sus scrofa*	4400	N	12.3950521
AA364_4492	Frantz et al 2019	*Sus scrofa*	4492	N	6.49942816
AA369_4492	Frantz et al 2019	*Sus scrofa*	4492	N	12.5607055
AA363_4492	Frantz et al 2019	*Sus scrofa*	4492	N	280.613315
AA254_4610	Frantz et al 2019	*Sus scrofa*	4610	N	1.60121592
AA507_4623	Frantz et al 2019	*Sus scrofa*	4623	Y	8.95942936
AA015_4700	Frantz et al 2019	*Sus scrofa*	4700	N	2.97291278
AL612_5100	Frantz et al 2019	*Sus scrofa*	5100	N	22.8975501
AA240_5112	Frantz et al 2019	*Sus scrofa*	5112	N	4.00258833
AL1108_5500	Frantz et al 2019	*Sus scrofa*	5500	N	1.06555107
AL1105_5500	Frantz et al 2019	*Sus scrofa*	5500	N	3.17703004
AL1106_5500	Frantz et al 2019	*Sus scrofa*	5500	N	30.9966893
AA506_6000	Frantz et al 2019	*Sus scrofa*	6000	N	1.49882622
AA495_6000	Frantz et al 2019	*Sus scrofa*	6000	N	1.60880034
AA500_6000	Frantz et al 2019	*Sus scrofa*	6000	N	2.35099019
AA503_6000	Frantz et al 2019	*Sus scrofa*	6000	N	3.77180521
AA493_6000	Frantz et al 2019	*Sus scrofa*	6000	N	5.19996388
AA497_6000	Frantz et al 2019	*Sus scrofa*	6000	N	5.99012821
AA017_6200	Frantz et al 2019	*Sus scrofa*	6200	N	1.21681815
AA140_6300	Frantz et al 2019	*Sus scrofa*	6300	N	2.68536688
BLT024_6500	Frantz et al 2019	*Sus scrofa*	6500	N	0.10257028
BLT014_6500	Frantz et al 2019	*Sus scrofa*	6500	N	1.04857642
VEM183_6500	Frantz et al 2019	*Sus scrofa*	6500	N	1.23228797
AA424_6500	Frantz et al 2019	*Sus scrofa*	6500	N	2.22975983
VEM184_6500	Frantz et al 2019	*Sus scrofa*	6500	N	2.45957985
AA158_6500	Frantz et al 2019	*Sus scrofa*	6500	N	2.64876904
AA068_6500	DQ873096	*Sus scrofa*	6500	N	3.02811052
AA179_6500	Frantz et al 2019	*Sus scrofa*	6500	N	4.32974177
VEM185_6500	Frantz et al 2019	*Sus scrofa*	6500	N	372.0434
AA384_6550	Frantz et al 2019	*Sus scrofa*	6550	N	1.44429062
AA385_6550	Frantz et al 2019	*Sus scrofa*	6550	N	5.97339433
AA346_6700	Frantz et al 2019	*Sus scrofa*	6700	N	1.67790285
AA217_6750	Frantz et al 2019	*Sus scrofa*	6750	N	1.07145007
AA211_6750	Frantz et al 2019	*Sus scrofa*	6750	N	1.17763198
AA327_6750	Frantz et al 2019	*Sus scrofa*	6750	N	1.23427436
AA219_6750	Frantz et al 2019	*Sus scrofa*	6750	N	2.12803226
AL1110_6750	Frantz et al 2019	*Sus scrofa*	6750	N	2.13892735
AA216_6750	Frantz et al 2019	*Sus scrofa*	6750	N	2.15247096
AA213_6750	Frantz et al 2019	*Sus scrofa*	6750	N	3.44272558
AL1111_6750	Frantz et al 2019	*Sus scrofa*	6750	N	27.9820021
AL1114_6800	Frantz et al 2019	*Sus scrofa*	6800	N	3.0010233
AA479_7000	Frantz et al 2019	*Sus scrofa*	7000	N	0.56985493
AA488_7000	Frantz et al 2019	*Sus scrofa*	7000	N	1.29073617
AA376_7000	Frantz et al 2019	*Sus scrofa*	7000	N	3.06988503
AA486_7000	Frantz et al 2019	*Sus scrofa*	7000	N	4.37278035
AA490_7000	Frantz et al 2019	*Sus scrofa*	7000	N	5.36104256
AA372_7000	Frantz et al 2019	*Sus scrofa*	7000	N	6.57858304
AA481_7000	Frantz et al 2019	*Sus scrofa*	7000	N	10.4545236
AA478_7000	Frantz et al 2019	*Sus scrofa*	7000	N	11.1772708
AA476_7000	Frantz et al 2019	*Sus scrofa*	7000	N	14.987901
AA382_7000	Frantz et al 2019	*Sus scrofa*	7000	N	19.5098417
AA489_7000	Frantz et al 2019	*Sus scrofa*	7000	N	20.5447541
AA483_7000	Frantz et al 2019	*Sus scrofa*	7000	N	25.1029314
AA477_7000	Frantz et al 2019	*Sus scrofa*	7000	N	26.3601998

AA484_7000	Frantz et al 2019	*Sus scrofa*	7000	N	26.5221212
AA480_7000	Frantz et al 2019	*Sus scrofa*	7000	N	71.8765425
AA287_7100	DQ873158	*Sus scrofa*	7100	N	2.0653103
AA288_7100	DQ873159	*Sus scrofa*	7100	N	151.330223
KD033_7100	Frantz et al 2019	*Sus scrofa*	7100	N	249.5155
KD037_7100	Frantz et al 2019	*Sus scrofa*	7100	N	341.078071
BLT010_7148	KJ416030.1	*Sus scrofa*	7148	Y	2.01041353
AA294_7200	Frantz et al 2019	*Sus scrofa*	7200	N	3.69716487
AA513_7418	Frantz et al 2019	*Sus scrofa*	7418	Y	1.44073918
AA241_7750	Frantz et al 2019	*Sus scrofa*	7550	N	2.85445133
AA351_7850	Frantz et al 2019	*Sus scrofa*	7850	N	3.70041534
AA353_7850	Frantz et al 2019	*Sus scrofa*	7850	N	4.68295913
AA354_7850	Frantz et al 2019	*Sus scrofa*	7850	N	7.40221513
AA274_8100	Frantz et al 2019	*Sus scrofa*	8100	N	1.17805333
AA405_8275	DQ873168	*Sus scrofa*	8275	Contextual	17.5112863
AA306_8500	Frantz et al 2019	*Sus scrofa*	8500	N	30.4119063
AL1103_8650	Frantz et al 2019	*Sus scrofa*	8650	N	2.66074761
AL1102_8650	Frantz et al 2019	*Sus scrofa*	8650	N	21.8981521
BLT025_8898	Frantz et al 2019	*Sus scrofa*	8898	Y	2.62848372
AA395_9250	Frantz et al 2019	*Sus scrofa*	9250	N	1.47041474
AA393_9250	Frantz et al 2019	*Sus scrofa*	9250	N	2.12062842
AA391_9250	Frantz et al 2019	*Sus scrofa*	9250	N	2.47727683
AA457_9268	DQ872999	*Sus scrofa*	9268	Contextual	30.3912599
AA458_9268	Frantz et al 2019	*Sus scrofa*	9268	Contextual	51.772347
AL720_9600	Frantz et al 2019	*Sus scrofa*	9600	N	13.8728706
AL726_9700	Frantz et al 2019	*Sus scrofa*	9700	N	3.35995907
AA515_9700	Frantz et al 2019	*Sus scrofa*	9700	N	3.8917715
AA459_9750	Frantz et al 2019	*Sus scrofa*	9750	Contextual	4.64160597
AL708_9775	Frantz et al 2019	*Sus scrofa*	9775	N	16.1917775
AL714_10200	Frantz et al 2019	*Sus scrofa*	10200	N	5.55179679
AL725_10200	Frantz et al 2019	*Sus scrofa*	10200	N	6.07542286
AL706_10200	Frantz et al 2019	*Sus scrofa*	10200	N	21.3153554
AL718_10200	Frantz et al 2019	*Sus scrofa*	10200	N	228.165352
GL210_10540	DQ873186	*Sus scrofa*	10540	N	1.05020165
GL980_10980	DQ873000	*Sus scrofa*	10980	N	30.7790285
AA523_11500	Frantz et al 2019	*Sus scrofa*	11500	N	27.9856739
AA522_13909	Frantz et al 2019	*Sus scrofa*	13909	Y	9.5477638
Japan_0	AB499824.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Sweden1_0	Arnason et al. 2007	*Canis lupus/Canis lupus familiaris*	0	-	-
D09_dingo_0	DQ480489.1	*Canis lupus/Canis lupus familiaris*	0	-	-
D07_dingo_0	DQ480491.1	*Canis lupus/Canis lupus familiaris*	0	-	-
D03_dingo_0	DQ480492.1	*Canis lupus/Canis lupus familiaris*	0	-	-
DQ480492.1_Jamthund_Sweden_D_0	DQ480492.1	*Canis lupus/Canis lupus familiaris*	0	-	-
D04_dingo_0	DQ480495.1	*Canis lupus/Canis lupus familiaris*	0	-	-
D01_A1b_0	DQ480499	*Canis lupus/Canis lupus familiaris*	0	-	-
D01_dingo_0	DQ480499.1	*Canis lupus/Canis lupus familiaris*	0	-	-
DQ480502.1_Jamthund_Sweden_D_0	DQ480502.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Spain_0	DQ480505.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Saudi_Arabia2_0	DQ480506.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Saudi_Arabia1_0	DQ480507.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada3_0	DQ480508.1	*Canis lupus/Canis lupus familiaris*	0	-	-
EU408288.1_NorwegianElkhound1P_D_0	EU408288.1	*Canis lupus/Canis lupus familiaris*	0	-	-
China2_0	EU442884	*Canis lupus/Canis lupus familiaris*	0	-	-
EU789662.1_E1_p77_E_0	EU789662.1	*Canis lupus/Canis lupus familiaris*	0	-	-
D96_A1a_0	FJ817358	*Canis lupus/Canis lupus familiaris*	0	-	-
D102_A1a_0	FJ817359	*Canis lupus/Canis lupus familiaris*	0	-	-
D97_A1a_0	FJ817362	*Canis lupus/Canis lupus familiaris*	0	-	-
D99_B_0	FJ817363	*Canis lupus/Canis lupus familiaris*	0	-	-
D98_B_0	FJ817364	*Canis lupus/Canis lupus familiaris*	0	-	-
Mongolia_0	GQ374438.1	*Canis lupus/Canis lupus familiaris*	0	-	-
JF342812.1_Alaskan_Malamute_A1a_0	JF342812.1	*Canis lupus/Canis lupus familiaris*	0	-	-
JF342877_Samoyed_A1a_0	JF342877.1	*Canis lupus/Canis lupus familiaris*	0	-	-
NGSD_JX08864_dingo_0	JX08864	*Canis lupus/Canis lupus familiaris*	0	-	-

Basenji_A1a_0	KF661036.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Dingo_A1b_0	KF661037.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Finland_Tha1_0	KF661038	*Canis lupus/Canis lupus familiaris*	0	-	-
Russia1_0	KF661039	*Canis lupus/Canis lupus familiaris*	0	-	-
Sweden2_Tha3_0	KF661040	*Canis lupus/Canis lupus familiaris*	0	-	-
China3_Tha4_0	KF661041	*Canis lupus/Canis lupus familiaris*	0	-	-
Israel1_Tha5_0	KF661042	*Canis lupus/Canis lupus familiaris*	0	-	-
India_Tha6_0	KF661043	*Canis lupus/Canis lupus familiaris*	0	-	-
Russia2_0	KF661044	*Canis lupus/Canis lupus familiaris*	0	-	-
Poland1_Tha8_0	KF661045	*Canis lupus/Canis lupus familiaris*	0	-	-
Ukraine_Tha10_0	KF661047	*Canis lupus/Canis lupus familiaris*	0	-	-
Italy_0	KF661048.1	*Canis lupus/Canis lupus familiaris*	0	-	-
Poland2_Tha12_0	KF661049	*Canis lupus/Canis lupus familiaris*	0	-	-
Oman_Tha13_0	KF661050	*Canis lupus/Canis lupus familiaris*	0	-	-
Iran_Tha14_0	KF661051	*Canis lupus/Canis lupus familiaris*	0	-	-
Sweden3_Tha15_0	KF661052	*Canis lupus/Canis lupus familiaris*	0	-	-
China4_0	KF661053	*Canis lupus/Canis lupus familiaris*	0	-	-
Croatia_Tha17_0	KF661054	*Canis lupus/Canis lupus familiaris*	0	-	-
Israel2_Tha18_0	KF661055	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada1_Tha19_0	KF661056	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada2_Tha20_0	KF661057	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska1_Tha21_0	KF661058	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada4_Tha22_0	KF661059	*Canis lupus/Canis lupus familiaris*	0	-	-
Mexico1_Tha23_0	KF661060	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada5_Tha24_0	KF661061	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada6_Tha25_0	KF661062	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada7_Tha27_0	KF661063	*Canis lupus/Canis lupus familiaris*	0	-	-
USA1_Tha27_0	KF661064	*Canis lupus/Canis lupus familiaris*	0	-	-
Mexico2_Tha28_0	KF661065	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska2_Tha29_0	KF661066	*Canis lupus/Canis lupus familiaris*	0	-	-
U.Arizona-2_Alaska_0	KF661066	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska3_Tha30_0	KF661067	*Canis lupus/Canis lupus familiaris*	0	-	-
U.Arizona-3_Alaska_0	KF661067	*Canis lupus/Canis lupus familiaris*	0	-	-
USA2_Tha31_0	KF661068	*Canis lupus/Canis lupus familiaris*	0	-	-
USA3_Tha32_0	KF661069	*Canis lupus/Canis lupus familiaris*	0	-	-
USA4_Tha33_0	KF661070	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska4_Tha34_0	KF661071	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska5_Tha35_0	KF661072	*Canis lupus/Canis lupus familiaris*	0	-	-
Alaska6_Tha36_0	KF661073	*Canis lupus/Canis lupus familiaris*	0	-	-
U.Arizona-6_Alaska_0	KF661073	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada8_Tha37_0	KF661074	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada9_Tha38_0	KF661075	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada10_Tha39_0	KF661076	*Canis lupus/Canis lupus familiaris*	0	-	-
Canada11_Tha40_0	KF661077	*Canis lupus/Canis lupus familiaris*	0	-	-
KJ139384.1_1643_TAR3_D_0	KJ139384.1	*Canis lupus/Canis lupus familiaris*	0	-	-
KJ139385.1_1646_TAR5_D_0	KJ139385.1	*Canis lupus/Canis lupus familiaris*	0	-	-
KJ139386.1_1662_TAR3_D_0	KJ139386.1	*Canis lupus/Canis lupus familiaris*	0	-	-
KJ139388.1_1695_CAL1_D_0	KJ139388.1	*Canis lupus/Canis lupus familiaris*	0	-	-
NC_002008.4_reference_0	NC_002008.4	*Canis lupus/Canis lupus familiaris*	0	-	-
China1_0	NC011218	*Canis lupus/Canis lupus familiaris*	0	-	-
D103_A1a_0	U96639	*Canis lupus/Canis lupus familiaris*	0	-	-
OL1600_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	31.95868954
OL1601_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	2.705446284
OL1602_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	6.826866742
OL1605_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	2.906259341
OL1606_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	13.89705267
OL1608_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	11.56579183
OL1609_Italy_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	18.12626293
A103_10002_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
A29_10100_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Alpine1_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Alpine2_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Alpine5_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
D02_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-

D05_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
D06_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
D08_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
D10_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
FraserIsland_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
GibsonDesert1_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
GibsonDesert2_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
GibsonDesert3_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
GibsonDesert4_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
GibsonDesert5_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Kimberley3_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Kimberley4_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
Kimberley5_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
ngsd1_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
ngsd2_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
SimpsonDesert1_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
SimpsonDesert2_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
SimpsonDesert3_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
SimpsonDesert4_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
SimpsonDesert5_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
OL1528_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	23.70006576
OL1530_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	1.659233574
OL1531_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	8.390865068
OL1532_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	4.421235129
OL1533_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	2.336880493
OL1534_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	2.86136187
OL1535_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	9.193758594
OL1537_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	15.34530998
OL1538_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	1.393674897
OL1540_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	1.347880672
OL1542_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	12.23004723
OL1543_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	3.036527769
OL1544_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	0.996233634
OL1545_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	7.64697794
OL1546_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	5.096670054
OL1547_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	0.953428589
OL1548_Indonesia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	8.241107192
AL3001_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	36.53942727
AL3003_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	46.07765887
AL3004_Russia_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	6.37370718
AL3005_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	3.312668141
AL3007_Russia_Xclade_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	54.40909906
AL3013_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	3.773061517
AL3014_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	12.14790459
AL3015_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	24.78334429
AL3020_Kazakhstan_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	6.155796018
AL3023_Russia_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	4.836192981
AL3027_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	4.156274287
AL3031_Russia_A1b_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	40.58067795
AL3032_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	12.94810785
AL3033_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	7.085490524
AL3052_Russia_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	4.633705984
AL3053_Russia_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	7.479225205
AL3054_Russia_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	5.267950021
AL3062_Russia_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	0.5392479225
AL3090_Russia_B_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	6.550726371
AL3299_Russia_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	7.133257607
OL1514_Peru_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	
OL1525_Peru_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	2.506426735
IN23_dingo_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	-
OL1515_Peru_A_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	16.23919412
GSD_19.5_Sisimiut_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	12295.7551
GSD_2.6_Sisimiut_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	17088.8223
GSD_51602_Tasiilaq_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	7425.2143

GSD_51603_Tasiilaq_Greenland_A1b_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	17355.9934
GSD_GS16_Ilulissat_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	553.0613
GSD_GS31_Ilulissat_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	179.8584
GSD_Mums_Aasiaat_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	292.8902
GSD_Pondus_Aasiaat_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	6065.4673
GSD_Q11_Qaanaaq_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	984.9097
GSD_QSON_Qaanaaq_Greenland_A1a_0	Unpublished	*Canis lupus/Canis lupus familiaris*	0	-	19213.3118
ChineseDog1_C_0	Wang et al. 2013	*Canis lupus/Canis lupus familiaris*	0	-	-
ChineseDog2_A1a_0	Wang et al. 2013	*Canis lupus/Canis lupus familiaris*	0	-	-
ChineseDog3_A1a_0	Wang et al. 2013	*Canis lupus/Canis lupus familiaris*	0	-	-
TRF.01.07_Monumentet_HallLand_GL_AD1985_A2a_20	Unpublished	*Canis lupus/Canis lupus familiaris*	20	N	4.0858
TRF.02.41_West_GL_AD1980_A1a_25	Unpublished	*Canis lupus/Canis lupus familiaris*	25	N	3.9800
TRF.01.35_Siorapaluk_Qaanaaq_GL_AD1977_A2a_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	5.4700
TRF.01.54_Siorapaluk_Greenland_AD1977_A2a_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	7.9186
TRF.01.55_Siorapaluk_Greenland_AD1977_A1a_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	8.8018
TRF.01.56_Siorapaluk_AD1977_A2a_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	13.4165
TRF.01.57_Siorapaluk_AD1977_A2a_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	10.4336
TRF.01.58_Siorapaluk_Qaanaaq_GL_AD1977_35	Unpublished	*Canis lupus/Canis lupus familiaris*	35	N	4.6263
TRF.02.40_West_GL_AD1962_A1a_45	Unpublished	*Canis lupus/Canis lupus familiaris*	45	N	6.0900
TRF.01.30_Qeqertarsuagg_GL_AD1950_A1a_50	Unpublished	*Canis lupus/Canis lupus familiaris*	50	N	19.1801
TRF.01.41_Hvidovre_AD1945_A1b_55	Unpublished	*Canis lupus/Canis lupus familiaris*	55	N	16.7096
TRF.02.27_Alaska_AD1945_55	Unpublished	*Canis lupus/Canis lupus familiaris*	55	N	4.9300
TRF.02.28_Alaska_AD1945_55	Unpublished	*Canis lupus/Canis lupus familiaris*	55	N	10.6900
TRF.02.29_Alaska_AD1945_Xclade_55	Unpublished	*Canis lupus/Canis lupus familiaris*	55	N	10.0700
TRF.02.16_St.LawrenceIsland_AD1939_A1a_60	Unpublished	*Canis lupus/Canis lupus familiaris*	60	N	23.9961
Canada_2040_A2a_70	Thalmann et al. 2013	*Canis lupus/Canis lupus familiaris*	70	N	-
Canada_2041_70	Thalmann et al. 2013	*Canis lupus/Canis lupus familiaris*	70	N	-
E_Greenland_2587_Greenland_A1b_70	Unpublished	*Canis lupus/Canis lupus familiaris*	70	N	397.0000
E_Greenland_3294_Greenland_A1b_70	Unpublished	*Canis lupus/Canis lupus familiaris*	70	N	108.0000
E_Greenland_35561_Greenland_A1b_70	Unpublished	*Canis lupus/Canis lupus familiaris*	70	N	437.0000
TRF.02.38_Tasiilaq_GL_AD1932_A1b_70	Unpublished	*Canis lupus/Canis lupus familiaris*	70	N	85.8600
W_Greenland_32987_A2a_70	Unpublished	*Canis lupus/Canis lupus familiaris*	70	N	211.0000
TRF.01.42_Scorsbysund_Greenland_AD1927_A1a_75	Unpublished	*Canis lupus/Canis lupus familiaris*	75	N	7.2359
TRF.02.14_Alaska_AD1926_A2a_75	Unpublished	*Canis lupus/Canis lupus familiaris*	75	N	21.5800
UA5231_Alaska_A2b_75	Unpublished	*Canis lupus/Canis lupus familiaris*	75	N	13.1
TRF.02.37_Tasiilaq_GL_AD1911_A1b_80	Unpublished	*Canis lupus/Canis lupus familiaris*	80	N	10.9400
W_Greenland_67941_A1a_80	Unpublished	*Canis lupus/Canis lupus familiaris*	80	N	79.0000
E_Greenland_32135_Greenland_90	Unpublished	*Canis lupus/Canis lupus familiaris*	90	N	31.0000
E_Greenland_32179_Greenland_90	Unpublished	*Canis lupus/Canis lupus familiaris*	90	N	12.0000
Canada_Sort_Nunavut_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	109.0000
Greenland_Obersten_2_A1a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	22.0000
KNK2643X1839_Greenland_A2b_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	-
M-14045_Greenland_A2a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	-
M-14445_Greenland_A2a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	-
N_Greenland_31867_A2a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	29.0000
TRF.01.06_HeklaHavn_GL_AD1895_A2a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	5.9871
TRF.01.46_SaundersIsland_GL_AD1903_A1a_100	Unpublished	*Canis lupus/Canis lupus familiaris*	100	N	23.4700
KNK2644X1119_Greenland_A2b_125	Unpublished	*Canis lupus/Canis lupus familiaris*	125	N	-
KNK2644X1120_Greenland_A2a_125	Unpublished	*Canis lupus/Canis lupus familiaris*	125	N	-
UA5072-7_Alaska_A2b_150	Unpublished	*Canis lupus/Canis lupus familiaris*	150	N	10.2
HJCL-9.07_Labrador_200	Unpublished	*Canis lupus/Canis lupus familiaris*	200	N	14.0
HJCL-9.14_Labrador_A2a_200	Unpublished	*Canis lupus/Canis lupus familiaris*	200	N	4.4
NHTN-1.2036Ha_NorthwestTerritory_A2a_250	Unpublished	*Canis lupus/Canis lupus familiaris*	250	N	2.4; 9.7
NHTN-1.2048H_NorthwestTerritory_A1b_250	Unpublished	*Canis lupus/Canis lupus familiaris*	250	N	28.3
2009-32-3316_Iceland_A2a_300	Unpublished	*Canis lupus/Canis lupus familiaris*	300	N	-
BELA-37369_Alaska_A1b_300	Unpublished	*Canis lupus/Canis lupus familiaris*	300	N	37.8
OKRN-1.TUR_NorthwestTerritory_A2a_300	Unpublished	*Canis lupus/Canis lupus familiaris*	300	N	6.0
TRF.01.11_Nugarsuk_GL_Thule_Greenland_A2a_300	Unpublished	*Canis lupus/Canis lupus familiaris*	300	N	5.0300
TRF.07.03_DMP_Labrador_CA_PostContact_Canada_A2a_300	Unpublished	*Canis lupus/Canis lupus familiaris*	300	N	4.7800
P91_2013KMG-GeoSociety-M4a_Greenland_A2a_350	Unpublished	*Canis lupus/Canis lupus familiaris*	350	N	0.8; 6.1
AL3130_Russia_A_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	5.020147068
AL3134_Russia_A_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	3.138040294
AL3137_Russia_A_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	3.928080349
OATI-1.F2-M1_NorthwestTerritory_A2b_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	5.9

OATI-1.H1R_NorthwestTerritory_A2a_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	4.5
SGFQ-1.H9-1_Nunavut_A2a_400	Unpublished	*Canis lupus/Canis lupus familiaris*	400	N	7.9
AL2784_USA_A1b_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	12.53153584
AL2788_USA_A1b_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	5.899742931
AL2791_USA_A_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	16.06486519
AL2794_USA_A1b_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	6.339212052
AL2795_USA_A1b_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	4.122197645
AL2797_USA_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	33.72947928
BELA-37371_Alaska_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	19.7
BELA-52965_Alaska_A2b_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	5.9
KKJG-1.H8-M1_Nunavut_A2a_500	Unpublished	*Canis lupus/Canis lupus familiaris*	500	N	-
09.P05.H2.894_Russia_A1b_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	-
30.KM90.2_2.626_Russia_A2b_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	-
31.KM90.2_2.307_Russia_A2b_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	-
KNK2643X1838_Greenland_A2a_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	-
TRF.01.01_Stormbugt_GL_Thule_Greenland_A2a_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	34.3700
TRF.01.08_Misigtoq_GL_Thule_Greenland_A1b_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	20.2000
TRF.01.09_Doedemandsbugt_GL_Thule_Greenland_A1a_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	6.2019
TRF.01.32_Kuk_CA_Thule_Canada_A2a_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	22.0100
TRF.01.34_Qilalukan_CA_Thule_Canada_A1b_600	Unpublished	*Canis lupus/Canis lupus familiaris*	600	N	17.2000
KNK492X33_Greenland_A2a_650	Unpublished	*Canis lupus/Canis lupus familiaris*	650	N	-
AL3239_USA_A_675	Unpublished	*Canis lupus/Canis lupus familiaris*	675	N	12.81150236
BELA-37374_Alaska_A2b_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	34.2
BELA-37375_Alaska_A1b_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	7.8
NAPI-2.14_NorthwestTerritory_A2b_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	4.4
NAPI-2.C70-10_NorthwestTerritory_A2b_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	5.1
NIHF-4.132C_Nunavut_A2a_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	14.3
NIHF-4.1C_Nunavut_A1b_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	68.9
NIHF-4.PLM_Nunavut_A2a_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	16.5
RBJR-1.642.56_Nunavut_A2a_700	Unpublished	*Canis lupus/Canis lupus familiaris*	700	N	20.4
KEDQ-2.M1_Nunavut_A2b_750	Unpublished	*Canis lupus/Canis lupus familiaris*	750	N	11.7
NCPF-1.8-20_NorthwestTerritory_A2a_750	Unpublished	*Canis lupus/Canis lupus familiaris*	750	N	3.6
OHRH-1.1462M2b_NorthwestTerritory_A1b_750	Unpublished	*Canis lupus/Canis lupus familiaris*	750	N	7.8; 9.5
SFFK-4.H4-H1_Nunavut_A1b_750	Unpublished	*Canis lupus/Canis lupus familiaris*	750	N	3.0
AL3206_Peru_A_800	Unpublished	*Canis lupus/Canis lupus familiaris*	800	N	30.53637831
AL2829_Peru_A_800	Unpublished	*Canis lupus/Canis lupus familiaris*	800	N	23.90291146
KKDO-2270_Nunavut_A2a_800	Unpublished	*Canis lupus/Canis lupus familiaris*	800	N	-
TRF.01.45_Inugsuk_GL_Thule_Greenland_A2a_800	Unpublished	*Canis lupus/Canis lupus familiaris*	800	N	26.8800
UA1-1941-1856_Alaska_A2b_800	Unpublished	*Canis lupus/Canis lupus familiaris*	800	N	8.7
CGG16_Yana_Russia_819	Unpublished	*Canis lupus/Canis lupus familiaris*	819	N	-
CGG31_MaliyLyakhov_Russia_827	Unpublished	*Canis lupus/Canis lupus familiaris*	827	N	-
5MT501_A2b_900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	900	N	76.37053865
AL3271_Russia_A_900	Unpublished	*Canis lupus/Canis lupus familiaris*	900	N	1.066778263
CK-H4-M2_Alaska_A2b_900	Unpublished	*Canis lupus/Canis lupus familiaris*	900	N	5.1
CK-H6-M5_Alaska_A2b_900	Unpublished	*Canis lupus/Canis lupus familiaris*	900	N	6.7
CK-H8-M7_Alaska_A1b_900	Unpublished	*Canis lupus/Canis lupus familiaris*	900	N	13.6
OP-B1-M2_Alaska_900	Unpublished	*Canis lupus/Canis lupus familiaris*	900	N	8.2
AL3223_USA_A_965	Unpublished	*Canis lupus/Canis lupus familiaris*	965	N	70.21372631
Argentina_A2b_1000	KF661084.1	*Canis lupus/Canis lupus familiaris*	1000	N	-
USA_A2b_1000	KF661086.1	*Canis lupus/Canis lupus familiaris*	1000	N	-
AM310A_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	102.7144736
AM310B_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	61.83386142
AM310C_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	47.30860286
AM474_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	47.62306451
May10_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	12.72218569
May2_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	61.54397083
May3_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	42.79273032
May4_A2b_1000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1000	N	63.31613559
01.P09.B1.1077_Chukotka_Russia_A1b_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	-
03.P04.H1.1024_Chukotka_Russia_Xclade_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	-
08.P05.H2.961_Russia_A1b_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	-
560028-28_Alaska_A2b_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	-
KBP2569.283_Russia_A2b_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	5.0
KP-1_Alaska_A2b_1000	Unpublished	*Canis lupus/Canis lupus familiaris*	1000	N	-

AL3226_USA_A_1035	Unpublished	*Canis lupus/Canis lupus familiaris*	1035	N	5.638070186
KBP3316.216_Russia_1100	Unpublished	*Canis lupus/Canis lupus familiaris*	1100	N	18.7
KBP3316.229_Russia_1100	Unpublished	*Canis lupus/Canis lupus familiaris*	1100	N	8.4
NIHF-4.9BSMb_Nunavut_A2b_1100	Unpublished	*Canis lupus/Canis lupus familiaris*	1100	N	6.7
NIHF-4.BSP_Nunavut_A2b_1100	Unpublished	*Canis lupus/Canis lupus familiaris*	1100	N	5.6
AL3170_Russia_A_1180	Unpublished	*Canis lupus/Canis lupus familiaris*	1180	N	40.1744485
5MT520_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	5.886829677
JBG11_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	162.6978538
JBG12_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	153.6255156
JBG13_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	157.7362348
JBG17_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	165.3036408
JBG19_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	129.5464817
JBG21_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	79.32217373
JBG24_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	154.409398
JBG26_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	137.4965027
JBG32_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	140.0304896
JBG35_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	25.7563819
JBG37_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	128.6252167
JBG41_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	168.990614
JBG42_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	158.2733903
JBG43_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	154.8202905
JBG45_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	154.8202905
JBG48_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	149.6821905
JBG5_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	122.7431099
JBG50_A2b_1200	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1200	N	50.2458301
DRG-99-0043_Alaska_1200	Unpublished	*Canis lupus/Canis lupus familiaris*	1200	N	3.1
WAL-B9-J19_Alaska_A2a_1200	Unpublished	*Canis lupus/Canis lupus familiaris*	1200	N	37.0
WAL-B9-J20_Alaska_A2a_1200	Unpublished	*Canis lupus/Canis lupus familiaris*	1200	N	15.7
AL3175_Turkey_A_1231	Unpublished	*Canis lupus/Canis lupus familiaris*	1231	N	11.90745501
AL3178_Turkey_A_1231	Unpublished	*Canis lupus/Canis lupus familiaris*	1231	N	4.932683685
5MT316_A2b_1250	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1250	N	39.31751061
AL3231_Mexico_A_1275	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1275	N	13.61714593
560028-29_Alaska_A2b_1280	Unpublished	*Canis lupus/Canis lupus familiaris*	1280	N	-
PRD1_A2b_1500	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1500	N	129.2258026
PRD10_A2b_1500	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1500	N	14.05033778
PRD9_A2b_1500	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1500	N	127.813535
PRW89_A2b_1500	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1500	N	4.785795421
35.KM90.2_5.873_Russia_A1a_1500	Unpublished	*Canis lupus/Canis lupus familiaris*	1500	N	-
AL2997_Russia_1525	Unpublished	*Canis lupus/Canis lupus familiaris*	1525	N	3.548633945
AL2922_Netherlands_A_1650	Unpublished	*Canis lupus/Canis lupus familiaris*	1650	N	30.64440725
AL2925_Netherlands_A_1650	Unpublished	*Canis lupus/Canis lupus familiaris*	1650	N	43.35583189
AL2803_USA_A_1700	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1700	N	10.04603336
AL2800_USA_A_1700	Unpublished	*Canis lupus/Canis lupus familiaris*	1700	N	4.229748311
AL2801_USA_A_1700	Unpublished	*Canis lupus/Canis lupus familiaris*	1700	N	7.559574341
AL2956_UK_A_1704	Unpublished	*Canis lupus/Canis lupus familiaris*	1704	N	17.3506905
CGG11_Aachim_Siberia_A1b_1720	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1720	Y	96.7221259
CGG10_Aachim_Siberia_A1b_1750	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1750	Y	28.83158965
ISM070_A2b_1750	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1750	N	26.36659293
AL2924_Netherlands_A_1750	Unpublished	*Canis lupus/Canis lupus familiaris*	1750	N	9.616548096
AL3198_USA_A_1800	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1800	N	4.788067197
OL4050_Spain_A_1800	Unpublished	*Canis lupus/Canis lupus familiaris*	1800	N	10.79512166
AL2754_USA_A_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	2.037663658
OSU611_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	13.16984516
OSU622_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	30.50002989
OSU624_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	89.69241346
OSU626_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	35.4464638
OSU628_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	29.37203324
OSU634_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	45.14933939
OSU638_A2b_1900	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	1900	N	61.43151791
560028-21_Alaska_A2b_1900	Unpublished	*Canis lupus/Canis lupus familiaris*	1900	N	-
AL3199_USA_A_2000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	2000	N	0.033777725
AL3202_USA_A_2000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	2000	N	0.261314043
AL3103_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	9.04537574
AL3107_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	4.63830932

AL3110_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	16.85938901
AL3111_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	8.601183715
AL3113_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	32.59771627
AL3114_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	26.48849166
AL3127_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	13.08596879
AL3144_Russia_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	0.9449393197
AL2966_UK_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	13.67639146
AL2975_UK_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	10.24905841
AL2022_Turkey_A_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	206.3075865
UA1-1939-1497-2_Alaska_A2b_2000	Unpublished	*Canis lupus/Canis lupus familiaris*	2000	N	6.8
SEL-33-0057b_Alaska_A2a_2100	Unpublished	*Canis lupus/Canis lupus familiaris*	2100	N	0.0; 6.4
AL3168_Russia_A_2185	Unpublished	*Canis lupus/Canis lupus familiaris*	2185	N	10.3029832
AL3161_Russia_A_2200	Unpublished	*Canis lupus/Canis lupus familiaris*	2200	N	23.27261314
AL3162_Russia_A_2200	Unpublished	*Canis lupus/Canis lupus familiaris*	2200	N	44.25264542
KDDQ-9_Nunavut_A2b_2200	Unpublished	*Canis lupus/Canis lupus familiaris*	2200	N	4.5
Cox6_A2b_2250	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	2250	N	16.97871704
AL2992_Russia_2400	Unpublished	*Canis lupus/Canis lupus familiaris*	2400	N	14.84133437
AL2834_Peru_A_2500	Unpublished	*Canis lupus/Canis lupus familiaris*	2500	N	10.81873618
OJRL-3.534_NorthwestTerritory_A2b_2500	Unpublished	*Canis lupus/Canis lupus familiaris*	2500	N	4.2
AL3120_Russia_A_2800	Unpublished	*Canis lupus/Canis lupus familiaris*	2800	N	17.0844742
AL3122_Russia_A_2800	Unpublished	*Canis lupus/Canis lupus familiaris*	2800	N	3.06606086
AL2365_Kazakhstan_A_3000	Unpublished	*Canis lupus/Canis lupus familiaris*	3000	N	5.884856818
AL3203_Czech.Republic_A_3086	Unpublished	*Canis lupus/Canis lupus familiaris*	3086	N	7.507263705
AL3167_Russia_A_3366	Unpublished	*Canis lupus/Canis lupus familiaris*	3366	N	71.04280505
TRF.01.23_Deering_Alaska_Ipiutak_A1a_3700	Unpublished	*Canis lupus/Canis lupus familiaris*	3700	N	18.0239
TRF.01.24_Deering_Alaska_Ipiutak_A1b_3700	Unpublished	*Canis lupus/Canis lupus familiaris*	3700	N	23.4741
TRF.01.27_Deering_Alaska_Ipiutak_3700	Unpublished	*Canis lupus/Canis lupus familiaris*	3700	N	7.1600
AL2356_Russia_A_3800	Unpublished	*Canis lupus/Canis lupus familiaris*	3800	N	1.307706104
CICVD_A2b_4000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	4000	N	22.52561727
AL3104_Russia_4000	Unpublished	*Canis lupus/Canis lupus familiaris*	4000	N	43.13385544
RBJU-1.COLDa_Nunavut_A2b_4000	Unpublished	*Canis lupus/Canis lupus familiaris*	4000	N	4.8; 2.3
AL3194_Canada_A_4157	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	4157	Y	116.2334549
AL2574_Turkmenistan_B_4350	Unpublished	*Canis lupus/Canis lupus familiaris*	4350	N	0.396245591
Kirschbaumhoehle_Germany_C_4700	Unpublished	*Canis lupus/Canis lupus familiaris*	4700	N	-
OL4162_Greece_A_4750	Unpublished	*Canis lupus/Canis lupus familiaris*	4750	N	11.59024332
Newgrange_Ireland_C_4800	Frantz et al 2016	*Canis lupus/Canis lupus familiaris*	4800	Y	-
P35_A2b_5000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	5000	N	3.871644646
P59_A2b_5000	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	5000	-	26.48329049
AC14_Kazakhstan_C_5350	Unpublished	*Canis lupus/Canis lupus familiaris*	5350	N	3.573563699
AL2989_Russia_A_5500	Unpublished	*Canis lupus/Canis lupus familiaris*	5500	N	4.935194596
AL2571_Iran_A_5800	Unpublished	*Canis lupus/Canis lupus familiaris*	5800	N	65.80181742
AL2998_Russia_A1b_6000	Unpublished	*Canis lupus/Canis lupus familiaris*	6000	N	34.69378849
AL2867_Germany_A_6300	Unpublished	*Canis lupus/Canis lupus familiaris*	6300	N	29.89017756
OL4222_Greece_D_6544	Unpublished	*Canis lupus/Canis lupus familiaris*	6544	N	2059.969152
AL2988_Russia_A_6750	Unpublished	*Canis lupus/Canis lupus familiaris*	6750	N	2.124887906
AL2990_Russia_6750	Unpublished	*Canis lupus/Canis lupus familiaris*	6750	N	3.143420817
AL2991_Russia_6750	Unpublished	*Canis lupus/Canis lupus familiaris*	6750	N	8.604471812
AL2994_Russia_A_6750	Unpublished	*Canis lupus/Canis lupus familiaris*	6750	N	2.15932325
AL2995_Russia_6750	Unpublished	*Canis lupus/Canis lupus familiaris*	6750	N	9.892449333
AL3118_Russia_A_6974	Unpublished	*Canis lupus/Canis lupus familiaris*	6974	N	31.25133018
Herxheim_Germany_C_7000	KX379529	*Canis lupus/Canis lupus familiaris*	7000	-	-
OL4061_Veretye_Russia_A_7000	Unpublished	*Canis lupus/Canis lupus familiaris*	7000	N	119.4855623
AL2942_Serbia_D_7000	Unpublished	*Canis lupus/Canis lupus familiaris*	7000	N	7.255873737
AL3185_Pietrele_Romania_B_7000	Unpublished	*Canis lupus/Canis lupus familiaris*	7000	N	51.96090154
OL4223_Baikal_Russia_A_7000	Unpublished	*Canis lupus/Canis lupus familiaris*	7000	N	511.0345549
OL4161_Greece_D_7500	Unpublished	*Canis lupus/Canis lupus familiaris*	7500	N	42.2953309
AL2978_Georgia_D_7550	Unpublished	*Canis lupus/Canis lupus familiaris*	7550	N	7.181502959
AL2642_Russia_A_7814	Unpublished	*Canis lupus/Canis lupus familiaris*	7814	N	1.669695702
OL4109_Serbia_D_8050	Unpublished	*Canis lupus/Canis lupus familiaris*	8050	N	29.36509834
ISM090_A2b_8385	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8385	Y	52.34136426
AL2806_USA_A_8450	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8450	N	21.55485144
USA_A2b_8500	KF661083.1	*Canis lupus/Canis lupus familiaris*	8500	N	-
CGG3_Zhokhov_Siberia_8558	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8558	Y	40.03258205
AL3251_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	3.686734023

AL3252_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	3.169725593
AL3253_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	4.564416811
AL3254_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	3.848568183
AL3255_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	1.205177258
AL3257_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	1.548335027
AL3263_Veretye_Russia_A_8750	Unpublished	*Canis lupus/Canis lupus familiaris*	8750	N	2.293358044
CGG8_Zhokhov_Siberia_A2a_8828	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8828	Y	31.15268727
CGG5_Zhokhov_Siberia_8855	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8855	Y	129.2819394
CGG6_Zhokhov_Siberia_A2a_8855	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8855	Y	129.2819394
CGG4_Zhokhov_Siberia_A2a_8883	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8883	Y	85.20930233
CGG1_Zhokhov_Siberia_A2a_8895	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8895	Y	99.49967119
CGG2_Zhokhov_Siberia_A2a_8935	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	8935	Y	69.40072936
CGG7_Zhokhov_Siberia_A2a_9135	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	9135	Y	145.7030549
CH1109_Noyen_France_D_9500	Unpublished	*Canis lupus/Canis lupus familiaris*	9500	N	0.9103246249
CGG9_Zhokhov_Siberia_A2a_9503	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	9503	Y	87.44449094
ISM256_A2b_9895	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	9895	Y	0.989956358
ISM357_A2b_9905	Ní Leathlobhair et al. 2018	*Canis lupus/Canis lupus familiaris*	9905	N	0.050277994
OL4070_Serbia_D_10450	Unpublished	*Canis lupus/Canis lupus familiaris*	10450	N	1.250493215
Germany_Th14_C_12500	KF661094.1	*Canis lupus/Canis lupus familiaris*	12500	N	-
AL2908_Germany_D_13168	Unpublished	*Canis lupus/Canis lupus familiaris*	13168	N	6.561606983
IRK_PlunkettCave_Ireland_13219	Unpublished	*Canis lupus/Canis lupus familiaris*	13219	N	13.71
CGG20_Nikita_Lake_Russia_13660	Unpublished	*Canis lupus/Canis lupus familiaris*	13660	N	-
CGG21_Berelekh_Russia_13660	Unpublished	*Canis lupus/Canis lupus familiaris*	13660	N	-
Tumat_Siberia_14164	Unpublished	*Canis lupus/Canis lupus familiaris*	14164	N	-
Switzerland3_Th15_14500	KF661080	*Canis lupus/Canis lupus familiaris*	14500	N	-
Switzerland1_Th7_14500	KF661087	*Canis lupus/Canis lupus familiaris*	14500	N	-
Switzerland2_Th11_14500	KF661091	*Canis lupus/Canis lupus familiaris*	14500	N	-
Russia_15000	KF661082.1	*Canis lupus/Canis lupus familiaris*	15000	N	-
CGG33_UlakhanSular_Russia_16864	Unpublished	*Canis lupus/Canis lupus familiaris*	16864	N	-
AL2660_Russia_A_LateUpperPal_17000	Unpublished	*Canis lupus/Canis lupus familiaris*	17000	N	7.031505948
Russia_18000	KF661081	*Canis lupus/Canis lupus familiaris*	18000	N	-
CGG17_Yana_Russia_28520	Unpublished	*Canis lupus/Canis lupus familiaris*	18520	N	-
CGG19_Yana_Russia_19710	Unpublished	*Canis lupus/Canis lupus familiaris*	19710	N	-
Alaska_Th10_20800	KF661090.1	*Canis lupus/Canis lupus familiaris*	20800	N	-
Russia_22000	KF661085.1	*Canis lupus/Canis lupus familiaris*	22000	N	-
Belgium_TH1_26000	KF661078	*Canis lupus/Canis lupus familiaris*	26000	N	8.3
Alaska_Th8_28000	KF661088	*Canis lupus/Canis lupus familiaris*	28000	N	-
CGG34_Badayarikha_Russia_29935	Unpublished	*Canis lupus/Canis lupus familiaris*	29935	N	-
Belgium_Th3_30000	KF661080	*Canis lupus/Canis lupus familiaris*	30000	N	-
CGG26_Yana_Russia_31351	Unpublished	*Canis lupus/Canis lupus familiaris*	31351	N	-
CGG22_Yana_Russia_31377	Unpublished	*Canis lupus/Canis lupus familiaris*	31377	N	-
CGG25_Yana_Russia_31424	Unpublished	*Canis lupus/Canis lupus familiaris*	31424	N	-
CGG28_Yana_Russia_31758	Unpublished	*Canis lupus/Canis lupus familiaris*	31758	N	-
CGG14_Yana_Russia_32734	Unpublished	*Canis lupus/Canis lupus familiaris*	32734	N	-
CGG27_Yana_Russia_32844	Unpublished	*Canis lupus/Canis lupus familiaris*	32844	N	-
CGG15_Yana_Russia_32926	Unpublished	*Canis lupus/Canis lupus familiaris*	32926	N	-
CGG23_Yana_Russia_330120	Unpublished	*Canis lupus/Canis lupus familiaris*	33020	N	-
Russia_33500	KF661092	*Canis lupus/Canis lupus familiaris*	33500	N	-
Taimyr1_Russia_35000	Skoglund et al 2015	*Canis lupus/Canis lupus familiaris*	35000	Y	-
AL2735_Russia_C_35000	Unpublished	*Canis lupus/Canis lupus familiaris*	35000	N	2.119985652
AL2736_Russia_A_35000	Unpublished	*Canis lupus/Canis lupus familiaris*	35000	N	3.307765887
AL2741_Russia_A_35000	Unpublished	*Canis lupus/Canis lupus familiaris*	35000	N	14.2909667
AL2746_Russia_A_35000	Unpublished	*Canis lupus/Canis lupus familiaris*	35000	N	1.359478687
CGG13_DuvannyiYar_Russia_48500	Unpublished	*Canis lupus/Canis lupus familiaris*	48500	N	-
CGG29_BungeToll_Russia_48500	Unpublished	*Canis lupus/Canis lupus familiaris*	48500	N	-
CGG32_Tirektyakh_Russia_48500	Unpublished	*Canis lupus/Canis lupus familiaris*	48500	N	-
AAL1-FIN_0	KF938342.1	*Ovis aries*	0	-	-
AFS1-IRN_0	SRS335703	*Ovis aries*	0	-	-
AFS2-IRN_0	SRS335685	*Ovis aries*	0	-	-
ALT1-CHN_0	KF938320.1	*Ovis aries*	0	-	-
ALTM1-RUS_0	Unpublished	*Ovis aries*	0	-	-
ALTM2-RUS_0	Unpublished	*Ovis aries*	0	-	-
ANDB1-RUS_0	KF938340.1	*Ovis aries*	0	-	-

ANDB2-RUS_0	Unpublished	*Ovis aries*	0	-	-
ANDB3-RUS_0	Unpublished	*Ovis aries*	0	-	-
APP1-ITA_0	KF302450.1	*Ovis aries*	0	-	-
APP2-ITA_0	KF302451.1	*Ovis aries*	0	-	-
ASF1-ISR_0	EF490456.1	*Ovis aries*	0	-	-
ASF2-ISR_0	HE577847.1	*Ovis aries*	0	-	-
ASF3-ISR_0	HE577848.1	*Ovis aries*	0	-	-
ASF4-ISR_0	HE577849.1	*Ovis aries*	0	-	-
ASF5-ISR_0	HE577850.1	*Ovis aries*	0	-	-
AWG1-CHN_0	KP998470.1	*Ovis aries*	0	-	-
AWS1-ISR_0	HM236182.1	*Ovis aries*	0	-	-
AWS2-ISR_0	SRS152492	*Ovis aries*	0	-	-
AWS3-ISR_0	SRS335723	*Ovis aries*	0	-	-
AWS4-ISR_0	SRS004311	*Ovis aries*	0	-	-
AWS5-TUR_0	SRS335726	*Ovis aries*	0	-	-
AWS6-TUR_0	SRS335704	*Ovis aries*	0	-	-
BDK1-RUS_0	KF938339.1	*Ovis aries*	0	-	-
BGE1-BGD_0	SRS335705	*Ovis aries*	0	-	-
BGE2-BGD_0	SRS335686	*Ovis aries*	0	-	-
BGL1-MAR_0	ERS154854	*Ovis aries*	0	-	-
BGL2-MAR_0	ERS154840	*Ovis aries*	0	-	-
BGL3-MAR_0	ERS154844	*Ovis aries*	0	-	-
BGL4-MAR_0	ERS154700	*Ovis aries*	0	-	-
BGL5-MAR_0	ERS154838	*Ovis aries*	0	-	-
BGL6-MAR_0	ERS154706	*Ovis aries*	0	-	-
BJD1-MAR_0	ERS154762	*Ovis aries*	0	-	-
BKFF1-RUS_0	Unpublished	*Ovis aries*	0	-	-
BKFF2-RUS_0	Unpublished	*Ovis aries*	0	-	-
BNT1-DEU_0	Unpublished	*Ovis aries*	0	-	-
BNT2-DEU_0	Unpublished	*Ovis aries*	0	-	-
BNT3-DEU_0	Unpublished	*Ovis aries*	0	-	-
BNT4-DEU_0	Unpublished	*Ovis aries*	0	-	-
BRK1-CHN_0	KF938321.1	*Ovis aries*	0	-	-
BSB1-CHN_0	KF938330.1	*Ovis aries*	0	-	-
BUUB1-RUS_0	Unpublished	*Ovis aries*	0	-	-
BUUB2-RUS_0	Unpublished	*Ovis aries*	0	-	-
BYB1-CHN_0	KF938331.1	*Ovis aries*	0	-	-
CBR1-DEU_0	Unpublished	*Ovis aries*	0	-	-
CBR2-DEU_0	Unpublished	*Ovis aries*	0	-	-
CBR3-DEU_0	Unpublished	*Ovis aries*	0	-	-
CBR4-DEU_0	Unpublished	*Ovis aries*	0	-	-
CCP1-TUR_0	SRS335708	*Ovis aries*	0	-	-
CHA1-IND_0	SRS335727	*Ovis aries*	0	-	-
CHA2-IND_0	SRS335689	*Ovis aries*	0	-	-
CHR1-FRA_0	Unpublished	*Ovis aries*	0	-	-
CHR2-FRA_0	Unpublished	*Ovis aries*	0	-	-
CHR3-FRA_0	Unpublished	*Ovis aries*	0	-	-
CHV1-GBR_0	SRS335729	*Ovis aries*	0	-	-
CHV2-GBR_0	SRS335690	*Ovis aries*	0	-	-
CMF1-UKR_0	KF938357.1	*Ovis aries*	0	-	-
COM1-ITA_0	KF302440.1	*Ovis aries*	0	-	-
COM2-ITA_0	KF302441.1	*Ovis aries*	0	-	-
COM3-ITA_0	KF302442.1	*Ovis aries*	0	-	-
COM4-ITA_0	KF302443.1	*Ovis aries*	0	-	-
COM5-ITA_0	KF302444.1	*Ovis aries*	0	-	-
COM6-ITA_0	KF302445.1	*Ovis aries*	0	-	-
COM7-ITA_0	KF302452.1	*Ovis aries*	0	-	-
COM8-ITA_0	KF302461.1	*Ovis aries*	0	-	-
COM9-ITA_0	KF302462.1	*Ovis aries*	0	-	-
CRS1-HRV_0	Unpublished	*Ovis aries*	0	-	-
CRS2-HRV_0	Unpublished	*Ovis aries*	0	-	-
DAGM1-RUS_0	Unpublished	*Ovis aries*	0	-	-
DAGM2-RUS_0	Unpublished	*Ovis aries*	0	-	-
DJL1-ISR_0	KF977845.1	*Ovis aries*	0	-	-

DLG1-CHN_0	KF938332.1	*Ovis aries*	0	-	-
DLG2-CHN_0	SRS1071053	*Ovis aries*	0	-	-
DMN01-MAR_0	ERS154835	*Ovis aries*	0	-	-
DMN02-MAR_0	ERS154824	*Ovis aries*	0	-	-
DMN03-MAR_0	ERS154822	*Ovis aries*	0	-	-
DMN04-MAR_0	ERS154813	*Ovis aries*	0	-	-
DMN05-MAR_0	ERS154784	*Ovis aries*	0	-	-
DMN06-MAR_0	ERS154786	*Ovis aries*	0	-	-
DMN07-MAR_0	ERS154799	*Ovis aries*	0	-	-
DMN08-MAR_0	ERS154798	*Ovis aries*	0	-	-
DMN09-MAR_0	ERS154812	*Ovis aries*	0	-	-
DMN10-MAR_0	ERS154785	*Ovis aries*	0	-	-
DMN11-MAR_0	ERS154767	*Ovis aries*	0	-	-
DMN12-MAR_0	ERS154774	*Ovis aries*	0	-	-
DMN13-MAR_0	ERS154787	*Ovis aries*	0	-	-
DMN14-MAR_0	ERS154783	*Ovis aries*	0	-	-
DMN15-MAR_0	ERS154768	*Ovis aries*	0	-	-
DMN16-MAR_0	ERS154744	*Ovis aries*	0	-	-
DMN17-MAR_0	ERS154735	*Ovis aries*	0	-	-
DMN18-MAR_0	ERS154765	*Ovis aries*	0	-	-
DMN19-MAR_0	ERS154766	*Ovis aries*	0	-	-
DMN20-MAR_0	ERS154758	*Ovis aries*	0	-	-
DMN21-MAR_0	ERS154756	*Ovis aries*	0	-	-
DMN22-MAR_0	ERS154743	*Ovis aries*	0	-	-
DMN23-MAR_0	ERS154733	*Ovis aries*	0	-	-
DMN24-MAR_0	ERS154725	*Ovis aries*	0	-	-
DMN25-MAR_0	ERS154717	*Ovis aries*	0	-	-
DMN26-MAR_0	ERS154711	*Ovis aries*	0	-	-
DMN27-MAR_0	ERS154814	*Ovis aries*	0	-	-
DMN28-MAR_0	ERS154800	*Ovis aries*	0	-	-
DMN29-MAR_0	ERS154775	*Ovis aries*	0	-	-
DMN30-MAR_0	ERS154734	*Ovis aries*	0	-	-
DSH1-GBR_0	SRS1709526	*Ovis aries*	0	-	-
DSH2-GBR_0	SRS1709528	*Ovis aries*	0	-	-
DSH3-GBR_0	SRS1709527	*Ovis aries*	0	-	-
DSH4-GBR_0	SRS1709561	*Ovis aries*	0	-	-
DWM1-GBR_0	SRS335709	*Ovis aries*	0	-	-
EDLB1-RUS_0	Unpublished	*Ovis aries*	0	-	-
EDLB2-RUS_0	Unpublished	*Ovis aries*	0	-	-
EFW1-DEU_0	Unpublished	*Ovis aries*	0	-	-
EFW2-DEU_0	Unpublished	*Ovis aries*	0	-	-
EFW3-DEU_0	Unpublished	*Ovis aries*	0	-	-
EFW4-DEU_0	Unpublished	*Ovis aries*	0	-	-
EMZ1-ETH_0	SRS335697	*Ovis aries*	0	-	-
FIN1-FIN_0	KF938354.1	*Ovis aries*	0	-	-
FIN2-FIN_0	KF938355.1	*Ovis aries*	0	-	-
FIN3-FIN_0	SRS335716	*Ovis aries*	0	-	-
FND1-FIN_0	EF490451.1	*Ovis aries*	0	-	-
FND2-FIN_0	EF490452.1	*Ovis aries*	0	-	-
FND3-FIN_0	EF490453.1	*Ovis aries*	0	-	-
FND4-FIN_0	EF490454.1	*Ovis aries*	0	-	-
FND5-FIN_0	EF490455.1	*Ovis aries*	0	-	-
GAR1-IND_0	SRS335735	*Ovis aries*	0	-	-
GAR2-IND_0	SRS335736	*Ovis aries*	0	-	-
GARZ01-CHN_0	KU681175	*Ovis aries*	0	-	-
GARZ02-CHN_0	KU681176	*Ovis aries*	0	-	-
GARZ03-CHN_0	KU681177	*Ovis aries*	0	-	-
GARZ04-CHN_0	KU681178	*Ovis aries*	0	-	-
GARZ05-CHN_0	KU681179	*Ovis aries*	0	-	-
GARZ06-CHN_0	KU681180	*Ovis aries*	0	-	-
GARZ07-CHN_0	KU681181	*Ovis aries*	0	-	-
GARZ08-CHN_0	KU681182	*Ovis aries*	0	-	-
GARZ09-CHN_0	KU681183	*Ovis aries*	0	-	-
GARZ10-CHN_0	KU681184	*Ovis aries*	0	-	-

GARZ11-CHN_0	KU681185	*Ovis aries*	0	-	-
GARZ12-CHN_0	KU681186	*Ovis aries*	0	-	-
GARZ13-CHN_0	KU681187	*Ovis aries*	0	-	-
GARZ14-CHN_0	KU681188	*Ovis aries*	0	-	-
GARZ15-CHN_0	KU681189	*Ovis aries*	0	-	-
GARZ16-CHN_0	KU681190	*Ovis aries*	0	-	-
GARZ17-CHN_0	KU681191	*Ovis aries*	0	-	-
GARZ18-CHN_0	KU681192	*Ovis aries*	0	-	-
GBHM1-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBHM2-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBHM3-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBHM4-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBRM1-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBRM2-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBRM3-DEU_0	Unpublished	*Ovis aries*	0	-	-
GBRM4-DEU_0	Unpublished	*Ovis aries*	0	-	-
GEN1-ITA_0	KF302448.1	*Ovis aries*	0	-	-
GEN2-ITA_0	KF302455.1	*Ovis aries*	0	-	-
GEN3-ITA_0	KF302457.1	*Ovis aries*	0	-	-
GHS1-DEU_0	Unpublished	*Ovis aries*	0	-	-
GHS2-DEU_0	Unpublished	*Ovis aries*	0	-	-
GHS3-DEU_0	Unpublished	*Ovis aries*	0	-	-
GHS4-DEU_0	Unpublished	*Ovis aries*	0	-	-
GLA1-AZE_0	KF938345.1	*Ovis aries*	0	-	-
GNJ1-CHN_0	KP998472.1	*Ovis aries*	0	-	-
GRZN1-RUS_0	Unpublished	*Ovis aries*	0	-	-
GRZN2-RUS_0	Unpublished	*Ovis aries*	0	-	-
GSTB1-TIB_0	KP981379.1	*Ovis aries*	0	-	-
GUR1-IDN_0	SRS335718	*Ovis aries*	0	-	-
GUR2-IDN_0	SRS335698	*Ovis aries*	0	-	-
HLB1-CHN_0	KF938327.1	*Ovis aries*	0	-	-
HMD01-IRQ_0	MF004242	*Ovis aries*	0	-	-
HMD02-IRQ_0	MF004243	*Ovis aries*	0	-	-
HMD03-IRQ_0	MF004244	*Ovis aries*	0	-	-
HTN1-CHN_0	KF938322.1	*Ovis aries*	0	-	-
HUB1-CHN_0	KP998471.1	*Ovis aries*	0	-	-
HZG1-CHN_0	KF938329.1	*Ovis aries*	0	-	-
IDF1-FRA_0	Unpublished	*Ovis aries*	0	-	-
IDF2-FRA_0	Unpublished	*Ovis aries*	0	-	-
IDF3-FRA_0	Unpublished	*Ovis aries*	0	-	-
JZG1-CHN_0	KF938328.1	*Ovis aries*	0	-	-
KALM1-RUS_0	Unpublished	*Ovis aries*	0	-	-
KALM2-RUS_0	Unpublished	*Ovis aries*	0	-	-
KARA1-UZB_0	KF938348.1	*Ovis aries*	0	-	-
KARA2-UZB_0	Unpublished	*Ovis aries*	0	-	-
KARA3-UZB_0	Unpublished	*Ovis aries*	0	-	-
KGZ1-KGZ_0	KF938334.1	*Ovis aries*	0	-	-
KLND1-RUS_0	KF938358.1	*Ovis aries*	0	-	-
KLND2-RUS_0	Unpublished	*Ovis aries*	0	-	-
KLND3-RUS_0	Unpublished	*Ovis aries*	0	-	-
KNG1-FIN_0	KF938353.1	*Ovis aries*	0	-	-
KORS1-HRV_0	Unpublished	*Ovis aries*	0	-	-
KORS2-HRV_0	Unpublished	*Ovis aries*	0	-	-
KRCH1-RUS_0	KF938351.1	*Ovis aries*	0	-	-
KRCH2-RUS_0	Unpublished	*Ovis aries*	0	-	-
KRCH3-RUS_0	Unpublished	*Ovis aries*	0	-	-
KRD01-IRQ_0	MF004245	*Ovis aries*	0	-	-
KRD02-IRQ_0	MF004246	*Ovis aries*	0	-	-
KRKS1-HRV_0	Unpublished	*Ovis aries*	0	-	-
KRKS2-HRV_0	Unpublished	*Ovis aries*	0	-	-
KRKS3-HRV_0	Unpublished	*Ovis aries*	0	-	-
KRS1-TUR_0	HM236176.1	*Ovis aries*	0	-	-
KRS2-TUR_0	HM236177.1	*Ovis aries*	0	-	-
KRS3-TUR_0	HM236178.1	*Ovis aries*	0	-	-

KRS4-TUR_0	SRS335719	*Ovis aries*	0	-	-
KRST1-SVN_0	Unpublished	*Ovis aries*	0	-	-
KRST2-SVN_0	Unpublished	*Ovis aries*	0	-	-
KRST3-SVN_0	Unpublished	*Ovis aries*	0	-	-
KUCH1-RUS_0	Unpublished	*Ovis aries*	0	-	-
KUCH2-RUS_0	Unpublished	*Ovis aries*	0	-	-
KUIB1-RUS_0	KF938346.1	*Ovis aries*	0	-	-
KUIB2-RUS_0	Unpublished	*Ovis aries*	0	-	-
KUIB3-RUS_0	Unpublished	*Ovis aries*	0	-	-
KZK1-KAZ_0	KF938333.1	*Ovis aries*	0	-	-
LAC1-FRA_0	KF302447.1	*Ovis aries*	0	-	-
LAC2-FRA_0	KF302453.1	*Ovis aries*	0	-	-
LAC3-FRA_0	KF302460.1	*Ovis aries*	0	-	-
LEZG1-RUS_0	Unpublished	*Ovis aries*	0	-	-
LEZG2-RUS_0	Unpublished	*Ovis aries*	0	-	-
LPN1-CHN_0	KF938323.1	*Ovis aries*	0	-	-
LZLT1-CHN_0	KF938335.1	*Ovis aries*	0	-	-
MANM1-RUS_0	Unpublished	*Ovis aries*	0	-	-
MANM2-RUS_0	Unpublished	*Ovis aries*	0	-	-
MBF1-CHN_0	KF938318.1	*Ovis aries*	0	-	-
MER1-AUS_0	HM236174.1	*Ovis aries*	0	-	-
MKR1-TUR_0	HM236179.1	*Ovis aries*	0	-	-
MKR2-TUR_0	HM236180.1	*Ovis aries*	0	-	-
MKR3-TUR_0	HM236181.1	*Ovis aries*	0	-	-
MLS1-DEU_0	AF010406.1	*Ovis aries*	0	-	-
MLS2-DEU_0	Unpublished	*Ovis aries*	0	-	-
MLS3-DEU_0	Unpublished	*Ovis aries*	0	-	-
MLS4-DEU_0	Unpublished	*Ovis aries*	0	-	-
MLS5-DEU_0	Unpublished	*Ovis aries*	0	-	-
MLS6-DEU_0	NC_001941.1	*Ovis aries*	0	-	-
MNG1-MNG_0	SRS1071018	*Ovis aries*	0	-	-
MNZ1-ITA_0	KF302446.1	*Ovis aries*	0	-	-
MNZ2-ITA_0	KF302459.1	*Ovis aries*	0	-	-
MZK1-AZE_0	KF938344.1	*Ovis aries*	0	-	-
NCSN1-RUS_0	Unpublished	*Ovis aries*	0	-	-
NCSN2-RUS_0	Unpublished	*Ovis aries*	0	-	-
NDZ1-TUR_0	SRS335721	*Ovis aries*	0	-	-
NDZ2-TUR_0	SRS335701	*Ovis aries*	0	-	-
NGW01-CHN_0	KU681193	*Ovis aries*	0	-	-
NGW02-CHN_0	KU681194	*Ovis aries*	0	-	-
NGW03-CHN_0	KU681195	*Ovis aries*	0	-	-
NGW04-CHN_0	KU681196	*Ovis aries*	0	-	-
NGW05-CHN_0	KU681197	*Ovis aries*	0	-	-
NGW06-CHN_0	KU681198	*Ovis aries*	0	-	-
NGW07-CHN_0	KU681199	*Ovis aries*	0	-	-
NGW08-CHN_0	KU681200	*Ovis aries*	0	-	-
NGW09-CHN_0	KU681201	*Ovis aries*	0	-	-
NGW10-CHN_0	KU681202	*Ovis aries*	0	-	-
NQA1-ZAF_0	SRS335740	*Ovis aries*	0	-	-
OPR1-RUS_0	KF938352.1	*Ovis aries*	0	-	-
OUD1-MAR_0	ERS154859	*Ovis aries*	0	-	-
OUD2-MAR_0	ERS154855	*Ovis aries*	0	-	-
OUD3-MAR_0	ERS154849	*Ovis aries*	0	-	-
OUD4-MAR_0	ERS154816	*Ovis aries*	0	-	-
OUD5-MAR_0	ERS154851	*Ovis aries*	0	-	-
OUD6-MAR_0	ERS154703	*Ovis aries*	0	-	-
OUD7-MAR_0	ERS154861	*Ovis aries*	0	-	-
OUD8-MAR_0	ERS154856	*Ovis aries*	0	-	-
OULA1-CHN_0	KU575248	*Ovis aries*	0	-	-
OULA2-CHN_0	KT148968	*Ovis aries*	0	-	-
OXD1-GBR_0	KF938359.1	*Ovis aries*	0	-	-
PDAL1-HRV_0	Unpublished	*Ovis aries*	0	-	-
PDAL2-HRV_0	Unpublished	*Ovis aries*	0	-	-
PKUP1-BIH_0	Unpublished	*Ovis aries*	0	-	-

PKUP2-BIH_0	Unpublished	*Ovis aries*	0	-	-
PMC1-DEU_0	Unpublished	*Ovis aries*	0	-	-
PMC2-DEU_0	Unpublished	*Ovis aries*	0	-	-
PMC3-DEU_0	Unpublished	*Ovis aries*	0	-	-
PPRI1-BIH_0	Unpublished	*Ovis aries*	0	-	-
PPRI2-BIH_0	Unpublished	*Ovis aries*	0	-	-
PSJE1-SRB_0	Unpublished	*Ovis aries*	0	-	-
PSJE2-SRB_0	Unpublished	*Ovis aries*	0	-	-
PSRB1-SRB_0	KF938347.1	*Ovis aries*	0	-	-
PSTO1-BIH_0	Unpublished	*Ovis aries*	0	-	-
PSTO2-BIH_0	Unpublished	*Ovis aries*	0	-	-
PTRA1-BIH_0	Unpublished	*Ovis aries*	0	-	-
PTRA2-BIH_0	Unpublished	*Ovis aries*	0	-	-
PTRA3-BIH_0	Unpublished	*Ovis aries*	0	-	-
PTRA4-BIH_0	Unpublished	*Ovis aries*	0	-	-
QHT1-CHN_0	KF938325.1	*Ovis aries*	0	-	-
QLW1-CHN_0	KP998473.1	*Ovis aries*	0	-	-
QRB1-CHN_0	KF938326.1	*Ovis aries*	0	-	-
RDA1-ZAF_0	SRS335722	*Ovis aries*	0	-	-
RDA2-ZAF_0	SRS335702	*Ovis aries*	0	-	-
RDR1-FRA_0	Unpublished	*Ovis aries*	0	-	-
RDR2-FRA_0	Unpublished	*Ovis aries*	0	-	-
RDR3-FRA_0	Unpublished	*Ovis aries*	0	-	-
RDR4-FRA_0	Unpublished	*Ovis aries*	0	-	-
RHN1-DEU_0	Unpublished	*Ovis aries*	0	-	-
RHN2-DEU_0	Unpublished	*Ovis aries*	0	-	-
RHN3-DEU_0	Unpublished	*Ovis aries*	0	-	-
RHN4-DEU_0	Unpublished	*Ovis aries*	0	-	-
RMB01-FRA_0	SRS1709705	*Ovis aries*	0	-	-
RMB02-FRA_0	SRS1709706	*Ovis aries*	0	-	-
RMB03-FRA_0	SRS1709707	*Ovis aries*	0	-	-
RMB04-FRA_0	SRS1709709	*Ovis aries*	0	-	-
RMB05-FRA_0	SRR4291268	*Ovis aries*	0	-	-
RMB06-FRA_0	SRR4290326	*Ovis aries*	0	-	-
RMB07-FRA_0	SRR4291301	*Ovis aries*	0	-	-
RMB08-FRA_0	SRR4291297	*Ovis aries*	0	-	-
RMB09-FRA_0	SRR4291284	*Ovis aries*	0	-	-
RMB10-FRA_0	SRR4291315	*Ovis aries*	0	-	-
RMNV1-RUS_0	Unpublished	*Ovis aries*	0	-	-
RMNV2-RUS_0	Unpublished	*Ovis aries*	0	-	-
RMNV3-RUS_0	Unpublished	*Ovis aries*	0	-	-
RMNV4-RUS_0	Unpublished	*Ovis aries*	0	-	-
RMNV5-RUS_0	Unpublished	*Ovis aries*	0	-	-
RMY1-AUS_0	HM236175.1	*Ovis aries*	0	-	-
RMY2-GBR_0	SRS335691	*Ovis aries*	0	-	-
RULH1-RUS_0	Unpublished	*Ovis aries*	0	-	-
RULH2-RUS_0	Unpublished	*Ovis aries*	0	-	-
RUV1-RUS_0	KF938356.1	*Ovis aries*	0	-	-
SALS1-RUS_0	Unpublished	*Ovis aries*	0	-	-
SALS2-RUS_0	Unpublished	*Ovis aries*	0	-	-
SAR01-MAR_0	ERS154843	*Ovis aries*	0	-	-
SAR02-MAR_0	ERS154823	*Ovis aries*	0	-	-
SAR03-MAR_0	ERS154771	*Ovis aries*	0	-	-
SAR04-MAR_0	ERS154763	*Ovis aries*	0	-	-
SAR05-MAR_0	ERS154752	*Ovis aries*	0	-	-
SAR06-MAR_0	ERS154761	*Ovis aries*	0	-	-
SAR07-MAR_0	ERS154760	*Ovis aries*	0	-	-
SAR08-MAR_0	ERS154759	*Ovis aries*	0	-	-
SAR09-MAR_0	ERS154754	*Ovis aries*	0	-	-
SAR10-MAR_0	ERS154753	*Ovis aries*	0	-	-
SAR11-MAR_0	ERS154747	*Ovis aries*	0	-	-
SAR12-MAR_0	ERS154738	*Ovis aries*	0	-	-
SAR13-MAR_0	ERS154745	*Ovis aries*	0	-	-
SAR14-MAR_0	ERS154740	*Ovis aries*	0	-	-

SAR15-MAR_0	ERS154737	*Ovis aries*	0	-	-
SAR16-MAR_0	ERS154731	*Ovis aries*	0	-	-
SAR17-MAR_0	ERS154730	*Ovis aries*	0	-	-
SAR18-MAR_0	ERS154728	*Ovis aries*	0	-	-
SAR19-MAR_0	ERS154722	*Ovis aries*	0	-	-
SAR20-MAR_0	ERS154720	*Ovis aries*	0	-	-
SAR21-MAR_0	ERS154716	*Ovis aries*	0	-	-
SAR22-MAR_0	ERS154701	*Ovis aries*	0	-	-
SAR23-MAR_0	ERS154764	*Ovis aries*	0	-	-
SAR24-MAR_0	ERS154721	*Ovis aries*	0	-	-
SAR25-MAR_0	ERS154727	*Ovis aries*	0	-	-
SAR26-MAR_0	ERS154746	*Ovis aries*	0	-	-
SAR27-MAR_0	ERS154729	*Ovis aries*	0	-	-
SBF1-GBR_0	Unpublished	*Ovis aries*	0	-	-
SBF2-GBR_0	Unpublished	*Ovis aries*	0	-	-
SBF3-GBR_0	Unpublished	*Ovis aries*	0	-	-
SBF4-GBR_0	Unpublished	*Ovis aries*	0	-	-
SDLT1-CHN_0	KP981380.1	*Ovis aries*	0	-	-
SHL1-ZAF_0	KF977846.1	*Ovis aries*	0	-	-
SKD1-Baltic_0	Unpublished	*Ovis aries*	0	-	-
SKD2-Baltic_0	Unpublished	*Ovis aries*	0	-	-
SKD3-Baltic_0	Unpublished	*Ovis aries*	0	-	-
SKZ1-TUR_0	SRS335711	*Ovis aries*	0	-	-
SKZ2-TUR_0	SRS335675	*Ovis aries*	0	-	-
SMS1-CHE_0	SRS335692	*Ovis aries*	0	-	-
SOA1-GBR_0	Unpublished	*Ovis aries*	0	-	-
SOA2-GBR_0	Unpublished	*Ovis aries*	0	-	-
SOA3-GBR_0	Unpublished	*Ovis aries*	0	-	-
SOP1-ITA_0	KF302449.1	*Ovis aries*	0	-	-
SOP2-ITA_0	KF302454.1	*Ovis aries*	0	-	-
SOP3-ITA_0	KF302456.1	*Ovis aries*	0	-	-
SOP4-ITA_0	KF302458.1	*Ovis aries*	0	-	-
SOVM1-RUS_0	Unpublished	*Ovis aries*	0	-	-
SOVM2-RUS_0	Unpublished	*Ovis aries*	0	-	-
SRP1-GBR_0	Unpublished	*Ovis aries*	0	-	-
SRP2-GBR_0	Unpublished	*Ovis aries*	0	-	-
SRP3-GBR_0	Unpublished	*Ovis aries*	0	-	-
SRP4-GBR_0	Unpublished	*Ovis aries*	0	-	-
STAV1-RUS_0	Unpublished	*Ovis aries*	0	-	-
STAV2-RUS_0	Unpublished	*Ovis aries*	0	-	-
STHL1-CHN_0	KP702285.1	*Ovis aries*	0	-	-
STHN1-MNG_0	KF977847.1	*Ovis aries*	0	-	-
SUF01-GBR_0	Unpublished	*Ovis aries*	0	-	-
SUF02-GBR_0	Unpublished	*Ovis aries*	0	-	-
SUF03-GBR_0	Unpublished	*Ovis aries*	0	-	-
SUF04-GBR_0	Unpublished	*Ovis aries*	0	-	-
SUM51-IDN_0	SRS335713	*Ovis aries*	0	-	-
SUM52-IDN_0	SRR501861	*Ovis aries*	0	-	-
SUN1-CHN_0	KF938317.1	*Ovis aries*	0	-	-
SWN1-POL_0	KF938349.1	*Ovis aries*	0	-	-
TAN1-CHN_0	KF938336.1	*Ovis aries*	0	-	-
TFB1-CHN_0	KF938324.1	*Ovis aries*	0	-	-
TIB1-TIB_0	SRS335677	*Ovis aries*	0	-	-
TIB2-TIB_0	SRS335695	*Ovis aries*	0	-	-
TKR1-CHN_0	KF938337.1	*Ovis aries*	0	-	-
TMD01-MAR_0	ERS154830	*Ovis aries*	0	-	-
TMD02-MAR_0	ERS154829	*Ovis aries*	0	-	-
TMD03-MAR_0	ERS154820	*Ovis aries*	0	-	-
TMD04-MAR_0	ERS154819	*Ovis aries*	0	-	-
TMD05-MAR_0	ERS154808	*Ovis aries*	0	-	-
TMD06-MAR_0	ERS154796	*Ovis aries*	0	-	-
TMD07-MAR_0	ERS154806	*Ovis aries*	0	-	-
TMD08-MAR_0	ERS154794	*Ovis aries*	0	-	-
TMD09-MAR_0	ERS154782	*Ovis aries*	0	-	-

TMD10-MAR_0	ERS154818	*Ovis aries*	0	-	-
TMD11-MAR_0	ERS154797	*Ovis aries*	0	-	-
TMD12-MAR_0	ERS154780	*Ovis aries*	0	-	-
TMD13-MAR_0	ERS154809	*Ovis aries*	0	-	-
TMD14-MAR_0	ERS154770	*Ovis aries*	0	-	-
TMD15-MAR_0	ERS154781	*Ovis aries*	0	-	-
TMD16-MAR_0	ERS154807	*Ovis aries*	0	-	-
TRWM1-GBR_0	SRS335715	*Ovis aries*	0	-	-
TUJ1-TUR_0	HM236183.1	*Ovis aries*	0	-	-
TUSH1-RUS_0	KF938343.1	*Ovis aries*	0	-	-
TUSH2-RUS_0	Unpublished	*Ovis aries*	0	-	-
TUSH3-RUS_0	Unpublished	*Ovis aries*	0	-	-
TUVA1-RUS_0	Unpublished	*Ovis aries*	0	-	-
TUVA2-RUS_0	Unpublished	*Ovis aries*	0	-	-
TXL1-NLD_0	KJ954145.1	*Ovis aries*	0	-	-
TXL2-NLD_0	Unpublished	*Ovis aries*	0	-	-
TXL3-NLD_0	Unpublished	*Ovis aries*	0	-	-
TXL4-NLD_0	Unpublished	*Ovis aries*	0	-	-
TXL5-NLD_0	Unpublished	*Ovis aries*	0	-	-
TZYG1-RUS_0	Unpublished	*Ovis aries*	0	-	-
TZYG2-RUS_0	Unpublished	*Ovis aries*	0	-	-
UDM1-RUS_0	KF938341.1	*Ovis aries*	0	-	-
UJQ1-CHN_0	KF938319.1	*Ovis aries*	0	-	-
UJQ2-CHN_0	KR868678.1	*Ovis aries*	0	-	-
VBN1-CHE_0	Unpublished	*Ovis aries*	0	-	-
VBN2-CHE_0	Unpublished	*Ovis aries*	0	-	-
VBN3-CHE_0	Unpublished	*Ovis aries*	0	-	-
VOLG1-RUS_0	Unpublished	*Ovis aries*	0	-	-
VOLG2-RUS_0	Unpublished	*Ovis aries*	0	-	-
WHSF1-GBR_0	SRS335714	*Ovis aries*	0	-	-
WRZ1-POL_0	KF938350.1	*Ovis aries*	0	-	-
YCH1-CHN_0	KF938338.1	*Ovis aries*	0	-	-
aSILG1-GBR_300	Unpublished	*Ovis aries*	300	-	-
OAA73-GBR_317	ERS426703	*Ovis aries*	317	-	-
OAA74-GBR_317	ERS426702	*Ovis aries*	317	-	-
aORKN2-GBR_800	Unpublished	*Ovis aries*	800	-	-
aHUNG1-GBR_967	Unpublished	*Ovis aries*	967	-	-
aGEO01-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO02-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO03-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO04-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO05-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO06-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO07-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO08-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO10-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO11-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aGEO13-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aKZB02-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aKZB10-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aKZB11-GEO_1000	Unpublished	*Ovis aries*	1000	-	-
aORKN1-GBR_1000	Unpublished	*Ovis aries*	1000	-	-
aKILP2-GBR_1150	Unpublished	*Ovis aries*	1150	-	-
aKILP3-GBR_1150	Unpublished	*Ovis aries*	1150	-	-
aDUBL1-IRL_1167	Unpublished	*Ovis aries*	1167	-	-
aDUBL2-IRL_1167	Unpublished	*Ovis aries*	1167	-	-
aDUBL3-IRL_1167	Unpublished	*Ovis aries*	1167	-	-
aYOQN1-ISR_1202	Unpublished	*Ovis aries*	1202	-	-
aYOQN4-ISR_1202	Unpublished	*Ovis aries*	1202	-	-
aBORN1-GBR_1400	Unpublished	*Ovis aries*	1400	-	-
aFRAN1-DEU_1700	Unpublished	*Ovis aries*	1700	-	-
aMAIN4-DEU_1700	Unpublished	*Ovis aries*	1700	-	-
aALZE1-DEU_1800	Unpublished	*Ovis aries*	1800	-	-
aMAIN1-DEU_1850	Unpublished	*Ovis aries*	1850	-	-

aMAIN2-DEU_1900	Unpublished	*Ovis aries*	1900	-	-
aMAIN3-DEU_1900	Unpublished	*Ovis aries*	1900	-	-
aDANE1-GBR_2325	Unpublished	*Ovis aries*	2325	-	-
aDANE2-GBR_2325	Unpublished	*Ovis aries*	2325	-	-
aDANE3-GBR_2325	Unpublished	*Ovis aries*	2325	-	-
aMIQN1-ISR_2600	Unpublished	*Ovis aries*	2600	-	-
aMIQN2-ISR_2600	Unpublished	*Ovis aries*	2600	-	-
aMIQN3-ISR_2600	Unpublished	*Ovis aries*	2600	-	-
aMIQN6-ISR_2600	Unpublished	*Ovis aries*	2600	-	-
aMASO1-ISR_2900	Unpublished	*Ovis aries*	2900	-	-
aHALL1-GBR_2940	Unpublished	*Ovis aries*	2940	-	-
aHALL2-GBR_2940	Unpublished	*Ovis aries*	2940	-	-
aHALL3-GBR_2940	Unpublished	*Ovis aries*	2940	-	-
aHALL4-GBR_2940	Unpublished	*Ovis aries*	2940	-	-
aHALL5-GBR_2940	Unpublished	*Ovis aries*	2940	-	-
aPOTT2-GBR_3000	Unpublished	*Ovis aries*	3000	-	-
aYOQN3-ISR_3825	Unpublished	*Ovis aries*	3825	-	-
aBROD1-GBR_4400	Unpublished	*Ovis aries*	4400	-	-
aBROD2-GBR_4400	Unpublished	*Ovis aries*	4400	-	-
aTARX3-MLT_4500	Unpublished	*Ovis aries*	4500	-	-
aSHIQ2-ISR_5850	Unpublished	*Ovis aries*	5850	-	-
aSHIQ4-ISR_5850	Unpublished	*Ovis aries*	5850	-	-
aHERX7-DEU_6160	Unpublished	*Ovis aries*	6160	-	-
aBLG04-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG06-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG07-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG08-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG09-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG10-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG11-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG13-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG15-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG17-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG18-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBLG20-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aBUBA1-SRB_7000	Unpublished	*Ovis aries*	7000	-	-
aMARM4-TUR_7062	Unpublished	*Ovis aries*	7062	-	-
aMARM5-TUR_7062	Unpublished	*Ovis aries*	7062	-	-
aMARM6-TUR_7062	Unpublished	*Ovis aries*	7062	-	-
aMARM7-TUR_7062	Unpublished	*Ovis aries*	7062	-	-
aMARM8-TUR_7062	Unpublished	*Ovis aries*	7062	-	-
aDZHU1-BGR_7100	Unpublished	*Ovis aries*	7100	-	-
aDZHU2-BGR_7100	Unpublished	*Ovis aries*	7100	-	-
aHERX3-DEU_7125	Unpublished	*Ovis aries*	7125	-	-
aHERX8-DEU_7125	Unpublished	*Ovis aries*	7125	-	-
aMNT01-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT02-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT04-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT05-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT06-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT07-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT09-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMNT11-TUR_7159	Unpublished	*Ovis aries*	7159	-	-
aMARM1-TUR_7200	Unpublished	*Ovis aries*	7200	-	-
aMARM3-TUR_7200	Unpublished	*Ovis aries*	7200	-	-
aSUBE2-TUR_7247	Unpublished	*Ovis aries*	7247	-	-
aSUBE3-TUR_7247	Unpublished	*Ovis aries*	7247	-	-
aCUKU1-TUR_8300	Unpublished	*Ovis aries*	8300	-	-
OARG138-2J_9625	Unpublished	*Ovis aries*	9625	-	-
OARG129-3D_9850	Unpublished	*Ovis aries*	9850	-	-
OARG130-3C_9850	Unpublished	*Ovis aries*	9850	-	-
OARG152-3D_9850	Unpublished	*Ovis aries*	9850	-	-
OARG153-3D_9850	Unpublished	*Ovis aries*	9850	-	-
OARG162-3D_9850	Unpublished	*Ovis aries*	9850	-	-

OARG165-3D_9850	Unpublished	*Ovis aries*	9850	-	-
aKORT1-TUR_9900	Unpublished	*Ovis aries*	9900	-	-
aKORT3-TUR_9900	Unpublished	*Ovis aries*	9900	-	-
OARG117-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG118-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG125-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG126-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG127-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG132-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG134-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG145-5_10150	Unpublished	*Ovis aries*	10150	-	-
OARG149-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG151-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG156-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG157-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG164-4_10150	Unpublished	*Ovis aries*	10150	-	-
OARG167-4_10150	Unpublished	*Ovis aries*	10150	-	-
OGM02-IRN_0	ERS154869	*Ovis gmelini*	0	-	-
OGM03-IRN_0	ERS154866	*Ovis gmelini*	0	-	-
OGM04-IRN_0	ERS154867	*Ovis gmelini*	0	-	-
OGM05-IRN_0	ERS154868	*Ovis gmelini*	0	-	-
OGM06-IRN_0	ERS154865	*Ovis gmelini*	0	-	-
OGM07-IRN_0	ERS154864	*Ovis gmelini*	0	-	-
OGM08-IRN_0	ERS154529	*Ovis gmelini*	0	-	-
OGM09-IRN_0	ERS154532	*Ovis gmelini*	0	-	-
OGM10-IRN_0	ERS154862	*Ovis gmelini*	0	-	-
OGM11-IRN_0	ERS154863	*Ovis gmelini*	0	-	-
OGM12-IRN_0	ERS154530	*Ovis gmelini*	0	-	-
OGM13-IRN_0	ERS154526	*Ovis gmelini*	0	-	-
OGM14-IRN_0	ERS154527	*Ovis gmelini*	0	-	-
OGM15-IRN_0	ERS154533	*Ovis gmelini*	0	-	-
OGM16-IRN_0	ERS154531	*Ovis gmelini*	0	-	-
OGM17-IRN_0	ERS419579	*Ovis gmelini*	0	-	-
OGM18-IRN_0	ERS419580	*Ovis gmelini*	0	-	-
OGM19-IRN_0	ERS239062	*Ovis gmelini*	0	-	-
OGM20-IRN_0	ERS239061	*Ovis gmelini*	0	-	-
OGM22-IRN_0	ERS239058	*Ovis gmelini*	0	-	-
OGM23-IRN_0	ERS239059	*Ovis gmelini*	0	-	-
OGM24-IRN_0	ERS239060	*Ovis gmelini*	0	-	-
OGM25-IRN_0	ERS239063	*Ovis gmelini*	0	-	-
OARG004-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG006-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG008-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG010-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG011-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG012-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG045-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG066-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG067-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG072-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG088-2B_9425	Unpublished	*Ovis gmelini aries*	9425	-	-
OARG019-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG021-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG022-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG028-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG049-2E_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG053-2D_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG064-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG082-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG083-2F_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG085-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG089-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG095-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG097-2J_9625	Unpublished	*Ovis gmelini aries*	9625	-	-

OARG098-2F_9625	Unpublished	*Ovis gmelini aries*	9625	-	-
OARG105-3C_9850	Unpublished	*Ovis gmelini aries*	9850	-	-
OARG106-3C_9850	Unpublished	*Ovis gmelini aries*	9850	-	-
OARG107-3C_9850	Unpublished	*Ovis gmelini aries*	9850	-	-
OARG111-3A_9850	Unpublished	*Ovis gmelini aries*	9850	-	-
OARG112-3C_9850	Unpublished	*Ovis gmelini aries*	9850	-	-
OARG040-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG041-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG075-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG077-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG114-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG120-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARG121-4_10150	Unpublished	*Ovis gmelini aries*	10150	-	-
OARM1-DEU_0	HM236184.1	*Ovis musimon*	0	-	-
OARM2-DEU_0	HM236185.1	*Ovis musimon*	0	-	-
OOP00-CYP_0	KF312238.1	*Ovis ophion*	0	-	-

Chapter 7, "Ancient pigs reveal a near complete genomic turnover following their introduction to Europe"

Published (2019) Proceedings of the National Academy of Sciences of the United States of America. doi:10.1073/pnas.1901169116

Statement of Contribution

Data
I organised and compiled the database used for the analysis (Table S1). I collated information for and wrote Archaeological Site Descriptions (Supplementary Information, pp. 2-50).

Analysis
I generated the maps seen in Fig 1 and Fig S6 demonstrating the biogeographical boundaries of these Asian, Near Eastern, European, and Y2 haplogroups. Maps were also drawn demonstrating the overall frequencies of these haplogroups in domestic pigs at various archaeological sites/locations between 8,000 BP and 5,100 BP; 5,099 BP and 180 BP; and in modern domestic pigs. I created the bar plots seen in Fig S5 that show the number of domestic pig and wild boar samples that possess different haplotypes.

Manuscript
I wrote the manuscript with LAFF, KJG, PR-C, MZ, JP, and GL.

Author list and affiliations

Authors

Laurent A. F. Frantz[a,b,1,2], James Haile[b,1], Audrey T. Lin[b,c,1], Amelie Scheu[d], Christina Geörg[d], Norbert Benecke[e], Michelle Alexander[f], Anna Linderholm[b,g], Victoria E. Mullin[h,i], Kevin G. Daly[i], Vincent M. Battista[j], Max Price[k], Kurt J. Gron[l], Panoraia Alexandri[m], Rose-Marie Arbogast[n], Benjamin Arbuckle[o], Adrian Bălăşescu[p] , Ross Barnett[l] , László Bartosiewicz[q] , Gennady Baryshnikov[r], Clive Bonsall[s], Dušan Borić[t] , Adina Boroneanț[p], Jelena Bulatović[u], Canan Çakirlar[v], José-Miguel Carretero[w], John Chapman[l], Mike Church[l], Richard Crooijmans[x], Bea De Cupere[y], Cleia Detry[z], Vesna Dimitrijevic[u] , Valentin Dumitraşcu[p], Louis du Plessis[c], Ceiridwen J. Edwards[aa], Cevdet Merih Erek[bb], Aslı Erim-Özdoğan[cc], Anton Ervynck[dd],

Domenico Fulgione[ee], Mihai Gligor[ff], Anders Götherström[gg], Lionel Gourichon[hh], Martien A.M. Groenen[x], Daniel Helmer[ii], Hitomi Hongo[jj], Liora K. Horwitz[kk], Evan K. Irving-Pease[b], Ophélie Lebrasseur[b,ll], Joséphine Lesur[mm], Caroline Malone[nn], Ninna Manaseryan[oo], Arkadiusz Marciniak[pp], Holley Martlew[qq], Marjan Mashkour[mm], Roger Matthews[rr], Giedre Motuzaite Matuzeviciute[ss], Sepideh Maziar[tt], Erik Meijaard[uu,vv,ww], Tom McGovern[xx], Hendrik-Jan Megens[x] , Rebecca Miller[yy,3], Azadeh Fatemeh Mohaseb[mm], Jörg Orschiedt[zz,aaa], David Orton[f], Anastasia Papathanasiou[bbb], Mike Parker Pearson[ccc], Ron Pinhasi[ddd], Darko Radmanovic[eee], François-Xavier Ricaut[fff], Mike Richards[ggg], Richard Sabin[hhh], Lucia Sarti[iii], Wolfram Schier[zz], Shiva Sheikhi[mm], Elisabeth Stephan[jjj], John R. Stewart[kkk], Simon Stoddart[lll], Antonio Tagliacozzo[mmm], Nenad Tasić[nnn], Katerina Trantalidou[bbb], Anne Tresset[mm,4], Cristina Valdiosera[ooo], Youri van den Hurk[v], Sophie Van Poucke[y], Jean-Denis Vigne[mm], Alexander Yanevich[ppp], Andrea Zeeb-Lanz[qqq], Alexandros Triantafyllidis[m], M. Thomas P. Gilbert[rrr,sss], Jörg Schibler[ttt], Peter Rowley-Conwy[l], Melinda Zeder[uuu], Joris Peters[vvv,www], Thomas Cucchi[mm], Daniel G. Bradley[i], Keith Dobney[ll,ggg,xxx], Joachim Burger[d], Allowen Evin[yyy], Linus Girdland-Flink[zzz], and Greger Larson[b,2]

Affiliations

a) School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
b) The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK
c) Department of Zoology, University of Oxford, Oxford, UK
d) Palaeogenetics Group, Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg-University Mainz, Mainz, Germany
e) German Archaeological Institute, Department of Natural Sciences, Berlin, Germany
f) University of York, BioArCh, Department of Archaeology, York UK
g) Department of Anthropology, Texas A&M University, College Station, USA
h) Department of Earth Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
i) Molecular Population Genetics, Smurfit Institute of Genetics Trinity College Dublin, Dublin, Ireland
j) Department of Anthropology, University of Michigan, Ann Arbor, USA
k) Massachusetts Institute of Technology, Department of Materials Science and Engineering, Cambridge, USA
l) Department of Archaeology, Durham University, Durham DH1 3LE, United Kingdom

m) mCNRS UMR 7044, Maison interuniversitaire des sciences de l'Homme, F-67083 Strasbourg Cedex, France
n) Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
o) Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
p) "Vasile Pârvan" Institute of Archaeology, Bucharest 010667, Romania
q) Osteoarchaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 106 91 Stockholm, Sweden
r) Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, St. Petersburg 199034, Russia
s) School of History, Classics and Archaeology, University of Edinburgh, Edinburgh EH8 9AG, UK
t) The Italian Academy for Advanced Studies in America, Columbia University, New York, NY 10027, USA
u) Laboratory for Bioarchaeology, Department of Archaeology, Faculty of Philosophy, University of Belgrade, 11000 Belgrade, Serbia
v) Institute of Archaeology, University of Groningen, 9712 ER, Groningen, The Netherlands
w) Laboratorio de Evolución Humana, Departamento de Historia, Geografía y Comunicación Universidad de Burgos, Burgos, Spain
x) Animal Breeding and Genomics Center, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
y) OD Earth and History of Life, Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium
z) Centro de Arqueologia da Universidade de Lisboa, Faculdade de Letras da Universidade de Lisboa, Alameda da Universidade, 1600-214 Lisboa, Portugal
aa) Department of Biological and Geographical Sciences, University of Huddersfield, Huddersfield HD1 3DH, United Kingdom
bb) Department of Archaeology, Gazi University, Ankara 06500, Turkey
cc) Department of Archaeology, Çanakkale Onsekiz Mart University, Çanakkale 17100, Turkey
dd) Flanders Heritage Agency, 1000 Brussels, Belgium
ee) Department of Biology, University of Naples Federico II, 80126 Napoli, Italy
ff) History, Archaeology and Museology Department, 1 Decembrie 1918 University, Alba Iulia 510009, Romania

gg) Department of Archaeology and Classical Studies, Stockholm University, SE-106 91 Stockholm, Sweden
hh) Université Côte d'Azur, CNRS, Cultures et Environnement, Préhistoire, Antiquité, Moyen Âge (UMR 7264), 06357 Nice, France
ii) CNRS, Archéorient (UMR 5133), Maison de l'Orient et de la Méditerranée, 69007 Lyon, France
jj) Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan
kk) National Natural History Collections, Faculty of Life Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
ll) Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool L69 7WZ, UK
mm) Unité Archéozoologie, Archéobotanique, Sociétés Pratiques et Environnements (AASPE), CNRS, Muséum National d'Histoire Naturelle, 75020 Paris, France
nn) School of Natural and Built Environment, Queen's University Belfast, Belfast BT9 5AG, UK
oo) Scientific Center of Zoology and Hydroecology, Institute of Zoology, Yerevan 0014, Armenia
pp) Institute of Archaeology, Adam Mickiewicz University, 61-712, Poznan´, Poland
qq) The Hellenic Archaeological Research Foundation, Tivoli House, Cheltenham GL50 2TD, UK
rr) Department of Archaeology, University of Reading, Reading RG6 6AB, UK ss) Lithuanian Institute of History, Vilnius University, LT-01513 Vilnius, Lithuania
tt) Institut für Archäologische Wissenschaften, Goethe University of Frankfurt, 60323 Frankfurt, Germany
uu) International Union for Conservation of Nature/Species Survival Commission Wild Pig Specialist Group, 15412 Jakarta, Indonesia
vv) Center of Excellence for Environmental Decisions, University of Queensland, St Lucia, QLD 4072, Australia
ww) Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, Marlowe Building, University of Kent, Canterbury, Kent CT2 7NR, UK
xx) Anthropology Department, Hunter College and Graduate Center, City University of New York, New York, NY 10065, USA
yy) Service de Préhistoire, Université de Liège, 4000 Liège, Belgium

zz) Institute of Prehistoric Archaeology, Free University of Berlin, 14195 Berlin, Germany
aaa) Curt-Engelhorn-Zentrum Archäometrie, 68159 Mannheim, Germany
bbb) Ephorate of Paleoanthropology and Speleology, Greek Ministry of Culture, 106 82 Athens, Greece
ccc) Institute of Archaeology, University College London, London WC1H 0PY, UK
ddd) Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria
eee) Museum of Vojvodina, 21101 Novi Sad, Serbia
fff) Laboratoire Évolution & Diversité Biologique-UMR 5174, Université de Toulouse Midi-Pyrénées, 31062 cedex 9 Toulouse, France
ggg) Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
hhh) Division of Vertebrates, Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK
iii) Dipartimento di Scienze storiche e dei Beni Culturali, University of Siena, 53100 Siena, Italy
jjj) Osteologie, Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart, 73728 Konstanz, Germany
kkk) Faculty of Science and Technology, Bournemouth University, Fern Barrow, Poole, Dorset BH12 5BB, UK
lll) Magdalene College, University of Cambridge, Cambridge CB3 0AG, UK
mmm) Sezione di Bioarcheologia, Museo delle Civiltà, 00144 Roma, Italy
nnn) Department of Archaeology, Faculty of Philosophy, University of Belgrade, 11000 Belgrade, Serbia
ooo) Department of Archaeology and History, Faculty of Humanities and Social Sciences, La Trobe University, Melbourne, MB 167, Australia
ppp) Institute of Archaeology of the National Academy of Sciences of Ukraine, 02000 Kiev, Ukraine
qqq) Generaldirektion Kulturelles Erbe Rheinland-Pfalz, Dir. Landesarchäologie, D–67346 Speyer, Germany
rrr) Natural History Museum of Denmark, University of Copenhagen, DK-1123 Copenhagen, Denmark
sss) University Museum, Norwegian University of Science and Technology, 7012 Trondheim, Norway
ttt) Integrative Prehistory and Archaeological Science, University of Basel, 4055 Basel, Switzerland

uuu) Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 37012
vvv) ArchaeoBioCenter and Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication and the History of Veterinary Medicine, Ludwig Maximilian University Munich, 80539 Munich, Germany
www) State Collection for Anthropology and Palaeoanatomy, Bavarian Natural History Collections, 80333 Munich, Germany
xxx) Department of Archaeology, School of Geosciences, University of Aberdeen, St. Mary's, Aberdeen AB24 3FUK, UK
yyy) Institut des Sciences de l'Evolution-Montpellier-UMR 5554-CNRS, IRD, Université de Montpellier, 34090 Montpellier, France
zzz) Research Centre in Evolutionary Anthropology and Palaeoecology, School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK

1) L.A.F.F., J.H., and A.T.L. contributed equally to this work.
2) Corresponding authors: laurent.frantz@qmul.ac.uk or greger.larson@arch.ox.ac.uk.
3) Deceased March 20, 2017.
4) Deceased January 17, 2019.

Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe

Laurent A. F. Frantz[a,b,1,2], James Haile[b,1], Audrey T. Lin[b,c,1], Amelie Scheu[d], Christina Geörg[d], Norbert Benecke[e], Michelle Alexander[f], Anna Linderholm[b,g], Victoria E. Mullin[h,i], Kevin G. Daly[i], Vincent M. Battista[j], Max Price[k], Kurt J. Gron[l], Panoraia Alexandri[m], Rose-Marie Arbogast[n], Benjamin Arbuckle[o], Adrian Bălăşescu[p], Ross Barnett[l], László Bartosiewicz[q], Gennady Baryshnikov[r], Clive Bonsall[s], Dušan Borić[t], Adina Boroneanţ[p], Jelena Bulatović[u], Canan Çakirlar[v], José-Miguel Carretero[w], John Chapman[l], Mike Church[l], Richard Crooijmans[x], Bea De Cupere[y], Cleia Detry[z], Vesna Dimitrijevic[u], Valentin Dumitraşcu[p], Louis du Plessis[c], Ceiridwen J. Edwards[aa], Cevdet Merih Erek[bb], Aslı Erim-Özdoğan[cc], Anton Ervynck[dd], Domenico Fulgione[ee], Mihai Gligor[ff], Anders Götherström[gg], Lionel Gourichon[hh], Martien A.M. Groenen[x], Daniel Helmer[ii], Hitomi Hongo[jj], Liora K. Horwitz[kk], Evan K. Irving-Pease[b], Ophélie Lebrasseur[b,ll], Joséphine Lesur[mm], Caroline Malone[nn], Ninna Manaseryan[oo], Arkadiusz Marciniak[pp], Holley Martlew[qq], Marjan Mashkour[mm], Roger Matthews[rr], Giedre Motuzaite Matuzeviciute[ss], Sepideh Maziar[tt], Erik Meijaard[uu,vv,ww], Tom McGovern[xx], Hendrik-Jan Megens[x], Rebecca Miller[yy,3], Azadeh Fatemeh Mohaseb[mm], Jörg Orschiedt[zz,aaa], David Orton[f], Anastasia Papathanasiou[bbb], Mike Parker Pearson[ccc], Ron Pinhasi[ddd], Darko Radmanović[eee], François-Xavier Ricaut[fff], Mike Richards[ggg], Richard Sabin[hhh], Lucia Sarti[iii], Wolfram Schier[zz], Shiva Sheikhi[mm], Elisabeth Stephan[jjj], John R. Stewart[kkk], Simon Stoddart[lll], Antonio Tagliacozzo[mmm], Nenad Tasić[nnn], Katerina Trantalidou[bbb], Anne Tresset[mm,4], Cristina Valdiosera[ooo], Youri van den Hurk[v], Sophie Van Poucke[y], Jean-Denis Vigne[mm], Alexander Yanevich[ppp], Andrea Zeeb-Lanz[qqq], Alexandros Triantafyllidis[m], M. Thomas P. Gilbert[rrr,sss], Jörg Schibler[ttt], Peter Rowley-Conwy[l], Melinda Zeder[uuu], Joris Peters[vvv,www], Thomas Cucchi[mm], Daniel G. Bradley[i], Keith Dobney[ll,ggg,xxx], Joachim Burger[d], Allowen Evin[yyy], Linus Girdland-Flink[zzz], and Greger Larson[b,2]

Edited by Dolores R. Piperno, Smithsonian Institution, Washington, DC, and approved June 24, 2019 (received for review February 8, 2019)

Archaeological evidence indicates that pig domestication had begun by ∼10,500 y before the present (BP) in the Near East, and mitochondrial DNA (mtDNA) suggests that pigs arrived in Europe alongside farmers ∼8,500 y BP. A few thousand years after the introduction of Near Eastern pigs into Europe, however, their characteristic mtDNA signature disappeared and was replaced by haplotypes associated with European wild boars. This turnover could be accounted for by substantial gene flow from local European wild boars, although it is also possible that European wild boars were domesticated independently without any genetic contribution from the Near East. To test these hypotheses, we obtained mtDNA sequences from 2,099 modern and ancient pig samples and 63 nuclear ancient genomes from Near Eastern and European pigs. Our analyses revealed that European domestic pigs dating from 7,100 to 6,000 y BP possessed both Near Eastern and European nuclear ancestry, while later pigs possessed no more than 4% Near Eastern ancestry, indicating that gene flow from European wild boars resulted in a near-complete disappearance of Near East ancestry. In addition, we demonstrate that a variant at a locus encoding black coat color likely originated in the Near East and persisted in European pigs. Altogether, our results indicate that while pigs were not independently domesticated in Europe, the vast majority of human-mediated selection over the past 5,000 y focused on the genomic fraction derived from the European wild boars, and not on the fraction that was selected by early Neolithic farmers over the first 2,500 y of the domestication process.

domestication | evolution | gene flow | Neolithic

Significance

Archaeological evidence indicates that domestic pigs arrived in Europe, alongside farmers from the Near East ∼8,500 y ago, yet mitochondrial genomes of modern European pigs are derived from European wild boars. To address this conundrum, we obtained mitochondrial and nuclear data from modern and ancient Near Eastern and European pigs. Our analyses indicate that, aside from a coat color gene, most Near Eastern ancestry in the genomes of European domestic pigs disappeared over 3,000 y as a result of interbreeding with local wild boars. This implies that pigs were not domesticated independently in Europe, yet the first 2,500 y of human-mediated selection applied by Near Eastern Neolithic farmers played little role in the development of modern European pigs.

The emergence of agricultural societies in the Near East at least 12,500 y before the present (BP) was followed by the westward dispersal of farmers into Europe beginning ∼8,500 y BP (1–4). This Neolithic expansion was characterized by the human-mediated dispersal of domesticated plants and animals, including cereals, pulses, sheep, goats, cattle, and pigs, all of which were derived from wild species indigenous to the Near East and Anatolia (5, 6). Given that the wild progenitors of modern domestic sheep and goats were never present in Europe, the presence of their remains in European archaeological sites

Author contributions: L.A.F.F., L.G.-F., and G.L. designed research; L.A.F.F., J.H., V.E.M., K.G.D., O.L., T.C., A. Evin, and L.G.-F. performed research; L.A.F.F., J.H., A.S., C.G., N.B., M.A., A.L., V.E.M., K.G.D., M.P., K.J.G., P.A., R.-M.A., B.A., A. Bălăşescu, R.B., L.B., G.B., C.B., D.B., A. Boroneant, J. Bulatović, C.Ç., J.-M.C., J.C., M.C., R.C., B.D.C., C.D., V. Dimitrijevic, V. Dumitraşcu, C.J.E., C.M.E., A.E.-Ö., A. Ervynck, D.F., M.G., A.G., L.G., M.A.M.G., D.H., H.H., L.K.H., O.L., J.L., C.M., N.M., A.M., H.M., M.M., R. Matthews, G.M.M., S.M., E.M., T.M., H.-J.M., R. Miller, A.M., J.O., D.O., A.P., M.P.P., R.P., D.R., F.-X.R., M.R., R.S., L.S., W.S., S. Sheikhi, E.S., J.R.S., S. Stoddart, A. Tagliacozzo, N.T., K.T., A. Tresset, C.V., Y.v.d.H., S.V.P., J.-D.V., A.Y., A.Z.-L., A. Triantafyllidis, M.T.P.G., J.S., P.R.-C., M.Z., J.P., T.C., D.G.B., K.D., J. Burger, A. Evin, L.G.-F., and G.L. contributed new reagents/analytic tools; L.A.F.F., A.T.L., V.M.B., L.d.P., and E.K.I.-P. analyzed data; and L.A.F.F., A.T.L., K.J.G., P.R.-C., M.Z., J.P., and G.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Data deposition: The reads from both ancient and modern samples have been deposited at the European Nucleotide Archive (ENA) (project no. PRJEB30282).

[1]L.A.F.F., J.H., and A.T.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: laurent.frantz@qmul.ac.uk or greger.larson@arch.ox.ac.uk.

[3]Deceased March 20, 2017.

[4]Deceased January 17, 2019.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1901169116/-/DCSupplemental.

almost certainly represents populations originally domesticated in Anatolia and the Near East. In the case of cattle and pigs, however, the widespread distribution of their wild progenitors across most of Eurasia complicates the classification of archaeological specimens as wild or domestic, and leaves open the possibility that these taxa were also independently domesticated in Europe. Consequently, the relative contribution of European wild boars populations to the gene pools of domestics introduced from the Near East remains contentious (7).

Traditional methods for distinguishing between wild and domestic pigs rely primarily on archaeological context and size differences (8) or are based on demographic profiling (9, 10). More recent methods have relied on the analysis of dental shape variation using geometric morphometrics (11, 12) and stable isotopes (13). Morphological analyses of archeological pig remains have indicated that the first domestic pigs introduced from the Near East were substantially smaller than European wild boars, something most clearly visible in tooth size (e.g., ref. 14). Dental development is generally unaffected by nutrition until extreme starvation approaches (15), and tooth size is slow to change. For example, Australian feral pigs whose ancestors have been living outside of anthropogenic contexts for as long as 2 centuries still possess the small tooth sizes of their domestic ancestors, even though their body size has substantially enlarged (16). In Europe, the earliest domesticated pigs (identified using tooth size) have been recovered from archaeological contexts associated with the earliest Neolithic farmers by ∼8,000 y BP (e.g., ref. 14), and these tooth size differences persist from prehistory to the present day (8, 17). Thus, the archaeological evidence implies that none of the *Sus scrofa* present in Europe before the arrival of Near Eastern farmers can be classified as domestic, indicating that European hunter-gatherers did not independently domesticate local wild boars.

Although the phenotype associated with Near Eastern domestic pigs does not appear to vary considerably following their introduction to Europe (18–20), there is substantial discontinuity with respect to their maternal (mitochondrial DNA [mtDNA]) ancestry. Ancient mtDNA analysis has shown that pigs of Near Eastern maternal ancestry occurred as far west as the Paris Basin (∼6,000 y BP) among early Neolithic European domestic pigs (21). By 5,900 y BP, however, these Near Eastern genetic signatures had been replaced by those of European wild boars (21), and it is possible that the Near Eastern ancestry also vanished from the nuclear genome of modern domestic pigs. A recent analysis of ∼37,000 single-nucleotide polymorphisms (SNPs) typed in modern pigs (22) was consistent with this hypothesis, but this study was likely underpowered due to ascertainment biases and a lack of ancient Near Eastern domestic and wild reference populations.

One possible mechanism to account for the apparent discontinuity between genotype and phenotype is gene flow from local European wild boars into the introduced domestic population. Domestic pigs have likely always interacted and interbred with wild populations, and this process has been suggested wherever domestic animals have arrived (e.g., ref. 23). Genetic introgression (including the mitochondrial genome) from local wild boars into the introduced domestic population potentially involved wild females being captured [perhaps as piglets during hunting as in modern New Guinea (24, 25)] and kept in farming settlements. Were these females allowed to reach sexual maturity and breed with male domestics, the offspring would possess mtDNA (and some nuclear ancestry) associated with local wild boars. Although perhaps initiated as an accident, if the offspring of the wild-caught females were perceived to possess superior traits, the acquisition of wild female piglets may have become a regular practice.

If this admixture was limited (at least initially), and the gene flow from wild boars did not substantially affect the phenotype of the domestic population, it is possible that modern domestic pigs retain a sufficient, yet undetected, fraction of Near Eastern ancestry that underlies domestic traits (26). This scenario of continuous gene flow with European wild boars predicts a gradual and incomplete genomic replacement. If pig domestication was a completely independent process, European pigs would derive exclusively from European wild boars, resulting in a sharp discontinuity of Near Eastern ancestry.

Here, we assessed whether modern domestic pig genomes retain a Near Eastern component that is essential for maintaining their domestic characteristics, and characterized the extent, speed, and mechanisms by which pigs acquired European wild boars ancestry. To do so, we obtained mitochondrial (including PCR data [n = 230] and next-generation sequencing (NGS) data [n = 327]) and nuclear data, including 2 high-coverage (>10-fold), 7 medium-coverage (1- to 10-fold), and 54 low-coverage (<1-fold) genomes from an assessment of >500 archeological pig remains (Dataset S1). Our dataset (including publicly available sequences) spans the past 14,000 y and includes a total of 2,099 samples from the Near East and Europe, including samples from contexts that precede and follow the origins of pig domestication.

Results and Discussion

A Neolithic Mitochondrial Turnover. Our mtDNA analysis revealed 2 broad groups: 1 from Western and Eastern Europe, including mt-Italian, mt-A, mt-C, and mt-Y2 haplogroups (Fig. 1*A* and *SI Appendix*, Figs. S7 and S8), and 1 from the Near East, including haplogroups mt-Y1 and mt-ArmT (Fig. 1*A* and *SI Appendix*, Figs. S7 and S8). These results substantiate previous findings that mt-Y1 and mt-ArmT are indigenous to the Near East, although mt-Y2, previously thought to be found exclusively in the Near East (21), also appears to be present in wild boars from the Balkans and northeast Italy (19, 27) (*SI Appendix*). In addition, the mt-Y1 signature, originally restricted to the Near East (Fig. 1*A*), was not only identified in early Neolithic contexts in the Near East and Europe but was also found in pigs that (based on context and traditional biometrical analysis) were assigned a domestic status (21, 28) (*SI Appendix*).

Altogether, this confirms that Near Eastern farmers brought domestic pigs possessing an mt-Y1 signature into Europe during the Neolithic expansion (21, 28). Our analysis of mtDNA data from 2,099 samples (557 newly generated data), including 1,318 ancient samples (262 of wild boars, 592 of domestic pigs, and 464 of unknown status) and 781 modern samples (467 of wild boars and 314 of domestic pigs), demonstrates that the first appearance of the mt-Y1 haplotype in our continental European dataset was ∼8,000 y ago in Neolithic Bulgarian pigs (Kovačevo: Kov18, Kov21), and its terminal appearance in a Neolithic context was ∼5,100 y ago in a Polish sample (AA134, Żegotki 2).

The few pigs possessing an mt-Y1 signature from post-Neolithic contexts were found mostly on islands beyond mainland Europe in southwestern Greece (4,350 to 3,250 y BP: MM495, MM486, MM303), in Crete (3,100 y BP), in Sardinia (∼3,750 y BP) (29), near Naples (∼800 y BP: VM_CM01, VM_CM02, VM_CM03), and in Corsica (modern noncommercial pigs) (21), as well as in Tuscany (∼800 y BP: VM_TM01) (Fig. 1 and *SI Appendix*). The persistence of the mt-Y1 signature within pigs on islands mimics the patterns seen in isolated island populations of both sheep and humans. For instance, sheep in Orkney and St. Kilda (30), and human populations in Sardinia (31), were not subjected to significant introgression from later migratory waves and, instead, possess a larger proportion of Anatolian/Near Eastern ancestry relative to their mainland counterparts.

Gene Flow and a Corresponding Near-Complete Nuclear Turnover. While these data confirm the existence of a complete turnover of mtDNA, this marker does not provide sufficient power to assess whether the turnover was the result of introgression with local female wild boars or the result of an indigenous domestication process (28). To address this issue, we sequenced 2 high-coverage, 7 medium-coverage, and 54 low-coverage ancient genomes spanning over 9,000 y. A neighbor-joining phylogenetic reconstruction of modern and ancient wild boars nuclear data reflects the distinct geographic partitioning of mtDNA data in western Eurasia (32). More specifically, distinct ancestries are present within ancient European and Near Eastern wild boars remains that predate

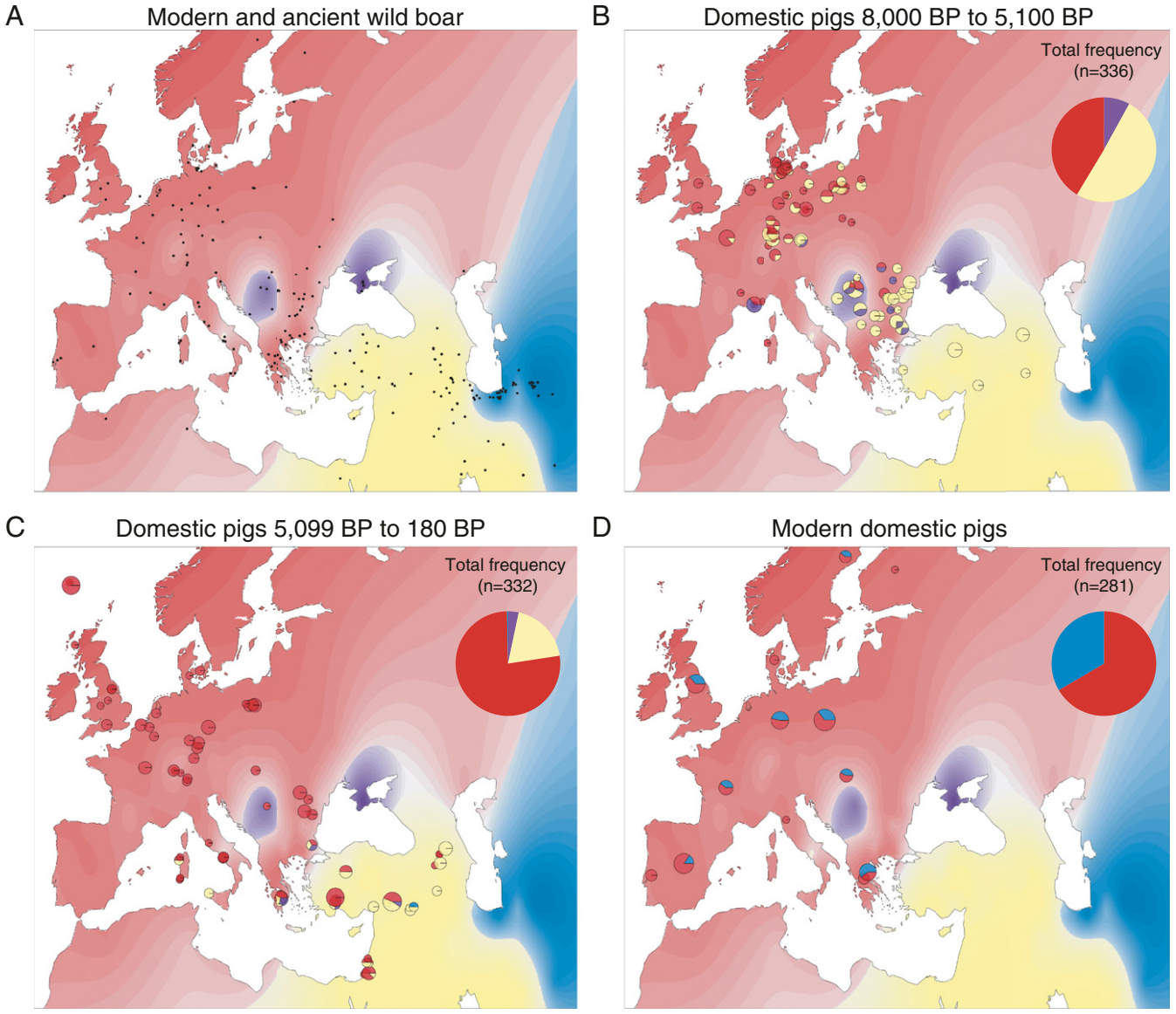


Fig. 1. (*A*) Map representing the distribution of East Asian (blue), Near Eastern (including haplogroups mt-Y1 and mt-ArmT; yellow), European (including haplogroups mt-Italian, mt-A, mt-C, and mt-Y2; red), and Y2 (purple) haplogroups in wild boars. Black dots represent the locations of 696 modern and ancient wild boar. Haplogroup assignments were used to interpolate the underlying color distribution, which demonstrates the biogeographical boundaries of these 3 general haplogroups. (*B*) Large pie chart in the upper right corner of the map represents overall frequencies of these haplogroups in domestic pigs. Small pie charts on the map show the frequencies at various archeological sites/locations between 8,000 y BP and 5,100 y BP (*B*), between 5,099 and 180 y BP [before the Industrial Revolution and the introduction of Asian pigs in Europe (35) (*C*), and in modern pigs (*D*)]. A few samples from our datasets have been excluded from these plots; more details are provided in *SI Appendix*, Figs. S6 and S7.

domestication (Fig. 2*A* and *SI Appendix*, Fig. S10). An ADMIXTURE analysis of 38 wild boars nuclear genomes, including an ancient wild boars from Aşıklı Höyük (~10,000 y BP, Turkey) reveals that modern wild individuals from Greece possess 33 to 38% Near Eastern nuclear ancestry, while those from Italy possess only 6 to 10% (Fig. 2*A*). The decreasing proportion of Near Eastern ancestry among wild boars from Greece to Italy most likely reflects admixture between wild populations from Anatolia into Greece and then into Italy (*SI Appendix*, Fig. S12). It is also possible, however, that a portion of the Anatolian ancestry found in Italian wild boars is the result of admixture from domestic pigs derived from the Near East into wild populations, instances of which have previously been shown to have occurred in northern Germany (33).

Additional ADMIXTURE analyses, including 111 genomes, clearly demonstrate that most modern domestic pigs (77 of 85) do not possess significant levels of Near Eastern ancestry (*SI Appendix*, Figs. S15 and S16). In fact, when modern European domestic pigs are treated as a single population, our haplotype-based analyses [GLOBETROTTER (34)] indicate that their overall Near Eastern ancestry is only ~4% (*SI Appendix*), and most of this Near Eastern signal is derived from a few modern breeds from Italy, Hungary, and Spain that possessed 1.7 to 6.4% Near Eastern nuclear ancestry (Fig. 2*B*). Interestingly, the majority of these breeds occur in regions of Europe where modern wild boar possess, on average, higher levels of Near Eastern ancestry (6 to 33%; Fig. 2 *A* and *B*), and, as opposed to many other European populations, these breeds were not mixed with Chinese pigs during breed improvement programs during the 19th century (35, 36) (*SI Appendix*, Figs. S15 and S16). It is therefore likely that the limited Near Eastern ancestral component detected in these samples was acquired through gene flow with local wild boars (in Italy or the Balkans), and maintained as a result of a lack of admixture with introduced Chinese pigs.

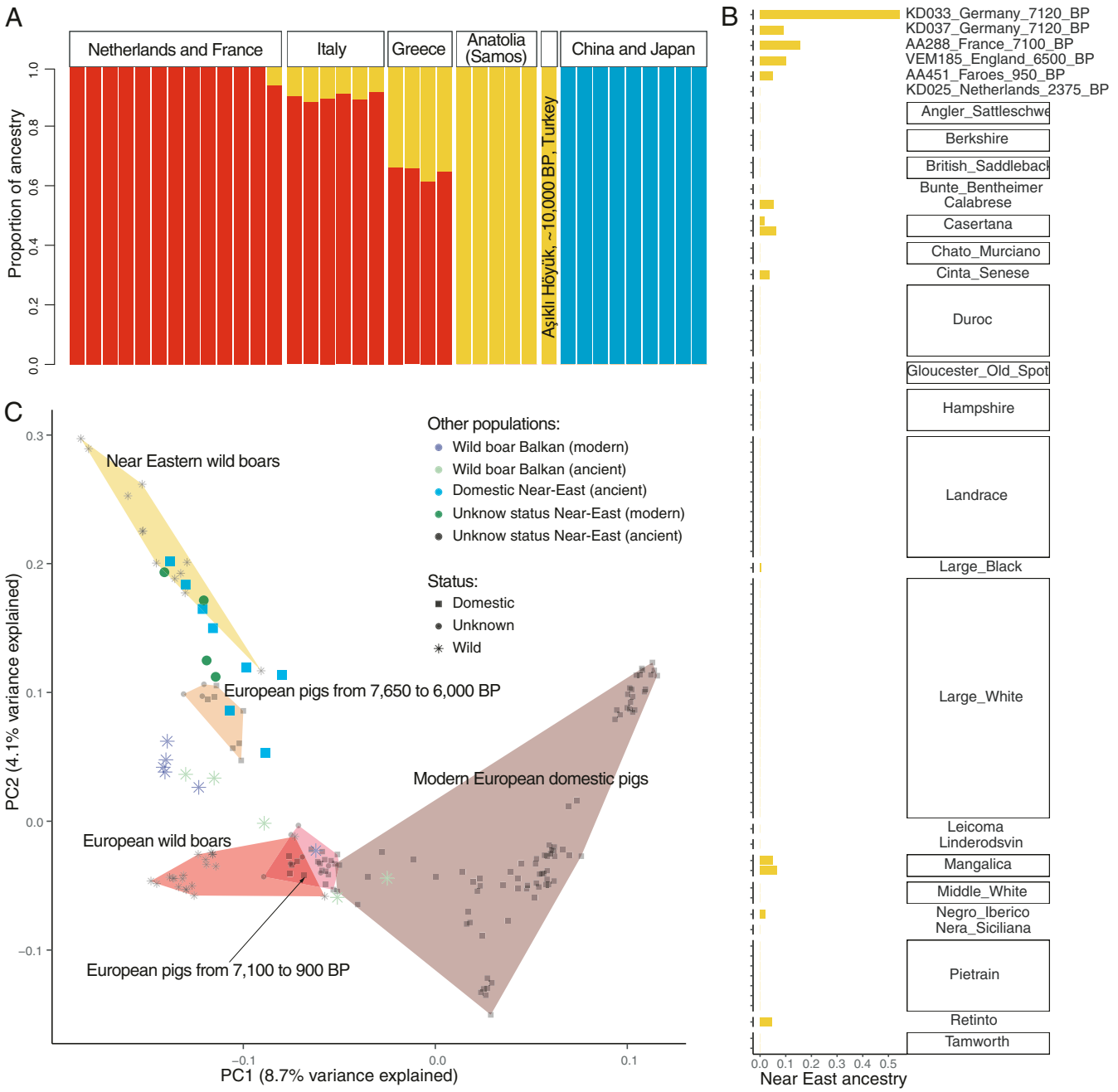


Fig. 2. (*A*) Bar plots representing the proportion of ancestry from Europe (red), the Near East (yellow), and East Asia (blue) in Eurasian wild boar genomes. (*B*) Bar plots depicting the proportion of Near Eastern ancestry in modern and ancient European domestic pigs. (*C*) PCA (excluding East Asian domestic pigs; *SI Appendix*, Fig. S14) showing the existence of 2 groups of ancient domestic pigs: 1 close to Near Eastern wild boar and 1 close to European wild boar.

We further assessed the degree of Near Eastern ancestry in archaeological pigs. Our ADMIXTURE analysis indicates that Bronze Age domestic pigs from western Iran (~4,300 y BP: AA363) and Armenia (~3,500 y BP: AA119) did not possess any European ancestry, and were exclusively derived from ancient Near Eastern wild boar (*SI Appendix*, Figs. S15 and S16). In Europe, 4 ancient high/medium-coverage domestic pigs did possess Near Eastern nuclear ancestry (Fig. 2*B*). Specifically, 2 early Neolithic samples from Herxheim, Germany (~7,100 y BP: KD033, KD037) possessed ~54% and ~9% Near Eastern ancestry, respectively; a domestic pig from la Baume d'Oulen, France (~7,100 y BP: AA288) possessed 15%; a Late Neolithic sample from Durrington Walls in Britain (~4,500 y BP: VEM185) possessed ~10%; and a 1,000-y-old Viking Age sample from the Faroe Islands (AA451) possessed only 5%. Of these, only the Herxheim sample (KD033), with ~54% Near Eastern ancestry, possessed the Near Eastern mt-Y1 haplotype (Dataset S1), and also had substantially more Near Eastern ancestry than any of the ancient or modern European wild boar (Fig. 2*B*). This is supported by outgroup f3-statistics analysis, which indicates that KD033 shares more drift with Near Eastern wild boar than any other ancient or modern pig genome (*SI Appendix*, Fig. S17), as well as significant D-statistics of the form D (outgroup, Near Eastern wild boar; European wild boar: KD033) ($Z << 3$; *SI Appendix*, Fig. S18). These results indicate that European wild boars were being incorporated into domestic populations relatively soon after the latter were introduced from the Near East.

To obtain a more precise temporal and geographic resolution of the disappearance of Near Eastern genomic signatures in Europe, we performed additional analyses of 54 low-coverage ancient genomes (<1-fold) that possessed sufficient data (>5,000 SNPs covered from a panel of ∼12 million SNPs; *SI Appendix*) to be confidently projected onto a principal component analysis (PCA) alongside both modern and (high- and medium-coverage) ancient genomes. We analyzed these data together with those of Asian wild and domestic pigs. In this analysis, principal component 1 (PC1) separated European and Asian pigs, while PC2 separated Near Eastern and European pigs (*SI Appendix*, Fig. S14). After removing Asian pigs, PC1 separated modern European domestic pigs from all other samples, while PC2 separated European from Near Eastern pigs (Fig. 2*C*). The separation between European domestic pigs and all other samples on PC1 is most likely the result of admixture between Asian and European breeds following breed improvement programs in the 19th century (35, 36) (*SI Appendix*, Figs. S15 and S16).

The PCA revealed 2 groups of ancient European pigs (including 25 previously identified as domestic using a combination of morphometric and contextual data and 10 with unknown status) (Fig. 2*C*). The first group consisted of 8 domestic pigs that are closer to Near Eastern wild boars and ancient Near Eastern domestic pigs (Fig. 2*C*). In all, this group comprised Neolithic pigs from contexts dating from 7,650 to 6,100 y BP, including the following: Madzhari, Northern Macedonia (∼7,650 y BP: BLT022, BLT023); Herxheim, Germany (7,100 y BP: KD033, KD032); Măgura, Romania (7,100 y BP: BLT010); Pločnik, Serbia (∼6,650 y BP: AA212); Vinča Belo Brdo, Serbia (∼6,500 y BP: BLT014); and Căscioarele, Romania (∼6,000 y BP: AA072). Interestingly, 7 of these samples also possessed the Near Eastern mt-Y1 haplogroup (AA212 is unknown) (Dataset S1). We also identified 3 samples from Buran-Kaya, Crimea (∼7,000 y BP: AA380, AA480, AA483) that also cluster close to Near Eastern wild boars, although they each possess the mt-Y2 haplotype and so are thought to be local wild boars (Dataset S1).

The second group of ancient European samples was closer to wild and modern domestic pigs from Europe and included samples that are mostly younger in age than the first group. This second group consisted of 18 domestic samples from overall more recent archaeological sites dating from 7,100 to 900 y BP, including the following: Herxheim, Germany (7,100 y BP: KD037); Oulens, France (∼7,100 y BP: AA288); Bozdia, Poland (∼6,700 y BP: AA346; ∼900 y BP: AA343, AA341); Durrington Walls, England (∼4,500 y BP: VEM183, VEM184, VEM185); Utrecht, The Netherlands (∼2,300 y BP: KD025; ∼700 y BP: KD024); Basel, Switzerland (∼2,000 y BP: AA266); Coppergate, England (∼1,800 y BP: AA301); Undir Junkariusfløtti, Faroe Islands (∼1,000 y BP: AA451, AA411, AA414, AA418, AA440); and Ciechrz, Poland (∼900 y BP: AA139). This group also comprised 7 ancient samples that could not be identified as either wild or domestic, including the following: la Grotte du Taï, France (∼7,100 y BP: AA294); Santa Maria in Selva, Italy (Late Neolithic: AA629); and El Portalón, Spain (∼5,400 y BP: AA513; ∼4,500 y BP: AA507; ∼3,600 y BP: AA512, AA511;∼900 y BP: AA513). Lastly, 2 ancient wild boars, 1 from Birsmatten-Basisgrotte, Switzerland (∼7,700 y BP: AA241) and 1 from Siniarzewo, Poland (∼2,900 y BP: LG507) were also found to fall closer to modern European wild boars. All of these samples possessed a European mtDNA signature (Dataset S1).

Collectively, these results reveal a fluctuating temporal pattern of Near Eastern genomic ancestry in western Eurasian domestic pigs, and the general trend shows that the samples closer in time and space to the source of the first Near Eastern pigs possessed a greater proportion of Near Eastern ancestry. In mainland Europe, domestic pigs from Neolithic sites situated around the Styrmon (e.g., Northern Macedonia), Danube (e.g., Romania), and Rhin (e.g., Germany) river systems in Germany, Romania, Macedonia, and Serbia possessed substantially more Near Eastern ancestry than is present in European wild boar (Fig. 2 *B* and *C*). The timing of the first (∼8,000 y BP) and last (∼5,100 y BP) appearances of Near Eastern mtDNA signatures in continental Europe [apart from 4 Italian suids from AD 1800 (37)] is coincident with our nuclear data, indicating that <3,000 y after domestic pigs were introduced, their Near Eastern ancestry (at both mitochondrial and nuclear levels) had all but vanished. The hybrid nature of the high-coverage genome from the Neolithic Herxheim pig in Germany (7,100 y BP: KD033; Fig. 2*B*) indicates that this disappearance was most likely gradual, and was the result of gene flow from European wild boar into the introduced Near Eastern domestic pig populations.

The Extent of Near Eastern Ancestry in Modern Domestic Pigs. To assess the threshold above which we could confidently identify Near Eastern ancestry in our ancient data, we simulated genomes with predefined Near Eastern ancestry proportions and analyzed the data using ADMIXTURE (38). We then used a binomial distribution to compute the probability of successfully detecting Near Eastern ancestry in 8 of 85 genomes (reflecting our modern data) (Fig. 2*B* and *SI Appendix*, Fig. S19). For admixture values ≥5%, the probability of observing only 8 genomes with Near Eastern ancestry is <1% (*SI Appendix*, Fig. S19*A*). This indicates that ADMIXTURE should detect significantly more pigs with Near Eastern ancestry if the genome of every modern domestic pig possessed a Near Eastern component ≥5%. Additionally, our simulations indicate that the GLOBETROTTER (34) analysis can accurately detect 4% Near Eastern ancestry (*SI Appendix*, Fig. S19*B*), which is less than what is present in modern Italian and Balkan wild boar. If a degree of Near Eastern ancestry was essential for the maintenance of the domestic phenotype in Europe, we would therefore predict that the underlying causative variants are present in no more than ∼4% of the genome.

To further explore this possibility, we investigated whether regions of modern domestic pig genomes reported to be subjected to positive selection (26) were more closely related to either Near Eastern or European wild boar. To do so, we first phased modern and high-coverage ancient genome data using shapeit (39). For each positively selected region, we computed the nucleotide distance between every pair of domestic and wild haplotypes. For each domestic pig haplotype, we computed the normalized difference between the nucleotide distance of the closest European haplotype and the closest Near Eastern wild boar haplotype. We then plotted the mean and SD of this statistic for each sweep region (*SI Appendix*). Our results show that a large majority of domestic pig haplotypes within these sweep regions share a closer genetic affinity to European wild boars than to Near Eastern wild boar (271 of 298; *SI Appendix*, Fig. S20). In fact, we did not identify a single region that was closer to Near Eastern wild boars (*SI Appendix*, Fig. S20). This suggests that the majority of human-mediated selection that took place after the arrival of pigs in Europe most likely did not target haplotypes of Near Eastern origin. We could not, however, distinguish between European and Near Eastern ancestry in ∼10 sweep regions. Given the bias toward modern European wild boar haplotypes in our dataset, it is possible that our analysis did not possess sufficient power to identify Near Eastern ancestry in those ∼10 regions. Doing so will require additional sequencing of modern and ancient Near Eastern pigs.

The Evolution and Dispersal of Black Coat Color. To further assess the potential relevance of Near Eastern ancestry to the genetic and phenotypic makeup of early and modern domestic pigs, we investigated the Melanocortin 1 Receptor (*MC1R*) gene. This gene has been shown to harbor functional mutations (linked to the loss of camouflage coat color) that are highly correlated with domestic status (*SI Appendix*). Our analyses of previously published and novel modern and ancient *MC1R* sequence data (269 domestic pigs and 46 wild boar) demonstrate that a specific nonsynonymously derived mutation [D124N (40)], which is associated with black (or black and white spotted) coat color in western Eurasian domestic pigs, is almost absent in both modern and ancient wild boars from the Near East and Europe (1 of 92; *SI Appendix*, Fig. S8). The only wild boar that possessed 1 copy of the derived allele originated from a population in The Netherlands that is known to have recently interbred

with domestic pigs (41). By characterizing this SNP in ancient domestic pigs (using NGS and PCR assays; Dataset S1), we identified 64 of 76 animals with at least 1 copy of the derived allele (the remaining 12 were homozygous for the wild type). Altogether, this suggests that while the ancestral allele at this locus cannot be used to unequivocally distinguish wild and domestic pigs, the derived allele is highly indicative of domestic status.

The earliest pigs that possessed the derived allele were found at Neolithic Ulucak Höyük in western Anatolia (~8,650 y BP: AL1102; ~8,250 y BP: Ulu48). The earliest European pigs that possess the derived allele are from Neolithic sites in Bulgaria (~7,500 y BP: Cav6, Kov19), Romania (~7,200 y BP: Uiv10), and Germany (~7,100 y BP: KD033, KD037). Further phylogenetic analysis of the ~100-kb region surrounding the *MC1R* gene indicated that 169 of 174 phased sequences, obtained from high-coverage modern and ancient domestic pigs that possessed the D124N allele, clustered in a monophyletic clade (*SI Appendix*, Fig. S9).

This result suggests that the D124N mutation found in Near Eastern and European pigs arose just once and was maintained, despite substantial gene flow with European wild boars. Interestingly, the nearest clade to this monophyletic cluster consisted of 2 haplotypes found in modern wild boar with European ancestry (The Netherlands) and Near Eastern ancestry (from Samos off the Anatolian west coast; *SI Appendix*, Fig. S9). This finding indicates that we do not possess the resolution to infer whether the D124N mutation (now fixed in many domestic breeds) first arose in the Near East or in Europe. Although we cannot definitively identify the geographic origin of the D124N mutation using phylogenetic analysis, the fact that it occurred in Anatolia before the arrival of domestic pigs into Europe, and that it likely arose only once, strongly suggests that this trait originated in Anatolia and was present in the first pigs that were transported into Europe.

Conclusion

Our results indicate that the Anatolian wild boars domesticated ~10,500 y ago were the ancestors of domestic pigs that were transported into Europe ~8,500 y BP. By the late Neolithic (5,000 y BP), the Near Eastern genomic proportion of domestic pigs in Europe had dropped to <50%, and the Near Eastern fraction is now 0 to 4% in modern European domestic pigs. This near-complete genomic replacement and gradual disappearance of Near Eastern ancestry occurred over 3 millennia in continental Europe and was the result of hybridization between Near Eastern domestic pigs and European wild boars. This further implies that European domestic pigs did not originate from an independent domestication process, but rather from the continuous management of herds that were interbred (however intentionally) with local wild boar. In Mediterranean regions, including Sardinia (42), Corsica (42), Spain (43), Greece (44), and Roman Italy (45), swineherd management often allowed for pigs to seasonally range freely away from human settlements. Combined with other traditions such as pig transhumance (42), these practices likely offered the opportunity for reciprocal gene flow between wild boar and managed pigs, although, at least in some regions, a clear size difference persisted throughout. Our results suggest that these management strategies may have been practiced in Europe from the first introduction of pigs in the Neolithic.

The introgression from European wild boars eroded the proportion of Near Eastern ancestry in European pigs to levels that are potentially below our detection threshold. As predicted by a model in which European pigs were not independently domesticated, we found the existence of a genetic variant leading to black coat color (within the *MC1R* gene) that was transferred from the Near East into Europe by early farmers, where it resisted introgression from wild boar. This finding suggests that other regions of the genome that govern domestic phenotypes (e.g., smaller size) may also have retained their Near Eastern ancestry, but our analyses indicate that these regions make up no more than 4% of the genome. In fact, we show that the vast majority of human-mediated selection over the past 5,000 y focused instead on the genomic fraction derived from the European wild boars, and not on genomic variants that were selected by Near Eastern Neolithic farmers during the first 2,500 y of the domestication process.

Previous coalsecent simulations have shown that a genomic replacement of this magnitude, as a result of introgression from a local population into an invading population is expected, so long as the incoming population is relatively small and strong barriers to interbreeding do not exist (46). The degree to which the Near Eastern fraction of the earliest domestic pigs in Europe has been erased from the genome of modern European pigs is unprecedented. Despite the fact that introgression has also been shown to be common (47, 48) between local wild populations and translocated domestic animals [e.g., cattle (49), horses (50), dogs (51), chickens (52), goats (5)] and plant species [e.g., grapes (53), apples (54), maize (55, 56)], pigs are the only species that has experienced a genomic turnover so substantial that their original ancestry is barely detectable within modern populations. This suggests that pigs experienced a significantly smaller degree of reproductive isolation from their wild European counterparts than did other dispersing domesticates that encountered closely related wild species in the regions into which they were introduced [e.g., cattle (49), dogs (51)].

Overall, our results suggest that domestication narratives are not as straightforward as a simple dispersal of fully domesticated plants and animals out of the area of initial domestication. Instead, domestication is a protracted process, a significant proportion of which takes place through continual admixture and human-mediated selection. These perspectives underscore the temporally dynamic nature of the relationship between humans and domestic taxa, and our increasing ability to monitor this process by analyzing ancient genomic data within the context of metrical, isotopic, and other analyses.

ACKNOWLEDGMENTS. We thank L. Orlando and J. Schraiber for their valuable comments and R. Redding, A. Dinu, A. Devault, A. Mahforozi, A. Hammon, A. McMahon, B. Wilkens, C. Commenge, C. Minniti, D. Kennet, D. Mol, D. Binder, E. Vika, G. Bayle, H.-J. Döhle, I. Dubset, I. Fiore, J. de Grossi Mazzorin, J. Enk, J. Glimmerveen, J. Courtin, J. L. Arsuaga, J. Lev-Tov, K. Aaris-Sørensen, K. Post, M. Özdoğan, M. Özbaşaran, P. Miracle, R. Portela-Miguez, R. Meadow, S. Davis, S. Bréhard, S. Hanik, U. Albarella, T. O'Connor, S. Hansen, M. Toderas, P. Bogucki, D. Lo Vetro, R. Andreescu, P. Mirea, D. Popovici, C. Micu, I. M. Chicideanu, P. Damian, C. Becker, R. Gleser, S. Hansen, R. Hofmann, R. Krauss, M. Lichardus-Itten, N. Müller-Scheessel, B. Weninger, and I. Österholm for providing material and support. We also thank the team working at the site of Aşıklı Höyük: M. Özbaşaran, H. Buitenhuis, N. Munro, N. Pöllath, and M. Stiner. For the Iranian material, we thank Eberhard Sauer, Jebreil Nokandeh (Iranian Cultural Heritage Handicraft and Tourism Organization) and European Research Council (ERC) Persia and Neighbours (ERC-2011-ADG_20110406). We also thank the staff of the Danish National High-Throughput Sequencing Centre for assistance in data generation. We acknowledge the University of Oxford Advanced Research Computing facility for providing computing time. L.A.F.F. was supported by the Wellcome Trust (Grant 210119/Z/18/Z) and by Wolfson College (University of Oxford). L.A.F.F., J.H., A. Linderholm, K.D., and G.L. were supported by the ERC (Grant ERC-2013-StG-337574-UNDEAD), Natural Environmental Research Council (Grants NE/K005243/1 and NE/K003259/1), or both. M.T.P.G. was supported by the ERC (Grant ERC-2015-CoG-681396–Extinction Genomics). D.B., V.E.M., and K.D. were supported by the ERC (Grant 295729-CodeX). A. Bălăşescu and V. Dumitrascu (Romanian archaeozoological research) were supported by the Romanian National Authority for Scientific Research (Project PN-III-P4-ID-PCE-2016-0676) T.M. was supported by the National Geographic Society, Icelandic Centre for Research (RANNÍS), and US National Science Foundation (NSF) (Grants 0732327, 1140106, 1119354, 1203823, 1203268, 1202692, 1249313, 0527732, 0638897, 0629500, 0947862, and 1446308). J.P. was supported by the General Directorate of Antiquities in Turkey and the Deutsche Forschungsgemeinschaft (Grant PE 424/10,1-3). C.V. and J.-M.C. were supported by the Atapuerca Project (CGL2015-65387-C3-2-P MINECO-FEDER, Ministerio de Economía y Competitividad, Spain), the Junta de Castilla y León, and the Fundación Atapuerca. L.d.P. was supported by the ERC (Grant ERC-2013-CoG-614725–PATHPHYLODYN). M.C. was supported by the NSF (Grants 0732327 and 1202692). M.G. was supported by a grant from the Romanian National Authority for Scientific Research (Project PN-II-RU-TE-2012-3-0461). V. Dimitrijevic was supported by the Ministry of Culture and Information and the Ministry of Education and Science of the Republic of Serbia (Project III 47001). B.A. was supported by the NSF (Grants BCS-0530699 and BCS-1311551), the National Geographic Society, and the American Research Institute in Turkey. C.D. was supported by the Fundação para a Ciência e Tecnologia (Grant SFRH/BPD/108326/2015). A.S. and C.G. were supported by the German Archaeological Institute, Berlin (Research Cluster 1, Project: Genetische Studien zur Zucht- und Nutzungsgeschichte der ältesten Wirtschaftshaustiere). L.G.-F. was supported by the British Academy and Leverhulme Trust (Grant SG143331). A.M. was supported by the National Science Centre, Poland (Grant 2017/25/B/HS3/01242).

[a]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom; [b]The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford OX1 3TG, United Kingdom; [c]Department of Zoology, University of Oxford, Oxford OX1 3SZ, United Kingdom; [d]Palaeogenetics Group, Institute of Organismic and Molecular Evolution, Johannes Gutenberg-University Mainz, D-55128 Mainz, Germany; [e]Department of Natural Sciences, German Archaeological Institute, 14195 Berlin, Germany; [f]BioArCh, Department of Archaeology, University of York, York YO10 5NG, United Kingdom; [g]Department of Anthropology, Texas A&M University, College Station, TX 77840; [h]Department of Earth Sciences, Natural History Museum, London SW7 5BD, United Kingdom; [i]Molecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland; [j]Department of Anthropology, University of Michigan, Ann Arbor, MI 48109; [k]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142; [l]Department of Archaeology, Durham University, Durham DH1 3LE, United Kingdom; [m]CNRS UMR 7044, Maison interuniversitaire des sciences de l'Homme, F-67083 Strasbourg Cedex, France; [n]Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; [o]Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [p]"Vasile Pârvan" Institute of Archaeology, Bucharest 010667, Romania; [q]Osteoarchaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 106 91 Stockholm, Sweden; [r]Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, St. Petersburg 199034, Russia; [s]School of History, Classics and Archaeology, University of Edinburgh, Edinburgh EH8 9AG, United Kingdom; [t]The Italian Academy for Advanced Studies in America, Columbia University, New York, NY 10027; [u]Laboratory for Bioarchaeology, Department of Archaeology, Faculty of Philosophy, University of Belgrade, 11000 Belgrade, Serbia; [v]Institute of Archaeology, University of Groningen, 9712 ER, Groningen, The Netherlands; [w]Laboratorio de Evolución Humana, Departamento de Historia, Geografía y Comunicación Universidad de Burgos, Burgos, Spain; [x]Animal Breeding and Genomics Center, Wageningen University and Research, 6708 PB Wageningen, The Netherlands; [y]OD Earth and History of Life, Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium; [z]Centro de Arqueologia da Universidade de Lisboa, Faculdade de Letras da Universidade de Lisboa, Alameda da Universidade, 1600-214 Lisboa, Portugal; [aa]Department of Biological and Geographical Sciences, University of Huddersfield, Huddersfield HD1 3DH, United Kingdom; [bb]Department of Archaeology, Gazi University, Ankara 06500, Turkey; [cc]Department of Archaeology, Çanakkale Onsekiz Mart University, Çanakkale 17100, Turkey; [dd]Flanders Heritage Agency, 1000 Brussels, Belgium; [ee]Department of Biology, University of Naples Federico II, 80126 Napoli, Italy; [ff]History, Archaeology and Museology Department, 1 Decembrie 1918 University, Alba Iulia 510009, Romania; [gg]Department of Archaeology and Classical Studies, Stockholm University, SE-106 91 Stockholm, Sweden; [hh]Université Côte d'Azur, CNRS, Cultures et Environnement, Préhistoire, Antiquité, Moyen Âge (UMR 7264), 06357 Nice, France; [ii]CNRS, Archéorient (UMR 5133), Maison de l'Orient et de la Méditerranée, 69007 Lyon, France; [jj]Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan; [kk]National Natural History Collections, Faculty of Life Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; [ll]Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool L69 7WZ, United Kingdom; [mm]Unité Archéozoologie, Archéobotanique, Sociétés Pratiques et Environnements (AASPE), CNRS, Muséum National d'Histoire Naturelle, 75020 Paris, France; [nn]School of Natural and Built Environment, Queen's University Belfast, Belfast BT9 5AG, United Kingdom; [oo]Scientific Center of Zoology and Hydroecology, Institute of Zoology, Yerevan 0014, Armenia; [pp]Institute of Archaeology, Adam Mickiewicz University, 61-712, Poznań, Poland; [qq]The Hellenic Archaeological Research Foundation, Tivoli House, Cheltenham GL50 2TD, United Kingdom; [rr]Department of Archaeology, University of Reading, Reading RG6 6AB, United Kingdom; [ss]Lithuanian Institute of History, Vilnius University, LT-01513 Vilnius, Lithuania; [tt]Institut für Archäologische Wissenschaften, Goethe University of Frankfurt, 60323 Frankfurt, Germany; [uu]International Union for Conservation of Nature/Species Survival Commission Wild Pig Specialist Group, 15412 Jakarta, Indonesia; [vv]Center of Excellence for Environmental Decisions, University of Queensland, St Lucia, QLD 4072, Australia; [ww]Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, Marlowe Building, University of Kent, Canterbury, Kent CT2 7NR; [xx]Anthropology Department, Hunter College and Graduate Center, City University of New York, New York, NY 10065; [yy]Service de Préhistoire, Université de Liège, 4000 Liège, Belgium; [zz]Institute of Prehistoric Archaeology, Free University of Berlin, 14195 Berlin, Germany; [aaa]Curt-Engelhorn-Zentrum Archäometrie, 68159 Mannheim, Germany; [bbb]Ephorate of Paleoanthropology and Speleology, Greek Ministry of Culture, 106 82 Athens, Greece; [ccc]Institute of Archaeology, University College London, London WC1H 0PY, United Kingdom; [ddd]Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria; [eee]Museum of Vojvodina, 21101 Novi Sad, Serbia; [fff]Laboratoire Évolution & Diversité Biologique-UMR 5174, Université de Toulouse Midi-Pyrénées, 31062 cedex 9 Toulouse, France; [ggg]Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; [hhh]Division of Vertebrates, Department of Life Sciences, The Natural History Museum, London SW7 5BD, United Kingdom; [iii]Dipartimento di Scienze storiche e dei Beni Culturali, University of Siena, 53100 Siena, Italy; [jjj]Osteologie, Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart, 73728 Konstanz, Germany; [kkk]Faculty of Science and Technology, Bournemouth University, Fern Barrow, Poole, Dorset BH12 5BB, United Kingdom; [lll]Magdalene College, University of Cambridge, Cambridge CB3 0AG, United Kingdom; [mmm]Sezione di Bioarcheologia, Museo delle Civiltà, 00144 Roma, Italy; [nnn]Department of Archaeology, Faculty of Philosophy, University of Belgrade, 11000 Belgrade, Serbia; [ooo]Department of Archaeology and History, Faculty of Humanities and Social Sciences, La Trobe University, Melbourne, MB 167, Australia; [ppp]Institute of Archaeology of the National Academy of Sciences of Ukraine, 02000 Kiev, Ukraine; [qqq]Generaldirektion Kulturelles Erbe Rheinland-Pfalz, Dir. Landesarchäologie, D–67346 Speyer, Germany; [rrr]Natural History Museum of Denmark, University of Copenhagen, DK-1123 Copenhagen, Denmark; [sss]University Museum, Norwegian University of Science and Technology, 7012 Trondheim, Norway; [ttt]Integrative Prehistory and Archaeological Science, University of Basel, 4055 Basel, Switzerland; [uuu]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 37012; [vvv]ArchaeoBioCenter and Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication and the History of Veterinary Medicine, Ludwig Maximilian University Munich, 80539 Munich, Germany; [www]State Collection for Anthropology and Palaeoanatomy, Bavarian Natural History Collections, 80333 Munich, Germany; [xxx]Department of Archaeology, School of Geosciences, University of Aberdeen, St. Mary's, Aberdeen AB24 3FUK; [yyy]Institut des Sciences de l'Evolution-Montpellier-UMR 5554-CNRS, IRD, Université de Montpellier, 34090 Montpellier, France; and [zzz]Research Centre in Evolutionary Anthropology and Palaeoecology, School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom

1. M. A. Zeder, "Out of the fertile crescent: The dispersal of domestic livestock through Europe and Africa" in *Human Dispersal and Species Movement: From Prehistory to the Present*, N. Boivin, R. Crassard, M. Petraglia, Eds. (Cambridge University Press, 2017), pp. 261–303.
2. Z. Hofmanová *et al*., Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6886–6891 (2016).
3. J. Conolly *et al*., Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. *J. Archaeol. Sci.* **38**, 538–545 (2011).
4. J. Peters, A. von den Driesch, D. Helmer, "The Upper Euphrates-Tigris Basin: Cradle of agro-pastoralism?" in *The First Steps of Animal Domestication: New Archaeozoological Approaches. Proceedings of the Ninth International Council of Archeozoology*, J.D. Vigne, J. Peters, D. Helmer, Eds. (Oxbow Books, Oxford, 2005), pp. 96–123.
5. K. G. Daly *et al*., Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).
6. S. Colledge, J. Conolly, K. Dobney, K. Manning, S. Shennan, *Origins and Spread of Domestic Animals in Southwest Asia and Europe* (Left Coast Press, 2013).
7. L. Frantz *et al*., The evolution of *Suidae*. *Annu. Rev. Anim. Biosci.* **4**, 61–85 (2016).
8. P. Rowley-Conwy, U. Albarella, K. Dobney, Distinguishing wild boar from domestic pigs in prehistory: A review of approaches and recent results. *J. World Prehist.* **25**, 1–44 (2012).
9. X. Lemoine, M. A. Zeder, K. J. Bishop, S. J. Rufolo, A new system for computing dentition-based age profiles in Sus scrofa. *J. Archaeol. Sci.* **47**, 179–193 (2014).
10. M. A. Zeder, Core questions in domestication research. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3191–3198 (2015).
11. A. Evin *et al*., The long and winding road: Identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743 (2013).
12. T. Cucchi, A. Hulme-Beaman, J. Yuan, K. Dobney, Early Neolithic pig domestication at Jiahu, Henan Province, China: Clues from molar shape analyses using geometric morphometric approaches. *J. Archaeol. Sci.* **38**, 11–22 (2011).
13. M. Balasse, *et al*., "Wild game or farm animal? Tracking human-pig relationships in ancient times through stable isotope analysis" in *Hybrid Communities Biosocial Approaches to Domestication and Other Trans-species Relationships*, C. Stépanoff, J.-D. Vigne, Eds. (Routledge Studies in Anthropology, Routledge, Abingdon, Oxon, UK, ed. 1, 2018), pp. 81–96.
14. A. Dinu, A. Boroneant, A. Balasescu, A. Soficaru, D. Miritoiu, Mesolithic and Neolithic pigs of the northern Balkans: Astragali vs. teeth as markers of domestication. *Mesolithic Miscellany* **19**, 7–12 (2008).
15. A. Legge, Practice with science: Molar tooth eruption ages in domestic, feral and wild pigs (Sus scrofa). *Int. J. Osteoarchaeol.* (2013) https://onlinelibrary.wiley.com/page/journal/10991212/homepage/News.html. Accessed 25 July 2019.
16. A. J. Legge, "Bone measurements and body weights from some Australian feral pigs" in *Economic Zooarchaeology: Studies in Hunting, Herding and Early Agriculture*, P. Rowley-Conwy, D. Sergeantson, P. Halstead, Eds. (Oxbow Books, Oxford, ed. 1, 2017).
17. S. Payne, G. Bull, Components of variation in measurements of pig bones and teeth, and the use of measurements to distinguish wild from domestic pig remains. *Archaeozoologia* **2**, 27–65 (1988).
18. C. Lega, P. Raia, L. Rook, D. Fulgione, Size matters: A comparative analysis of pig domestication. *Holocene* **26**, 327–332 (2016).
19. A. Evin *et al*., Unravelling the complexity of domestication: A case study using morphometrics and ancient DNA analyses of archaeological pigs from Romania. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130616 (2015).

20. A. Evin *et al*., Phenotype and animal domestication: A study of dental variation between domestic, wild, captive, hybrid and insular Sus scrofa. *BMC Evol. Biol.* **15**, 6 (2015).
21. G. Larson *et al*., Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15276–15281 (2007).
22. A. Manunza *et al*., A high throughput genotyping approach reveals distinctive autosomal genetic signatures for European and Near Eastern wild boar. *PLoS One* **8**, e55891 (2013).
23. K. J. Gron, L. Sørensen, Cultural and economic negotiation: A new perspective on the Neolithic transition of Southern Scandinavia. *Antiquity* **92**, 958–974 (2018).
24. P. Sillitoe, "Pigs in the New Guinea Highlands: An ethnographic example" in *Pigs and Humans: 10,000 Years of Interaction*, U. Albarella, K. Dobney, A. Ervynck, Eds. (Oxford University Press, Oxford, UK, 2007), vol. 10, pp. 330–356.
25. J. Studer, D. Pillonel, "Traditional pig butchery by the Yali people of West Papua (Irian Jaya): An ethnographic and archaeozoological example" in *Pigs and Humans: 10,000 Years of Interaction*, U. Albarella, K. Dobney, A. Ervynck, Eds. (Oxford University Press, Oxford, UK, 2007), pp. 308–329.
26. L. A. F. Frantz *et al*., Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148 (2015).
27. S. Vai *et al*., The Biarzo case in northern Italy: Is the temporal dynamic of swine mitochondrial DNA lineages in Europe related to domestication? *Sci. Rep.* **5**, 16514 (2015).
28. C. Ottoni *et al*., Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol. Biol. Evol.* **30**, 824–832 (2013).
29. C. Lega *et al*., Like a pig out of water: Seaborne spread of domestic pigs in Southern Italy and Sardinia during the Bronze and iron ages. *Heredity* **118**, 154–159 (2017).
30. B. Chessa *et al*., Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**, 532–536 (2009).
31. W. Haak *et al*., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
32. G. Larson *et al*., Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618–1621 (2005).
33. B. Krause-Kyora *et al*., Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nat. Commun.* **4**, 2348 (2013).
34. G. Hellenthal *et al*., A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
35. S. White, From globalized pig breeds to capitalist pigs: A study in animal cultures and evolutionary history. *Environ. Hist. Durh. N. C.* **16**, 94–120 (2011).
36. M. Bosse *et al*., Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392 (2014).
37. V. Maselli *et al*., Southern Italian wild boar population, hotspot of genetic diversity. *Hystrix It. J. Mammal*, 10.4404/hystrix-27.2-1148927 (2016).
38. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
39. O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
40. M. Fang, G. Larson, H. S. Ribeiro, N. Li, L. Andersson, Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* **5**, e1000341 (2009).
41. D. J. Goedbloed *et al*., Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol. Ecol.* **22**, 856–866 (2013).
42. U. Albarella, F. Manconi, A. Trentacoste, "A week on the plateau: Pig husbandry, mobility and resource exploitation in central Sardinia" in *Ethnozooarchaeology. The Present and Past of Human-Animal Relationships*, U. Albarella, A. Trentacoste, Eds. (Oxbow Books, Oxford, 2011), pp. 143–159.
43. A. Hadjikoumis, "The origins and evolution of pig domestication in prehistoric Spain" PhD thesis, University of Sheffield, Sheffield, UK (2010). http://etheses.whiterose.ac.uk/23773/. Accessed 16 May 2019.
44. P. Halstead, V. Isaakidou, "A Pig Fed by Hand is Worth Two in the Bush" in *Ethnozooarchaeology: The Present and Past of Human-Animal Relationships*, U. Albarella, A. Trentacoste, Eds. (Oxbow Books, Oxford, 2011), pp. 160–174.
45. M. MacKinnon, High on the hog: Linking zooarchaeological, literary, and artistic data for pig breeds in Roman Italy. *Am. J. Archaeol.* **105**, 649–673 (2001).
46. M. Currat, M. Ruedi, R. J. Petit, L. Excoffier, The hidden side of invasions: Massive introgression by local genes. *Evolution* **62**, 1908–1920 (2008).
47. G. Larson, J. Burger, A population genetics view of animal domestication. *Trends Genet.* **29**, 197–205 (2013).
48. L. A. F. Frantz, G. Larson, "A genetic perspective on the domestication continuum" in *Hybrid Communities*, C. Stépanoff, J.-D. Vigne, Eds. (Routledge Studies in Anthropology, Routledge, Abingdon, Oxon, UK, ed. 1, 2018), vol. 46, pp 23–37.
49. S. D. E. Park *et al*., Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
50. V. Warmuth *et al*., Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8202–8206 (2012).
51. M. Ní Leathlobhair *et al*., The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).
52. J. Eriksson *et al*., Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* **4**, e1000010 (2008).
53. S. Myles *et al*., Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3530–3535 (2011).
54. A. Cornille *et al*., New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* **8**, e1002703 (2012).
55. L. Kistler *et al*., Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**, 1309–1313 (2018).
56. R. R. da Fonseca *et al*., The origin and evolution of maize in the Southwestern United States. *Nat. Plants* **1**, 14003 (2015).

Supplementary Information for

Ancient pig genomes reveal a near complete turnover following their introduction to Europe

Laurent A.F. Frantz, James Haile, Audrey T. Lin, Amelie Scheu, Christina Geörg, Norbert Benecke, Michelle Alexander,Anna Linderholm Victoria E. Mullin Kevin G. Daly, Vincent M. Battista, Max Price, Rose-Marie Arbogast, Benjamin Arbuckle, Adrian Bălăşescu, Ross Barnett, László Bartosiewicz, Gennady Baryshnikov, Clive Bonsall, Dušan Borić, Adina Boroneanț, Jelena Bulatović, Canan Çakirlar, José-Miguel Carretero, John Chapman, Mike Church, Richard Crooijmans, Bea De Cupere, Cleia Detry, Vesna Dimitrijevic, Valentin Dumitrascu, Ceiridwen Edwards, Merih Erek, Aslı Erim-Özdoğan, Anton Ervynck, Domenico Fulgione, Mihai Gligor, Anders Götherström, Lionel Gourichon, Martien Groenen, Daniel Helmer, Hitomi Hongo, Liora K. Horwitz, Evan K. Irving-Pease, Ophélie Lebrasseur Joséphine Lesur, Caroline Malone, Ninna Manaseryan, Arkadiusz Marciniak, Holley Martlew, Marjan Mashkour, Roger Matthews, Giedre Matuzeviciute, Sepideh Maziar, Erik Meijaard, Tom McGovern, Hendrik-Jan Megens, Rebecca Miller, Azadeh Mohaseb, Jörg Orschiedt, D. Orton, Anastasia Papathanasiou, Mike Parker Pearson, Ron Pinhasi, Darko Radmanović, François-Xavier Ricaut, Mike Richards, Richard Sabin, Lucia Sarti, Wolfram Schier, Shiva Sheikhi, Elisabeth Stephan, John R. Stewart, Simon Stoddart, Antonio Tagliacozzo, Nenad Tasić, Katerina Trantalidou, Anne Tresset, Cristina Valdiosera, Youri van den Hurk, Sophie Van Poucke, Jean-Denis Vigne, Alexander Yanevich, Andrea Zeeb-Lanz, M. Thomas P. Gilbert Jörg Schibler, Melinda Zeder, Joris Peters Thomas Cucchi, Daniel G. Bradley, Keith Dobney, Joachim Burger, Allowen Evin, Linus Girdland-Flink, and Greger Larson

Corresponding authors: Laurent A. F. Frantz – laurent.frantz@qmul.ac.uk ; Greger Larson – greger.larson@arch.ox.ac.uk

This PDF file includes:

Materials and Methods
Figs. S1 to S22

Other Supporting Online Material for this manuscript includes the following:

Tables S1, and S2 (Excel)

Archaeological Site Descriptions **3**

Ancient DNA **53**
Data generation - mtDNA - Durham *54*
Data generation - MC1R/mtDNA *55*
Data generation - Nuclear and mtDNA capture - Oxford *57*
Data generation - Nuclear - Stockholm *57*

Data Processing **58**
Sexing *58*
MapDamage *58*
Publicly available sequences *59*
High coverage ancient genomes *59*
Ascertainment and ancient genome quality *59*
Low coverage data 60

mtDNA analyses **60**
Filtering 60
Building a database for haplogroup assignment 60
Phylogenetic analysis and reference sequences 60
A method for rapid haplogroup assignment 61
Low coverage data 62
mtDNA lineage geographic distribution (Figure 1) 62
Samples excluded from Figure 1 62

Nuclear DNA analyses **64**
MC1R and domestic/wild status 64
Ancestry analysis 65
Neighbour Joining tree 65
Modelling population history of wild boars *66*
Low-coverage data - testing PCA *66*
PCA *67*
ADMIXTURE *67*
F3 & D-statistics *68*
Phasing, Chromopainter and GLOBETROTTER *69*
ADMIXTURE simulations *70*
GLOBETROTTER simulations *70*
Sweep ancestry *71*
X chromosome *71*
Supplementary Tables **72**
Supplementary Figures **73**

Archaeological Site Descriptions

Abri Fuchskirche I near Allendorf (Thuringia), Germany
Contact: Norbert Benecke
Abri Fuchskirche I is a rockshelter site near Allendorf in Thuringia, Germany. Occupation layers are of the late glacial period (Bølling – Allerød); there are four 14C dates on bones of elk, wild horse, and aurochs (1). One wild boar specimen (GL115) was analyzed.

Abu Ghosh, Israel
Contact: Liora Kolska Horwitz
Abu Ghosh is situated in the Judean Hills, ca. 12 km west of the city of Jerusalem. Several excavations at the site since the 1950s have documented remains of a mid-Pre-Pottery Neolithic B (PPNB; 10,100-9,250 cal BP) village, containing rectangular houses with plaster floors, installations, intra-mural human burials and a rich assemblage of lithic artifacts, groundstone vessels and fauna (2); (3) and papers therein). The single sample (AA042) analyzed in this study from this site is wild boar.

Acemhöyük, Turkey
Contact: Benjamin Arbuckle
Acemhöyük is a large mound located at an elevation of 950 m on the alluvial fan of the Melendiz river near the Tüz Gölü (Salt Lake) in Central Turkey. The mound is approximately 800 x 600 m in dimension rising 20 m above the surrounding Yesilova plain and represents the remains of a large, fortified Bronze Age urban center. The site has been excavated since 1962 by Dr. N. Özgüç, Ankara University in association with the Turkish Historical Society and the General Directorate of Antiquities and Museums and by Prof. Dr. A. Öztan of the same institution since 1989 (4); (5). Acemhöyük consists of 12 levels with deposits representing Chalcolithic to Medieval occupations.

Four *Sus* samples (all representing small sized and therefore probably domestic pigs) were submitted for ancient DNA analysis. Two of these samples (AA525 and AA526) derive from deposits stratigraphically assigned to level XI (associated radiocarbon date c. 4,000 cal BP), representing the Early Bronze Age; the remaining two specimens (AA324 and AA325) derive from level III representing the Assyrian Colony period of the Middle Bronze Age (3,700s cal BP).

The animal economy at MBA Acemhöyük is dominated by domestic sheep, goats and cattle. Domestic pigs are present in MBA levels II and III where they represent c. 10-20% of the identified remains depending on context (6). The Acemhöyük faunal assemblage is curated on site at the Acemhöyük dighouse, Yesilova, Aksaray, Turkey. Specimens were exported with the permission of the Konya Museum and the Turkish Ministry of Culture and Tourism.

Ajvide, Sweden
Contact: Göran Burenhult
The Stone-Age settlement and burial ground at Ajvide in Eksta Parish is situated on the south-west coast of the island of Gotland in the Baltic Sea. During the period of use, c. 5,000-4,300 cal BP, the site was located at a lagoon by the sea. Today, the site is situated on a slope between 12 and 17 m above the sea level. It belongs to the Pitted-Ware Culture in the late Middle-Neolithic period of the Swedish Stone Age.

The site covers an area of c. 200,000 m^2, and has been partially excavated 1983-2009. The excavated area is c. 3,000 m^2, and so far 85 skeleton graves have been documented. Thanks to the limestone-rich soil, the skeletons, as well as the animal bones, are exceptionally well preserved and both osteological and DNA analyses have been successfully performed. The settlement predates the burial ground, and was occupied between c. 5,000-4,750 cal BP. The earliest graves can be dated to c. 4,750 cal BP, and the burial ground was used until c. 4,300 cal BP (7); (8); (9). The samples from the site include AA289, AA290, AA291, AA292, and AA293.

Alba Iulia-Lumea Nouă, Romania
Contact: Mihai Gligor, John Chapman
The Neolithic and Eneolithic settlement at Alba Iulia-Lumea Nouă was discovered by chance in the early 1940s, when the area was subject to public utility works. The site is located on the second terrace of the Mureş river in the north-eastern part of the city of Alba Iulia, Transylvania (Romania), at 240-246 m altitude, and covers some 40 ha. In the past 15 years, continuing residential development of the city has resulted in several rescue excavations, which have provided evidence of habitation by Vinča and Lumea Nouă cultures (Middle Neolithic), Foeni and Petreşti cultures (Early and Middle Eneolithic). The site is considered one of the most important settlements in Transylvania on the basis of the quality of painted pottery and funerary complexes with commingled human remains. The six animal bone samples analyzed in this study (AA384, AA385, AA386, AA387, AA388, and AA389) belongs to the Early Eneolithic Foeni period (5,770-5,550 BP) (10); (11); (12); (13); (14).

Albertfalva Roman vicus (Budapest), Hungary
Contact: Keith Dobney, László Bartosiewicz
The Albertfalva Roman vicus was located in Pannonia province on the right bank of the Danube in the southern outskirts of present-day Budapest, Hungary (15). This section of the river once formed the eastern limes of the Roman Empire. Approximately one quarter (25.9%) of the 1,888 identifiable bones (NISP) previously published by Bökönyi (16) originated from domestic pig. The material was dominated by the remains of domesticates, including a relatively high proportion of horse, explained by the presence of Roman cavalry units nearby. The two samples from this site (AA408 and AA409) date to the 2/3rd century AD.

Alepotrypa Cave, Greece
Contact: Anastasia Papathanasiou, Mike Richards
Alepotrypa Cave is located at Diros Bay, Lakonia, Greece. The cave is about 300 m long and it is situated 50 m from the present Mediterranean shoreline, in an arid and rocky limestone environment. It has been excavated since 1970 and it is dated from 8,000 to 5,200 BP, from the Early to the Final Neolithic. Artefacts include a variety of pottery, lithic tools, grindstones, personal decoration items, figurines and food remains consisting of cultivated cereal, legume, and fruit remains, a large number of animal bones from domesticated species, and to a lesser degree of wild plant and animal resources, fish and shells. Stable isotope analysis suggests a primarily agricultural diet with emphasis on plant resources. The cave also yielded a large human skeletal assemblage of primary and secondary disposals and evidence of rich ritualistic expression, including massive concentrations of deliberately broken pots

possibly associated with mortuary practices (17); (18). One specimen (GL1081) analyzed dates to the Middle Neolithic.

Altenburg-Rheinau, Germany
Contact: Joris Peters
Altenburg-Rheinau is a La Tène culture site founded in the middle of the 4th century BP its abandonment likely coincided with the Roman conquest of the northern alpine forelands 2,015 BP. Site occupation therefore essentially coincided with the La Tène D cultural period. Excavations in the 1970s produced a large faunal assemblage comprising more than 10,000 remains of pigs, the second most important taxon in terms of numbers after cattle (19). Ten specimens from this site (LG161, LG162, LG577, LG578, LG579, LG580, LG582, LG583, LG584, and LG585) date to the Iron Age.

Arbon Bleiche 3, Switzerland
Contact: Jörg Schibler
Arbon Bleiche 3 is a Neolithic lakeshore site in Switzerland situated at the southern shore of Lake Constance. The site is known since more than 130 years but was excavated for the first time in 1983 (46 m^2) but to a larger extent (1,050 m^2) between 1993 and 1995 (20). Drilling exploration revealed that at least one third and at most half of the Neolithic village was excavated. As typical for wetland sites Arbon Bleiche 3 is dated by dendrochronology. The settlement period spans in total 15 years, most probably between 5,384 BP and 5,370 BP. During the excavation (1993 and 1995) 71,472 animal bones were excavated, and 32,261 could be identified to the species level (21). 23 specimens (AA181, AA182, AA183, AA184, AA185, AA186, AA187, AA189, AA191, AA192, AA193, AA195, AA255, AA256, AA257, AA258, AA259, AA260, AA261, AA262, AA263, AA264, AA265) from this site date from the Late Neolithic to the Bronze Age. The bones were provided by Amt für Archäologie Thurgau for this study.

Aşağı Pınar, Turkey
Contact: Norbert Benecke
Aşağı Pınar (lower well) is the name of Middle to Late Neolithic *tell* (8,200-6,300 cal BP) in the eastern part of Turkish Thrace close to Kırklareli. Its location alone already makes it predestined to mirror the relationships between the two areas, and the archaeological finds indeed resemble both, Anatolian and Balkan Cultures (22). The domestic faunal assemblage reveals a quite stable relative composition throughout the settlement history with a slight dominance of sheep and goat, followed by cattle and only few pigs (up to 7%) (23). The 15 samples analyzed here represent both the Middle and Late Neolithic horizons and date to 7,500-7,000 cal BP (AP23, AP24, AP25, AP26, AP27, AP28, AP29, AP30, AP31, AP32, AP33, AP34, AP35, AP36, AP37). One domestic sample (AP26) turned out to come from an only around 2,500 year old Hellenistic intrusion (KIA-42161: 2415 ± 30 BP). One wild boar and eight domestic pigs were successfully analyzed for a partial mtDNA fragment.

Aşıklı Höyük, Turkey
Contact: Mihriban Özbaşaran, Joris Peters
Aceramic Neolithic Aşıklı Höyük is a 16 m high settlement mound located next to the Melendiz River in east Central Anatolia, the volcanic region named historically as Cappadocia. Site habitation started prior to 10,300 BP and came to an end after 9,500

BP. Stratigraphic analysis allows defining five occupation levels with multiple building phases, distinguished by their architecture and material culture. This archaeological sequence is pivotal to our understanding of the process of Neolithization in Central Anatolia. Of particular interest from a faunal viewpoint is the discovery of extensive dung layers at the site, implying that people held animals captive in the village from the earliest habitation onwards. Zooarchaeological analyses point to intensifying management of sheep and goat with time, alongside fishing, fowling, and hunting: Besides aurochs and red deer preferred game also included wild boar (24); (25); (26). 14 samples (AL643, AL644, AL645, AL646, AL647, AL706, AL708, AL712, AL714, AL718, AL720, AL724, AL725, AL726) date to the Aceramic Neolithic.

Atij, Syria
Contact: Melinda Zeder
Atij dates to c. 4,700-4,400 BP, and is located on the middle Khabur River in northeastern Syria (27). The number of large silos and other storage facilities encountered here suggests that that may have served as entrepots involved in the coordination and transshipment of agro-pastoral products (28); (29). One suid sample (AA051) dates to the Ninevite V period.

Augusta Raurica, Switzerland
Contact: Jörg Schibler
Augusta Raurica is a roman town about 12 km east of Basel, Switzerland. The town was founded in the early years of the 1st century AD and was settled as a civil town until the 4th century AD. Animal bones are systematically identified during the 1950s and 1960s by Elisabeth Schmid. This archaeozoological data are published in Schibler and Furger (30), but other archaeozoological data have been published elsewhere (e.g. (31); (32)). Two samples (AA268 and AA269) dates to the Roman period. The bones were provided by Museum Augusta Raurica for this study.

Bademağacı Höyük, Turkey
Contact: Bea De Cupere
Bademağacı Höyük is situated in the south of the Lake District (SW Turkey), about 50 km north of Antalya. This oval shaped mound lies at an altitude of 780 m a.s.l. on a small plain surrounded on all sides by mountains. The Early Bronze Age (EBA II) is evidenced by three architectural levels and a stone paved terrace and covers an Early Neolithic occupation (9th millennium BP) (33). 5 samples (AA305, AA306, AA307, AA309, AA310) date to the Neolithic.

Balanbash (Krasny Yar), Russia
Contact: Gennady Baryshnikov
Balanbash settlement is situated at Belaja River (Volga River basin) near Krasny Yar village, close to Sterlitamak city, in Russia. It is dated by the middle part of Bronze Age (34). The bone material of *Sus scrofa domesticus* (ZIN 18451) came from K. Salnikov excavation in 1934. It comprises 43 bone fragments from 7 adult specimens (calculated by G. Baryshnikov). One sample (LG236) was analyzed that dates to the Iron Age.

Basel-Gasfabrik, Switzerland
Contact: Jörg Schibler

Basel-Gasfabrik is an urban centre site with two cemeteries of the late Iron Age period. The site is situated at the southern Rhine shore in Basel. Most of the archaeological material are dated between 2,200-2,150-2,080 BP and belong to the La Tène periods C2 and D1. Several excavations between 1911 and today took place. The urban site was unfortified and covered an area of about 15 ha that offered space for several 100 people (35). Two domestic samples (AA266 and AA267) analyzed date to the late Iron Age. The bones were provided by Archäologische Bodenforschung Basel-Stadt for this study.

Bercy, France
Contact: Anne Tresset
Bercy (Paris, France) is a site located on the right bank of the Seine River, downstream from the confluence with the Marne River. The site corresponds to a hamlet set on the riverbank. The main period of occupation is dated to ca. 6,000 cal BP and belongs to the Chasséen septentrional culture (regional Middle Neolithic). The faunal assemblage is dominated by domesticates. The suid bones included in the present study came from this cultural horizon and accounted for ca. 10-20% of the identified mammal remains depending on the layer considered (36). 14 samples (AA492, AA493, AA494, AA495, AA496, AA498, AA499, AA500, AA501, AA502, AA503, AA504, AA505, AA506) date to the French middle Neolithic.

Birsmatten-Basisgrotte, Switzerland
Contact: Jörg Schibler
Birsmatten-Basisgrotte is a Mesolithic rock shelter in the north-west of Switzerland about 18 km south of Basel in the Village Nenzlingen. The site was excavated by Carl Lüdin between 1940 and 1945 and 1955-56 under the direction of Hans-Georg Bandi. The excavated material was divided in five horizons (I to V). All five horizons were typologically dated to the Mesolithic. 14C samples were analyzed in the early 1960s and published by Gfeller (37). The calibrated (Oxcal) dates show for the horizons V to III Mesolithic values of the 9th and 10th millennium cal BP, whereas for the horizon II Mesolithic and Neolithic and for horizon I only Neolithic dates are found. Elisabeth Schmid (38) has identified the animal bones and mentioned only for the horizon I the possibility that domestic pig could be present due to the smaller sizes of the bones. Eleven samples (AA241, AA242, AA243, AA245, AA263, AA264, AA265, AA267, AA268, AA269, AA270) date from the Mesolithic to the Neolithic.The bones were provided by Archäologie Baselland and Naturhistorisches Museum Bern for this study.

Blätterhöhle Hagen, Germany
Contact: Jörg Orschiedt
The Blätterhöhle site is located in Hagen in North Rhine-Westphalia, Germany. Archaeological excavations in the cave and rock shelter site started in 2006 and research is ongoing. The rock shelter contains a stratigraphy from the late Mesolithic to the late Palaeolithic. Finds include stone artefacts, faunal and human remains as well as several hearths with activity zones. In the cave, the stratigraphy shows two main layers. The upper layer contains human remains from the late Neolithic (according to local chronology), and the lower layer early Mesolithic human remains. Faunal remains are present in all layers. Radiocarbon dates are between 11,200 and 10,600 cal BP for the early Mesolithic, and 5,900 and 5,000 cal BP for the Late Neolithic. Neolithic collective burials in caves and rock shelters are known for the 6th

millennium BP in other areas (Belgium, British Isles, Ireland) but not for the Westphalian mid-range mountain area. The Blätterhöhle provides the first clue that this burial practice. Although from bioturbated sediments and without anatomical context the very well preserved human remains, provide a rich source of information.

Ancient DNA sampling showed results both for Mesolithic and Neolithic remains. The results suggest that the Mesolithic population of the Blätterhöhle represents a typical hunter-gatherer population (mitochondrial haplogroups U5/U4), whereas the Neolithic population seems to be an admixture of hunter-gatherer (haplogroup U5) and farmer lineages (haplogroup H and others). Additionally the analysis of stable Isotopes (13C/15N/34S) was carried out in order to reconstruct long-term diets. The results show three distinct clusters with significant differences. Terrestrial diet was evident for both the Mesolithic and a Neolithic group, but the diet of one Neolithic group was based on the consumption of freshwater fish. This group consists of people exclusively with haplogroup U5. This result demonstrates a "non-Neolithic" way of life in the 6th millennium BP (39); (40); (41). Three specimens (Hag1, Hag2, Hag3) date to the the Mesolithic.

Bodzia 1, Poland
Contact: Arkadiusz Marciniak
The site in Kujavia, in the Polish lowlands was discovered in 2000 and excavated in the 2000s as a part of large rescue excavation project preceding the construction the A1 motorway. An area of nearly 2 ha was exposed, and 7 cultural levels spanning from the Neolithic to the early modern period were distinguished. The most distinct were LBK and Late Lengyel settlements represented by longhouses and more than 200 different character features. A large cemetery with elaborated chamber-like graves and rich grave goods dated back to the 10th-11th centuries AD have no parallels in Europe. Out of Four samples were dated to LBK (AA154, AA336, AA337, AA339), four to Neolithic (AA330, AA334, AA335, AA338), one to Bronze Age (AA331) and ten to Early Medieval period (AA149, AA155, AA332, AA333, AA340, AA341, AA342, AA343, AA344, AA345) (42); (43). LG750 dates to the late 9th millennium BP.

Boulancourt, France
Contact: Adrian Bălăşescu, Jean Denis Vigne
The site of "Le Chatelet" (Seine-et-Marne, France), located on the western edge of the Gâtinais Beauceron plateau, is a fortified hill settlement (44) occupied at the end of the Late Bronze Age IIIb. It has produced more than 33,000 faunal remains of which almost 8,000 have been determined. The choice of particular products of stock-breeding for quality meat consumption and the diversity of hunting, fishing and trapping suggest a privileged social milieu, and fits the aristocratic nature of the site. The analysis of the fauna reveals a strong predominance of the remains of domestic species, within which the pig represents more than 60% in the number of remains (45). Five pig mandibles (AA296, AA297, AA298, AA299, AA300) were analyzed in this study.

Bozejewice 28, Poland
Contact: Arkadiusz Marciniak
The site was discovered in the 1990s during the rescue excavation project preceding the construction of the Yamal pipeline from Siberia to western Europe. Remains of

three cultures were unearthed: (a) Late Lengyel trapezoidal longhouse, (b) Bell Beaker burial, and (c) Lusatian culture settlement (46). Four samples (LG761, LG762, LG763, LG764) date to the Neolithic.

Bruchsal-Aue, Germany
Contact: Elisabeth Stephan
This Michelsberg enclosure located northeast of Karlsruhe in southwest Germany was excavated between 1987 and 1993. The site consisted of three ditches with numerous gateways, partially encircling a crest of a hill which was nearly devoid of finds, probably because of erosional processes. The ditch fillings were abundant with ceramics and animal bones and a segment of the outer ditch contained several aurochs skulls, deposited vessels and burials at the bottom and alongside the ditch, indicating potentially an additional function of parts of this site as sanctuary. Based on typological and chronological analysis of the pottery and radiocarbon measurements of bones the earthwork is dated from c. 6,250-5,650 BP. Nine specimens of domestic pigs were analyzed (LG602, LG603, LG604, LG606, LG608, LG609, LG611, LG618, LG619).

Bundsø, Denmark
Contact: Kim Aaris-Sørensen, Peter Rowley-Conwy
Bundsø is a Neolithic settlement situated on the Island of Als. Bundsø is a causewayed camp defined as a site demarcated by one or more parallel rows of long or short ditches interrupted by causeways and flanked on the inner side by a palisade. A feature shared by most causewayed camps is their position on projecting landscape formations. Bundsø is found on a flat, low-lying ground on the Flintholm peninsula, protruding into a now-drained fjord. A rich deposit of settlement material has been excavated on the bank of the peninsula, mainly dating from MN AIII, with a smaller part from MN AV. The excavation in 1980–81 located a double row of ditches running across the neck of the peninsula, covered by layers with settlement material from the MN AIII. Human skulls have been found among the settlement remains on various occasions, for example found outside the ditches, embedded in waste from later settlements (47).

Pigs excavated from Bundsø come from the early part of the Middle Neolithic, specifically Funnel Beaker Culture (5,300-4,500 BP), and forms the type assemblage of the MN A III period. The site comprised a 200 x 100 m area of cultural debris; no house plans discerned although burnt clay daub suggests they were present. Many finds of cultivated cereals and domestic animal bones attest to a predominantly agricultural economy. The earliest domestic pig samples found from southern Scandinavia date to the Middle Neolithic (48). One domestic pig specimen (AA015) was analyzed.

Bunnik Odijk, Utrecht, Netherlands
Contact: Youri van der Hurk
In 2000, near Bunnik Odijk, Utrecht the remains of a settlement were discovered. Several postholes and ditches were found, including those of a what appears to have been a Middle Iron Age granary. Pottery remains suggest that the site was occupied during the Middle Iron Age (2,500-2,250 BP) and during the Late Roman/Early Medieval period (4-7th century AD). The sample taken from this site (KD025) dates to the Middle Iron Age (49).

Buran-Kaya IV, Ukraine
Contact: Alexandr Yanevich
Buran-Kaya IV is located in the South Ukraine in the mountainous part of the Crimean peninsula at a distance 30 km northeast from Simferopol. The rock shelter site is situated 8 m up the bank of a small stream at 326 m in altitude above sea level. Structures of the habitation of the Neolithic layers of the site are represented only by ash and charcoal lenses from season fireplaces. Late Neolithic layers have AMS-dates 7,800-7,300 BP. The status of the pigs from Buran-Kaya IV is unclear. Archaeological context of the Crimean Neolithic allows to assume, that there were wild pigs, but the presence of the domesticated individuals cannot be excluded. 29 samples (AA372, AA373, AA375, AA376, AA377, AA378, AA380, AA381, AA382, AA476, AA477, AA478, AA479, AA480, AA481, AA482, AA483, AA484, AA485, AA486, AA487, AA488, AA489, AA490) were analyzed.

The Y2 haplogroup pigs have Balkanian origin: they came to Crimea from Balkans in the Final Pleistocene. Status of the pigs from Buran-Kaya IV and Crimean Neolithic in general is unclear. The lack of the sedentism of the Neolithic population and evidence of the developed hunters-gatherers economy allows to assume, that there were wild pigs.

Buxheim, Germany
Contact: Joris Peters
Buxheim, a Late Neolithic settlement of the Münchshöfen culture, was inhabited between ca. 6,500-6,150 BP. Excavations between 1996 and 1998 revealed a diverse faunal assemblage dominated by domestic animals but with a considerable contribution of game species (c. 30%), including red deer, wild boar, roe deer, brown bear and beaver. Pigs are the second most important domestic species. Faunal analysis has been carried out by C. Kitagawa and J. Peters, but the results are not published yet. Two domestic samples (LG177 and LG558) date to the Late Neolithic.

Cabeço da Amoreira, Portugal
Contact: Cleia Detry
Cabeço da Amoreira is part of the same complex of the Muge shell middens, and is located in the left bank of the Muge river, in front of C. Arruda. It was a smaller site only excavated since the 1930s, and continues to be excavated today. Occupied by the last hunter-gatherers, the only domestic species is the dog represented by scarce fragments of bones (50); the suids (AA350, AA351, AA354) have been all classified has wild boar (51), and date to the Mesolithic.

Cabeço da Arruda, Portugal
Contact: Cleia Detry
Cabeço da Arruda is an archaeological site part of a complex of Mesolithic shell middens located in the Tagus estuary (Portugal), that were occupied from 8,200 cal BP to 7,600 cal BP. The sites were produced by the last hunter-gatherers in the Iberian Peninsula which had no domestic species except for dogs (51). Therefore the suids from this site (AA352 and AA353) can clearly be classified as wild boar .

Çadır Höyük, Turkey
Contact: Benjamin Arbuckle

Çadır Höyük is a 32 m high mound located in the Kanak Su basin situated in the Yozgat Province of north central Turkey. The modestly sized 240 x 185 m mound boasts an occupational history spanning 6,000 years (ca. 7,200 BP to 13th century AD). This sequence has been explored since 1993 by the Çadır Höyük Archaeological Project led by Ron Gorny and more recently Greg MacMahon and Sharon Steadman (52); (53).

Four *Sus* specimens from Çadır Höyük were submitted for ancient DNA analysis. Three of these specimens (AL1105, AL1106, AL1108) derive from Late Chalcolithic deposits dated to the late 6th millennium BP. One specimen (AL1107) was recovered from a locus assigned to the Iron Age. Pigs represent 16-30% of the faunal assemblages from multiple phases of the Late Chalcolithic and 5% of the mammalian remains from the Iron Age at Çadır. The vast majority of pig remains are small-bodied, domestic animals (54). The Çadır Höyük faunal assemblage is curated on site at the Çadır Höyük dighouse, Peyniryemez, Sorgun, Turkey. Specimens were exported with the permission of the Yozgat Museum and the Turkish Ministry of Culture and Tourism.

Çamlıbel Tarlası, Turkey
Contact: László Bartosiewicz
The Late Chalcolithic (5,590–5,470 cal BP) settlement of Çamlıbel Tarlası was located on the Anatolian plateau near Boğazkale, 200 km east of Ankara. The small rural settlement, built on rolling hills consisted of stone buildings. Evidence of early copper smelting was identified at the site (55). Pig remains made up one third (33.6%) of the bone material consisting of 864 identifiable specimens originating overwhelmingly from domesticates (56). Eight specimens were analyzed: BLT011, BLT012, BLT015, BLT016, BLT017, BLT018, BLT019, and BLT020.

Cârlomanesti - La Arman, Romania
Contact: Adrian Bălăşescu
The site at Cârlomanesti - *La Arman* (settlement and necropolis) is located on a small plateau on the valleys of Buzău River and its tributary, the Nișcov River, on the southwestern part of the village of Cârlomăneşti (Buzău county). Both the settlement and the necropolis were dated to the Middle Bronze Age (Monteoru culture), with features attributed to the settlement at times overlapping directly the burials. The remains of an Early Medieval settlement (6th-7th centuries AD) overlap and cut into the Middle Bronze Age occupations (57). Middle Bronze Age faunal remains consist predominantly of domesticate bones, with only very few wild species represented. The present sample (AA078) from a pig mandible originates from the Monteoru feature labelled C4 (58).

Casa Solima (Troina), Italy
Contact: Simon Stoddart
Casa Solima was a hut preserved under colluvium on a south facing terrace at 636 m a.s.l., sheltered to the north by a flysch outcrop, in the southern approaches to the Nebrodi mountains in northeastern Sicily, dating to the Serraferlicchio phase of the Chalcolithic (c. 5,500 BP). As such, in a tradition of research that concentrates on the funerary, it represents very rare settlement evidence containing highly fragmented food refuse where sheep (62%) and cattle (16%) were dominant, but pig (11%) was also present including samples contributing to this paper. Amongst the cereals einkorn

and hulled barley were the most frequent, as well as fruits and oil rich seeds. The settlement was probably more extensive, but only preserved in this one distinct geomorphological location (59). Four samples from this site were analyzed: AA270, AA271, AA272, AA273.

Căscioarele A, Romania
Contact: Adrian Bălășescu
Ostrovel point is *tell* with a 103 x 57 m diameter with 5 m of archaeological deposits, located on an island on Cătălui Lake, Călărași County. The prehistoric settlement (60) is among the few intensely studied at that time, giving us a glimpse on the evolution of the fauna and the animal economy. The lowermost level of the *tell* is the Spanţov stage of the Boian culture, followed by massive debris of a Gumelniţa A2 settlement. The next level, entirely excavated, is dated in the Gumelniţa B1 phase (61). The pig bones originate in the Gumelniţa A2 (6,500-6,200 BP) level (5 remains from domestic pig - AA067, AA069, AA071, AA072, AA073; and 2 from wild boar - AA070, GL566).

Čavdar, Bulgaria
Contact: Norbert Benecke
The early Neolithic settlement of Čavdar is located east of the Bulgarian capital of Sofia in the southern part of a narrow basin between Sredna Gora and the Balkan mountains (62). The faunal assemblage of Čavdar is characterised by 90% of domesticated animals, slightly dominated by sheep and goat. Most of the adult pigs represent wild boar (63). The six *Sus* remains analyzed in this study (four domestic, two wild, one probably wild - Cav1, Cav2, Cav3, Cav4, Cav5, Cav6) come from Early Neolithic phases (Karanovo I/Kremikovci I) and date to approximately 8,000-8,500 cal BP.

Cave of Cyclops (Youra), Greece
Contact: Katerina Trantalidou
The rocky island of Youra, whose total surface area is 11 km², lies at the edge of two of the most fish productive areas in the Aegean, a crucial parameter for a period of increasing exploitation of marine resources as a response to rising sea-levels. Excavations (1992-1996) at the Cave of Cyclops (37°27′N, 24°10′E) attested human presence, mainly, from the Mesolithic to the Final Neolithic period (ca. 10,600-5,500 BP) as well as during the first centuries AD. The buck of organic remains (in terms of density per volume) and the bone implements offer a rare glimpse into the subsistence patterns of Mesolithic - EN fishing/bird hunting communities. Engaging in fishing suggests also knowledge of seafaring, while the material culture (e.g. the stone artefacts) evidenced the existence of exchanging networks on both sides of the Aegean Sea and the Islands.

In the lowest deposits of the Cave, suids were the only mammalian vertebrates to be found. However, from Upper Mesolithic onward, they did not exceed the 1% among Ungulates (in terms of NISP). Suids, which could have circulated freely on the island could have provided an alternative source of protein when fish and birds were not accessible.

The bone sample (LG481), mainly that of the Lower Mesolithic, evidenced the following characteristics and interaction with humans: 1) their diet relied on vegetal

resources; 2) the comparison of the height from all periods showed that swine ranged from medium-sized (LM) to small sized (UM) -similar morphometrics have been attested for the suids of Maroulas on Kythnos, Cyclades (10,930-10,410 cal BP); 3) there were twice as many females as there were male animals; 4) the slaughter pattern affected the younger animals (97% were killed up to two years old); 5) the common effects of domestication, such as the shortening of skull, can be seen from the EN period onward. Two specimens (LG481, LG482) produced Y2 and Y1 haplotypes, and have ages corresponding to the Upper Mesolithic and Early Neolithic cultural periods (64); (65); (66); (67); (68); (69).

The site was excavated by Professor Ad. Sampson, University of the Aegean.

Çayönü, Turkey
Contact: Hitomi Hongo
Çayönü Tepesi is located on one of the small tributaries of the Tigris at an elevation of about 800 m at the foot of the Taurus Mountains in the plain of Ergani, 60 km north of Diyarbakır in southeastern Turkey. The occupation of the site spans about 3,000 years (c. 12,000-8,500 cal BP), from Late PPNA into the early Pottery Neolithic.
The Prepottery Neolithic period at Çayönü is divided into six subphases, with each period named for characteristic architectural types. The Round Building subphase (ca. 10,200-9,400 uncal BP) and the early Grill Building subphase (9,400-9,200 uncal BP) are contemporary with the Prepottery Neolithic A (PPNA) period. The late Grill Subphase (9,200-9,100? uncal BP) and the Channelled Building Subphase (9,100-9,000 uncal. BP) belong to the Early PPNB, followed by the Cobble Paved Building Subphase (9,000-8,600? Uncal BP, Middle PPNB), the Cell Building Subphase (8,600-8,300 uncal BP, Late PPNB) and the Large Room Building Subphase (8,200-8,000? uncal BP. The Prepottery sequence is immediately followed by early Pottery Neolithic Phase, but only a small area of this phase has been excavated. The Pottery Neolithic layer is partially disturbed by the Early Bronze Age graves and Early Iron Age buildings (70); (71); (72).

Sample AA514 comes from a mixed context that mainly belongs to Cell/Large Room subphase, but with some pottery was also found. Sample AA515 came from a context of late Cell Subphase. Sample LG459 was also analyzed, and dates to 7,000-6,000 BC. Although these samples came from a feature that has been described as the late/final PPNB/PN mixed (thus the date of c.7500-6000 cal BC). However, Çayönü's PN contexts is often mixed with Iron Age material, therefore it cannot be excluded that this sample possibly belonged to the Iron Age unless it is directly dated.

Cazis-Petrushügel, Switzerland
Contact: Jörg Schibler
Cazis-Petrushügel is a Neolithic dryland site that is dated to the later Horgen Culture (4,800 BP-4,900 BP) by 14C-dating. The site is situated in the alpine region of the canton Grison 3 km northeast of the village Thusis 740 m over sea level. The site was excavated 1938-39 (Walo Burkart) and 1981-82 (Margarita Primas). A small sample of animal bones were identified and published in (73). The specimen analyzed AA074.

Cencelle, Italy

Contact: Claudia Minniti
Excavations in area III of the town of Leopoli-Cencelle (ancient Centumcellae) was carried out in the 1990s by the Ėcole Françoise of Rome under the direction of François Bougard (74). The town was built in 854 AD at the behest of Pope Leo IV and in 1416 it was listed as destroyed and uninhabited in an inventory of papal territories. The area III is located on the slope of the hill at the eastern urban gate. Excavations have brought to light several buildings with different functions, whose life has developed into a narrow chronological range. Although evidence from the late 9th century AD has been identified, the most consistent is dated to the second half of the 14th century AD. A huge quantity of animal remains (>10,000 fragments) were found, among them pigs are largely dominant (75). The samples analyzed are: GL326, GL345, GL367, GL372, GL438, GL456, GL464.

Chagar Bazar, Syria
Contact: Augusta McMahon
Chagar Bazar lies on the west bank of the Wadi Khanzir (Pig Wadi), a tributary of the Khabur River in the Jezirah region of northeast Syria. The mounded portion of the site is approximately 12 ha, with a maximum height of 21 m above plain level. It was occupied from the 8th through early 4th millennia BP. The domestic pig specimen analyzed (AA053) comes from domestic rubbish dating to the final occupation in the early 4th millennium BP, when the site was an administrative node within a small regional state, the Kingdom of Upper Mesopotamia.

Chamalevri, Greece
Contact: Anastasia Papathanasiou, Mike Richards
Chamalevri is a site in the Bolanis field in the village of Chamalevri, located not far from the coast and 12 km east of Rethymnon in west Crete. The finds date from Middle Minoan IA (ca. 4,160-4,000 BP) to Late Minoan IIIC1 (ca. 3,190-3,130 BP). Chamalevri is of great importance because a Middle Minoan aromatic workshop was discovered at the site through organic residue analysis carried out on Middle Minoan IA sherds by the late Prof. Curt Beck, (Vassar College, USA). The site was excavated in 1992 by Dr. Maria Andreadaki-Vlazaki, General Secretary, Hellenic Ministry of Culture. The two samples (GL1086, GL1087) from this site dates to 1,775-1,650 BC.

Chedworth Roman villa, England
Contact: Andy Hammon
Chedworth Roman villa is a major Romano-British complex situated at the head of a small valley that leads directly into the valley of the River Coln in the Cotswolds. The building began as a simple villa in the first half of the 1st century AD, partially burnt down in the late 2nd century AD and was immediately rebuilt and extended. By the 4th century AD, a garden court had been created with the addition of more rooms and verandas. The villa continued to increase in luxury with a larger bath suite, Turkish-style baths and a water cistern, which was turned into an elaborate shrine. The dining rooms were enlarged and improved and up until the late 4th century the villa complex became steadily more opulent. It was eventually abandoned in the early 5th century AD (76). The one specimen from this site (GL527) is a wild boar.

Ciechrz 25, Poland
Contact: Arkadiusz Marciniak

Cierchz 25 is a Neolithic and Early Iron Age site in Kujavia in the Polish lowlands. The site was discovered in the 1990s during the rescue excavation project preceding the construction of the Yamal pipeline from Siberia to western Europe. The major discoveries at this partially excavated site comprised the Globular Amphora Culture settlement represented by pits and postholes as well as the Corded Ware Culture burial ground. Four specimens provided successful amplification of their DNA, two of them represented Neolithic (AA136, AA138), one Early Iron Age (AA137) and one Early Medieval period (AA139) (77).

Climente II, Romania
Contact: Clive Bonsall, Adina Boroneanț
Climente II (44.59001° N, 22.25763° E) is a cave site in the upper gorge of the Iron Gates region, in southwest Romania. The cave was excavated in 1968-9. Human remains were recovered from contexts identified as "Clisurean" (Final Epigravettian), along with ~6,000 chipped stone artifacts, bone tools including awls, arrowheads and a fragment of a harpoon, and shell and animal tooth ornaments. 14C dates for the Climente II main burial and animal remains from the "Clisurean" layer fall in the Bølling-Allerød Interstadial and the early Younger Dryas, between ~14,460-12,780 cal BP (78); (79). Comparison of the Climente II archaeological inventory with that from later, fisher-hunter-gatherer settlements in the Iron Gates indicates continuity of mortuary ritual, lithic tradition and subsistence practices from the Lateglacial into the Early Holocene. The two samples in this study (both wild boar - LG411_R1, LG412_R1) originate from the alleged "Clisurean" cultural layer (80); (81).

Coppergate, England
Contact: Terry O'Connor
Coppergate, York, is located between the historic centre of the city and the River Foss, to the south-east of the Roman fortress. Excavations in 1976-81 revealed 9th to 14th century settlement at urban density, based on four tenement plots set out in the early 10th century and most intensively occupied through the 10th to 11th centuries. Later levels include extensive refuse spreads, and the stratigraphy was truncated by modern construction. The specimens reported here (AA301, AA302) are from a securely late 12th century refuse spread (82).

Dęby 29, Poland
Contact: Arkadiusz Marciniak
This is Mesolithic hunter-gatherer site of the Janisławice culture in Kujavia, located in the Polish lowlands. The site yielded over 3,100 items made of chocolate flint. Specimens were imported from its outcrops located in the Holy-cross Mountains in southern Poland, ca. 250 km from the Dęby site (83); (84). The one sample analyzed (LG723) dates to the LBK period.

Derekh Namir, Israel
Contact: Liora Kolska Horwitz
Derekh Namir was a short-lived late Chalcolithic village (late 7th–early 6th millennium BP), located on the outskirts of the city of Tel Aviv in central Israel along a fluvial channel that led into the Ayalon River. Architecture was ephemeral but included several deep and narrow human-made shafts as well as a constructed water-well head, Based on examination of the animal bones, flint tools and grinding stones found at the site, the subsistence of the Chalcolithic inhabitants was based on a mixture of

agriculture, caprine and cattle husbandry, as well as hunting and fishing. The pig remains analyzed in this study (AA039 and AA044) are domestic Near Eastern pigs.

Direkli Cave, Turkey
Contact: Benjamin Arbuckle
Direkli Cave is located in the central Taurus mountains of southern Turkey at elevation of 1,100 meters a.s.l. in the province of Kahramanmaras. Excavated by Dr. C.M. Erek since 2007, the site has revealed a prehistoric sequence dating to Epipaleolithic period. Radiocarbon dates for the Epipaleolithic levels place the occupation at the Terminal Pleistocene between 14,000-11,000 cal BP (OxA-34195; OS-79556; OS-79558; Beta-276742; Beta-414912; Beta-414913). These dates are supported by a lithic assemblage dominated of microliths with parallels to the Natufian industry of the Levant.

The remains of hearth features, round structures, baked clay figurines, as well as a human burial have been uncovered in these levels. Based on analysis of the faunal remains the cave was used as a seasonal (summer/fall) campsite associated with hunter-gatherers in the region exploiting upland resources including primarily wild goats (38%) and secondarily deer (4.1%) and also large numbers of tortoise remains (35%). Wild boar represent less than one percent of the recovered faunal assemblage (N=6). Three specimens (AA522, AA523, AA524) recovered from the 2013 excavation season were identified as wild boar (based on the age of the deposits), *Sus scrofa*, and were submitted for ancient DNA analysis (DM2992 maxilla; DM2401 second phalanx [GL=28.9mm, Bp=20.7; SD=17.2; Bd=17.9]; DM3757 tarsal III). The Direkli Cave faunal assemblage is curated on site at the project dighouse, Döngel, Kahramanmaras, Turkey. Specimens were exported with the permission of the Kahramanmaras Museum and the Turkish Ministry of Culture and Tourism (85); (86).

Drama-Gerena and *Drama-Merdžumekja, Bulgaria*
Contact: Norbert Benecke
The two prehistoric settlements, Gerena and Merdžumekja, are located at the Southern and Northern bank of the river Kalnica close to the village of Drama in Thrace, Southeastern Bulgaria. The older settlement Drama-Gerena dates to the Middle Neolithic with three defined settlement phases (A-C) which can be paralleled with Karanovo II-IV. The seven animal samples analyzed here (Dra14, Dra15, Dra16, Dra17, Dra18, Dra19, Dra20) come from the older layers A and B dating to around 7,500-7,200 cal BP. The faunal assemblage is dominated by domesticated animals, especially by sheep and goat (up to 77% in phase A). Pigs are particularly rare with only about 7% among all domestic remains (87). Four domestic of the seven *Sus* remains analyzed provided successful amplification of their partial mtDNA.

Drama-Merdžumekja is a Late Neolithic settlement only 300 m apart from Drama-Gerena. Its most significant cultural horizons can be paralleled with Karanovo V-VI (6,800-6,400 cal BP). Just as in Gerena, the faunal assemblage is of mostly domesticated origin (94%) but is dominated by cows (around half of the remains), followed by the small ruminants. Pigs only played a minor role (around 10%). However, compared to Merdžumekja, wild boar are more frequently found among the game animals, pointing towards a significant environmental change in the area (87). Of the ten domesticated animals analyzed here (Dra31, Dra32, Dra33, Dra38, Dra41,

Dra45, Dra46, Dra47, Dra48, Dra49), four could be successfully analyzed for their partial mtDNA.

Durostorum-Ostrov Farm 4, Romania
Contact: Adrian Bălășescu
Durostorum-Ostrov Farm 4 site is located at 132 km by the DN3A road from Bucharest to Constanța, in southern Drobrudja. The Roman site (dated to the 2th -3th centuries AD) is considered to be part of the civilian crafts settlement that had close economic and possibly administrative ties with the ancient (Roman) town of Durostorum, located some 3.5 km west (88). The present sample (AA084) is a domestic pig bone comes from Pit 33 (89).

Durrington Walls, England
Contact: Umberto Albarella
Durrington Walls is a large Neolithic henge 3 km northeast of Stonehenge. Its encircling bank and ditch, 440 m in diameter, were constructed over the remains of a large Late Neolithic village consisting of many square-plan houses and a series of ceremonial timber circles. Large quantities of pig and cattle bones were recovered during excavations in 1967 and 2004-2007, most of these from pits, middens and postholes as well as house floors (90); (91). The village was in use in the decades around 4,500 BP, replaced by the henge earthworks around 4,450 BP. Its period of occupation coincides with the building of Stonehenge during its second stage when the iconic stone circle broadly took the form in which it appears today. Durrington Walls Neolithic village is likely to have been occupied by the builders of Stonehenge and has produced much evidence of feasting, especially during the winter season (92); (93); (94); (95). The samples analyzed are: VEM183, VEM184, VEM185.

Eggolsheim, Germany
Contact: Joris Peters
The site of Eggolsheim is located in the so-called Magna Germania, a region encompassing north-central Europe and situated north of the Roman frontier. Inhabited by a Germanic tribe, site occupation lasted from the 2nd until the 5th century AD. Contributing slightly less than 50% of the total livestock sample identified, raising pigs must have been a major economic activity at the site. Its husbandry certainly benefited from the extensive woodlands characterising the local landscape and site catchment, populated in late Antiquity by varied large game species including aurochs, European bison, elk, red deer, brown bear, and wild boar (96). The specimens analyzed (LG638, LG639, LG640, LG641, LG643, LG6414) date between the 4th to the 6th century AD.

Eilsleben, Germany
Contact: Hans-Jürgen Döhle
Eilsleben is situated 25 km southeast from Magdeburg, the capital of the federal state of Saxony-Anhalt. It is an enclosed settlement of LBK (LBK enclosure) dated ca. 7,500 up to 6,900 BP (all pig samples are LBK). Excavations of the former Landesmuseum für Vorgeschichte (County Museum of Prehistory) Halle were carried out between 1974-1990. The area excavated/investigated is 15 ha in total. Moreover further neolithic settlement traces were detected coming from: Stichbandkeramik, Rössener Kultur, Bernburger Kultur as well as burials from Aunjetitzer Kultur (97). The samples analyzed (AA460, AA461, AA462) date to the LBK era.

El Portalón (Portalon Cueva Mayor), Spain
Contact: José-Miguel Carretero, Cristina Valdiosera
The El Portalón is an Holocene archaeological site in the Sierra de Atapuerca, a hill located 15 km east of the city of Burgos (Spain). The Sierra forms part of the foothills of the Iberian mountain range in the area known as the "Corredor de la Bureba", which connects the northern Iberian Meseta and the Ebro river valley. Within the Sierra, there is a karst system reaching 3,700 m in length, known as the Cueva Mayor-Cueva del Silo Karst system. Inside this system, numerous archaeo-paleontological sites have been discovered with records from the Pleistocene (98) and the Holocene (99).

Among the Holocene archaeological sites found in the Cueva Mayor-Cueva del Silo, the site of El Portalón, located at its current entrance, is of particular interest due to the large stratigraphic sequence (exceeds 10 m of potential) starting in the Late Pleistocene and showing evidence of human occupation throughout the Holocene. Its complex formation is the result of a combination of natural processes, domestic anthropogenic activities, farming and funerary ritual practices (100); (101); (102). A reported detailed radiocarbon dates for the entire stratigraphy range from 30,000 BP to 1,000 BP and the Holocene phase of El Portalón, includes Mesolithic, Neolithic, Chalcolithic, Bronze Age, Iron Age, Roman and Medieval periods of human occupation (99); (102). Previous studies have successfully retrieved DNA from both humans (103); (104); (105) and faunal (106); (107) remains from the El Portalón site. The samples analyzed (AA509, AA510, AA511, AA512, AA513) date from the late Neolithic to the Medieval period.

Erbaba Höyük, Turkey
Contact: Benjamin Arbuckle
Erbaba Höyük is a low mound site located on alluvial deposits on the east side of Lake Beysehir, approximately 10 km northwest of the city of Beysehir, southwest Turkey at an elevation of 1,150 m. It was excavated by J. Bordaz as part of the Beysehir-Sugla Project during four seasons from 1969 to 1977 (108); (109); (110). The site represents the remains of a small Pottery Neolithic village settlement of approximately 0.5 ha dating to 8,500-8,000 BP. Ceramic and architectural parallels link the lower stratigraphic level (level III) of Erbaba to the upper levels at levels at Çatalhöyük (VIII-0), whereas the upper layers of the site (II and I) have close ceramic and figurine parallels with the late Neolithic levels at Hacilar VI-IX (111).

A large faunal assemblage has been studied by Perkins (112), Makarewicz (113), and Arbuckle (6) and is thought to represent the remains of an animal economy focused on the husbandry of domestic sheep, goat, and cattle. Pigs represent 5-10% of the faunal assemblage and based on large body size and limited evidence for juveniles are thought to represent Anatolian wild boar (*Sus scrofa)* (114). The lack of domestic pigs in the Pottery Neolithic at Erbaba is somewhat surprising but fits within a regional pattern characteristic of central Anatolia in which domestic pigs are absent from the region for the entirety of the Aceramic and Pottery Neolithic periods (115). Eight *Sus* specimens (AA099, AA102, AA105, AA106, AA107, AA108, AA109, AA110) were submitted for ancient DNA analysis. The Erbaba Höyük faunal assemblage is curated at the Peabody Museum, Harvard University.

Exedra of Crypta Balbi, Italy
Contact: Jacopo de Grossi Mazzorin, Claudia Minniti
Excavations at the exedra of Crypta Balbi (Rome) were conducted by the *Soprintendenza Archeologica di Roma* in the 1990s under the direction of Lucia Saguì (University of Rome La Sapienza (116)).The Crypta Balbi is a portico of a Roman theatre (Balbo theatre) that was originally built under Augustus in the 1st century BC and it remained functional until its abandonment in the 5th century AD (116). In the following centuries, the area was mainly used for rubbish disposal and the exedra became occupied by a midden (116); (117). This was largely composed of material probably deriving from the nearby monastic complex of San Lorenzo in Pallacinis. In addition to the huge number of animal bones (>15,000 fragments), the midden produced a large variety of different finds, ranging from pottery to coins, metals and glass remains (118); Saguì (117). The material (GL707, GL723), is dated between the 7th and the 10th century AD.

Faimingen, Germany
Contact: Joris Peters
Faimingen, Roman Phoebiana, was a Roman military station north of the Danube River erected ca. 70 AD. It secured the river's left bank and was essential for supplying soldiers and settlers during territorial expansion further north. Phoebiana thus developed into a major traffic junction for the transportation of persons and goods along and across the Danube. Roads leaving Faimingen also linked up with the Via Claudia Augusta, the major road connecting the Province Rhaetia with northern Italy and the valley of the Po River across the Alps. Site occupation at Phoebiana came to an end by the middle of the 3rd century AD, when raids by Germanic tribes forced troops and civilians to abandon Roman territory north of the Danube and withdrew south of the river, which became the new border. The large faunal assemblage still awaits documentation (119). The samples analyzed include LG654, LG655, LG657, and LG658.

Flynderhage, Denmark
Contact: Kim Aaris-Sørensen, Peter Rowley-Conwy
Flynderhage is a large Mesolithic Ertebølle site on the interior part of Norsminde Fjord, on the east coast of Jutland. Flynderhage is the largest type of coastal settlement referred to as Køkkenmøddinger, which are characterised by a dominant shell refuse mixed with cultural debris such as flint, bone, antler, charcoal, ceramics, ash, firecracked stones ('potboilers'), and features such as hearths, pits, stake-holes and graves (120). The midden comprises mainly oyster shells, so the fjord was more saline in the Late Mesolithic than it is today. Adjacent to other large and small Ertebølle shell middens including Norsminde. Few features were found but the midden contained many sherds of pointed-based Ertebølle ceramics and bones of wild mammals. No domestic mammals were present except dog. The three samples (AA017, GL429, GL907) analyzed date to the Late Mesolithic and of are of wild boar.

Fontbrégoua, France
Contact: Lionel Gourichon
Fontbrégoua is a large cave located in the Provence region, southeastern France, ca. 50 km from the Mediterranean coastline. Mainly occupied during the Early and Middle Neolithic (ca. 7,400-5,600 BP), the site yielded rich material culture attributed

to the Late Cardial and Chassey cultures (121); (122). 20 of the 36 selected specimens provided successful results from the DNA analysis. Among the 6 specimens dated to the Cardial period (ca. 7,300-7,050 BP), 5 belong to domestic pigs (LG188, LG194, LG195, LG197, LG200), and the other is wild boar (LG210). Regarding the following phases (Post-Cardial and Chassey), the animals studied were mostly identified as domestic (LG204, LG211). The other samples analyzed include: LG360, LG361, LG363, LG367, LG369, LG370, LG370).

Franchthi Cave, Greece
Contact: Anastasia Papathanasiou, Mike Richards
Franchthi cave is located at south-western Argolid, Greece, at the tip of a limestone peninsula, overlooking a coastal plain. It is about 120 meters long and it was used for approximately 25–30,000 years spanning the Upper Palaeolithic to the end of the Neolithic (123). The site from the Early Neolithic includes also Paralia, an adjusted, open air settlement. It was excavated in the 1970s and 1980s by Thomas Jacobsen and Indiana University and has yielded evidence for occupation and mortuary ritual as implied by a well preserved human osteological assemblage, a wide spectrum of plant and animal remains, stone tools as well as pottery in the Neolithic. It also yielded evidence for seafaring in the Paleolithic and the Mesolithic as implied by the obsidian from Melos Island and by the numerous tuna vertebrae. The two samples analyzed are GL1081 and GL1082.

Friesack, Germany
Contact: Bernhard Gramsch
Friesack is located in northern Germany, Havelland district of Brandenberg state, and contains 2 Mesolithic marshland sites, Friesack 4 and Friesack 27. The sites are in the westernmost part of the Warsaw-Berlin ice-marginal valley, and filled with fluvioglacial sand and peat. A multi-stratified sequence of sediments consisting of peat muds, sands, and humose sands indicates many Mesolithic occupations from the middle Pre-boreal (~9,700 uncal BP) to the middle Atlantic (~6,600 uncal BP). Flint, stone, wood, bark, and zooarchaeological artefacts of bone, antler, teeth, have been found, in addition to bast including fragments of nets, strings and ropes (124). The wild boar samples analyzed are GL122 and GL538.

Frilli W, Italy
Contact: Domenico Lo Vetro, Lucia Sarti
Frilli W is a multiphase site, part of a broad settlement unit that is positioned in the east of the area of Sesto Fiorentino. It is located in the upper part of the alluvial plain close to the foothill side of Monte Morello. It is a multi-stratified site with evidence spanning from Roman to Bell Beaker period. The Roman, Iron Age, Middle Bronze Age, Early Bronze Age and Bell Beaker phases are separated by alluvial layers and spread over a wide Settlement Unit. Samples are related to Middle Bronze Age dwelling activities. (125). Five wild boar specimens were analysed: AA630, AA631, AA632, AA633, AA634.

Ganj Dareh, Iran
Contact: Melinda Zeder

Ganj Dareh is a small aceramic Neolithic upland (1,400 m) site located in Iranian Kurdistan consisting of five building levels occupied over a relatively brief period (c. 2-300 years) at about 9,900 cal BP (126); (127). Demographic evidence points to initial management of morphologically wild goats at the site. One specimen (AA046) dates to the Pre-Pottery Neolithic.

Glaumbaer III (Skagafjordur), Iceland
Contact: Mike Church
The 2 pig bones sampled (AA400, AA401) were recovered from midden material associated with the occupation of a Norse farmstead. The midden was dated by 14C and tephra to between 984-1104 cal AD (128).

Göbekli Tepe, Turkey
Contact: Joris Peters
Göbekli Tepe is an aceramic Neolithic site dating to the early 12th millennium BP and 11th millennium BP. The site is not only outstanding for its unique stone architecture with conspicuous T-shaped pillars, but also because of its various animal and other motifs carved on the pillars that are illustrative of a complex belief system (129); (130). The faunal assemblage hitherto studied comprises more than 60 wild mammal and bird taxa. So far, Göbekli Tepe's fauna did not provide indications of efforts to stable and/or manage ungulates, which is in contrast to the site's wider surroundings, where management and early domestication of ungulates are well underway by the second half of the 11th millennium BP, e.g. at Çayönü (131) and Nevalı Çori (132). The two specimens analyzed are AA317 and AA318.

Gohar Tepe, Iran
Contact: Marjan Mashkour
Gohar Tepe is located in the Mazandaran Province in northern Iran, southern Caspian Sea (36°40'42"N 53°24'07"E). The site was excavated by A. Mahforouzi (ICCHTO). The site was occupied during the Bronze and Iron Age, however Late Chalcolithic, (middle of 6th millennium BP) remains also very occasionally found. The settlement reached more than 10 ha in the Middle Bronze Age, and some parts were used as a cemetery during the Late Bronze Age. Various sources of subsistence based mainly on agro-pastoralism with additional input of wild marine and terrestrial resources (133); (134). The specimens analyzed are AA027, and AA028, which were determined to be wild boar.

Gomolava, Serbia
Contact: David Orton
Gomolava is a large multi-period *tell* settlement on the left bank of the river Sava in Vojvodina, Serbia, near the village of Hrtkovci and around 60 km west of Belgrade. Around 5,000 m^2 were excavated out of an estimated original area of 13,800 m^2 [(135)], much of which was lost to river erosion prior to excavation. The specimens included here are from the first phase, Gomolava I, belonging to the late Neolithic/Copper Age Vinča group. Three subphases were identified (135) and have subsequently been radiocarbon dated (136); (137):
Ia (Vinča B2-C; c.6,900-6,800 cal BP)
Ia-b (Vinča C; c.6,850-6,750 cal BP)
Ib (Vinča D; c.6,800-6,600 cal BP)

Occupation was probably continuous throughout this date range. Where Gomolava Ia is characterised mainly by large pits, Gomolava Iab features large houses recognised from post-holes, and Gomolava Ib has a large number of smaller but densely packed burnt houses. Gomolava Ib also features the only known intra-settlement cemetery from the Vinča period, probably representing the latest phase of Neolithic activity at the site.

Animal bones have been studied from two areas, Blok I (138) and Blok VII (139). In each case wild animals represented between a third and a half of the identified mammal bone assemblage, with (domestic) cattle, red deer, and pigs being the three most important taxa. Of the pigs, the relative numbers of identified as wild and as domestic vary considerably between the two studies, but each form appears to have been present in roughly equal quantities overall. In Blok VII, wild pigs appear to have been preferentially deposited in pits. Specimens AL1062, AL1068, and AL1115 are domestic pigs; AL1063, AL1070, AL1110, AL1112 are wild boar; and the other undetermined specimens are: AL1109, AL1111, AL1113, and AL1114.

Gordion, Turkey
Contact: Melinda Zeder
Gordion is a large multi-component site located in Central Anatolia 80 m southwest of Ankara. Capital of the Iron Age Phrygian Empire (3,200-2,550 BP) the site also has levels dating back to the Late Bronze Age Hittite period up through the Roman era with a major Galatian occupation of the site (140). Pigs were a particularly important component of the animal economy at Gordion in the Middle and Late Phrygian periods (2,800-2,330 BP) (141). Six specimens (two Late Bronze - AA087, AA089, two Late Phrygian - AA088, AA090, one Hellenistic - AA091, one III-I century BC - AA087) identified as domestic pig on the basis of size and demographic profiles were successfully analyzed.

Griesstetten, Germany
Contact: Joris Peters
Griesstetten is a final Neolithic settlement pertaining to the Cham culture dating 5,500-4,700 BP. Excavated between 1988 and 1991 the site produced a large assemblage of faunal remains. The dominant species are cattle and red deer, followed by pigs. Other game species exploited for food include roe deer, aurochs, wild boar, brown bear, and beaver. Two distinct size groups can be observed in the *Sus* assemblage, a smaller one considered domestic and a larger one considered wild (142). One wild boar was analyzed in this study (LG622).

Grotta del Santuario della Madonna, Italy
Contact: Antonio Tagliacozzo
The cave is located at Praia a Mare (CS), in Northern Calabria, Southern Italy. The archaeological deposit, over 8 m thick, evidencing human frequentation from the Upper Palaeolithic to the Middle Ages. The Mesolithic layers are characterized by the presence of hearth structures, shell middens, lithic industry and bone remains of wild mammals, tortoises and malacofauna. The lithic assemblage, with very low microlithic component, is known as Undifferentiated Epipalaeolithic. The new dates frame the Mesolithic levels between 10,430 and 9,040 cal. BP. Only a tooth of wild boar (GL1000) from Mesolithic layers were successfully analyzed in this study.

Grotta Verde (Alghero, Sardinia), Italy
Contact: Domenico Fulgione, Barbara Wilkens
Grotta Verde (Alghero, Sardinia Italy) is located in the northwestern part of Sardinia. The entrance of the cave is situated 75 m above present day sea level, on the steep east slope of Capo Caccia, which is a coastal promontory (containing several karstic caves) next to the bay of Porto Conte. Underwater excavations in the chamber revealed a series of human burials, along a wall of the chamber, in the natural niches and in rock cavities. On the walls, graffiti as well as other undisputed signs of a human presence are still recognizable. The sample AL612 collected in Grotta Verde, dates back to 5,500-4,700 cal BP, and is a humerus of a domestic pig.

Gudnja, Croatia
Contact: Preston Miracle
Within the East Adriatic, the first traces of the Neolithic by impressed Ware pottery and domesticated animals are found in caves on the Adriatic islands just before 8,000 BP (143). Gudnja is a cavern site within the base of the Pelješac Peninsula, Dalmatia, in the southern end of the Adriatic. Gudnja culture, distinguished by Impressed Ware pottery, is stratigraphically well-defined, and a completely self-contained cultural phenomenon in Neolithic Dalmatia, as characterised by unique ceramographic and iconographic properties. Neolithic Gudnja is dated to 9,170 ± 70 BP, [GrN-10315; 1 sd range 8,160-7,920 BP] (144); (145); (143). The specimens analyzed (LG343_R1, LG344_R1, LG346_R1, LG347_R1) date to the Neolithic.

Guran Tepe, Iran
Contact: Melinda Zeder
Guran Tepe is a Neolithic age site located in upland (c. 700 m) elevations in Luristan, Iran (146). Animal economy in Aceramic levels at the site (c. 9,000 cal BP) is dominated by managed goats with a mix of wild animals, including wild boar, while ceramic Neolithic age levels at the site (c. 8,500 cal BP) have evidence of the introduction of domestic sheep and an increased emphasis on the hunting of gazelle (147). Two specimens (AA123, AA124) from ceramic Neolithic levels were analyzed here, both presumed to be wild.

Gürcütepe, Turkey
Contact: Joris Peters
Gürcütepe is located south to Göbekli Tepe, in the Harran plain, at the altitude of about 450 m a.s.l.. A total of 5 pig specimens (AA319, AA320, AA321, AA322, AA323) from Gürcütepe, dated to the 10th millennium BP, were analyzed in the present study, providing no successful amplification of DNA (148); (149).

Haftavan Tepe, Iran
Contact: Azadeh Mohaseb and Marjan Mashkour
Haftavan Tepe is located in the middle of the small plain of Salmas, in the northwestern part of Lake Urmia in the province of Western Azerbaijan in Iran. Haftavan Tepe is one of the largest protohistoric urban sites in the region, excavated during the ’70s. Architectural remains from the Early Bronze Age (4,300-3,800 BP according to the original dating of the site) through to the beginning of the Islamic period were uncovered and show strong ties with Caucasus. A very large assemblage of animal bones was studied within a PhD project by F.A. Mohaseb and supervised by M. Mashkour that constitutes a reliable source of information in particular for Bronze

Age and Iron Age subsistence economy and pig domestication in this part of the Iranian Plateau (150–152). The samples analyzed include the following: AA360, AA361, AA362, LG799.

Hallan Çemi, Turkey
Contact: Melinda Zeder
Hallan Çemi is an Epipaleolithic (c. 11,600 cal BP) site in the Batman drainage basin of southeastern Anatolia (153). Among the earliest sedentary communities in the region occupied year-round, the site consisted of an estimated 10-15 round houses arrayed around a central activity area. Despite the site's relatively brief occupation (200-300 years), salvage excavations yielded an exceptionally large faunal assemblage (estimated at approximately 2 metric tons), consisting of a diverse array of large and small game species. Initial arguments for early pig management at the site (154) are currently under review. The following specimens were analyzed: AA047, AA048, AA063, AA066, AA100, AA101, AA103, AA104, AA180, BLT001, BLT002, BLT003.

Hârşova-tell, Romania
Contact: Adrian Bălăşescu
Hârşova-*tell* is located on a terrace on the eastern bank of the Danube River in southeast Romania (Constanţa County). One of the largest *tells* in Europe, the thickness of the stratigraphic sequence has ca. 12 m. After to the development of the *tell*, the Danube moved its course northwards, substantially eroding the settlement. The stratigraphic sequence is very complex, ranging from the Late Neolithic (Hamangia and Boian cultures) to the Late Eneolithic (Gumelniţa and Cernavodă I cultures) (155); (156). The present bone sample (*Sus* sp., AA081) comes from the Cernavodă I levels (5,700-5,300 cal BP - (157)).

Hassek Höyük, Turkey
Contact: Bea de Cupere
Hassek Höyük. Hassek Höyük is located in the Karababa Basin in the Urfa and Adiyaman provinces in southeastern Anatolia. Altitudes range between 400 and 600 m. Studies of the faunal material has been published (158); (159); (160). Based upon stratigraphical sequencing of the sites and AMS radiocarbon dating, pig samples from Hassek Höyük were dated to the Chalcolithic and Early Bronze Age. A total of 25 specimens from Hassek Höyük were analyzed, of which three (AA312, AA313, AA314) yielded positive and reproducible amplifications of DNA (161); (158); (159); (160).

Herxheim, Germany
Contact: Andrea Zeeb-Lanz, Rose-Marie Arbogast
The early Neolithic site of Herxheim, (Rhineland-Palatinate, Germany), occupied between 7,300 and 6,950 BP, has left vestiges (pits and houses) of a occupation attributed to Linearband culture surrounded by a double earthwork. Herxheim was "normally" inhabited during the 350 years of the Bandkeramik culture left of the river Rhine, but at the end of its existence the site drastically changed its function and became a central ritual place where more than 1,000 humans were sacrificed and their skeletons smashed into small fragments. Together with equally destroyed high quality pottery, stone implements and grinding stones the rests were then disposed of in huge concentrations in a double ditch around the settlement; the parallel ditches had been

dug to receive the rests of the rituals which lack any tradition in Bandkeramik behaviour and were also not repeated in later cultures.

The faunal remains from the different features amount to a little less than 10,500 pieces. Remains of pig form an important part of these assemblages, and represent most often the most intensely exploited domestic animal after cattle. Boar hunting was also regularly practiced, as attested by the consistent. contribution of this animal in all assemblages. These characteristics reflect the key role occupied by the exploitation of the suinae in the economy of this early neolithic occupation. The Neolithic specimens analyzed are KD032, KD033, KD034.

Homolka, Czech Republic
Contact: Richard Meadow, Peter I. Bogucki
Homolka is a fortified hilltop settlement of the Řivnáč Culture in central Bohemia located 21 km northwest of Prague. It was excavated in 1929-31 and 1960-61. The Řivnáč Culture is dated ca. 5,100-4,800 BP. Homolka is a rounded hill of Algonkian filite, a soft and flaky shale with occasional outgroups of quartz. It which rises to the 20 m high above the narrow valley east of the village of Stehelceves in the county of Kladno. The floor of the valley is rich with fertile alluvial deposits. The top of Homolka resembles a roughly pointed horseshoe, the widest point approximately 100 m long and a width of 60 m. Two successive palisade systems enclosed an area of approximately one hectare that contained about two dozen small post structures surrounding an open area. Pits and postholes yielded samples of animal bones and artifacts. Animal bones found at the site include the following: domesticated -- cattle, pig, sheep, dog, and possibly cat; wild: red and roe deer, red fox, beaver, badger, cat, otter, hare, bear, horse; birds: goose, crane, black cock (synsacrum of a large gallinaceous bird); miscellaneous: frog, hamster (162); (163); (164). The specimens analyzed from this site include: GL37and GL38.

Houten-Hoogdijk, Utrecht, Netherlands
Contact: Youri van den Hurk
During excavations in Houten-Hoogdijk, Utrecht in 1999-2000, eight farmhouses were discovered. These farmhouses were dated to the 12th-13th century AD. The faunal remains, as well as the pottery and metal finds, suggest that the farmers were not particularly rich. The sample taken from this site (KD024) also dates to the 12th-13th century AD (165).

Hrísheimar, Iceland
Contact: Mike Church
The 16 pig bones (SI003, SI004, SI006, SI007, SI008, SI009, SI010, SI011, SI012, SI013, SI014, SI015, SI016, SI017, SI018, SI019) were recovered from a sheet midden, associated with turf buildings of Viking age. The midden immediately overlay the Landnám tephra (dated to 871±2 cal AD - (166) and was in turn overlain by V-Sv tephra (dated to c. 940 cal AD - (167)), producing a date range of 870-940 cal AD for the material. This range corresponded very well with 14C dating of bone material from the midden layers (168).

Icoana, Romania
Contact: Adina Boroneanț, Clive Bonsall

Icoana (44.65167° N, 22.30056° E) in the Iron Gates gorge in southwest Romania was an open-air site located on a narrow strip of land bordering the Danube. As with most sites in the area the site was flooded in 1971 during construction of the Iron Gates I dam and hydropower plant. The site was pluristratified but the Holocene deposits had been extensively eroded by the Danube prior to archaeological investigation, leaving traces of an Early Neolithic Starčevo-Criș culture (mainly pit features and sunken huts) and Mesolithic occupation (trapezoidal- and rectangular-plan dwellings). The faunal remains originated mainly from the identified features rather than the so-called cultural layers. The samples included in the present study were assigned stratigraphically to the Middle Mesolithic, radiocarbon dates for this phase ranging between ~10,800–9,500 cal BP (169). The two directly dated samples (BLT006/ICO-07, BLT008/ICO-09; (170)) fall into this time range (80); (81). The other samples analyzed (BLT007, BLT029, BLT029) also date to this period.

Jarmo, Iraq
Contact: Max Price
Qalat Jarmo is a 1.3 ha site located in the foothills of the Zagros Mountains, near the town of Chemchemal in northern Iraq. The occupation of Jarmo straddles the final PPNB and early Pottery Neolithic (late 10th to early 9th millennia cal BP). The site provides some of the earliest evidence for domestic pigs in the Zagros region (171); (172); (173); (174). The specimens analyzed are AA057, AA060, and AA061.

Kanlıgeçit, Turkey
Contact: Norbert Benecke
Kanlıgeçit is a multiphased Bronze Age tell in the eastern part of Turkish Thrace close to Kırklareli and southeast of the Neolithic tell of Aşağı Pınar. In contrast to Aşağı Pınar, cattle are dominating the faunal assemblage of the Bronze Age settlement, followed by goat/sheep and pig (one out of four each) (23). The five wild boar (Kan14, Kan15, Kan16, Kan17, Kan18) and six domestic pigs (Kan8, Kan9, Kan10, Kan11, Kan12, Kan13) analyzed here are from Early Bronze Age Layers (II-III) . All but one wild boar could be successfully analyzed for their partial mtDNA.

Karlburg, Germany
Contact: Joris Peters
Early medieval Karlburg was one of the castles granted by King Pippin to the bishopric of Würzburg founded in 741-742 AD. Occupation lasted until the 13th century AD. Excavations were carried out in the castle and associated valley settlement. Faunal analysis revealed differences between the two contexts. Obviously, the castle's inhabitants preferred the meat of juvenile pigs and cattle as well as poultry and venison (red deer), whereas the farming community in the valley had to be content primarily with the meat of aged cattle, pigs, and even horses, although forbidden by the church (175). Three domestic pigs (LG624, LG625, LG626) from the Late Medieval period were analyzed here.

Khatunarkh, Armenia
Contact: Ninna Manaseryan
Khatunarkh is a site in Etchmiadzin region dated by the 7th and beginning of the 6th millennium BP. The excavations are made by R. Torosyan. Khatunarkh is located in the Etchmiadzin region, the Ararat plain. Echmiadzin (Vagharshapat) is the 4th-largest city in Armenia and the most populous municipal community of Armavir Province,

by about 18 km (11 mi) west of the capital Yerevan, and 10 km (6 mi) north of the closed Turkish-Armenian border (176). One wild boar specimen (AA113) was analyzed.

Klimonas, Cyprus
Contact: Jean-Denis Vigne, Thomas Cucchi
Klimonas (Ayios Tychonas, Limassol District; 137 m a.s.l.) is an open-air site located on a sea facing hillslope overlooking a small valley abundant in high quality flints, less than 3 km from the South modern seashore of Cyprus. The huge material culture documentation is homogeneous and refers to a regional culture dating to the Late PPNA, principally characterized by the absence of bidirectional knapping. More than 20 radiocarbon dating (charcoal and charred seed) indicate human occupation developed here during around 10,800 cal. BP. In terms of buildings, the most striking discovery was a 10 m circular semi-embedded feature, which was interpreted as a communal building, similar to the largest ones discovered in the PPNA villages of Southeast Anatolia and the Euphrates valley (177); (178). The rest of the 20 domestic buildings of smaller size, made out of earth and trassed into the hillside, were also excavated. Together with many other aspects of the material culture and the micromorphological analyses of the mud used for the walls and floors of the buildings, scarce archaeobotanical data indicated that the Klimonas villagers were cultivating cereals, especially emmer wheat (*Triticum dicoccoides/dicoccum*), likely introduced from the continent. More than 25,000 animal remains indicate that people were hunting the only large game living on the island at that time: the small wild boar (*Sus scrofa* sp.; used in the present study), which represents more than 95% of the identified bone remains. Domestic dogs (*Canis familiaris*) and commensal cats *(Felis s. lybica*), the latter for catching mice, were living in the village.

The wild boar which was the main source of hunted animal food at Klimonas has been introduced to Cyprus ca. 12,500 cal. BP by the Epipaleolithic hunters, likely in the aim to restock the island game resources short after the extinction of the Pleistocene endemic dwarf hippo fauna (179); (180). The suspected origin is the North Levant / Southwest Anatolia. It rapidly proliferated on the island, evolving toward an insular lineage 15-20% smaller in size as the contemporaneous wild boars on the nearby continent. They were the main source of animal food until the introduction of the Mesopotamian fallow deer and the domestic ruminants, at the beginning of the 10th millennium BP (Middle PPNB), nearly 1,000 years after the occupation of the PPNA Klimonas village (181). Two Y1 wild boar specimens (AL1024 and AL1031) dating to the PPNA were analyzed. As for AL1017, AL1024, AL1024, and AL1026, it is unknown if they are wild or domestic.

Klingenberg-Schlossberg and Hetzenberg, Germany
Contact: Elisabeth Stephan
Klingenberg-Schlossberg and Hetzenberg are Michelsberg enclosures near Heilbronn in the Neckar valley in southwest Germany. Excavated in 1986-1987, the Klingenberg enclosure has two ditches across a loess promontory, traces of a palisade inside the inner ditch, remains of burnt superstructure in both ditches, numerous pits both inside and outside (182). More than 90 AMS-dates of human and animal bones, botanical remains, and charcoal suggest that the site was occupied from 40th–39th centuries to the mid-37th century cal BC (183). Eight specimens of domestic pigs (LG143, LG145, LG592, LG593, LG595, LG598, LG599, LG600) were analyzed.

The excavations of the Hetzenberg enclosure in 1966, 1989 and 1990 revealed three parallel ditches, but due to erosional processes no other features are preserved. According to radiocarbon dates and archaeological finds these enclosure was constructed earlier than the Klingenberg earthwork and used from c. 6,300-6,000 BP (182). Four specimens of domestic pigs (LG147, LG169, LG171, LG173) were analyzed.

Kohne Tepesi, Iran
Contact: Sepideh Maziar Marjan Mashkour, Shiva Sheikhi Seno, F.A. Mohaseb
Kohne Tepesi is located in the southern part of the Araxes River, in northwestern Iran. Despite its small size (0.2 ha), it contains 6 m deposits of Kura-Araxes culture (Early Bronze Age), and two burial tombs. This phase based on C14 dating, dated to the 4,636-4,339 cal BP (184). A garrison was built later on top of the Kura-Araxes remains, which dated to the Parthian period. The Archaeozoological material was studied by S. Sheikhi Seno, M. Mashkour and A. Mohaseb (in prep).

The suid assemblage of Kohneh Tepesi because of their cultural and chronological relation were analyzed both metrically and genetically (partly exploited through GMM in (161). All of the seven specimens (AA363, AA364, AA365, AA366, AA367, AA368, AA369) selected for analysis in this paper are allocated to the domestic pig and all belong to the Early Bronze Age levels. It is not known if specimens LG217, LG219, LG221, LG222 are wild boar or domestic pig.

Kolosy tér 6, (Aquincum / Budapest), Hungary
Contact: László Bartosiewicz
The Roman period settlement at Kolosy tér 6 (Budapest, Hungary), fell within the southern periphery of Aquincum, the capital of Pannonia province. It was located in the broad floodplain on the right bank of the Danube (185). Excavations brought to light floor remains in the southern section of a small building. They were found underneath a thick Roman period destruction layer dated to the 2nd–3rd century AD (186). The function of the building remains unclear. The building pre-dates the Roman finds and is thus attributable to the Celtic remains below. No zooarchaeological analysis has been carried out at the site. For this study, two domestic pigs have been analyzed (AA407 and AA410).

Koprivec, Bulgaria
Contact: Norbert Benecke
The Neolithic settlement of Koprivec lays southeast of the eponymous contemporary village at the river Baniski Lom in Northern Bulgaria (187). The two wild boar samples (Kop8, Kop9) successfully analyzed here for their partial mtDNA come from early Neolithic layers of the settlement dating to between 8,200-7,800 cal BP. They are part of a faunal assemblage that is vastly dominated by domestics (more than 80% throughout all phases), but revealed only 1-13% of domestic pig remains (63).

Kovačevo, Bulgaria
Contact: Norbert Benecke
The Bulgarian site of Kovačevo is located in the Middle Struma Valley at the foot of the Pirin Mountains close to the eponymous contemporary village. Interestingly, the Early Neolithic phases reveal Near Eastern elements additionally to Anatolian

relations (188). The domestic faunal spectrum is dominated by sheep and goat (65%), followed by pigs (almost 21%) and only few cattle (14%). Out of the 13 *Sus* remains analyzed here (Kov13, Kov14, Kov15, Kov16, Kov17, Kov18, Kov19, Kov20, Kov21, Kov29, Kov30, Kov36, Kov39) from Early Neolithic phases (8,200-7,600 cal BP), five were successfully analyzed for a partial mtDNA fragment, two of them represent wild boar.

Kuczkowo 5, Poland
Contact: Arkadiusz Marciniak
Kuczkowo 5 is a Neolithic site in Kujavia in the Polish lowlands. The site was discovered in the 1990s during the rescue excavation project preceding the construction of the Yamal pipeline from Siberia to western Europe. Remains of three Neolithic cultures were unearthed comprising: (a) LBK longhouses, (b) Late Lengyel cattle burial, and (c) round barrow burial with a palisade ditch of the Corded Ware Culture (189); (190). Six domestic pig samples from the Neolithic were analyzed: LG733, LG735, LG736, LG737, LG738, LG739.

Künzing-Unternberg, Germany
Contact: Joris Peters
Künzing-Unternberg is a middle Neolithic ditch earthwork dating ca. 6,900-6,500 cal BP. Excavated in the 1980s, the site produced a large faunal assemblage dominated by remains of red deer and cattle, while pig ranks third (191). Apart from red deer game comprised roe deer, wild boar, aurochs, brown bear, wild horse and elk, i.e. taxa indicative of extensive woodland habitat in the site catchment. Comparison with other broadly contemporaneous Neolithic *Sus* populations classified as wild or domestic using dental and post-cranial size does not point to admixture between the two forms (191), despite the fact that it could have been expected (192). Thirteen specimens were analyzed: LG148, LG153, LG155, LG156, LG159, LG174, LG255, LG256, LG628, LG269, LG633, LG634, LG635.

Kush, Ras al-Khaimah, United Arab Emirates
Contact: Derek Kennett
Kush is a small (c. 1 ha) archaeological tell close to the shore in the Shimal area of northern Ras al-Khaimah (UAE). It is located on the Arabian/Persian Gulf about 80 km inside the Straits of Hormuz close to a key maritime trade route that has operated since at least the Bronze Age. The site has been partially excavated and has yielded an occupation sequence dating from the Sasanian period (c. 5th AD) to about the mid 13th century AD (193). One domestic pig specimen (AA019) analyzed is from the Sasnian culture.

Latronico (Potenza, Basilicata), Italy
Contact: Domenico Fulgione, Barbara Wilkens
Latronico (Potenza, Italy) is a cave that opens in a travertine bench referable to the middle upper Pleistocene. The caves are located at the left side of the upper valley of the River Sinni, at about three kilometers from the town of Latronico, in the Province of Potenza (Basilicata, Italy). Evidence of Mesolithic, Castelnovian occupation has been found here. Using radiocarbon dating, four eras can be assigned to this cave, ranging from 7800±90 BP (R-449) to 7400±90 BP (R-447). The clearest and most complete stratigraphy is that of sectors III and V (grotta 3), which show the cultural

succession from the Mesolithic to the Metal Age. The sample of the present study comes from Metal Age and is a first phalanx (AL616) of a presumably wild boar.

Lchashen, Armenia
Contact: Ninna Manaseryan
Lchashen are burials close to villages Lchashen and Chkalovka. These burials are situated on the drained land in 150-200 m far from the Lake Sevan shore line, and are dated to the 4th millennium BP. Part of these burials are also dated to the Early Iron Age (the first half of the 3rd millennium BP). The excavations are made by archaeologists A. Mnatsakanyan and L. Petrosyan (129). Five specimens were analyzed: AA111, AA112, AA115, AA116, AA119.

Le Taï (Grotte du Tai), France
Contact: Stéphanie Bréhard, Jean-Denis Vigne, Allowen Evin
The site of Le Taï is located at Remoulins, in the Gard department (southern France). It is located about 125 m .a.s.l at the bottom of a ravine with steep slope in the Gardon valley, very close to the Rhône River. Intensive archaeological research have been conducted recently (194). The excavations documented a long stratigraphic sequence from Early Neolithic to Early Bronze Age. The Early Neolithic phase belongs to the Epicardial culture. Currently available radiocarbon dates indicate a continuous occupation between c. 7,250 and 7,100 cal BP. During the Epicardial phase, the site was intensively used, with occupations both inside the cave and outside. Given the diversity and density of archaeological structures (hearths, pits and post-holes) and the characteristics of the faunal, ceramic and macro-lithic remains, it is likely that the site was occupied permanently. The excavation of the Epicardial levels provided a large sample of well-preserved faunal remains: the number of identified specimens (NISP) is 3,100 for mammals. The site is characterized by a high proportion of lagomorph remains (which are food refuse) but if we exclude these small wild mammals, the faunal spectrum is greatly dominated by domestic sheep/goats and cattle (91% of the NISP). And caprines, mostly sheep, are clearly the main domestic species (71% of the NISP). Suids and wild carnivores have been identified (respectively 5% and 4% of the NISP) but they did not play an important role in the economic system. *Sus* remains mostly belong to young or subadult individuals so determine if they correspond to wild or domestic animals was not possible. The specimens examined for this site date to the Early Neolithic (AA294, AA295).

Leilan, Syria
Contact: Melinda Zeder
Leilan is large urban settlement with occupations dating to the mid-late 5th millennium and the 4th millennium BP. Early Bronze Age-Leilan consisted of an upper town with large public buildings and a surrounding lower town (195). Domestic pigs were an important component of the subsistence economy of the site. One specimen from (AA120, c. 4,300 BP) level of the lower town was analyzed (28). Two specimens from somewhat earlier levels of the upper town (AA122, AA329, c. 4,500 BP) was also analyzed.

Londerzeel, Belgium
Contact: Anton Ervynck
The castle of Londerzeel is now located in the centre of a village in the former duchy of Brabant (Flanders, Belgium). This medieval feudal stronghold originated as a

motte-and-bailey castle and later was transformed into a brick castle, belonging to the noble house of Vianden (196). From this period (end 13th- early 14th c. AD), a large collection of consumption remains has been recovered (197), including the domestic pig bones (AA280, AA281) analyzed in this study.

Ludwinowo 7, Poland
Contact: Arkadiusz Marciniak
The site Ludwinowo 7, located in Kujavia within the Polish lowlands was discovered in 2000 and excavated in the 2000s as a part of large rescue excavation project preceding the construction the A1 motorway. Its major part comprise a large LBK settlement which size is estimated to be 9.5 ha. Altogether, 14 longhouses accompanied by numerous features, including two wells and two graves, were discovered (198). Three Neolithic domestic pig specimens were analyzed (LG724, LG726, LG727).

Luncaviţa-Cetăţuie, Romania
Contact: Adrian Bălăşescu
Luncaviţa-Cetăţuie is a *tell*-settlement on a terrace near the Luncaviţa stream, 4.5 km southeast of the Luncaviţa village (Tulcea County). The stratigraphic sequence (with a thickness of 3.5 m) consists of mainly of six Gumelniţa A (phases A1 and A2) levels (199). Scarce finds were attributed to the Hallstatt period and five burials to the Late Medieval (18th century (200)). The faunal assemblage of the studied Gumelnița A2 level (ca. 6,490-6,263 cal BP) comprise both domestic and wild species, almost equally represented (201); (157) (LG800, LG802, LG803, LG804, LG805, LG806, LG807, LG808, LG809, LG811, LG812, LG813).

Madzhari, Republic of North Macedonia
Contact: Lionel Gourichon
Tumba Madzhari (or Madžari) is an open-air Neolithic site situated in the north-east neighbourhood of Skopje (Republic of North Macedonia). Recent re-excavations of the site indicate a continuous occupation during the first part of the 6th millennium (8,060-7,500 cal BP), i.e. at the end of the Early Neolithic and the very beginning of the Middle Neolithic (202). The two specimens (BLT022, BLT023) analyzed belong to domestic pigs (as identified by biometric data in comparison to samples from other Macedonian sites) with a Y1 Near Eastern signature.

Măgura, Romania
Contact Adrian Bălăşescu
The locality Măgura-Boldul lui Moș Ivănuș is part of the Măgura Buduiasca, being its earliest Early Neolithic occupational phase (ca. 8,064-7,746 cal BP). It was suggested that the earliest Neolithic communities settled first at Boldul lui Moş Ivănuş, and in time, the settlement developed at a larger scale within the Buduiasca zone. It is so far the earliest dated Early Neolithic site in southeast Romania. Two wild boar bones, from pit-features Starčevo- Criș C51 (BLT009) and C57 (LG432) were analyzed. The latter was dated ca. 8,060-7,903 cal BP (203).

Măgura-Buduiasca is located in southern Romania, 10 km north of Alexandria (Teleorman County), in the vicinity of Măgura village on the Teleorman River lower terrace. The stratigraphic sequence of the site comprises several Neolithic occupations, overlapped in some areas by remains from later occupations (e.g. Bronze

Age, Iron Age, Migration Period, and Middle Age). Archaeological excavations confirmed the following Neolithic stratigraphy: Early Neolithic (Starčevo-Criş culture), Middle Neolithic (Dudeşti culture), and Late Neolithic (Vădastra culture), covering a time span between ca. 8,064 and 7,078 cal BP. The analyzed pig bone from C22 (and attributed to the Late Neolithic Vădastra culture - BLT010) is dated 7,318-7,078 cal BP (204); (203).

Marmaray -Yenikapı
Contact: Canan Çakirlar
This is a Neolithic (ca. 8,500-7,000 cal BP) site located on the western coast of Istanbul, situated underneath a Byzantine harbor close to the southern end of the Bosphorus Channel. The settlement and its environs were investigated through extensive rescue excavations led by the Istanbul Archaeological Museum. Although the pottery and other material cultural remains display features similar to key Neolithic sites such as Fikirtepe on the eastern coast of Istanbul, burial practices show higher diversity at Marmaray-Yenikapi, and evidence for cereal cultivation is unequivocal, thanks to partial wet-preservation. Sheep, goat, cattle, and pig appear in morphologically domesticated form. It should be noted excavated areas are huge, and stratigraphic studies are still ongoing (205). One domestic pig specimen was analyzed (AL1104).

Mashnaqa, Syria
Contact: Melinda Zeder
Mashnaqa is a small settlement in the middle Khabur basin of northeastern Syria with occupations dating from the 6th to the 5th millennium BP (206); (207). One specimen (AA052) dating to Late Ubaid levels (6,600-2400 BP), presumed to be domestic pig, was analyzed.

Matarrah, Iraq
Contact: Max Price
Tell Matarrah is a Pottery Neolithic site (mid-late 9th Millennium cal BP), with similarities to the Hassuna tradition. It is located on the Piedmont plain adjacent to the foothills of the Zagros Mountains, near modern-day Kirkuk in northern Iraq. A small number of pig remains (although constituting ca. 25% of the assemblage) all appear to be morphologically domestic (208); (171); (174). Specimens AA058 and AA062 were analyzed.

Montaigle, Belgium
Contact: Bea De Cupere
The site of Montaigle is located on a promontory at the confluence of the Molignée and Flavion rivers, about 25 km south of Namur and 8 km to the east of the Meuse river. Below the ruins of a late 13th-16th century castle, human presence has been attested for the Iron Age (around 450 BC), followed by a late Roman level (270 AD to 5th century) which corresponds to the occupation of a small military garrison (209); (210). One domestic specimen (AA279) was analyzed.

Mount Petrino (Mondragone, Caserta), Italy
Contact: Domenico Fulgione, Barbara Wilkens
Mount Petrino (Mondragone, Caserta Italy) site is located on the small terraces of the upper slopes of the South-East side of Monte Petrino, at an altitude of about 190-212

m.a.s.l.. Archaeological excavations started in the 80s, when numerous ceramic fragments were found. A sample used in this study (coded AL1116) was dated 3,600-3,200 BP in the stratigraphic/cultural Bronze Age. It is a lower molar M2 of *Sus scrofa domesticus* from the mandible of a small cooked specimen. Other samples included in this style are of domestic and undetermined status (AL1124, AL1125, AL1126). Other domesticated species found are ox (*Bos taurus*), goat and sheep (*Capra hircus* and *Ovis aries*). Interestingly, this site shows a clear separation between wild and domestic forms.

Mureybet, Syria
Contact: Lionel Gourichon
Tell Mureybet was an archaeological mound situated on the left bank of the Euphrates River, ca. 100 km south from the Syrian-Turkish border, and excavated by J. Cauvin in the early 1970s before the creation of Assad Lake in Syria. The stratigraphy covers a long sequence ranging from the end of the 13th millennium to the 10th millennium BP, i.e. a large part of the neolithization process in the Near East. According to the botanical studies cultivation of morphologically wild cereals was attested since the 12th millennium BP while hunting of large and small game, including a great species diversity of mammals and birds was the main economy until the latest occupations. The two pig specimens (AA355, AA356) analyzed come from the Pre-Pottery Neolithic A (PPNA) levels, dated approximately to the 12th millennium BP (211); (212).

Narva I, Estonia
Contact: Gennady Baryshnikov
Narva I (Riigiküla I in Estonia) is the open site, one of three Neolithic sites (Narva I, II, III) located on the sand dune on the left bank of Narva River, 8-9 km northwards of the city of Narva, in Estonia. Narva I is situated at height of 8.5 m above the river. It is dated by 5th millennium BP (213). Fossils of *Sus scrofa* have been recorded in Narva (layers II and III, early Holocene), including 71 bone fragments (23.1% of all ungulate remains) and in Narva (layer I, middle Holocene), including 140 bone fragments (46.7% of ungulate remains) (214). Collection ZIN 25885 (LG234) from Narva I involves bone material from N.N. Gurina excavation in 1951-1953. It contains 493 bones from 8 adult individuals (calculated by G. Baryshnikov). The sample taken from a humerus (ZIN 25885-65) has been provided for the analysis.

Nivå (Nivaa), Denmark
Contact: Kim Aaris-Sørensen, Peter Rowley-Conwy
Nivå is a multiple Late Kongemose and Early Ertebølle settlement and activity areas on a former east-facing fjord in northeast Zealand, just north of the better-known Vedbæk Fjord. The major settlement of Nivå 10 produced the sites of 4 small huts, as well as 6 inhumation and 3 cremation burials, suggesting the presence of a cemetary like that at Vedbæk. If the economy was similar to Vedbæk, subsistence would have been dominated by small marine fish. No shell middens were present. The sample analyzed (AA014) dates from the Late Mesolithic period.

North Sea, Netherlands
Contact: Dick Mol
The southern North Sea is a rich source of fossil remains dating from the entire Pleistocene (e.g. (215); (216); (217); (218); (219)). Fishermen have been finding

fossils in their nets since at least 1874 (220). In the beginning, localities near the Brown Bank, half-way the Dutch IJmuiden and the English Lowestoft, were recognized as being sources of proliferous finds, but later, many other fossil-rich sites have become known. Smaller vessels that fish within the 12-mile zone of the Netherlands also brought many skeletal remains ashore. These fossils originated from an area north of the mouth of the Westerschelde estuary, offshore from the provinces of Zeeland and South Holland. In the early 1990s these smaller ships started fishing in the so-called Eurogeul, the 28 m deep shipping lane leading toward the Rotterdam harbour. This appears to be the most productive locality nowadays, together with the region south of it. Now we have an enormous wealth of material from the area of the North Sea that has become known as the Southern Bight. The area southwest of the Brown Bank and the surroundings of the locality called "De Stekels" (52°27'05" N. lat. and 03°10'08" E. long.) is well-known for an Early Holocene fauna including remains of humans and wild boar.

The fossil record is dominated by a Last Interglacial Fauna including terrestrial mammal species such as *Stephanorhinus kirchbergensis*, *Elephas antiquus*, and *Hippopotamus incognitus* of which the radiocarbon date is >45,000 BP followed by a Late Pleistocene Fauna with 14C dates ranging between 24,670 BP and 48,400 BP. This terrestrial fauna includes species such as *Ursus sp*., *Castor fiber*, *Crocuta spelaea*, *Panthera leo spelaea*, *Megaloceros giganteus*, *Rangifer tarandus*, *Bison priscus*, *Ovibos moschatus*, *Equus caballus*, *Canis lupus*, *Coelodonta antiquitatis*, *Mammuthus primigenius*, and *Lepus sp*. Remains of an Early Holocene terrestrial fauna has been recorded including the following species: *Castor fiber, Cervus elaphus, Alces alces, Capreolus capreolus, Canis familiaris, Homo sapiens, Lutra lutra, and Sus scrofa*. This Early Holocene fauna with 14C dates in the range from 7,780 to 8,780 BP indicate that the southern bight of the North Sea between the British Islands and the continent of Europe was still dry land and the area was inhabited by humans hunting on, among others, *Sus scrofa* (221); (222). The specimens analyzed are GL208, GL209, GL210, GL211, GL212, GL515, GL516.

Noyen (Haut-des-Nachère), France
Contact: Jean-Denis Vigne
The open-air site of the Haut-des-Nachères, at Noyen-sur-Seine (Seine-et-Marne), is located on the bank of the Seine river, in the center of the Paris Basin, at c. 60 m a.s.l. Besides the rests of a large Middle Neolithic Michelsberg fortified occupation, the excavations conducted between 1983 and 1987 provided five main loci with stratified Mesolithic deposits, totaling nearly 1,000 m² spread across an exposed area of approximately 3 ha (223). These deposits were composed of peat and sandy stratigraphic units accumulated in fluvial channels which were used for rejecting domestic refuses by human settlements which were located on the nearby banks and which are nowadays completely eroded. The excavations provided more than 7,000 animal bones remains referable to both hunting and fishing, worked objects in bone and wood (including a dugout canoe), evidence of wickerwork, human remains as well as a sparse and "atypical" lithic industry. Two topo-chronological systems could be distinguished: the oldest one was radiocarbon dated (wood; N=8) to between 9,190 and 7,970 cal BP (N=6) and attributed to the Middle Mesolithic, and the most recent to the north, dated to between 8,060 and 6,995 cal BP and assigned to the Late/Final Mesolithic with Montbani bladelets. Although they are still incompletely published, these data represent one of the richest and best quality documentation for the Middle

and Late Mesolithic of the Northwest Europe (224). The specimens analyzed (AA357, AA358, AA359, GL431, LG370) date to the Middle Mesolithic.

The rich collection of 7,200 vertebrate remains (5,350 identifiable) also represents a reference collection, remarkable not only for the its state of preservation, but also for the care in which it was collected (large volumes of wet-sieved sediment). During the seemingly year-round Middle Mesolithic occupations, red deer (*Cervus elaphus*) was the main prey species (43% NISP), followed by wild boar (*Sus scrofa*; 27%; used in the present study), aurochs (*Bos primigenius*) and roe deer (*Capreolus capreolus*). Patterns of Late/Final Mesolithic faunal remains depart significantly from those of the Middle Mesolithic as wild boar come to represent 70% of the prey signals. Hunting practices targeted females with their young. Seasonality data is consistent with the results obtained from fish remains (mostly pikes and eels) in the same deposits and indicates a small number of temporally specific hunting episodes most likely situated at the end of the summer (225); (226); (224).

Okolište, Bosnia-Herzegovina
Contact: Norbert Benecke
Okolište is a late Neolithic tell in Central Bosnia (Bosnia-Herzegovina) belonging to Butmir Culture. It was located directly at the river Bosna, which, together with the river Neretva, was connecting the Central Danubian with the Adriatic area. Most of the animal remains found during the excavations of 2002 and 2004 are of domestic origin (82%), which in turn are largely dominated by cattle (227). Out of the only 47 domestic *Sus* remains described, twelve samples (Oko1, Oko2, Oko3, Oko4, Oko12, Oko13, Oko14, Oko15, Oko16, Oko17, Oko18, Oko19) dating to around 5,100-5,000 were analyzed within this study. Five of them (Oko1, Oko3, Oko13, Oko15, Oko16) provided successful PCR amplification of their partial mtDNA.

Oudenburg, Belgium
Contact: Anton Ervynck
The castellum of Oudenburg is a Roman military fortification located on a sand ridge overlooking the Flemish coastal plain (Flanders, Belgium). Its history and food economy are well-known through recent excavations (228); (229). The domestic pig remains used (AA282, AA283, AA284, AA285, AA286) for this study derive from a dark earth layer, deposited (and reworked) during late Roman and early medieval times. Most of the material, however, dates back to the Roman period (2nd to early 5th c. AD).

Oulen (Baume d'Oulen ou d'Oullins), France
Contact: Jean-Denis Vigne, Daniel Helmer
The site of Oulen (Garn, Gard) is mostly composed of a vast cave porch (50 m wide, 15 m high; 220 m a.s.l.), dominating the right bank of the Ardèche River, near the lower Rhône Valley. A thick stratigraphy has been deposited between the Late Mousterian and the Neolithic times. The excavation of the Neolithic deposits has been processed during the eighties (230); (231). They revealed a thick, rich and rather well preserved deposits (fireplaces, pole holes, abundant domestic refuses) dating to the early Neolithic Cardial (layer 6) and Epicardial phases (layer 5), respectively dated to c. 7,300 and 6,900 cal BP, topped by a far less well preserved and rich succession of layers dating to the Chasséan Middle Neolithic and Ferrière Late Neolithic (5th to 4th millennia cal BC; 18 radiocarbon dates on bone collagen, Vigne & Helmer, unpubl.).

These deposits result from the sub-permanent, then seasonal occupations of the porch of the cave by groups of caprine herders (232).

With nearly 4,800 identified specimens, the faunal remains of the Cardial/Epicardial Neolithic deposits of Oulen represents one of the three most important archaeozoological collections in the French Midi for this period. It is highly dominated by domestic sheep and goat (58%), domestic cattle and suids (used in the present study) representing 16% each. The unpublished osteometric study demonstrated that most of the suids were small sized (thus attributed to domestic ones), some larger individuals being considered as hunted wild boar (Vigne & Helmer unpubl.). Two specimens were analyzed (AA287, AA288) date to the Early Neolithic.

Ovčarovo-gorata, Bulgaria
Contact: Norbert Benecke
Ovčarovo-gorata is an Early Neolithic *tell* in Northeastern Bulgaria close to Târgoviște belonging to Ovčarovo-Samovodene-Culture (paralleled with Karanovo II). It was part of a probably densely populated settlement chamber continuously inhabited from the Neolithic to Copper Age (62). The *tell* itself dates to between 7,700-7,500 cal BP and can be paralleled with Karanovo II. The faunal remains represent typical food deposits of a settlement with only around 6% of wild animals. Cattle are dominating the domestic faunal record (72%) followed by sheep and goat (20%) and pigs (8%). One (Ovc1) of the two (Ovc12) domestic pig samples analyzed here provided successful PCR amplification of its partial mtDNA, and was typed as Y1.

Palaikastro, Greece
Contact: Anastasia Papathanasiou, Mike Richards
Palaikastro is located on the coast of east Crete (about 145 km from Herakleion via Agios Nikolaous). It is the site of a large and wealthy Minoan town first excavated by British archaeologists Bosanquet and Dawkins (1902-1906). From the 1960s to the present, work has continued under the auspices of the British School at Athens. The total area uncovered is about one-third of a square kilometre but remote sensing indicates there is a great deal more to be found in the south and east, including the possibility of a palace. A multitude of important artefacts have been found at the site, including Linear A tablets and an exquisite ivory sculpture of a bull leaper. It was a wealthy town with fine houses and a drainage system. It was inhabited from the middle of the third millennium until post-3,450 BP, the time of the destruction of most Minoan settlements, although with its living standards at a much reduced level. Three specimens (GL1078, GL1079, GL1080) that date to the Bronze Age were analyzed.

Palegawra, Iraq
Contact: Max Price
Palegawra Cave is located in the Zagros foothills, just to the east of Jarmo. The site was occupied in the Epipaleolithic, with material culture attributed to the Zarzian tradition (Ca. 16,000?-11,500? cal BP). All the suids found were hunted wild boar (233). One wild boar specimen (AA059) from the Zarzian culture was analyzed, and dates to approximately 12,200-9,600 BP.

Pendimoun, France

Contact: Lionel Gourichon
The rock-shelter of Pendimoun is located on the French Riviera about 700 m a.s.l. and 4 km from the actual Mediterranean seashore. This place was regularly used by human groups from the Mesolithic (Sauveterrian) to the 5th millennium BP and is one of the earliest Neolithic settlements in Western Mediterranean (early 8th millennium BP, Impressa Ware culture (234). Suid remains are relatively scarce along the Neolithic sequence and are generally highly fragmented (122). Among the 3 specimens analyzed, all coming from the late phase of the Early Neolithic (Square Mouth Culture culture, ca. 7,200-6,950 BP), only one (LG215) gave genetic data which indicate a ANC-Italy haplogroup pig.

Pietrele, Romania
Contact: Norbert Benecke
The settlement mound of Măgura Gorgana lays close to the village Pietrele in Southern Romania next to the Danube river. While the settlement surrounding the *tell* was inhabited already during the Late Neolithic Boian culture, the tell itself revealed finds from Copper Age Gumelniţa Culture (235). During the Copper Age phase, domestic and wild pigs dominate the faunal record (around 75% and 56%, respectively). The samples analyzed here stem exclusively from the Copper Age *tell* itself and date to around 6,450-6,250 cal BP. Eight of ten samples were successfully analyzed for a partial mtDNA fragment, four of them represent wild boar (Pie1, Pie2, Pie3, Pie4), and the others domestic pig (Pie12, Pie13, Pie14, Pie15, Pie16).

Pločnik, Serbia
Contact: Jelena Bulatović, David Orton
Pločnik is very important site located on the left bank of the Toplica river in south Serbia. Based on its very abundant and specific ceramic finds, the late phase of the Late Neolithic Vinča culture was named after Pločnik (236). Besides numerous copper artefacts, traces of early metallurgical activity are found across the deposit of the settlement, as well as a unique find of in situ metallurgical workshop for casting and/or repair of metal tools in the central Balkans (237). This settlement with c. 3.6 m thick cultural layer (238) was occupied for c. 600 years, between 7,200-6,650 cal BP (136); (237). In total, 18 specimens of domestic pigs were analyzed (AA204, AA206, AA209, AA210, AA211, AA212, AA213, AA215, AA216, AA217, AA218, AA219, AA220, AA221, AA222, AA223, AA224, AA225).

Popeşti, Romania
Contact: Valentin Dumitrascu
The site at Popești is located at a distance of approximately 25 km southwest of Bucharest, on the right bank of the Argeş River (Giurgiu County). The complex stratigraphy ranges from Bronze Age to La Tène period (239). The late Bronze levels and are not separated by a sterile layer from the ones above them: the ceramic layer of the Zimnicea-Plovdiv type (about 14th or 13th centuries BC) and the Fundeni-Govora Ceramic Layer dated by radiocarbon before 3,400 BP. Two samples from domestic pigs were analyzed from the last layer. An Early Hallstatt layer (dated back to 10th century BC) overlaps a layer of building debris, pre-Basarabi and Novaci pottery (about 9th century BC). The Middle Hallstatt consists mainly of sporadic architectural remains and Basarabi type fragmentary ceramics (about 8th century BC). A wild boar mandible was analyzed from this context. The La Tène is represented by 5 successive levels, dated between approximately 2,150 BP and year 2 (or 4) AD. The settlement is

urban and is considered to be a regional center of power. A domestic pig mandible (AA083) was analyzed from a level dated around 2,100 BP (240). Five specimens were analyzed (AA079, AA080, AA082, AA083, GL150).

Pupićina Cave, Croatia
Contact: Preston Miracle
Pupićina Cave is situated at 220 m a.s.l. in the lower reaches of Vela Draga canyon in north-eastern Istria, Croatia. The cave is funnel-shaped; the southeast-facing entrance is about 20 m wide and 8 m high. Thick deposits revealed deep stratigraphic sequence with archaeological layers spanning from the late glacial period to the late Roman period (241). Eleven specimens (AA456, AA457, AA458, AA459, AA471, AA472, AA473, AA474, AA475, GL726, GL980) were analyzed.

Racot 18, Poland
Contact: Arkadiusz Marciniak
The Late Lengyel settlement in the western part of the Polish lowlands south of Poznań excavated in the years 1984-1987. The excavated part revealed 14 trapezoidal longhouses accompanied by numerous external and internal pits and one burial. The site extends over an area of ca. 3 ha with an estimated number of 40-48 to 48-56 longhouses (242). Three domestic specimens analyzed (LG711, LG712, LG721) date to the Neolithic.

Regensburg-Harting, Germany
Contact: Joris Peters
The remains collected at Regensburg-Harting originate from a Roman villa rustica that was built towards the end of the 1st century AD and abandoned in the 4th century AD. Analysis of the faunal remains has been concluded but the faunal results still await publication (243). Five specimens were analyzed, of which four are domestic pigs (LG164, LG165, LG589, LG590) and one is unknown (LG588).

Rosstal, Germany
Contact: Joris Peters
Erected in Carolingian times, Rosstal castle was inhabited from the early 9th until the 11th century AD. Its fauna is largely dominated by the remains of pigs. Compared to contemporaneous *Sus* populations elsewhere in Bavaria, the pigs from Rosstal are large-sized animals, with shoulder heights ranging between ca. 70 and 94 cm. Pig husbandry likely benefited from the extensive deciduous forests occurring in the region. Provided silvopasturing was practiced by the site inhabitants, one explanation for the large size of the pigs could be admixture with wild boar (244). Six early Medieval suids were analyzed (LG646, LG647, LG649, LG650, LG651, LG652).

Roucadour, France
Contact: Josephine Lesur, Anne Tresset, Jean-Denis Vigne
Roucadour (Lot, France) is a site located south-west of the Central Massif, in a sinkhole at 346 m a.s.l.. It has been occupied during a long period, from the Neolithic up to the Iron Age. The suid bones included in this study come from a Middle Neolithic horizon belonging to the Chasséen culture and dated between the end of the 7th millennium and the beginning of the 6th millennium cal BP. The faunal assemblage is strongly dominated by wild species, which is somehow surprising in a site from the

Chasséen culture. Suid remains account for ca. 26% of the identified mammal remains (245). Three wild boar (GL321, GL470, GL485) specimens were analyzed.

Rottenburg-Siebenlinden, Germany
Contact: Elisabeth Stephan
Rottenburg-Siebenlinden is a Mesolithic open-air site on the banks of the river Neckar about 50 km south of Stuttgart in southwest Germany. Excavated extensively between 1990 and 2004 a site of about 480 m^2 was investigated and several thousand lithic artefacts as well as wild animal remains were recovered from four different Mesolithic layers. Layer II dates to the Late Mesolithic (c. 8,500-8,100 cal BP, Atlantic period), and layer IIIo, III, and IV to the Middle Mesolithic (III: Beuronien C 9,400-9,100 cal BP, Late Boreal period; IV: Beuronien B 10,100-9,700 cal BP, Early Boreal period). One wild boar specimen (LG166) from layer IV was analyzed.

Sagalassos, Turkey
Contact: Bea De Cupere
Sagalassos is an urban site in southwestern Turkey near the village of Ağlasun (Burdur province), in a region known as Pisidia in ancient times. The town was built on a south-facing hill slope in the western portion of the Taurus mountain range and mainly inhabited during the Roman-Early Byzantine period. The nearby fortified village of Düzen Tepe dates to the Classical/Hellenistic period (2,500-2,200 BP) (246). Specimens dating to the Imperial period to the Early Byzantine were analyzed (AA030, AA031, AA032, AA033, AA034, AA035, AA036, AA037, AA038).

Samovodene, Bulgaria
Contact: Norbert Benecke
Samovodene is a Neolithic *tell* in Bulgaria North of Veliko Tarnovo at the river Jantra. The earliest settlement phase is paralleled with Early Neolithic Karanovo II (7,800-7,500 cal BP) and belongs to Ovčarovo-Samovodene-Culture (187). This early phase represented by one wild boar and one domesticated pig sample in the present study. We also analyzed two wild boar from the early middle Neolithic period (around 7,500 cal BP, paralleled with Karanovo II/III) and three late Neolithic wild boar (paralleled with Karanovo IV, 7,300-6,900 cal BP). Analyses of the faunal remains revealed a focus on cattle farming (more than 80%) and a only minor role of domestic pigs (63). Four (Sam5, Sam9, Sam10, Sam18) out of the seven samples (Sam6, Sam19, Sam20) provided successful amplification of the partial mtDNA fragment, including the domestic pig (Sam5).

Šandalja II
Contact: Preston Miracle
Šandalja II (44° 52' 57" N, 13° 53' 48" E) is a “fossil” cave, located at the southern tip of Istrian Peninsula in the northern part of the eastern Adriatic coast at an absolute elevation of 72 m/a.s.l. (247); (248). The deposits were divided into layers A – i. Layer A was of the Holocene age, layer Bg has dates from both the Early Holocene and the Late Glacial interstadial, while the rest of the layers were deposited during the Late Pleistocene (249). One specimen (LG335_R1) was analyzed.

Santa Maria in Selva, Italy
Contact: Domenico Fulgione

The site of Santa Maria in Selva is located near the city of Macerata, in the center of Italy, on the top of a hill. In it were found various findings of domestic animals (bred) and various objects including vases with a spool handle and with curled margins in the guise of cornets, a truncated cone shell with a wide mouthpiece adorned entirely by triangular incisions. fragments of copper confirm the dating of this facies to the final Neolithic. In the village were also used domestic cereals that were also found in the lining of huts made of clay. the suids are attributable to domesticated forms (250). One wild boar specimen (AA629) dating to the Late Neolithic was analyzed.

Schela Cladovei, Romania
Contact: Clive Bonsall
Schela Cladovei (44.62612° N, 22.60591° E) is a large, open-air settlement in the Iron Gates section of the Danube Valley in southwest Romania. The site occupies an Early Holocene terrace adjacent to the Danube, ~7 km downriver from the Iron Gates I dam. Discovered in 1964, early investigations (1965-1968, 1982-1991) were undertaken by Vasile Boroneanţ. Since 1992 the excavations have been conducted mainly as a joint Romanian-British research project, co-directed by V. Boroneanţ (later A. Boroneanţ) and C. Bonsall. In the excavated areas of the site the archaeological remains relate mainly to Late Mesolithic (~9,200-8,300 cal BP) and Early Neolithic (~8,000-7,600 cal BP) occupations (80), but there is sporadic evidence of later (Bronze Age, Iron Age and Medieval) activity. The 5 samples (AA402, AA403, AA404, AA405, AA406) included in this paper were excavated from Area VI in 1992-1996 by V. Boroneanţ and C. Bonsall, and they were submitted for aDNA analysis by Alexandru Dinu without the details of their archaeological contexts, which makes chronological placements uncertain for those samples without direct AMS ^{14}C dates.Of the two directly dated samples, AA406 (2-sigma calibrated range 9,247-8,649 cal BP) is Late Mesolithic and AA402 (2-sigma calibrated range 7,756-7,505 cal BP) is Early Neolithic (251)

Shan-Koba (Abri Šan-Koba), Ukraine
Contact: Norbert Benecke, Gennady Baryshnikov
Shan-Koba shelter is located in south-western, mountainous, Crimea, in the Baidar valley. Excavations by G.A. Bonch-Osmolovsky in 1928 and by S.N. Bibikov in 1935-1936 revealed 6 cultural layers. It was inhabited during Mesolithic and Neolithic times. Layers 1a is dated by Neolithic, layers 2-4 – Mesolithic, layers 5 and 6 – Final Paleolithic (252). Archaeozoological investigations showed that the faunal remains mostly constitute food waste of human hunters (253). Two Mesolithic (around 7,900-9,500 cal BP - San3, San6) and three Neolithic (around 7,300-7,700 cal BP - San7, San8, San9) of the nine wild boar (San1, San2, San3, San4, San5, San6, San7, San8, San9) in this study provided successful PCR amplification of their partial mtDNA. *Sus scrofa* is represented by 101 bone fragments belonging to 8 individuals, from layers 4-6 (San2), and by 482 remains belonging to 10 individuals, from layers 2-3 (San3, San6) (254). Bone material recovered by Bibikov excavation is kept in Zoological Institute, Russian Academy of Sciences in St. Petersburg (ZIN).

Siniarzewo 1, Poland
Contact: Arkadiusz Marciniak
Siniarzewo 1 is a Late Lengyel settlement in Kujavia in the Polish lowlands. The site was discovered in the 1990s during the rescue excavation project preceding the construction of the Yamal pipeline from Siberia to western Europe. The discovered

remains comprised a trapezoidal longhouse with accompanying pits in addition to a pig burial in one of the pits. Remains of Bronze Age settlement were also revealed. Out of the specimens that provided successful amplification of their DNA, two of them represented Late Lengyel (AA140, AA141) while remaining are Bronze Age (AA144, AA145, LG507, LG508, LG513, LG514) (255).

Sirkeli Höyük, Turkey
Contact: Joris Peters
Sirkeli Höyük is located in Cilicia, the south coastal region near Adana in present-day Turkey. The site was inhabited from the Chalcolithic period till the Iron Age (6,000-2,500 BP). Because of the semi-arid climatic conditions prevailing in the region, however, pig husbandry was not that important in the site's economy, but relative frequencies increase throughout site occupation. The game species most commonly hunted by the site inhabitants was Mesopotamian fallow deer, but from time to time wild boar was killed as well (256). Four Iron Age specimens (LG522, LG524, LG527, LG529) were analyzed, of which three are domestic.

Sofia-Slatina, Bulgaria
Contact: Norbert Benecke
Sofia-Slatina is a Neolithic settlement south of the former village of Slatina, which is now a district of the Bulgarian capital of Sofia. Excavations started in 1985 as a rescue excavation and revealed finds dating back up to the Early Neolithic.The four domestic pigs analyzed here come from horizons belonging to Kremikovci Culture (paralleled with Karanovo II) and date to 7,750-7,600 cal BP. The faunal spectrum of the site is dominated by domestics with pigs playing only a minor role (below 10%) (63). Three of the four samples (Sof1, Sof2, Sof3, Sof4) were successfully analyzed for a partial mtDNA fragment.

Soroca III near Soroka, Moldova
Contact: Norbert Benecke
Soroca III is a Mesolithic/Neolithic settlement on the left bank of the Dnestr river in Moldova, dating to 8^{th}–7^{th} millennium BP (257). One Neolithic wild boar specimen (GL126) was analyzed.

Shpan-Koba near Lesnoe Kipčak (Crimea), Ukraine
Contact: Alexandr Yanevich, Norbert Benecke
This site is a rockshelter at a distance 15 km east from Simferopol in the mountainous part of the Crimean peninsula at an altitude of 800 m above sea level, on the south-eastern edge of the mountain range Dolgorukovskaya Yayla. It was discovered in 1925 by A.S. Trusova. Excavations by A.A.Yanevich in 1988-1989 (258) revealed there 6 cultural layers with occupation during the Mesolithic and Neolithic periods (Preboreal-Atlantic). There are 12 14C dates (c. 12,500-9,000 cal BP) (259). Cultural layers are left during seasonal visits of hunting groups in the warm season. Two wild boar specimens (GL128, GL536) were analyzed.

Spazzavento, Italy
Contact: Domenico Lo Vetro, Lucia Sarti
The Neolithic Spazzavento area is part of the larger Spazzavento-Neto-Via Leopardi Settlement Unit in the Sesto Fiorentino area of Italy (260). The site is part of a broad settlement unit which is formed by excavation areas: Neto-via Verga, Neto di Bolasse

and Spazzavento. The unit, positioned in the west of Sesto Fiorentino territory, is located on a sort of morphological top formed by depositional accumulations of two alluvial fans. The stratigraphic sequence as a whole covers a wide chronological range which indicates that the area was occupied from the 5th millennium until the Roman period (261); (262). 3 Neolithic wild boar samples (AA635-AA637) could not be used in the analysis because the haplogroup tests failed.

Staraya Ladoga, Russia
Contact: Gennady Baryshnikov
Staraya Ladoga, is an ancient city in the lower stream of Volkhov River near Ladoga Lake, Russia. Bone material has been collected there during N. Repnikov excavation in 1909-1913, from Zemlianoe gorodishche, which is characterized by several temporal levels dated by the second half of 8th century to 10th century (263). Collection of *Sus scrofa domestica* (ZIN 18757) comprises 394 bone remains belonging to 13 adult individuals (calculated by G. Baryshnikov). The sample taken from tibia ZIN 18757-105 (LG239) has been given for the analysis.

Stuttgart-Mühlhausen, Germany
Contact: Elisabeth Stephan
Stuttgart-Mühlhausen 'Viesenhäuser Hof' is located on a loess ridge in the Neckar river basin in the north of Stuttgart in southwest Germany. The excavations between 1931 and 1993 have revealed remains of the oldest, older and younger Linearbandkeramik as well as the Middle Neolithic, the Middle and Late Bronze Age, and the Early Iron Age (Late Hallstatt/Early Latène). During the LBK a settlement and large cemeteries existed. Two specimens of domestic pigs from the LBK settlement were analyzed (LG103, LG167).

Suberde Höyük, Turkey
Contact: Benjamin Arbuckle
Suberde Höyük is located in the intermontaine Sugla basin in southwest Turkey (264) on a limestone ridge along the margin of Lake Sugla at an elevation of 1,070 m. Suberde was first recorded in a survey by R. Solecki (265) and was excavated for two seasons in 1964 and 1965 by Jacques Bordaz as part of the Beysehir-Sugla Project. Bordaz described four stratigraphic levels at the site with levels I-III consisting of cultural remains, while level IV consisted of sterile sediments. Levels II and III consist of the remains of plastered floors and rectilinear mudbrick architecture including storage and bench features. Pottery is rare or absent in these levels which are thought to represent the remains of a small, 0.5 ha, village settlement dating to the latest part of the Aceramic Neolithic Pottery (264). Level I represents a mixed surface layer consisting of Roman to Islamic period burials, and a small number of Neolithic, Chalcolithic, and Bronze Age sherds. The prehistoric levels of the site are radiocarbon dated to 9,500-9,000 BP (266). Ten *Sus* specimens from these prehistoric levels were submitted for ancient DNA analysis, and 9 were analyzed (AA390, AA391, AA392, AA393, AA394, AA395, AA396, AA397, AA398).

The faunal assemblage from Suberde is dominated by the remains of morphologically wild, but probably managed, sheep and goats followed by pigs which represent 13% of the mammalian remains (267). Wild cattle and red deer are present in significant but relatively small numbers. The pig remains from Suberde are interpreted as representing Anatolian wild boar. Recent archaeozoological analysis of the *Sus*

remains has shown that the Suberde pigs are comparable in postcranial breadth and depth measurements (and therefore body size) to prehistoric wild boar documented at Göbekli Tepe, Erbaba Höyük, Çatalhöyük, and Asıklı Höyük (268). Only a small portion of the original Suberde faunal assemblage remains following its export from Turkey by Dexter Perkins in the 1960s (see (266)). The remaining assemblage is curated in the Zooarchaeology Laboratory, Department of Anthropology, University of North Carolina at Chapel Hill.

Surb nahatak (Sevcar), Armenia
Contact: Ninna Manaseryan
Surb nahatak (Sevcar) is a fortress is located on the southern outskirts of the village of Sevkar (Ijevan district). It dates from the 2,700-2,500 BP. The excavations were performed by archaeologist S. Yesayan (176). One domestic specimen (AA118) was analyzed.

Swinegate, England
Contact: Terry O'Connor
The site in Swinegate, York, is located at the heart of the medieval city, in the southernmost quarter of the Roman fortress. The medieval occupation consisted in the main of small properties clustered around St Benet's churchyard, now extinct. The specimens reported here derive from mid-14th century activity (269). Two specimens (AA303 and AA304) were analyzed.

Tel Bet Shean, Israel
Contact: Liora Kolska Horwitz
Tel Bet Shean is favorably situated in the central Jordan Valley, at the intersection of trade routes running north-east from the coast into the hinterland. The main occupation levels date to the Late Bronze Age II through to the end of the Iron Age (ca. late 3,600-3,400 BP to 2,700-2,500 BP), although earlier remains dating to the Early Bronze Age have been found at the site as well as fragmentary evidence for occupation in the Roman, Byzantine and Islamic periods (270). Pigs occur in low frequencies throughout the sequence. The two specimens (AA041, AA043) examined in this study derive from the Late Bronze Age II strata (ca. 3,550-3,400 BP) and represent domestic swine (271).

Tel es-Safi/Gath, Israel
Contact: Liora Kolska Horwitz
Tel es-Safi/Gath is located ca. 53 km east of the city of Ashkelon in the central coastal plain of Israel. Current excavations at the *tell* have demonstrated that it was occupied from the Early Bronze Age to the Ottoman period (272). Human activity at the site peaked during the Iron Age (3,200-2,700 BP), when it was one of the main Philistine cities. The domestic pig specimen (AA043) examined here, derives from these layers. Notably, the Philistine occupation was accompanied by the introduction of pigs from the Aegean, who were characterized by European haplotypes and subsequently altered the genotypic profile of local swine populations (273). Two specimens (AA041, AA043) were analyzed.

Tel Miqne-Ekron, Israel
Contact: Liora Kolska Horwitz

The *tell* is located on the Israeli coastal plain (Shephelah) ca. 35 km south-west of Jerusalem. It sits on trade routes going north-east from the coast into the hinterland. The site was excavated under the direction of Trude Dothan (Hebrew University) and Seymour Gitin (274) beginning in 1981 through 1996 (274); (275). Tel Miqne-Ekron has been identified with biblical Ekron, one of the five Philistine cities that existed on the central Israeli coastal plain in the Late Bronze Age II through to the end of the Iron Age (ca. 3,600-3,400 BP to 2,700-2,500 BP). Scanty earlier remains dating to the Chalcolithic and Early Bronze Age have been found at the site as well as fragmentary evidence for occupation in the Roman, Byzantine and Islamic periods.
At its zenith in the Iron Age, site size is estimated as 20 ha, with a lower (elite) area of 40-65 acres and an upper area, ca. 10 acres. The site experienced a complex history of destruction, abandonment and re-occupation. Architectural finds include a fortification system (Iron Age I), residential and industrial areas, and off-site agricultural installations. Philistine material culture attests to Aegean and Cypriot affinities, as well as interactions between the Philistines, Israelites, Phoenicians, Assyrians, and Egyptians.

Faunal assemblages from different parts of the *tell* have been studied by the late Brian Hesse, Justin Lev-Tov and Edward Maher (276); (277); (278); (279). In almost all periods, domestic caprines formed the mainstay of the economy and were closely followed by cattle and, at the height of Philistine rule, also by pigs whose frequency drops off in the late Iron Age. Remains of domestic equids, dogs and wild animals are negligible in all periods but are often found in ritual contexts (278). Sheep-goat proportions change over time, with a predominance of sheep in the late Iron Age indicative of wool-production. Mortality patterns also change and are interpreted by Lev-Tov (277) as evidence for a shift from a local, household-oriented caprine production system focused on production for meat and secondary products, to a market-oriented system geared primarily toward secondary products in the late Iron Age. Three specimens (AA276, AA277, AA278) dating to the Iron Age were analyzed.

Tell Aqab, Syria
Contact: László Bartosiewicz
The two pig samples (BLT004, BLT005) from Tell Aqab, Syria presumably come from Middle Halaf Period layers. They included a mandibula fragment and a proximal radius fragment. This is a *tell* site in Jezirah Province, NE Syria. A single Middle Halaf sample has been 14C dated (Aqab 212, AA-30498) to 7,579–7,350 BP (280).

Tell Brak, Syria
Contact: Roger Matthews
Tell Brak is located in the Upper Khabur region of northeast Syria, approximately halfway between the modern towns of Hassake and Qamishli. Tell Brak, ancient Nagar, is one of the most important and long-lived archaeological sites of Upper Mesopotamia, with occupation spanning ca 8,000 to 3,200 BP. Major features of its history including evidence for early urbanisation at ca. 5,500 BP, a multi-phase Eye Temple containing thousands of stone ‘eye idols’, major administrative and religious buildings of the 5th millennium BP, and a Mitanni palace on top of the mound at ca. 3,500 BP. Archaeological evidence indicates that the site was a major focus of trade, administration and trans-regional engagement across Upper Mesopotamia and well beyond, taking advantage of its prime location at a transection of east-west and north-

south trade routes. Nine suid specimens (AA020, AA021, AA023, AA093, AA094, AA095, AA096, AA097, AA098) were analyzed.

Tell-Afis Saraqeb, Syria
Contact: Domenico Fulgione, Barbara Wilkens
Tell-Afis Saraqeb, Syria is an archaeological site in the district of Idlib. Excavations of the massive city wall in the lower town and a large sacred building of the *in antis* type on the acropolis, prove that this site was flourishing between the second and third Iron Age. The sample from this site (piece of skull, coded AL613 *domesticus*) was excavated in 2000-2001, in an area that shed more light on the first Iron Age, by discovering many other hunting species, besides wild boar, such as roe deer, gazelle, hare, fox and leopard. During this excavation, also domestic fauna was found, such as sheep, goat, cattle and, even though less frequent, pig. The latter species is less common in the Bronze Age.

Thebes, Greece
Contact: Vassilis Aravantinos, Efrossini Vika
Thebes is found in the region of Boeotia, in the centre of mainland Greece. The region consists of two extensive plains crossed by rivers and surrounded by mountains. Large, freshwater lakes and many natural springs were present in the area. Natural geomorphology and climate were not significantly different in ancient times. The alluvial soils were fertile, suitable for crops and livestock, thus the area was beneficial for human settlement. Several important Bronze Age settlements were located in Boeotia, culminating in the emergence of a Mycenaean palatial building in the Late Bronze Age.

The samples come from the same site and represent the two phases of its use. The site is an Early Helladic building of the characteristic for the time ellipsoidal architectural form, that was destroyed for reasons unknown to research so far. The debris together with a mass burial that was placed over the ruins, was covered under a mud brick tumulus and left to abandonment, until the next period, the Middle Helladic, when it was used as a burial ground. The study of the ceramic assemblage from the site reveals the co-existence of a local ceramic tradition together with shapes of the "Anatolianising" type (Lefkandi I group), with both categories being produced locally (281); (282); (283). One Early Bronze Age specimen (AA370) and one Middle Bronze Age specimen (AA371) was analyzed.

Tmbatir (Achajur), Armenia
Contact: Ninna Manaseryan
Tmbatir (Achajur) is located on the North West boundary of villages Achajur (Ijevan district). It dates from the 2,600-2,4000 BP. The excavations were performed by archaeologist S. Yesayan (176). One Iron Age specimen (AA117) was analyzed.

Tol e Bashi, Iran
Contact: Marjan Mashkour
Tol e Bashi (Iran) is a Late Neolithic in the Province of Fars in Southwest of Iran. The site was excavated by Kamyar Abdi, Susan Pollock and Reinhard Bernbeck. Archaeozoological studies were performed by M. Mashkour and Salvador Bailon (284). Tol e Bashi has a highly pastoral economy. The pig remains are very scarce

in the assemblage and do not seem to derive from the domestic stock. One specimen from possibly Neolithic contexts (AA025) was analyzed.

Trelleborg, Denmark
Contact: Kim Aaris-Sørensen, Peter Rowley-Conwy
Trelleborg is a Viking-period geometric ring fort on the island of Sjælland. The military fortress was constructed using timbers felled in AD 980, probably under the orders of King Harold Bluetooth. It is a causewayed camp situated on a plateau of morainic clay projecting between the rivers Tude Å and Vårby Å in southeastern Zealand. Its position on a peninsula between the two great rivers was probably intended to dominate passage through the Great Belt. It is 137 m in diameter and had four entrances on the cardinal points. Transverse roads connecting opposing entrances divided the interior into four quadrants, each of which contained four major halls arranged in a square. The 16 halls would have provided accommodation for over 1,000 people. 125 pits were found during excavation of the Viking fortress in 1934–42, some of which were later identified as ditches distributed in two parallel rows (285). The various features contained finds from all periods of the Middle Neolithic Funnel-Beaker Culture (286); (47). Two domestic pig specimens (AA016, GL306) were analyzed.

Trou Al'Wesse, Belgium
Contact: John Stewart
Trou Al'Wesse is a cave site located in the valley of the Hoyoux river, a tributary of the Meuse, adjacent to the Belgian Ardennes. The slope deposits on the terrace immediately outside the cave entrance have been excavated revealing several facies of C14-dated Mesolithic (Couches 4b-α, 4b-β, 4b-γ, 5a), transitional Mesolithic/Neolithic (LaH), and Neolithic (Couche 4a) occupations. Seven samples were analyzed from layers 4b-α (1 sample dated to ca. 10,000 BP), 4a (5 samples dated to ca. 5,500 BP), and 2 (1 sample undated) (287).

The following were successful:
TAW 2004 G12 piece 56 (AA128) – Comes from Couche 4a
TAW 2004 G12 piece 48 (AA129) – Comes from Couche 4a
TAW 03 I9 piece 29 (AA131) – Comes from Couche 2
TAW 2004 H7 piece 33 (AA126) – Comes from Couche 4a

Tsakaektsy, Armenia
Contact: Ninna Manaseryan
Tsakaektsy is a rural settlement of the XI-XIY centuries on the ravine of the Agstev river. It is a part of the Ijevan-Taush historical-ethnographic zone of Armenia. The settlement is located 5 km to the east of the village of Sari Glukh (15 km east of the city Ijevan). The area of the settlement is 900 m^2. The excavations are made by H. Yesayan (the son of S. Yesayan)

Six species of animals have been identified there from osteological materials excavated at Tsakaektsy. The results of the analysis returned more than 250 identifiable bone fragments. The fragments show that 97% of them belong to domestic animals (cattle, sheep, pig, and horse). The wild fauna are represented by the roe deer and the turtle. One suid sample (AA114) was analyzed.

Twann-Bahnhof, Switzerland
Contact: Jörg Schibler
Twann-Bahnhof is a Neolithic lakeshore site situated at the northern shore of lake Bienne. The site was excavated between 1974 and 1976. An area of 160 m of 14.5 m of the site had to be excavated because of the construction of a small highway. Within this area a rich stratigraphy of 13 settlement layers was found. 10 settlement layers belong to the Cortaillod Culture, these layers are dated between 5,838 BP and 5,532 BP by dendrochronology. Three settlement layers belong to the Horgen Culture and are dated between 5,405 BP and 5,072 BP. The used sample bone belongs to the youngest layer of the Horgen Culture that is dated to 5,093-5,072 BP. In total 203,743 animal bones were identified from the Cortaillod layers (288); (289); (290). Only 8,085 animal bones were identified from the Horgen layers (291). One Neolithic domestic specimen (LG691) was analyzed. The bones were provided by Archäologischer Dienst des Kantons Bern for this study.

Uivar, Romania
Contact: Wolfram Schier
Gomila is a late Neolithic to early Eneolithic tell close to the village of Uivar, around 40 km southwest of Timișoara in Southwest Romania. It has been surveyed and excavated 1998-2009 by a Romanian-German team and yielded finds from mostly middle to late Neolithic Szakálhát and Vinča (phase C) and early Copper Age Tiszapolgár Culture (292). While domesticated animals were dominating the faunal record of all Vinča phases, especially the number of cattle remains decreased until the final C2 phase from around 70% to 50%. Increased hunting activities and focus on the small domestics pigs, sheep and goat point towards climatic changes at the end of the Neolithic (293). Two wild (Uiv9, Uiv11) and one domestic (Uiv10) *Sus* remains were successfully analyzed for a partial mtDNA fragment. They were excavated in 2008 and come from middle Neolithic Szakálhát / Vinča phase B, more specifically from building phases 5a (7,205-7,090 cal BP (95%), (292), and 4a/3d (7,090-6,990 cal BP (95%) (292)). The third pig sample, belonging to the same building phase 4a/3d, was dated directly (KIA-42156: 6,560 ± 35 BP). This date, attributable to the later Early Neolithic, suggests a rediposited older bone, although no Early Neolithic (Starčevo) settlement could be identified at Uivar.

Ulucak Höyük, Turkey
Contact: Özlem Çevik Canan Çakırlar
Ulucak Höyük is favorably situated along one of the main arteries between the Aegean coast and inland Anatolia in the Izmir province. Well-preserved Neolithic deposits in Levels IV and V represent a material culture akin both with the Lake District and the Greek Neolithic (294), whereas Level VI is unique in the region with pre-pottery deposits on red-plastered floors and material culture with direct links to sites in Syria, Jordan and SE Turkey (295). Pigs in these layers dating to 8,800-8,500 cal BP are morphologically domestic (295); (296). 17 pig specimens that were associated with levels IV and V date from 8,400-7,900 BP (296). Five of these samples, all identified as domestic pigs (AA327, Ulu1, Ulu48, AL1102, AL1103), were successfully analyzed previously (161).

Umm Qseir, Syria
Contact: Melinda Zeder

Umm Qseir is a small settlement on the middle Khabur Basin. Halafian (8th millennium BP) levels at the site, dating to about 7,900-7,000 BP, represent a year-round settlement occupied by a small number of families engaged in mixed agro-pastoral pursuits that included management of domestic sheep, goats, and pigs, supplemented by wild game (297). A later 6th millennium BP Uruk period, occupation is more ephemeral and may represent the encampment of both local and non-local people possibly from southern Mesopotamia. Two specimens (AA049, AA050) from Uruk age deposits were analyzed.

Undir Junkarinsfløtti, Faroe Islands
Contact: Mike Church
The 28 pig bones were recovered from a series of midden and occupation layers, sampled from the coastal erosion section cutting through the extensive farm-mound of Undir Junkarinsfløtti (298). 3 main phases have been identified at the site: 1) UJF1 represents the basal midden layers which contains material dating from the initial Viking landnám in the Faroes, as well as later material from 11th and 12th centuries cal AD that were mixed in with these basal layers (14C dated to 880-1,220 cal AD); 2) two later phases (UJF2 & UJF3) were 14C dated to 1,010-1,220 cal AD. Unusually in North Atlantic zooarchaeological assemblages, pig bones were recovered in significant numbers from all 3 phases, and the pigs were kept both on the settlement and the wider outfield (299); (300). All 37 (AA411, AA412, AA413, AA414, AA415, AA416, AA417, AA418, AA432, AA433, AA434, AA435, AA436, AA437, AA438, AA439, AA440, AA441, AA442, AA443, AA444, AA445, AA446, AA447, AA448, AA449, AA450, AA451, SF010, SF011, SF012, SF013, SF014, SF015, SF016, SF017, SF028) specimens analyzed are of domestic pigs.

Untervaz-Haselbodensenke, Switzerland
Contact: Jörg Schibler
Untervaz-Haselbodensenke is a Neolithic dryland site that was excavated 2001 by the archaeological service of canton Grison (301). Three different Neolithic settlement layers were found. The sample bone belongs to the youngest settlement that is dated typological (by ceramic pot shred) to the Corded Ware Culture (4,700-4,500 BP). The site is situated near the border between the canton Grison and St. Gallen in the region of the upper Rhine valley. The animal bones of the site were identified by S. Braschler. Only 167 bone fragment out of a total of 974 could be identified because the bones were highly fragmented. Four specimens (AA075, AA076, AA077, ASPig15), both wild and domestic, were analyzed.

Vărăşti, Romania
Contact: Adrian Bălășescu
Vărăşti is located on the small island of Grădiştea Ulmilor (nowadays disappeared) on Lake Boian, Vărăşti commune, Călărași County. The faunal remains come from a Boian culture context, the Vidra phase (6,700-6,500 BP), named the settlement Boian A by the archaeologist Eugen Comşa (302). The palaeofaunistic remains, not very numerous (260 determined), come from a closed complex (303). One *Sus* species specimen dating to the Middle/Late Neolithic - Eneolithic (AA068) was analyzed in this study.

Verga 3 & Verga 5 (Neto-via Verga), Italy
Contact: Domenico Lo Vetro / Lucia Sarti

Verga 3 and Verga 5 are part of the Neto-Via Verga area within the greater Spazzavento–Neto–Via Leopardi Settlement Unit in the Sesto Fiorentino area of Italy (260). This is a multiphase Neolithic and Copper Age settlement, and is part of a broad settlement unit formed by excavation areas: Neto-via Verga, Neto di Bolasse and Spazzavento. The unit, positioned in the west of Sesto Fiorentino territory, is located on a sort of morphological top formed by depositional accumulations of two alluvial fans. The stratigraphic sequence as a whole covers a wide chronological range which indicates that the area was occupied from the beginning of the 5th millennium until the Roman period. In particular, Via Verga is a multistratified spread in 5 excavation areas: Roman, Bronze Age, Bell Beaker, Copper Age, Neolithic phases are separated by alluvial layers.

The samples are related to area 3 and Area 5 which are referred to Neolithic phase (orizzonte 7) at the end of 5th millennium AD (304); (262); (305). 3 Neolithic wild boar samples (AA638, AA639, AA640) could not be used in the analysis because the haplogroup tests failed.

Via Neruda, Italy
Contact: Domenico Lo Vetro, Lucia Sarti
The archaeological site of Via Neruda is located in Sesto Fiorentino, Italy. The site dates to the Multiphase Copper and Bronze Age (306). The site, positioned in the East of Sesto Fiorentino territory, is located in the upper part of the alluvial plain close to the foothill side of Monte Morello. Two different layers (layers F and G) testify dwelling activities within an abandoned palaeoriver bed occurred during Early Bronze Age. The samples are related to layers F and G (262). The 6 wild boar samples (AA623, AA624, AA625, AA626, AA627, AA628) could not be used in the analysis because the haplogroup tests failed.

Vinča Belo Brdo, Serbia
Contact: Vesna Dimitrijević
Vinča – Belo Brdo (44º 77’ N, 20º 62’ E) is situated on the right bank of the Danube, 14 km southeast of Belgrade, Serbia. It is the eponymous site for the Late Neolithic/Eneolithic Vinča culture group. Excavations started in 1908, continued sporadically until 1936, followed up in 1978–84, resumed in 1998 (307), and are currently ongoing. Sequence of cultural levels, more than 9 m deep, is mostly built up of Vinča culture deposits. At the lowermost level, a Middle Neolithic Starčevo collective burial is found, while at the top of the sequence, there are Late Eneolithic and Bronze Age deposits, as well as the late medieval necropolis. Vinča Belo Brdo is a *tell*-type site, embracing multiple horizons of large settlements composed of rectangular wattle and daub houses. Material culture is best known for elaborated pottery and anthropomorphic figurines. In the sample of animal remains from the upper levels of the Vinča culture sequence (excavation campaigns 1998-2003), pig is, after cattle, the second best represented domestic species, while wild pig is also frequent among hunted animals (308). Forty pig specimens (AA156, AA157, AA158, AA159, AA160, AA168, AA169, AA170, AA171, AA172, AA173, AA174, AA175, AA176, AA177, AA178, AA179, AA420, AA421, AA422, AA423, AA424, AA425, AA426, AA427, AA428, AA429, AA430, AA431, BLT013, BLT014, BLT021, LG058, LG180, LG181, LG183, LG184, LG669, LG679, LG681) were analyzed.

Vitănești-Măgurice, Romania
Contact: Adrian Bălășescu
Vitănești-Măgurice is *tell* settlement located in a swampy area on the Teleorman valley, 7 km north of Alexandria (Teleorman County). The 5.5 m thick occupational layers were attributed to the Eneolithic Gumelniţa culture (phases A1, A2 and B1, (309). The two wild boar samples (LG423, LG426) come from the Gumelniţa A2 level (ca. 6,449-6,162 cal BP) (310); (311).

Vlasac, Serbia
Contact: Vesna Dimitrijević, Dušan Borić
Vlasac (44.53368° N, 22.05032° E) is a large open-air settlement in the upper gorge of the Iron Gates region, on the right (Serbian) bank of the Danube. The site was investigated in 1970-1 by D. Srejović and Z. Letica who examined an area of 640 m^2 along the river bank below 70 m a.s.l., which was subsequently submerged beneath the reservoir created by the Iron Gates I dam (312). Further excavations were undertaken between 2006-9 by Dušan Borić, who examined a further 326 m^2 upslope of the area excavated in 1970-1 (313). The site was occupied at various times during the Mesolithic and early Neolithic, though the majority of the 14C dates fall in the later part of the time range between ~9,200 and 7,800 cal BP. Of the 5 samples (BLT024, BLT025, BLT026, BLT0247, LG263) from Vlasac included in the present study, three (BLT025, BLT026, LG263) have been AMS 14C dated to between ~9,150 and 8,700 cal BP (170), corresponding to the early part of the Late Mesolithic (80) (81).

Voldtofte, Denmark
Contact: Kim Aaris-Sørensen, Peter Rowley-Conwy
Voldtofte is a large Early Bronze Age settlement in southwestern Fyn, located approximately halfway between Assens and Harby on West Funen. It is adjacent to "Lusehøj", one of Denmark's largest Late Bronze Age burial mounds. Excavations that took place between 1976 and 1977 yielded a wealth of material. Some possible house sites were found, in addition to nearly 100 pits containing thousands of potsherds and flints. The cereal grain assemblage was dominated by cultivated barley. Out of 3,397 bone fragments collected, 1,359 could be identified to the species. 82% of the material were of cattle (*Bos taurus domesticus*), 11% pig (*Sus scrofa*), 5% sheep (*Ovis aries*), 2% horse (*Equus caballus*), 0.2% dog (*Canis familiaris*), and 0.1% red deer (*Cervus elaphus*). Immediately striking is the almost total lack of wild game as well as the two predominant groups of cattle and pig (314). One domestic specimen (AA013) was analyzed.

Wroxeter, England
Contact: Andy Hammon
Wroxeter, *Viroconium Cornoviorum*, developed out of a civil settlement adjacent to a 1st century AD legionary fortress between about 90-150 AD to become the tribal capital of the Cornovii. It grew from the 2nd century AD to be the fourth largest town in Roman Britain, with a defined area of 78 ha. Following periods of growth and relative decline in the 2nd, 3rd and 4th centuries AD, an unusual resurgence of activity took place after AD 400. The life of the later "sub-Roman" town may have continued to about AD 650; with a phase of building in timber after the mid-6th century AD. It has been conjectured that Wroxeter had become the centre of a British, or possibly Irish, leader (315); (316). Two wild boar specimens (GL528, GL804) were analyzed.

Zambujal, Portugal
Contact: Simon Davis
Zambujal is located in the Estremadura region of central Portugal. Like most Portuguese Chalcolithic sites they are fortified settlements, dating to around 4,600-3,800 BP. The occupation at Zambujal spanned the whole Chalcolithic period, and pig is the most common taxon represented at this site (317). Five Chalcolithic specimens from this site were analyzed: GL333, GL454, GL465, GL805, and GL840.

Żegotki 2, Poland
Contact: Arkadiusz Marciniak
The site located in Kujavia, within the Polish lowlands. The site was discovered in the 1990s during the rescue excavation project preceding the construction of the Yamal pipeline from Siberia to western Europe. The major discoveries at this partially excavated site comprised the TRB and Early Medieval settlements. One TRB domestic specimen (AA134) provided successful amplification of their DNA (318). Specimens AA135 and LG497, also domestic pigs, were also analyzed.

Zelgno, Poland
Contact: Arkadiusz Marciniak
This Late Lengyel site is in Chelmno Land within the Polish lowlands. It was excavated in 2000 and 2001 as a part of large rescue excavations project preceding the construction the A1 motorway. The unearthed elements of the settlement comprise 10 trapezoidal houses accompanied by numerous pit of different character, in addition to one inhumation burial (46). Two LBK domestic pig samples were analyzed (LG730, LG732).

Zürich-Mozartstrasse and Zürich-Seefeld, Switzerland
Contact: Jörg Schibler
Zürich-Mozartstrasse and Zürich-Seefeld are two lakeshore sites at the outflow of lake Zürich in Switzerland. Today both sites are situated within the city of Zürich. In both sites a stratigraphy of several settlement layers is found. All settlements are dated by dendrochronology between the beginning of the 6th millennium BP and the beginning of the 4th millennium BP. Only two settlement layers are dated to the early Bronze Age; all other layers (Seedeld: 15 layers; Mozartstrasse: 12 layers) are Neolithic. Archaeozoological results of the animal bone from all settlements are published in (319). New interpretation of the archaeological features and the dendrochronology are published by (320). According to (320), all settlement layers 4 are dated to 3,612-3,595 BC, all settlement layers 3 to 3,126-3,098 BC and all settlement layers 2 to 2,625 -2,568 BC. The bones were provided by Kantonsarchäologie Zürich for this study.

Fourteen specimens were analyzed from Zürich-Mozartstrasse: AA200, AA201, AA202, AA240, AA244, AA246, AA247, AA248, AA249, AA250, AA251, AA252, AA253, AA254; and one specimen (AA197) was analyzed from Zürich-Seefeld.

Ancient DNA

Samples spanning a wide range of time and location, of both wild and domestic individuals, were sourced for this project (Table S1). We produced data using Sanger

sequencing (MC1R and control region of mtDNA) and with NGS in various lab, including Durham (UK), Mainz (Germany), Oxford (UK) and Stockholm (Sweden

Data generation - mtDNA - Durham

(LG/AL, SF001, SF002, SF003, SF004, SF005, SF006, SF007, SF008, SF009, SF010, SF011, SF012, SF013, SF014, SF015, SF016, SF017, SF018, SF019, SF020, SF021, SF022, SF024, SF025, SF026, SF027, SF028, SI003, SI004, SI006, SI007, SI008, SI009, SI010, SI011, SI012, SI013, SI014, SI015, SI016, SI017, SI018, SI019 in Table S1)

Samples were extracted in a dedicated ancient DNA facility at Durham University using the protocol described in (161). In brief, ll equipment and work surfaces were cleaned before and after each use with a dilute solution of bleach (5-10%) followed by water and ethanol (99%). Pipettes and plastic racks were UV-irradiated in a cross-linker (254 nm wavelength) prior to and after use. Pre- and post-PCR laboratories are physically isolated and access to the pre-PCR laboratories is restricted to Ancient DNA lab users only. Ancient DNA lab users wear clean lab coats, double set of gloves (nitrile and latex) and over-shoes in order to avoid introducing contamination from post-PCR areas.

The ancient pig remains were prepared for DNA extraction by removing an approximately two millimeter layer of the outer bone surface by abrasion using a Dremel drill with clean cut-off wheels (Dremel no 409), targeting compact cortical bone or dental dentine. The bone was then pulverized in a Micro-dismembrator (Sartorious-Stedim Biotech), followed by collection in 15ml Grainer tubes. Milling containers and grinding balls were subsequently suspended and cleaned in 1% virkon and rinsed in absolute ethanol.

Bone powder (100-400mg) was digested in 0.425 M EDTA, 0.05% SDS, 0.05 M Tris-HCI and 0.333 mg/mL proteinase K and incubated overnight (18-24 hours) on a rotator at 50°C, or until fully dissolved. The digestion buffer, excluding proteinase K, was UV-irradiated (254 nm wavelength) for an hour in a dedicated cross-linker prior to use. 2mL of extract solution was then concentrated in a Millipore Amicon Ultra-4 30 KDa MWCO (Millipore) to a final volume of 100µL. The concentrated extract was purified using silica spin-columns (QIAquick PCR Purification Kit, Qiagen) following manufacturers recommendations, except that the final elution step was performed twice to produce a final volume of 100µL. One in five or one in ten negative
extraction controls were performed alongside the ancient bone samples. All extraction blank controls were negative when PCR screened for the ANC1 fragment.

PCRs were setup in 25µL reactions using 1.25U Taq GOLD (Applied Biosystems), 1x Gold buffer (Applied Biosystems), 2.5mM MgCl2 , 0.5µg/µL BSA (Bovine Serum Albumine), 200µM of each dNTP, 0.8µM of each forward and reverse primers, and 2µL of aDNA extract. We used PCR primers ANC1 (321). One PCR negative control was included for every 5-8 aDNA template PCRs. PCR cycling conditions were 95°C for 5min, 50 cycles of 94°C for 45 sec, 54°C for 45 sec and 72°C for 45 sec, followed by 72°C for 10 min. PCR products were stored at -20°C.

An initial PCR using the ANC1 primers was performed in order to screen the extracts for preserved DNA. Successful amplifications were Sanger sequenced on the Applied Biosystems 3730 DNA Analyser at the DNA sequencing service in the School of Biological and Biomedical Sciences at Durham University. Once preserved samples were identified we used 5bp 5'-tagged PCR primers (322) to re-amplify the ANC1 fragment. In both instances PCR products were visualized on agarose gel and stained with GelRed (Biotium), and then pooled by eye into approximately equimolar concentrations using a reference series of PCR products previously quantified on the Qubit fluorometer; approximately 12µg/µL of each PCR product was used for the final pool. The pooled 5' tagged PCR products were then concentrated using an Amicon Ultra-4 30KDa MWCO filter column to a final volume of 100µL. The concentrated amplicon pool was subsequently purified using the QIAquick PCR Purification Kit following manufacturers recommendations, except that the final elute volume was 80µL.

Next generation sequencing of LG samples

The concentrated PCR amplicon pool was then built into a paired-end library (Paired-End DNA Sample Prep Kit, Illumina) following manufacturers guidelines and subsequently sequenced on the Illumina GAII platform at the Department of Biology at Copenhagen University. Illumina's Genome Analyzer Sequencing Control Software (SCS) v2.4 was used for base calling. A custom written PERL script (Rasmussen, M., University of Copenhagen) was used to filter out sequences containing the 5' tag label and to mate paired-end reads into single lines containing both
forward and reverse 5' tag label information. A second custom written PERL script was used to write a single fasta file for each tag label/amplicon. The resulting fasta files were assembled into contigs against a reference sequence (EU333163) in Geneious Pro 5.4.3. Assembly was performed using total quality score to call the best base (any base with a quality <20, equivalent to PHRED scores, was called as N and subsequently excluded from further analysis). At least one hundred sub-clones per sample were obtained for each re-sequenced PCR product, although we reached an average of several thousand copies per PCR amplicon. Nucleotide positions that could not be resolved despite the deep coverage were discarded from further analysis and called according the IUPAC nucleotide code.

Sanger sequencing

Sequencing for all other samples was carried out at the University of Oxford, Department of Zoology, Sequencing Facility on an ABI 3100.

Data generation - MC1R/mtDNA

Mainz (AP23, AP24, AP26, AP27, AP28, AP29, AP31, AP32, AP37, Aru24, Aru25, Aru10, Aru11, Aru12, Aru13, Aru7, Cav1, Cav2, Cav3, Cav4, Cav5, Cav6, Dra14, Dra15, Dra18, Dra20, Dra31, Dra32, Dra33, Dra38, Hag1, Hag2, Hag3, Kan10, Kan11, Kan12, Kan13, Kan14, Kan15, Kan16, Kan18, Kan8, Kan9, Kop8, Kop9, Kov21, Kov14, Kov16, Kov18, Kov19, Mal9, Mal12, Men4, Men5, Oko13, Oko15, Oko16, Oko1, Oko3, Ovc1, Pie13, Pie14, Pie15, Pie16, Pie1, Pie3, Pie4, Pie5, Sam10, Sam18, Sam5, Sam9, San7, San9, San1, San3, San8, Sof1, Sof2, Sof3, Tac14, Tac15, Tac16, Tac17, Tac10, Tac13, Tac6, Tac8, Uiv10, Uiv11, Uiv9, Ulu1, Ulu20, Ulu24, Ulu27, Ulu48, Ulu49 in Table S1)

Samples were processed in dedicated ancient DNA facilities at Johannes Gutenberg-University Mainz under strict rules for contamination prevention and using methods amplification and sequencing methods already described in (161). 510 bp of the mitochondrial HVSI were amplified using six overlapping primer pairs:
U15516_5'CTATCTTTAAAACAAAAAAACCCATAA3' and
L15620_5'TACATGCTTATATGCATGGGGACT3',
U15613_5'ARCCCTATGTACGTCGTGCATTA3' and
L15704_5'GCATGTTGACTGGARTTATTTGGAC3',
U15697_5'CATATYATTATTGATCGTACATAGCACA3' and
L15787_5'AAGAGGGATCCCTGCCAAG3',
U15775_5'AAYTACCATGCCGCGTGAAA3' and
L15864_5'GTTCTTACTTCAGGACCATCTCACC3',
U15852_5'TGGGGGTTTCTATTGATGAACTTTA3' and
L15941_5'TATGTGTGAGCATGGGCTGATTA3', and
U15932_5'CCCTTAAATAAGACATCTCGATGGA3' and
L16027_5'TTTGACTGTGTTAGGGCCTTTGA3'.

Primers were hybridised at 57°C. Codon 22 and 124 of the MC1R locus were amplified using the primer pairs MC1R_U44 5'AGGCTGCTGGCTTCCCTCAG3' and MC1R_L87 5'CCCGTCTGGTTGGTCTGGTT3' and MC1R_U234 5'TGGTGCAGCAGCTGGACAA3' and MC1R_L246 5'TGGAGCCGCAGATGAGCA3', respectively. MC1R_U234 was prolonged by an artificial tail (MC1R_U234_tail 5'AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATGGTGCAGCAGCTGGACAA3') (Binladen et al. 2007) to allow capillary sequencing of the only 11 bp long target sequence. U41_tail 5'CCACGTCGTGAAAGTCTGACAAT3' was used as a matching sequencing primer. Primers were hybridised at 57°C. Betaine solution 5M (Sigma Aldrich) was added to each PCR at a final concentration of 1M.

Data Authenticity of MC1R PCR data

Two independent DNA extractions were performed per sample. Both extractions had to yield a minimum of 1 successful PCR amplifications, with a minimum number of successful PCRs per sample of 3. A marker was only accepted as homozygote if all amplifications showed a homozygote signal. A marker was only accepted as heterozygote if the heterozygote signal was present at least once in each of the extractions.
MC1R results from 3 samples analysed for the MC1R loci were integrated to the dataset even though the limited amount of available DNA extract did not allow to fulfil the defined stringent authenticity criteria above: Only two sequences retrieved from the same extract of two pig samples (Cav5 and Cav6) are homozygote for not having the CC insertion at codon 22. One pig sample (Kan8) is homozygote in 1 amplification from each of the 2 extractions for having the CC insertion at codon 22. We integrated those sample to our dataset since i) extraction and PCR blank control show no evidence for contamination in any of our experiments, and ii) the data comes

from samples that yielded well-preserved and fully replicated mitochondrial DNA sequences without any evidence for in-sample or cross-contamination.

Data generation - Nuclear and mtDNA capture - Oxford

(all AL and AA samples in Table S1)

DNA extraction

DNA extractions were performed in dedicated ancient DNA (aDNA) facilities at the PalaeoBARN (University of Oxford). DNA was extracted from pig teeth or bone samples (Table S1) in a dedicated ancient DNA laboratory using the appropriate sterile techniques and equipment. Extraction was carried out using the Dabney extraction protocol (323) but with a 30 minutes pre-digest stage (324).

Library Preparation & Sequencing

Illumina libraries were built following (325), but with the addition of a six base-pair barcode added to the IS1_adapter.P5 and IS3_adapter.P5+P7 adapter pair. The libraries were then amplified on an Applied Biosystems StepOnePlus Real-Time PCR system to check that library building was successful, and to determine the minimum number of cycles to use during the indexing amplification PCR reaction. A six base-pair barcode was used during the indexing amplification reaction resulting in each library being double-barcoded with an "internal adapter" directing adjacent to the ancient DNA strand and which would be the first bases sequenced, and a traditional external barcode that would be sequenced during Illumina barcode sequencing.

Six libraries (KD033 KD037, KD025, VEM185, AA119, AA133) were additionally partially USER treated for three hours (326) before whole genome sequencing (see Table S1)

Up to 200 libraries with unique barcode combinations were pooled at equimolar levels (as determined by an Agilent Technologies 2200 TapeStation) and an 80bp run was carried out on an Illumina 2500 sequencer. Selected sample extracts were then built into libraries with a single barcode, and then sequenced individually at greater depth.We selected 12 high quality libraries for deep sequencing (Table S1). Each library was sequenced separately on multiple Hi-Seq 2500/4000 lanes (Table S1). The resulting reads were aligned to the reference genome as detailed above.

mtDNA Capture

Library pools containing up to 200 libraries were captured using MYcroarray mitochondrial MYbaits with 24 hours hybridisation at 56C, and following the MYbaits manual V3.

Data generation - Nuclear - Stockholm

(all BLT samples in Table S1)

DNA extraction

DNA extractions were performed in dedicated ancient DNA (aDNA) facilities at the Archaeological Research Laboratory, University of Stockholm, Sweden. Approximately 100-200 mg bone powder was excised per sample using a dremel drill at the lowest possible rpm. The work surface (a dead-air, UV-enabled cabinet) was thoroughly cleaned after sampling each specimen. DNA extractions were carried out

in batches of eight including one extraction blank. The bone powder was first incubated (washed) three times for 15 min under constant rotation in 500 µL 1M sodium phosphate buffer, pH 6, in order to minimise the amount of microbial contaminant DNA and to enrich for endogenous DNA(327). Buffers were replaced after each wash step and pellets were washed with 1 mL water after the final wash step. DNA was extracted following protocols published previously(328). In summary, bone powder was mixed with lysis buffer containing 0.45M EDTA (pH 8), 1M Urea, and 100 µg proteinase K, and incubated with constant agitation for at least 18 h at 37 °C. DNA was concentrated on Amicon Ultra-4 columns (Merck Millipore) and purified on MinElute silica columns (Qiagen) following manufacturer recommendations but with the addition of a second wash step. Purified DNA was eluted in 65 µL EB buffer (Qiagen).

Library Preparation & Sequencing

Double-stranded sequencing libraries were prepared from 20 µL extracted DNA following (325) but with the modification that we added 3U USER enzyme (NEB) and increased the incubation time to 3hrs at 37°C for the blunt-end repair step. The reason for doing so was to remove deaminated cytosine bases (329). The libraries were subsequently quantified using qPCR, PCR amplified, pooled at equimolar concentrations, and purified as described previously in (104). Library pools were shotgun-sequenced on the Illumina HiSeq-2500 platform at the SciLife DNA sequencing facility, Stockholm.

Data Processing

Sequence reads were aligned using Burrows-Wheeler Algorithm (BWA) version 0.7.5a-r405 (330) to Ssc.10.2, with default parameters apart from disabling the seed option ("-l 1024") (331). Samtools version 0.1.19-96b5f2294a (332) was then utilised to remove duplicates. BAM files from different sequencing lanes were merged using the MergeSamFiles tool from Picard v1.129 (http://broadinstitute.github.io/picard/), FilterUniqueSAMCons (333) was then used to remove duplicates. Depth of coverage was computed using Bedtools (334). We report alignment stats of ancient samples (including those that were only screened) in Table S1.

Sexing

We used a read depth based method based on (335), comparing alignment of sequencing reads post filtering for a mapping quality q30 to the X chromosome and chromosome 1.

MapDamage

Molecular damage was assessed using MapDamage2.0 using default parameters (336) (Figure S1). Most samples display clear signs of deamination (Figure S1), except for those that were USER treated (all BLT samples, KD033 KD037, KD025, VEM185, AA119, AA133).

Publicly available sequences

We obtained genome-wide genotype information for 149 high quality modern genomes from (337, 338) Table S2). These include 23 wild boars from Europe and the Near East: France (1), Netherlands (12), Switzerland (1), Italy (5) and Greece (4) as well as 3 from the Near East: Armenia (1) and the island of Samos (2; Near Eastern wild boars (339)), 10 wild boars from East Asia, as well as as 85 European domestic pigs that represent a broad sampling of non-commercial pigs, and commercial pigs and 24 East Asian domestic breeds. We used the genome of a *Sus verrucosus* as an outgroup (340).

High coverage ancient genomes

We used samtools 'mpileup' (0.1.19; (332) to call SNPs with default settings. Pileup files were further filtered, for each sample, using the following criteria:

- Strict minimum DoC >= 3
- Excluded all sites in region of high DoC (top 5%; taking into account sex for X chromosome genotyping)
- Excluded all sites within 3bp of an indel
- Minimum Phred >= 20
- Minimum fraction of reads supporting heterozygous >= 0.3 (value set to 0 if male for chromosome X genotyping).
- We also discarded the first and last 8bp of each read to avoid incorporating errors from deaminated sites

Ascertainment and ancient genome quality

We used only high quality ancient genome: AA288, AA363, AA133, AL718, KD025, AA119, KD033, AA451, AA349, VEM185, KD037 together with the modern genomes for SNPs ascertainment (to be called in low coverage ancient data for PCA projection, see below). In order to assess the quality of our genotyping and decide which ancient genome to include in the ascertainment we computed transversion/transition ratio at homozygous and heterozygous non reference SNPs (Figure S2). We expect this ratio to approach 0.5 for both heterozygous and homozygous. Lower genome coverage, however, clearly showed a much lower ratio for heterozygous sites suggestive that more damage were incorporated in their genotype (damage most likely incorporated as heterozygous calls). In order to limit the incorporation of deamination in our ascertainment we only used the genome of: AA119, AA133, AA349, KD025, KD033, KD037, and VEM185 (Table S1; Figure S2).

Genotypes called from these genomes were merged together with those of modern genomes using BedTools (334). Ascertainment was done without the outgroup, all non bi-allelic markers were also removed as well as markers with more than 20% missing data. This resulted in 12,522,931 SNPs on autosomal chromosomes and 213,104 SNPs on X the chromosome.

Low coverage data

We also used 54 ancient samples (Table S1) in our PCA (see below). Genotypes in these samples were called by randomly sampling a single read of 30 base pair minimum and with a mapping quality (MAQ) and base quality (BQ) of at

least 30 at each covered position in the genome, excluding bases within 5bp of the start and end of a read.

mtDNA analyses

Filtering

For the mtDNA we generated majority consensus (using reads with BQ>=20 and MAPQ>=30) sequence for all samples that had at least 2x average coverage (111 samples; Table S1) excluding bases within 5bp of the start and end of a read.

Building a database for haplogroup assignment

Our goal here is to build a robust method to assigned haplogroup to low coverage mtDNA data (Table S1) using phylogenetics. To do so we first build a data-base of high quality modern and ancient mtDNA.

We obtained the full mitogenome of 77 pigs previously published in (337, 338) (Figure S3) and combined this data with 111 newly sequenced mitogenomes (including all pigs with coverage greater or equal than 2x; Table S1). Altogether these data set contain 188 full genomes from European and Asian wild and domestic pigs.

Phylogenetic analysis and reference sequences

We used RAxML (version 8)(341) to build a maximum likelihood tree, with 100 bootstrap replicates under a GTR-GAMMA model. The resulting tree is presented in Figure S3. In the tree we find a high support (bootstrap value >=90 in all case but root of Y1 which is >70) for the 6 major mtDNA haplgroups in *S. scrofa*, namely European (Eu), Y1, Y2, ArmT, Italian (It) and Asian (As; Larson *et al.* 2005, 2007; Ottoni *et al.* 2012) (Figure S3). We then selected two high quality genome sequences as reference for each haplogroup:

Haplogroup As - found primarily in East Asian pigs (but also in Iran; Figure 1a). We selected WB30U09 and WB29U14, two high quality sequences from Chinese wild boars (Bosse *et al.* 2014; Frantz *et al.* 2015) as reference sequences.

Haplogroup Eu (including ANC-Aside and ANC-Cside) - found primarily in European populations, this haplogroup is found in the vast majority of ancient and modern European wild boars as well as a majority of European breeds (Figure 1a; (321, 342)). We selected WB21F03 and WB25U11, two high quality sequences from Dutch and French wild boars respectively (337) as reference sequences.

Haplogroup It – found primarily in Italian and northern balkan wild boars (Figure 1a). We selected WB42M09 (wild boar from Italy) and AA457 an ancient wild boar from Croatia that was previously assigned to this haplogroup based on its control region (GL979 in (321).

Haplogroup Y1 - found primarily in Near Eastern (e.g. Turkey) wild boars and ancient European domestic pigs (321, 342). We selected WB33U05, a wild boar from the Island of Samos (337) as the region is known for its high frequency of Y1 haplogroup (339).We also selected the sequence of Turkish wild boar (AL718;

coverage >220x) as this sample was previously assigned to Y1 based on its control region (321).

Haplogroup Y2 - Found in ancient wild boars from the south of Europe (161, 321). In Figure S3 this clade (in purple) can be identified by the sequence BLT025 (control region published as LG260 in (161), from the site of Vlasaac (Serbia). We selected two high coverage reference genome for this clade, AA479 (>228x) and AA480 (>68x) two ancient Ukrainian wild boars (Figure S3).

Haplogroup ArmT (including Arm2T and Arm1T)- Found mainly in wild and domestic pigs from the Near East (e.g. Armenia, Iran, Turkey) as well as in ancient European domestics (161, 321). In Figure S3 this clade (in green) can be identified by the sequence AA364 (control region published as LG419 in (161), from the site of Kohneh Tepesi (Iran) and AA119 (control region published as GL896 in (321) from the site of Lchashen in Armenia (Figure S3) . We selected two high coverage reference genome for this clade, AA363 (same site as AA364; >280x) and AL1107 (>29x) from the site of Cadir Hoyuk in Turkey (Figure S3; Table S1).

Haplogroup NE1 - Found uniquely in modern wild boars from Iran, Iraq, Southern Levant and South Caucasus (161). This haplogroup was never found in modern or ancient domestic pigs from either Near East or Europe (161). The sole full mtDNA genome available representative of NE1 was GL943 (pink), obtained from an egyptian wild boar (Figure S3; Table S1).

A method for rapid haplogroup assignment

We used the sequences described above to rapidly assigned haplogroup to all other mtDNA sequences (partial or full). To do so, we first build a maximum likelihood phylogenetic tree (using RAxML) based on the references sequences (see above) and a query sequence (the sequence we want to assign to a haplogroup). We then compute the pairwise distance between our query and each reference haplogroup sequences in the phylogeny using Dendropy (version 4) (343). We then assign the haplogroup of the reference sequence that was the closest to our query sequence.

We tested the performance of our simple method using 182 sequences with known haplogroup (based on phylogenetic analysis; see above; NE1 was excluded given we only had one sequence). We recovered the right haplogroup for every sequences, demonstrating that our method provide a robust assignment for high quality data. We then tested how missing data affected our success rate. To do so, we randomly selected 250, 400, 800, 1600 and 3200bp (not necessarily consecutive) from each 182 sequences with known haplogroup. We repeated this 100 times per length / genome combination and assigned a haplogroup to these sequences with various degree of missing data. Our method recovered the right at rate of 83.8%, 92.7%, 98.2%, 99.7%, and 99.8% for sequences of 250, 400, 800, 1600 and 3200bp respectively. Our recovery rate per clade was also high (>90% for each clade based on 800bp; Figure S4). Based on this analysis we decided to use 800bp as a minimum length for haplogroup assignment for our low coverage data.

We further tested this method on empirical data. To do so, we compared mtDNA haplogroup for 116 samples as inferred using PCR (Table S1) to our NGS based

assignment. Our method retrieved the right haplogroup for 107 out of 116 samples for which we had both PCR and NGS based haplogroup assignment (>92% success rate; Table S1). Lastly, we checked for mismatch between PCR and NGS sequence data in sample with >1x coverage genome wide (29 samples in total). Although one sample low coverage sample (~1%) had up to 4% mismatch between NGS and PCR these were mostly transition and transversion rate of mismatch were below 1% in all samples (Fig. S5).

Low coverage data

Many sequences did not pass the stringent criteria of 10x depth of coverage (Table S1). Nevertheless, as we shown above, only few reads should provide enough resolution for a robust haplogroup assignment. We thus, lowered our requirement from 10 to 1 reads with MAQP>=30 and BQ>=30 and created a sequence consensus for all samples that did not pass our filtering criteria detailed above. All samples that yielded at least 800bp of coverage, based on this filtering, were assigned a haplogroup using the approach described above.

mtDNA lineage geographic distribution (Figure 1)

We grouped wild boars haplogroups, from our study (Table S1) and from publically available data (Table S1) into four meta groups, including Europe (mt-Eu, and mt-Italy), Y2, Near East (mt-ArmT and mt-Y1) and Asian (As) as previous studies demonstrated that Eu and Italian haplogroups were mostly restricted to European wild boars while mt-ArmT and mt-Y1 were restricted to wild boars found in the Near East (161, 344) (also see Figure S6-7). Aside from a few examples of wild boars carrying mt-Y2 (6/99) haplogroup in the Near East we found that most Near Eastern wild boar carried either mt-Y1 (54/99) or mt-ArmT (39/99) (Figure S6-7).

To further test this phylogeographic signal, haplogroup membership was then transposed to create an ancestry matrix as in (345). This matrix was then plotted onto a map with a tessellated projection, using the R package "tess3r" (346–348). This projection demonstrate a highly pronounced population structuring between Near East and European wild boars (Figure 1a). Following haplogroup assignment of both low and high coverage data, as well as from PCR results we plotted meta group assignment of domestic pigs on the wild boar projection (Figure 1b,c,d).

Samples excluded from Figure 1

EU haplotype in wild boar found in the Near East

ARD3 and ARD4 (349) are modern wild boar with mt-Eu haplotype (ANC-Aside) found in Iran. GL141 is from Armenia; GL753 is from Iran; GL766 and GL779 are from Iraq yet possesses a EU haplotype. This sample was excluded from analysis previously (321) (although GL141 was reported as haplotype A) and was justified as: "*the turnover of domestic pigs in Armenia from domestic pigs of Near Eastern origin to those of European origin before the end of the Iron Age in Romania suggests that pigs domesticated in Europe were transported east into the Near East before or during the Iron Age. This eastward movement of European domestic pigs increases the likelihood of discovering European signatures in samples that superficially appear to be modern wild boar from the Near East, but are in fact feral descendants of domesticates introduced from Europe during or after the Iron Age.*"

GL935 is a wild boar that was found in the north slope of the West Caucasus, and GL944 is a wild boar from Georgia. Because both samples have mt-Eu haplotypes and are from the Near East, we have excluded them from this analysis for the same reason as explained previously (321). AA315 and AA316 are wild boars from Gobekli Tepe, Turkey. The PCR worked failed and there was low NGS coverage (0.98x). Our NGS analysis showed it as mt-Eu, but phylogenetic reconstruction showed that the haplotype is even more basal than mt-Eu, suggesting a potential new haplogroup. We were unable to definitively classify the samples. Also the age of the sample is unknown insecure/direct dating is necessary.

AA308 is a wild boar from Bademagaci, Turkey, dating to 7,000-6,000 BC. The PCR work failed and there was low NGS coverage (0.48x). The NGS analysis showed it as EU, but phylogenetic reconstruction showed that the haplotype call is even more basal than EU, suggesting a potential new haplogroup. We were unable to definitively classify the sample. AA399 is a wild boar from Suberde Hoyuk, Turkey, dating to 7,500-7,000 BC. The PCR work failed and there was low NGS coverage (0.24x). The NGS work showed it as mt-Eu, but phylogenetic reconstruction showed that the haplotype call is even more basal than EU, suggesting a potential new haplogroup. We were unable to definitively classify the sample. Tp1 and WBTR_514, two wild boar samples from Turkey, also have mt-Eu (specifically ANC-Aside) haplotype (161).

Y1 haplotype in wild boar found in Europe

AA41 was found in Siniarzewo 1 site, Poland, dating from 4,100-4,400 BC. AA379 was found in Buran-Kaya 4 site in Ukraine, dating from 7613-7484 cal. BP (OxA-24692). LG610 was found in Bruschal, Germany, dating from 4,200-3,400 BC. . LG829 is also a wild boar found at the same site, except the sample is older, dating to 5,200-4,500 BC. Oko16 was found in Okoliste, Bosnia-Herzegovina, dating to 5,100-5,000 BC.

AA466 has an unclear status. It is assumed to be from a Mesolithic context, although some Neolithic context and domestic pigs were also found at the same site, Bern-Birsmatte, Switzerland. LG595 was found in Klingenberg, Germany, dating to Late Neolithic contexts, 4,000-3,650 BC. AA159 was found in Vinča Belo Brdo, Serbia, dating to Neolithic contexts, Mid 5th Millennium BC Uiv11 was found in Uivar, Romania, dating to the Late-early Neolithic, 5520 ± 26 calBC (KIA-42156: 6560 ± 35 BP). AA497 was found in France, with French middle Neolithic (Northern Chasseen culture) contexts, dating to the early 4th millennium BC. All of these wild boars are found at sites on which domestic pigs were also found - as previous studies suggested, this is likely the result of domestic to wild admixture (350, 351).

Asian haplotype found in wild boar in Europe

DQ379240 (352) was excluded as it is an Asian haplotype in Belgian wild boar. Kbl020 (accession JF774408; (339)), a Greek wild boar sample was excluded as although the paper gave the sample an Asiatic haplotype call, 2 unique mutations suggests that it is an unknown haplogroup.

Nuclear DNA analyses

MC1R and domestic/wild status

We tested the usability of a non-synonymous mutation (D124N that is the result of a G->A mutation at the second codon position) in the *MC1R* gene (leading to black coat colour and loss of camouflage coat colour (353)) as a diagnostic of wild/domestic status. We first computed the frequency of this mutation in a total 222 modern wild and domestic pigs, including genomes used in this study as well as MC1R sequences obtained from (353). We found that the frequency of the derived allele (A) in European domestic (EUD) was >85% (Figure S8). If we take into account breeds that are known to have been selected to have camouflage coat colour (*e.g.* mangalitsa) or breeds that have other derived coat colour phenotypes (*e.g.* red coat colour such as Duroc) (353), the frequency increased to >95%. In contrast, the frequency of the ancestral allele (G) in European wild boar (EUW) was >98% (Figure S8). In fact out of 29 wild boars, we found only one sample (Dutch wild boar) with an heterozygous genotype. We also note that none of the three boars from the Near East and four wild boars from the Balkans possessed the derived allele (Figure S8).

We enriched this data with information from NGS and PCR based assay of ancient samples (see Table S1). For the NGS we counted the number of high quality read (MAPQ>= 30 and BQ>=30) for each allele to obtain a consensus (see Table S1).

This allowed us to 1) assess the robustness of this diagnostic mutation to discriminate between wild and domestic 2) determine the status of samples that could not be classified as wild or domestic based on morphology/context alone. Similarly to the modern data, we found that all ancient wild boars (17) that had been sampled in this study possessed the ancestral allele (G). This included wild boars from Mesolithic Germany, Neolithic/Chalcolithic Ukraine, Romania, Bulgaria, and Turkey, as well as Bronze age Turkey.

Of the 76 domestic samples that we typed for this allele, 57 were homozygous for the derived allele (A) and 7 heterozygous while 12 were homozygous for the ancestral allele (G). Homozygous derived and heterozygous (which express dominant black phenotype) were from a variety of sites, including 43 from Norse period sites in Iceland and Faroe, 13 from a Neolithic/Chalcolithic context in Turkey, England, Romania, Bulgaria, Germany, and Georgia, 6 from a Bronze age context in Turkey, England, Armenia, and Georgia, and from later period (Medieval and Hellenistic) from Turkey and the Netherlands. Of the 12 samples that were homozygous ancestral, 6 were from Neolithic/Chalcolithic context including sites in Georgia, Romania and Turkey and 6 were from Bronze Age sites in Turkey and Romania.

Altogether, these results demonstrate a clear division between wild and domestic samples based on this diagnostic mutation. The dominant black phenotype seem variable in domestic pigs from the Neolithic/Chalcolithic (~68% in Europe and Near East) and in the Bronze Age (~50% in Europe and Near East), but becomes fixed in later period. In addition, we found only one heterozygote wild boar out of 29 modern and 17 ancient samples. This fits very well with previous studies that found the dominant black allele in wild boars at a frequency of 5% in Greece (354) and less than 1% in Luxembourg (355). Altogether this suggests that while the ancestral allele

cannot distinguish between wild and domestic pigs, at least in the ancient data, the derived allele is highly diagnostic of a domestic status. Using this rationale, we infer the status of two samples, for which we had genome-wide data, AA288, a sample from Neolithic France, and AA363, a sample from Bronze Age Armenia.

Given that both European and Near East domestic pigs have the derived D124N allele we investigated its origin. To do so we built an ultrametric tree (using phangorn (356)) of haplotypes from a 50kb region around the *MC1R* gene using the phase of all our high ancient coverage genome data and modern data (see methods section *Chromopainter and GLOBETROTTER* for details about the phasing). Interestingly besides a few examples, most haplotypes (149/153) that had the derived allele formed a monophyletic group (Figure S9). This group included all ancient European and Near East domestic pigs as well as the majority of modern domestic pigs. This suggests that the allele shared by the majority of these pigs was the result of a single mutation. Interestingly, the wild boar haplotypes that were the closest to this group (besides the haplotype from the Dutch wild boar that possess the derived allele and which is most likely the result of recent introgression from domestic into wild (357)) were found in samples from both Near East (WB33U04) and Europe (WB22F02, WB22F03).

Thus, while we cannot definitely ascertain that this haplotype originated in the Near East, it seems plausible, that this haplotype, on which the derived allele first appeared, originated in the Near East. Especially given that an early Bronze age Iranian (AA363, 4468-4279 BP) and a Bronze Age Armenian sample (AA119) share the exact same haplotype as many ancient and modern domestic pigs from Europe. In addition, we also found multiple domestic pigs from Neolithic Anatolia that carry the derived allele (see above; Table S1). Alternatively, it is possible that the mutation was brought into the Near East from Europe, this, however, seems unlikely as European were only brought into the Near East later during the Iron Age (161).

Ancestry analysis

Neighbour Joining tree

For this analysis, we used all ancient and modern high and medium coverage data including the ancient genomes: AA288, AA363, AA133, AL718, KD025, AA119, KD033, AA451, AA349, VEM185 (Table S1). We used plink (358) to compute an Identity By State (IBS) matrix based on our high genome coverage (Table S1) and modern genomes (Table S2). This matrix was used to build a neighbour joining tree (NJ) using the R package “ape” (359). To compute support for each node we bootstrapped our ped file 100 times and re-computed a NJ tree each time (Figure S10). For this analysis we used only transversion removed linked SNPs by enforcing a distance of 1,000 bp between SNPs using plink. This resulted in a set of 1,055,138 SNPs.

Modelling population history of wild boars

For this analysis, we used all wild boar ancient/modern, high/medium coverage data including the ancient genome: AA133, AL718, AA349 (Table S1) and the same 1,055,138 SNPs as for the NJ analysis. In order to understand the ancestry of

domestic pigs, we first established the relationship between wild boar populations in Western Eurasia. Based on our NJ tree above and mtDNA analyses (Figure 1a; Figure S7&9), we define three populations, Near Eastern (NEW), European (EUW; including wild boars from France and the Netherlands; Table S2), Balkan (BLW) and Italy (ITW; including Swiss wild boars) and Asian (ASW) as suggested by our mtDNA analysis. We then ran ADMIXTURE with K=1 to K=10. We used a 10 fold cross validation approach to assess the best K value (Figure S11a). Our cross validation suggested that the best K values were 2 and 3. At K=2 ADMIXTURE splits Western populations (EUW, NEW, BLW, ITW) with Eastern populations (ASW; Figure S11b). Interestingly, this analysis suggests that NEW share some ancestry with ASW, which is also suggested by the fact that Iranian populations of wild boars share the same mtDNA haplogroup as ASW (Figure 1a). At K=3, ADMIXTURE splits NEW and EUW population (red and yellow ancestry in Figure S11c). This analysis indicates that BLW and ITW populations are a mixture of NEW and EUW ancestry, with the former being ~30% NEW and the latter ~10% NEW. Ths analysis also demonstrates that NEW ancestry has been stable for at least ~10000 years, as demonstrated by the result of AL718, a wild boar from Aşıklı Höyük (Figure S11c).

To better understand patterns of gene-flow between NEW, BLW and ITW, we used *qpGraph* to model relationship between Wild boar populations (360). We divided our wild boars into 5 populations based on Figure S10, NE2 (Iran and Armenia), NE1 (Turkey and Samos), BLW (Balkan), ITW (Italy and Switzerland), EUW (rest of Europe). We used *S. verrucosus* as an outgroup. We first fitted a simple model with no admixture (Figure S12a). This model left 11 outliers (Figure S12a). We added admixture from NE1 into BLW (as suggested in Figure ADMXITUREa). This model (Figure S12b) left a few outliers involving the Italian wild boars and Near Eastern populations. This analysis support the above results that the wild boars from the Balkans (BLW) are a mixture of NEW and EUW and further indicate that the best source for the NEW ancestry in NEW is the NE2 population (Turkey and Samos). Adding admixture from the Balkan into Italy removed these outliers (Figure S12c). This further support the result of the ADMIXTURE analysis (Figure S11).

Low-coverage data - testing PCA

Here we want to project low coverage data for which we have only few SNPs covered. To do so, we first selected all our low coverage samples that had at least 10 covered position from the ~12M autosomal SNPs we identified using high coverage samples (see *Ascertainment and ancient genome quality* section above). We then randomly selected a maximum of 100,000 SNPs per sample, that were covered by at least one read (see Table S1 - used-PCA, for a list of samples; note that the majority of samples had less than 100k SNPs covered from our 12M SNP panel). We then merged these lists of SNPs among our low coverage samples - this resulted in 1,534,393 SNPs.

We then assessed the accuracy of smartPCA (361) when projecting low coverage sample using lsqproject=YES option and these 1,534,393 SNPs. To do so, we estimated the position, on the PC space, of sub-sampled high coverage samples. We sub-sampled 10, 50, 100, 200, 500, 1000, 2000, 5000 and 10000 SNPs 1,000 times for each sample and projected these sub-sampled genome on the PCA (PC1 and PC2). For each sample we then computed the distance of the projected data and the expected

position of the full data on each PC. We then divided the difference by the total size of each PC and estimated the mean combining both PC1 and PC2. This analysis was run without the East Asian pigs.
Our results show that, using these 1.5M SNPs the PCA projections are highly accurate when a sample only 5,000 callable SNPs (Mean Error of ~1%; Figure S13). We thus chose to use all samples with at least 5,000 SNPs to be projected on the PCA (Figure S14; Table S1).

PCA

Based on the results above, we used every sample that had at least 5,000 SNPs covered (out of the 1,534,393 SNPs). This resulted in a total of 203 samples, including 54 low coverage genomes. We split the populations into ASD (Asian domestic), ASW (Asian wild), BLW (Balkan wild), BLW-A (Balkan wid ancient), EUD (European domestic), EUD-A (European domestic ancient), EUU-A (European undermined [domestic/wild] ancient), EUW (European wild), EUW-A (European wild ancient), NED-A (Near East domestic ancient), NEU-A (Near East undermined [domestic/wild] ancient), NEW (Near East wild), and NEW-A (Near East wild ancient). On Figure S14a, PC1 separates European and Asian pigs, and PC2 Near East and European pigs. On Figure S14b we find that NEW(-A), NED-A and NEU-A are separated from EUW, EUD(-A), BLW on PC2. Most EUD and EUD-A are clearly clustering with EUW rather than NEW. Some EUD-A from the Neolithic, however, (blue circles) are clearly intermediary between EUW and NEW, suggesting they share ancestry from both NEW and EUW (this includes KD033 as indicated Figure S10 and Figure S15 and Figure S16). PC1 on Figure S14b, however, most likely represent ascertainment bias in European domestic and Asian strong asian admixture (338, 362, 363)(Figure S15-5) - as it splits EUD from from the rest of the populations.

ADMIXTURE

For this analysis, we used all wild boar ancient/modern, high/medium coverage data including the ancient genome: AA288, AA363, AA133, AL718, KD025, AA119, KD033, AA451, AA349, VEM185 (Table S1) and the same 1,055,138 SNPs as for the NJ analysis to compute ancestry proportion using *ADMIXTURE* (364). We used a 10 fold cross validation approach to assess the best K value (Figure S15a). The best K value was 4. Interestingly at K=4, ADMIXTURE does not separate EUW and NEW populations (Figure S15b). Instead it separates between breeds of domestic pigs, in particular, between Large White and other breeds. This is most likely an effect of ascertainment bias given the large number of EUD in this analysis and/or potentially influenced by strong drift (due to artificial selection) in domestic population (337, 365). NEW and EUW are separated at K=6 (Figure S15c). This analysis recapitulate our finding discussed above (Figure S11). We found that the two high coverage Near Eastern domestic pigs (NED) from the Bronze Age of Armenia (AA119) and Iran (A363) possessed 100% NEW ancestry.

We found Near East ancestry in 8 modern EUD samples. These include breeds from Hungary (Mangalitsa [6.4-4.9%]), Italy (Casertana [6.2-1.7%] and Cinta Senese [~3.6%], and Calabrese [~5.2%]) and Spain (Negro Iberico [~1.9%] and Retinto [4.5%]). NEW ancestry in these populations, however, was always lower than in modern ITW (~6.5-8.5%) or than in modern BLW (30-36%).

With the exception of KD025, a sample from the Netherlands, we found NEW ancestry in all ancient domestic pigs from Europe. This includes a medieval pig from the Faroe island (AA451; ~5%), a Neolithic pig from Britain (VEM185; ~10%), and two Neolithic pigs from Germany (KD037: ~9% and KD033: ~54%). Interesting, KD033 also had a Near Eastern haplogroup (Y1; Table S1).

We also ran *ADMIXTURE* on a subset of our data - in particular we pruned out European domestic pigs (EUD) as they make up most of the data set keeping at most 2 genome per breed (Figure S16). We used a 10 fold cross validation approach to assess the best K value (Figure S16a). Our cross validation suggested that the best K value is 3. At K=3 ADMIXTURE identify three populations roughly corresponding to 1) Asian pigs (ASW/ASD; blue in Figure S16a), European and Near East wild boars (NEW/EUW; red in Figure S16a) and European domestic pigs (EUD; pink in Figure S16a). This shows that the result presented in Figure S15 (splitting EUD populations before NEW/EUW) analysis above (with full data set) was indeed affected by an over representation of EUD in the data.

F3 & D-statistics

For this analysis, we used all wild boar ancient/modern, high/medium coverage data including the ancient genome: AA288, AA363, AA133, AL718, KD025, AA119, KD033, AA451, AA349, VEM185 (Table S1) and the same 1,055,138 SNPs as for the NJ analysis. Figure S10 suggests that most EUD are equally related to both NEW and EUW (EUD is outgroup to EUW and NEW). In addition, Figure S15b suggests that EUW is more closely related to NEW than EUD. Here we used outgroup-f3 to further assess whether this is the result of the know admixture from ASD (here we use ASW as proxy for ASD to avoid issues with gene-flow EUD->ASD) into EUD during the 19th century (338, 362, 363). Outgroup-f3 were computed as (*S. verrucosus* [outgroup species]; [ASW/EUW/NEW], EUD). This statistics measure shared drift between two populations (here ASW/EUW/NEW and EUD). The higher the value the more closely related the two populations tested. We plotted every possible combination of this test for each EUD genome (Figure S17). Our results for show that every EUD genome (except KD033, a Neolithic domestic pig from Germany) is more closely related to EUW than to NEW and ASW. However, we can see that many of the modern EUD have higher f3 value with ASW than ancient domestics (EUD-A) or than breeds such as Mangalitsa (MA01F20 and MA01F18 in Figure S17) a traditional breed from Hungary that is known to be non admixed with Asian pigs but also seem to possess NEW ancestry (Figure S15). Altogether this indicates that the placement of EUD in Figure S10 is most likely an artefact of strong admixture between ASD and EUD.

We further used D-statistics to assess whether admixture from ASD would affect our power to detect any remaining NEW ancestry in the genome of modern domestic pigs. To do so, we computed every combination of D(S.verrucosus, [NEW/ITW/BLW/ASW]; EUD, EUW) for each EUD genome. Positive values means admixture from NEW/ITW/BLW/ASW into EUW and negative values indicate admixture into EUD. In Figure S18 we plotted the distribution of Z values of this statistic (plotting all possible combinations). We found that, most EUD genome have high negative Z values (Z<<-4), when using ASW, indicative of gene-flow from Asian pigs (Figure S18). These values were mirrored by high positive Z when using NEW/ITW/BLW (Z>>4). This, however, is not the result of admixture from NEW

into EUW since the MRCA of EUW/EUD but instead most likely the result of the admixture from Asian pigs into EUD, which makes EUW and NEW artificially more related relative to NEW/EUD. This indicates that a D-statistic test is unlikely to be able to assess whether most EUD have any remaining NEW ancestry or if the NEW ancestry can be distinguished from admixture with ITW/BLW.

In some cases, however, traditional breeds such as Mangalitsa (MA in Figure S18) show non-significant values ($3<|Z|<3$) for admixture from both ASW and NEW. In fact, the only genome that we find with significant evidence for admixture from NEW was KD033, a Neolithic domestic pig from Germany (Figure S18), which has according to our ADMIXTURE analysis possess ~54% Near Eastern ancestry (Figure S15b). All other ancient genomes seem to possess admixture from ITW rather than BLW or NEW. Altogether this suggest that the NEW ancestry in modern EUD found in our ADMIXTURE analysis is likely due to admixture from European wild boars that possess some degree of NEW such as in Italy or in the Balkan (Figure S11c).

Phasing

We used the approach implemented in *GLOBETROTTER (366)* to model the ancestry of EUD as a function of EUW, NEW and ASW. To do so, we first phased all using *SHAPEIT2* (367) with default parameters and using the recombination map from (368). We used all the wild boar ancient/modern, high/medium coverage data including the ancient genome: AA288, AA363, AA133, AL718, KD025, AA119, KD033, AA451, AA349, VEM185 (Table S1). We restricted this analysis to transversion (~4M SNPs).

Chromopainter and GLOBETROTTER

We then used *ChromoPainterV2 (369)* to paint the genome of all EUD (recipients) using NEW, ASW and EUW as donors. To do so, we first estimated switch parameter over global mutation rates by taking weighted average values (weighted over number of SNPs per chromosome) of the final value given by running *ChromoPainterV2* using 10 E-M iterations (-i 10 -in -iM) using all EUD samples on chromosome 1 and 2. We used these values (-n 7249.57 -M 0.00784444) to paint the chromosome of EUD ("target") using NEW, ASW and EUW as donors. We also painted every NEW, ASW and EUW genome using the same set of donors and combined the resulting files for *GLOBETROTTER*. We used these painting profile to compute to admixture proportion of all EUD genomes (without inferring admixture date: num.mixing.iterations=0) as in (370). *GLOBETROTTER* found that the EUD ancestry can be modelled as ~4% NEW, ~26% ASW and, ~60% EUD. The ASD proportion is very close to previous estimates (362, 371, 372).

ADMIXTURE simulations

Here we want to estimate the minimum amount of NEW that we could detect using *ADMIXTURE* in modern European domestic. To do so, we simulated admixed genomes. To simulate an admixed genome, we mixed a random NEW and EUW genome as described in (373). Starting at the beginning of a chromosome we sampled a genotype from a random NEW genome with a probability of α or from EUW with a probability of 1-α. We note that α is the NEW ancestry in the simulated genome. Along the chromosome, at each SNP we re-sampled a new ancestry (either NEW or EUW) with a probability of $1-e^{-\lambda g}$ where λ is the number of generation since

admixture and *g* the distance in Morgan of the whole block. We used the recombination map described in (368). We simulated 1000 genome for each possible combination of values, with λ: 100, 500, 1000, 2000, 4000 and with α: 0.01-0.1, 0.15.

For each simulated genome, we then estimated NEW ancestry using all NEW and EUW samples with ADMIXTURE (K=2). We then estimated, for each combination of λ and α, the rate at which ADMIXTURE detected NEW ancestry (with ancestry > 0.00010). We then estimated the probability of observing 8 out of 85 genomes (modern data estimate) with NEW ancestry given the rate for each combination of λ and α using a binomial distribution. Our results indicate that the number of EUD with NEW ancestry is compatible with 4-5% NEW ancestry in modern European pigs (Figure S19a).

GLOBETROTTER simulations

As for *ADMIXTURE*, we estimated the minimum amount of NEW ancestry that our *GLOBETROTTER* analysis could detect in EUD. To do so, we simulated admixed haplotypes as described above (170 for 85 EUD) with the same range of λ and α as above. We joined two haplotypes (sampled from the same pair of NEW and EUW) to create a diploid individual and painted each diploid individual using NEW and EUW as donors, excluding the sample that were used for the simulation. We then painted every NEW and EUW genome using the leave one out approach described in (366). Briefly, to ensure that we painted the simulated genome and surrogate genome with the same number of donors, we painted each NEW/EUW with all but 2 randomly selected sample (NEW or EUW). Each painting profile were then used to estimate NEW and EUW ancestry proportion using *GLOBETROTTER* (without LD curve fitting; num.mixing.iterations=0). We repeated this 10 times for each λ and α combination and estimated the proportion of runs for which *GLOBETROTTER* detected NEW ancestry. Our results indicate that *GLOBETROTTER* cannot detect NEW ancestry below 4% and can detect NEW 80% of the time when above 6% (Figure S19b). This suggests that our result from *GLOBETROTTER* (~4% NEW ancestry in EUD overall) is accurate.

Sweep ancestry

Here we want to assess whether selection, during the evolutionary history has acted on haplotypes that were of Near Eastern origin. To do so, we used previously detected selective sweeps regions that had a *p-value* < 0.01 (299 sweep regions in total (337)). For each region, we extracted 25Kb around the peak (as in (337); 50kb in total) and computed Kimura-2-parameter model corrected nucleotide distances between every pairs of haplotypes using the R package ape (359) (see Figure S21 for examples of UPGMA trees computed using this distance). Foreach EUD haplotype we then computed *D*, the normalised difference between the nucleotide distance of this haplotype to the closest EUW haplotype and to the closest NEW as:

$$D = (d_{NEW} - d_{EUW}) / (d_{EUW} + d_{NEW})$$

Where d_{EUW} is the distance to the closest EUW haplotype and d_{NEW} is the distance to the closest NEW haplotype. *D* value range between -1 and 1 with positive value

meaning that the haplotype is closer to EUW and negative value meaning it is closer to NEW. To plot these results we computed mean and standard deviation of *D* over all EUD. The results are presented in Figure S20. This figures shows that none of the sweeps found in (337) affected haplotypes that are of clear NEW ancestry. In fact we found only 9/299 regions in which D was on average lower than 0, with no region with both mean-standard deviation were lower than 0. On the other hand we found 270 regions with mean greater than 0 and 158 regions in which both mean and mean-standard deviation were greater than 0.

X chromosome

A recent study identified a region on the X chromosome with extremely low recombination rate (374). This large ~50Mb region displayed a very peculiar pattern of divergence where *S. scrofa* from South China clustered together with other *Sus* species from ISEA. This pattern was suggested to be the result of either ancient admixture and/or retaining ancestral state due to low recombination rate. Here we assessed whether this low recombination rate could have allowed EUD to retained NEW ancestry. We first assessed whether this region has either low or high divergence between wild boar populations. To do so, we computed d_{xy} (375) between these three populations in 500kb bins across chromosome 1 and X (Figure S22). We found that the region, in between 50 and 100Mb on chromosome X, described by (374), indeed has a very low divergence rate, between all three populations, compared to other region. As opposed to wild populations in China, however, this region does not seem to display relatively higher divergence rate between NEW and EUW.

We then build NJ trees using identity by state with data from i) the entire chromosome X, ii) the entire chromosome X except for the low recombining region, iii) only the low recombining region. To limit the impact of missing data we used the same set as for the *ADMIXTURE* analysis (see above). We found no difference between the trees build using data in any of these regions (Figure S23) and those built using autosomal chromosomes.

Supplementary Tables

Table S1.

Table describing data used in this paper. Extract No. / Lab code: Code used in this paper. **Previous Extraction Code:** Extract code used in previous publications. **Accession of mtDNA data:** Accession code for mtDNA data if previously published (and provided in the reference). **Source:** Author that provided the material (see site description for more information). **USER treatment:** Y=library was user treated; N=library was not used treated. **Haplogroup based on NGS data:** haplogroup call based on methodology described above. **Control haplotype and Control haplogroup**: haplotype and haplogroup based on control region. **Combined haplogroup**: consensus haplogroup if data from NGS and control region available (control region haplogroup was used if disagreement). **Clade**: Clade used in Figure 1. **Mismatch NGS / control haplogroup:** 0=both method agree; 1=methods disagree. **Capture Status**: indicate whether the library was capture for mtDNA. **SNPs Covered:** Number of autosomal SNP covered. **used-PCA:** indicates whether a sample was used in PCA. **use-other-nuc:** indicates whether a sample was used in other nuclear analyses. **Group:** meta group to which the sample belong: ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, BLW-A=Balkan wild ancient, EUD=European domestic, EUD-A= European domestic ancient, EUU-A=European unknown (either wild or domestic) ancient, EUW=European wild (including Italian), EUW-A=European wild (including Italian), NED-A= Near Eastern domestic ancient, NEU= Near Eastern unknown (either wild or domestic) ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient. **Number of reads MC1R (A allele):** number of reads mapped at the D124N variant that possess the A allele. **Number of reads MC1R (G allele):** number of reads mapped at the D124N variant that possess the G allele. **Genotype MC1R:** Predicted MC1R genotype. **Status based on Zooarcheolgy:** Domestic, wild or feral status based on information from context and morphology (see site description for more information). **Status based on MC1R and Zooracheology**: consensus status based on both MC1R and zooarcheology call.

Table S2.

Summary of the modern genome used in the analysis.

Supplementary Figures

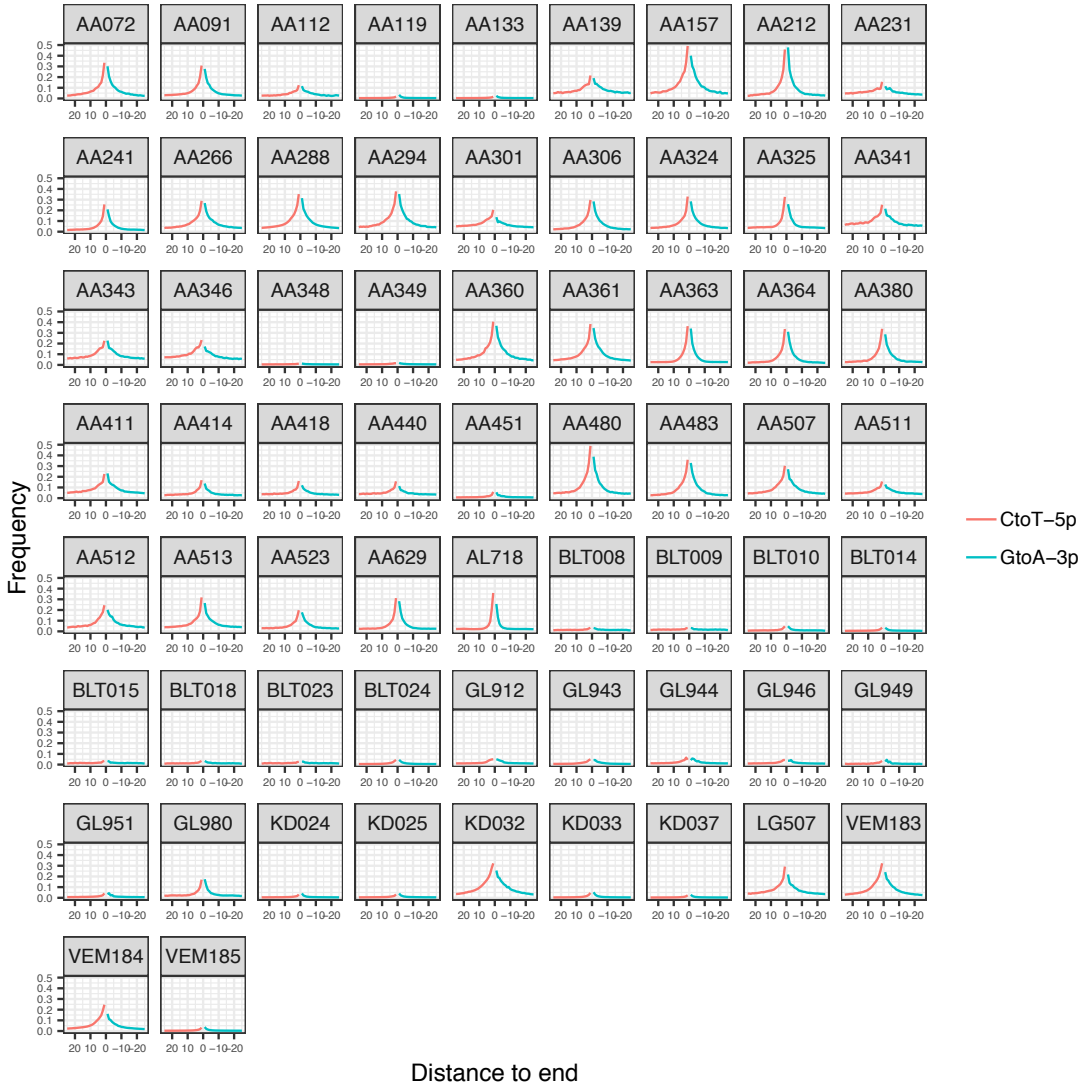


Fig. S1.

Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in nuclear genome analyses. Libraries starting with BLT, as well as KD033 KD037, KD025, VEM185, AA119, and AA133 were all USER treated.

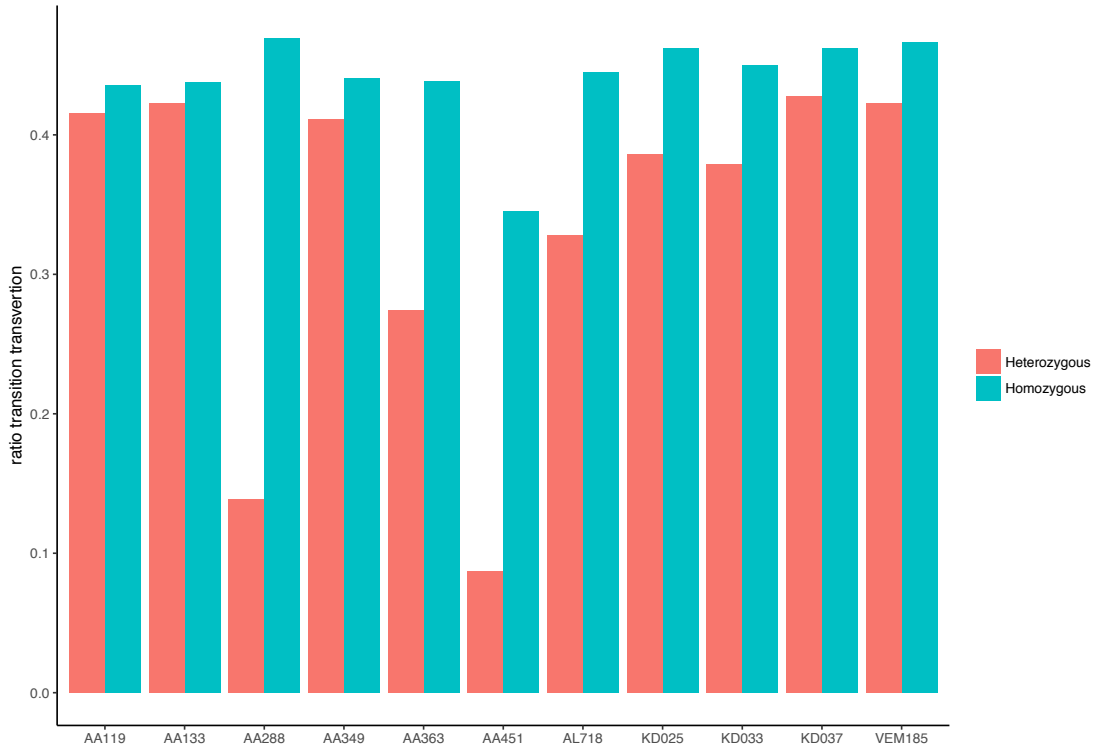


Fig. S2.
Ratio of transversion/transition at heterozygous sites for genome with >3x coverage.

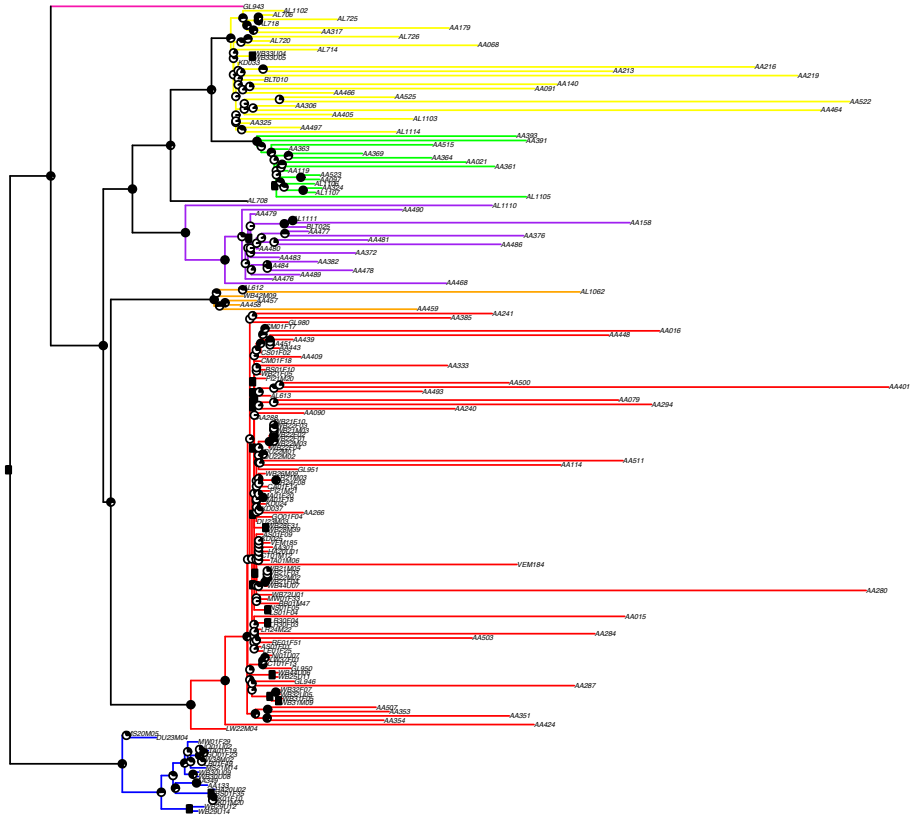


Fig. S3.
Maximum likelihood tree of 188 mitogenomes including both ancient and modern sequences. All ancient ancient genomes were at least 2x coverage. This tree depicts the relationship between the major haplogroup used in this study: As (blue), Y1 (yellow), Y2 (purple), ArmT (green), Eu (red), Italy (orange) and NE1 in pink. Black circle at each node represent bootstrap support (e.g. full black circle = 100 bootstrap support).

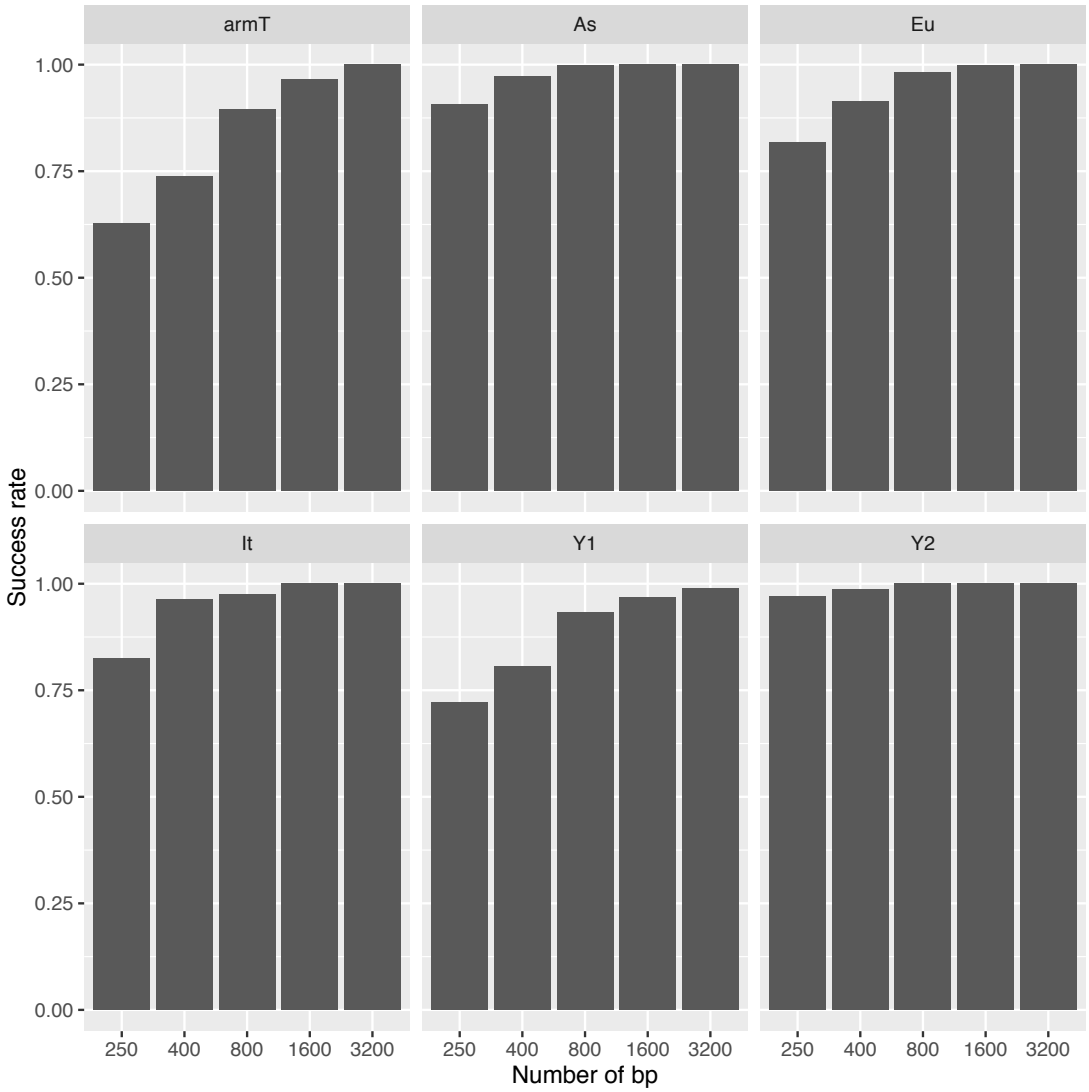


Fig. S4.
Success rate of our method to assign low coverage genome to specific haplogroup for different level of coverage (Number of base pair covered in the reference mitogenome). This was obtain by subsampling genome with known haplogroup.

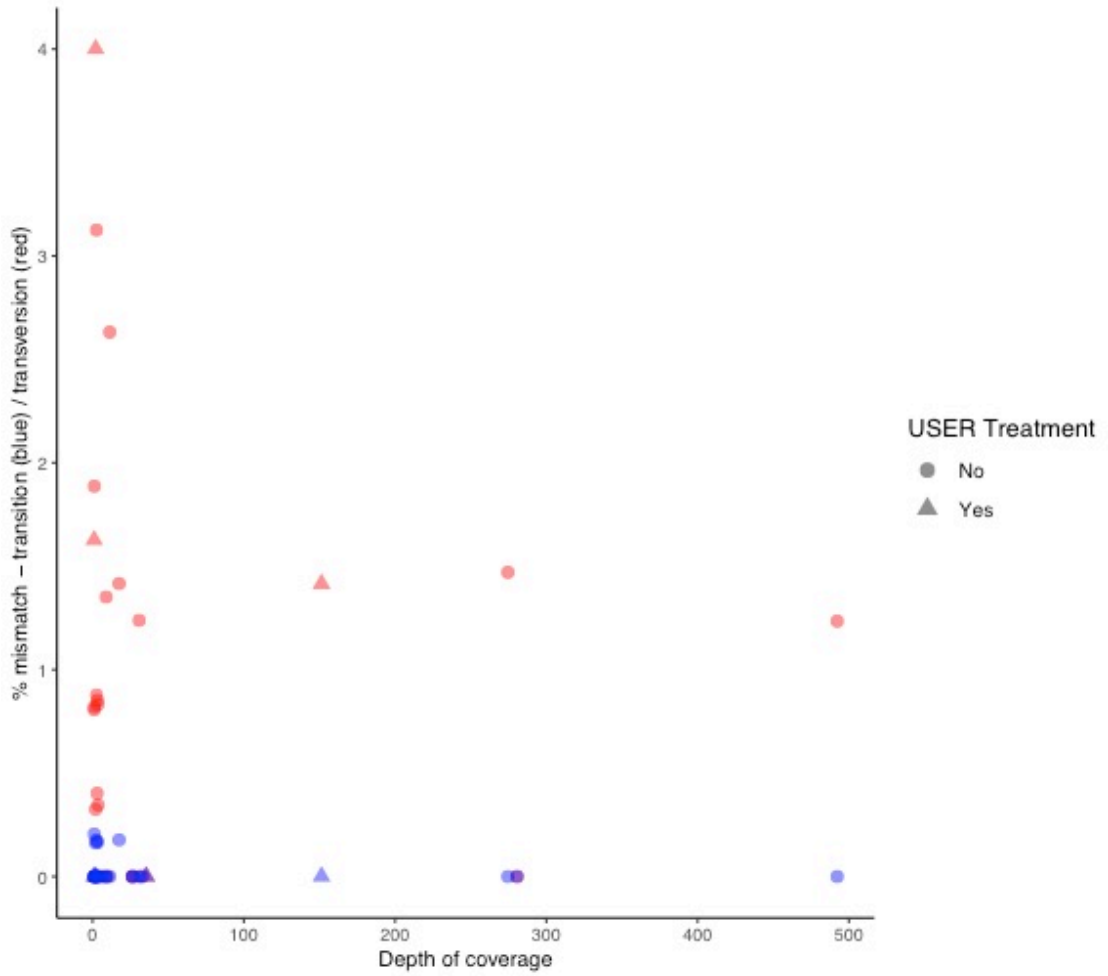


Fig. S5.
This plot represents the % of mismatch between sanger sequenced and NGS (depth of coverage on the x-axis) mitochondrial control haplotypes in 29 samples for which we possessed both data.

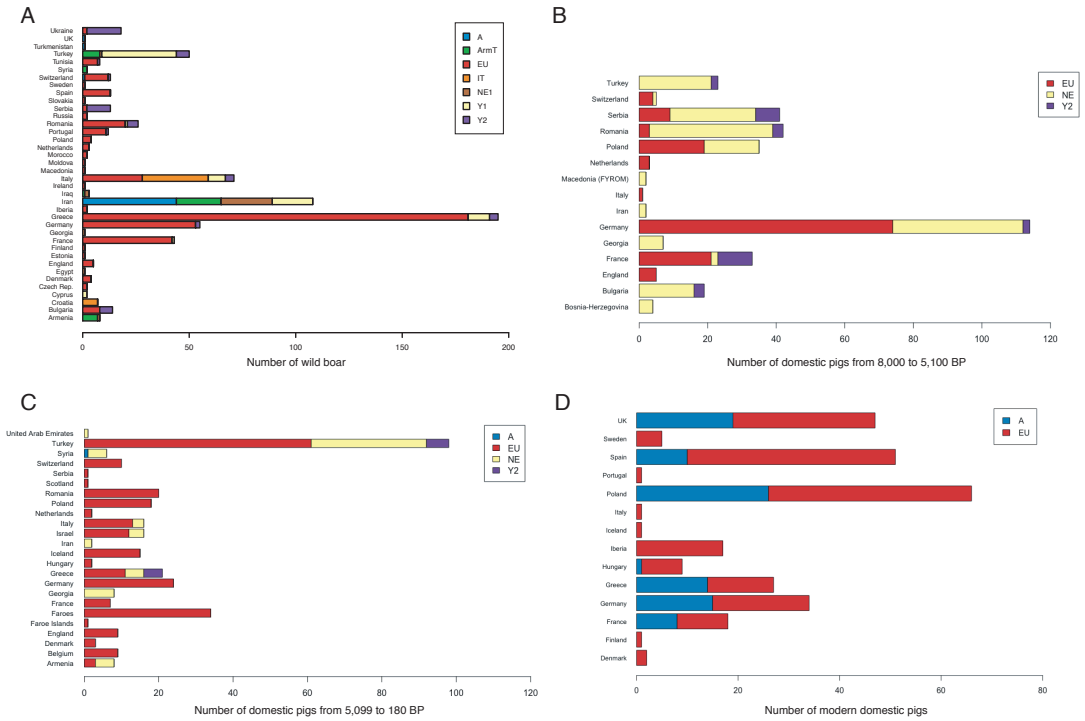


Fig. S6.
The bar plots show the number of domestic pig/wild boar samples that possess that possess As (blue), Y1 (yellow), Eu (red), Y2 (purple), ArmT (green), NE1 (brown), and Italy (orange) haplotypes divided different categories as seen in **Fig. 1** and **Fig. S7**. **A.** modern wild boar. **B.** domestic pigs between 8,000 BP and 5,100 BP. **C.** domestic pigs between 5,099 and 180 BP **D.** modern domestic pigs.

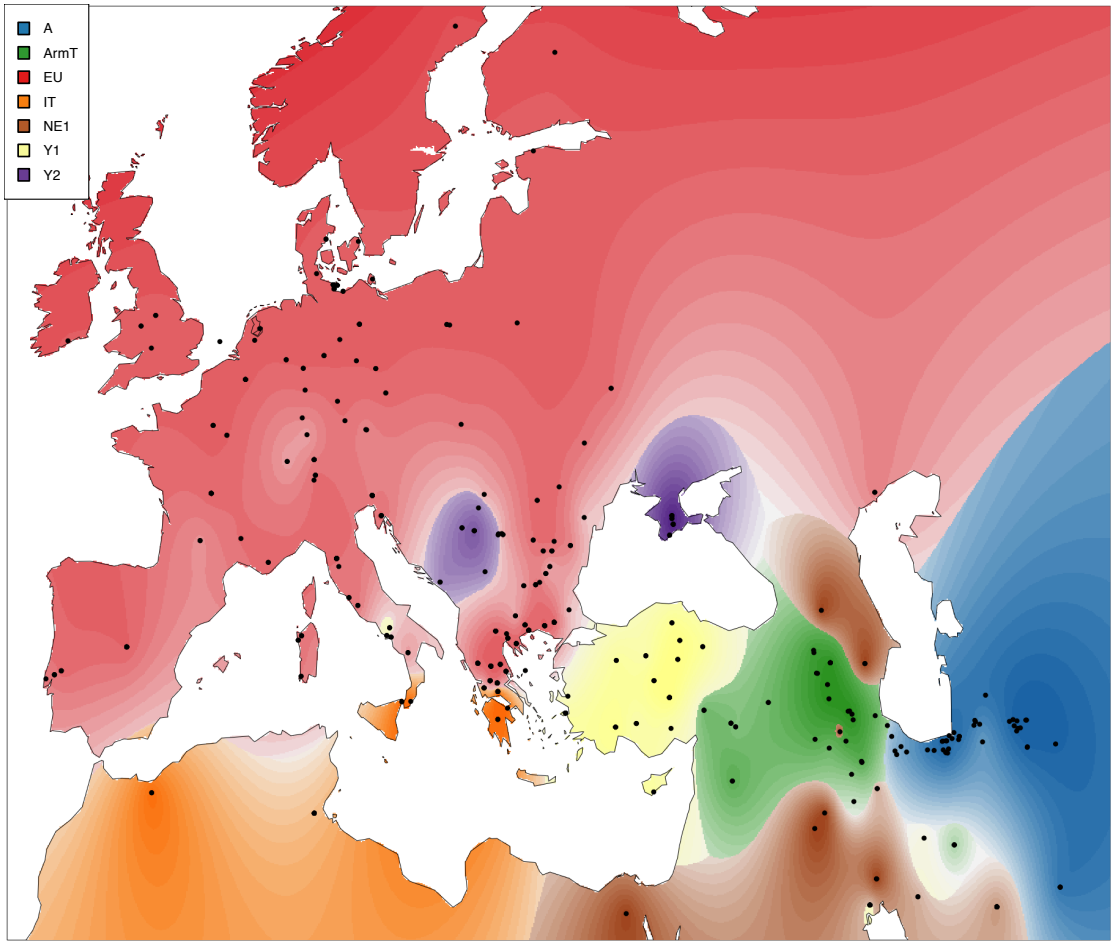


Fig. S7.
The black dots on the map represent locations of 696 modern and ancient wild boar that possess As (blue), Y1 (yellow), Eu (red), Y2 (purple), ArmT (green), NE1 (brown), and Italy (orange) haplotypes. The haplotype assignments were used to interpolate the underlying color distribution which demonstrates the biogeographical boundaries of these haplotypes. The bar plots that show the number of wild boar samples per country are in **Fig. S6A.**

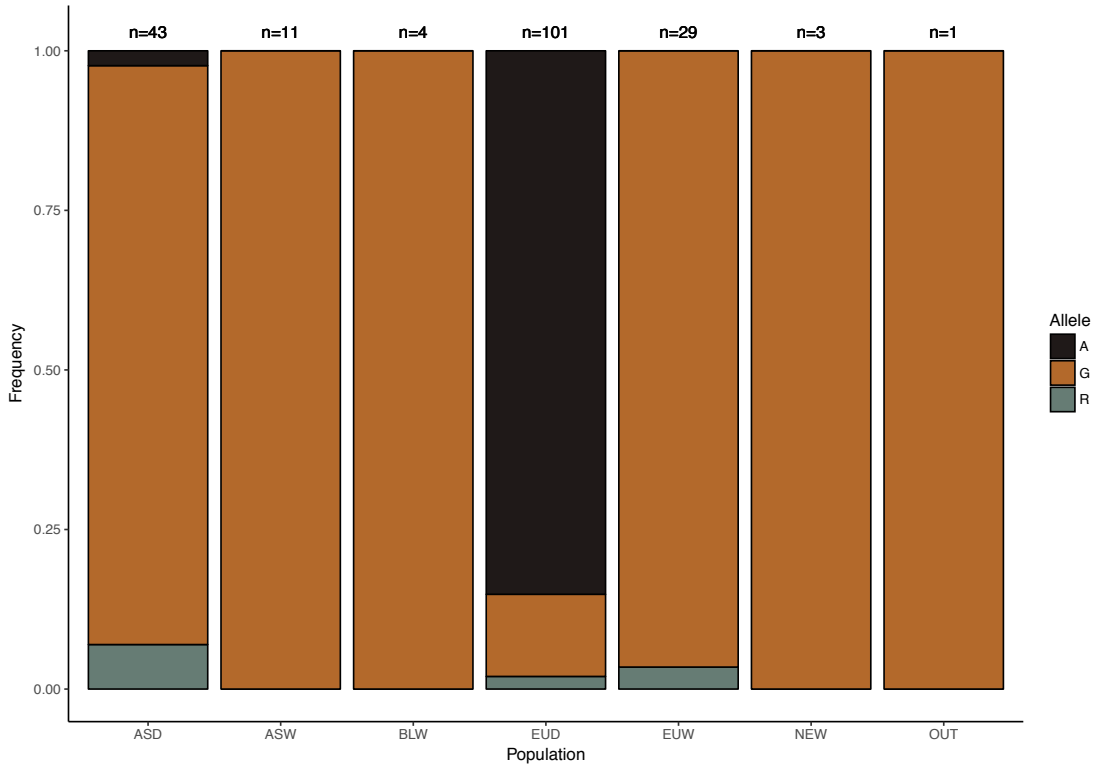


Fig. S8.
Frequency of three possible genotypes in the MC1R gene. G=ancestral allele (camouflage coat colour) and A=derived allele (black coat colour). Population codes: ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, EUD=European domestic, EUW=European wild (including Italian), NEW=Near Eastern wild, OUT=outgroup (S. verrucosus).

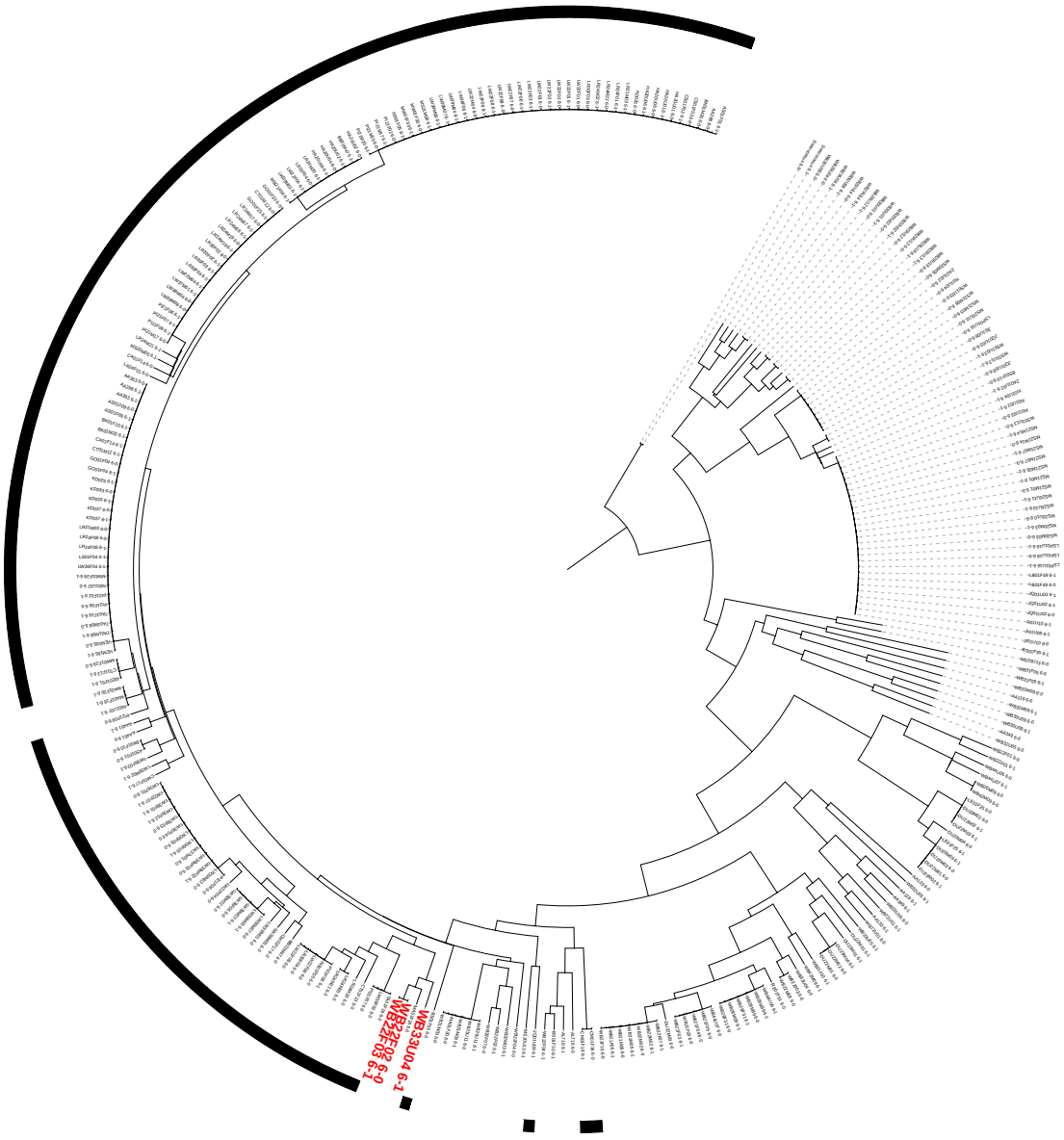


Fig. S9.
Ultrametric tree (UPGMA) based on a 50kb haplotypes around the *MC1R* gene. Black colour stripe represent the samples that possess the A derived allele (black coat colour). Red labels represent the three wild boar haplotypes that are closest to the monophyletic group that possess the derived allele.

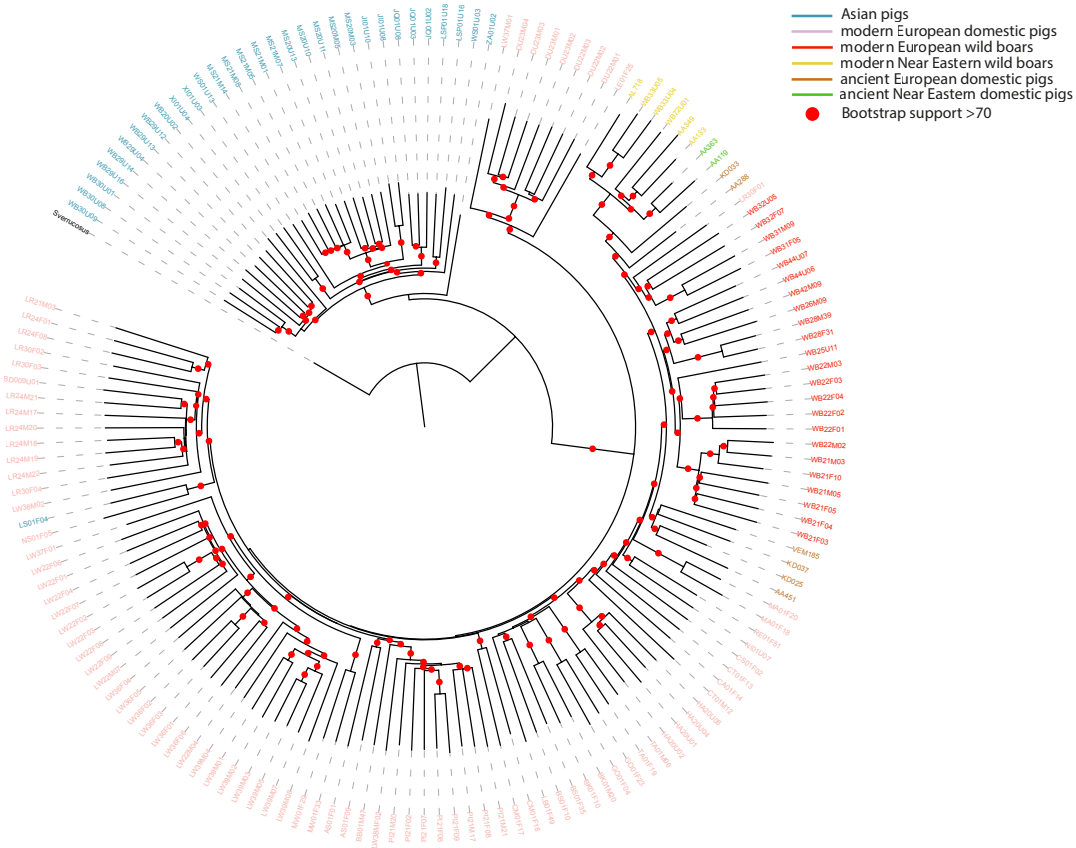


Fig. S10.
Neighbour joining tree based on high genome coverage ancient samples. Node labels represent bootstrap support. This figure shows that KD033 and AA288, both of which have substantial Near East ancestry (see Fig. S15; Fig. 2B) cluster with Near Eastern wild boars.

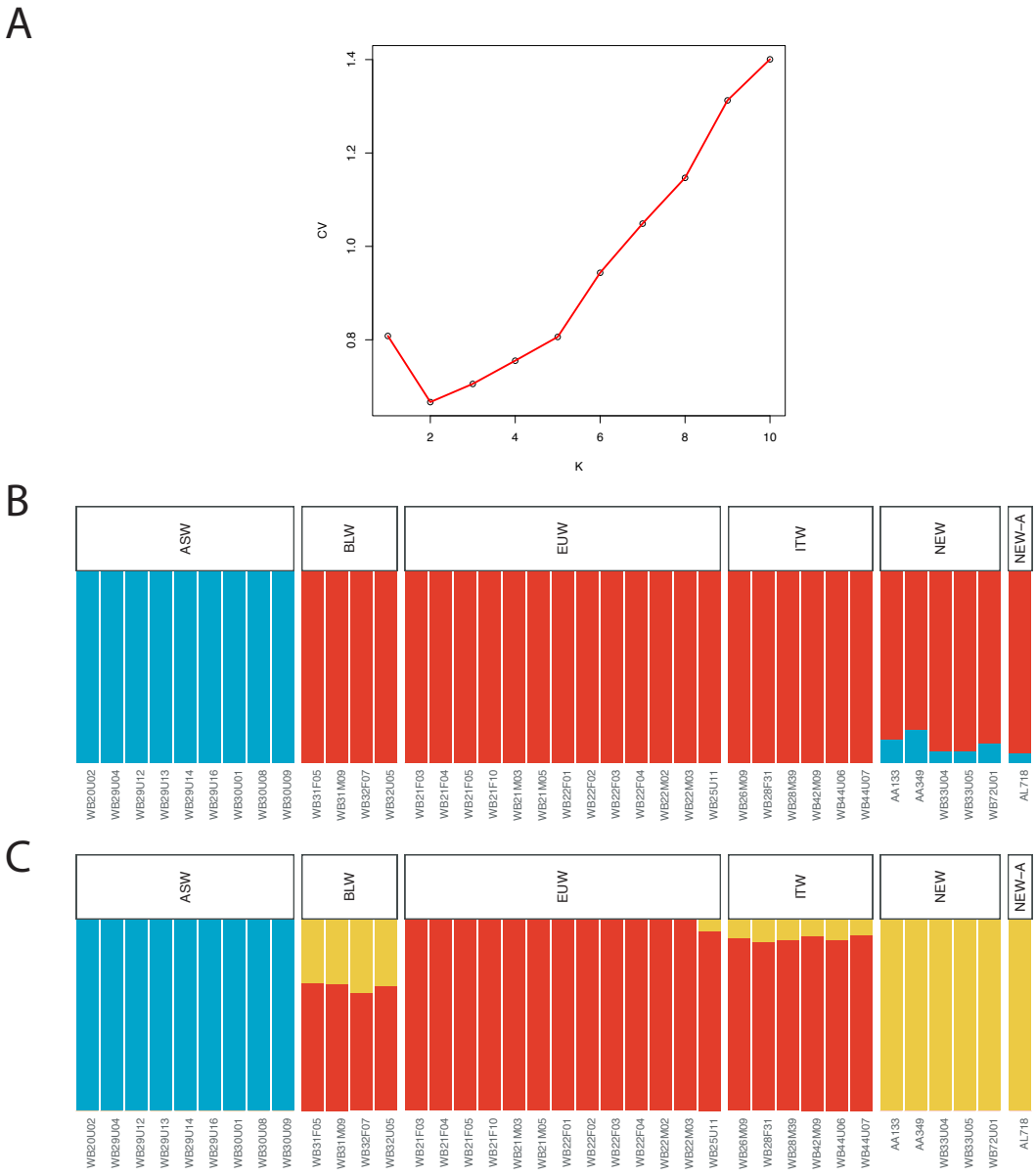
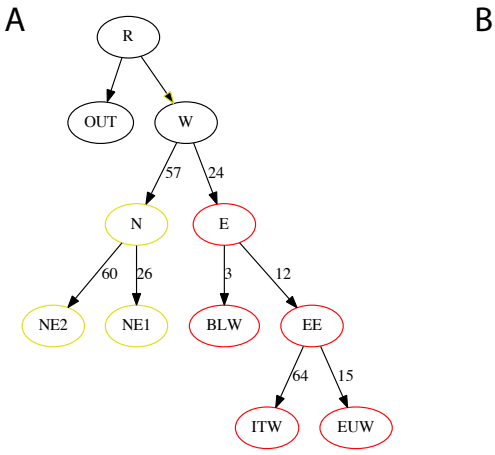
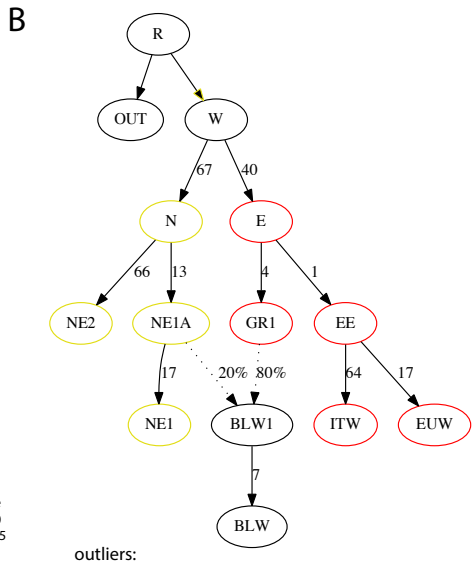


Fig. S11.
Results of ADMIXTURE analysis for wild boar genomes. **A.** Cross validation for different K values. **B.** Results for K=2. **C.** Results for K=3. ASW=Asian wild, BLW=Balkan wild, EUW=European wild, ITW= Italian wild, ITW-A= Italian wild ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient.



outliers:

	f4			fitted	estim	diff	std. err	Z score
OUT	NE2	ITW	BLW	0.000000	0.017129	0.017129	0.004127	4.150
OUT	NE2	EUW	BLW	0.000000	0.014034	0.014034	0.003612	3.885
OUT	ITW	NE2	NE1	0.000000	0.017525	0.017525	0.004425	3.961
OUT	ITW	NE2	BLW	0.024235	0.045152	0.020917	0.005386	3.883
OUT	EUW	NE2	EUW	0.050777	0.067098	0.016321	0.005439	3.001
OUT	NE1	EUW	BLW	0.000000	0.017049	0.017049	0.002950	5.780
NE2	EUW	ITW	BLW	-0.011806	-0.028878	-0.017072	0.004931	-3.462
NE2	EUW	EUW	BLW	-0.026543	-0.044327	-0.017784	0.005668	-3.138
NE2	NE1	ITW	EUW	0.000000	-0.010708	-0.010708	0.003277	-3.268
ITW	NE1	EUW	BLW	0.011806	0.023546	0.011740	0.003654	3.213
EUW	NE1	EUW	BLW	0.026543	0.047342	0.020799	0.005160	4.031



outliers:

	f4			fitted	estim	diff	std. err	Z score
OUT	ITW	NE2	NE1	0.000000	0.017525	0.017525	0.004425	3.961
NE2	NE1	ITW	EUW	0.000000	-0.010708	-0.010708	0.003277	-3.268

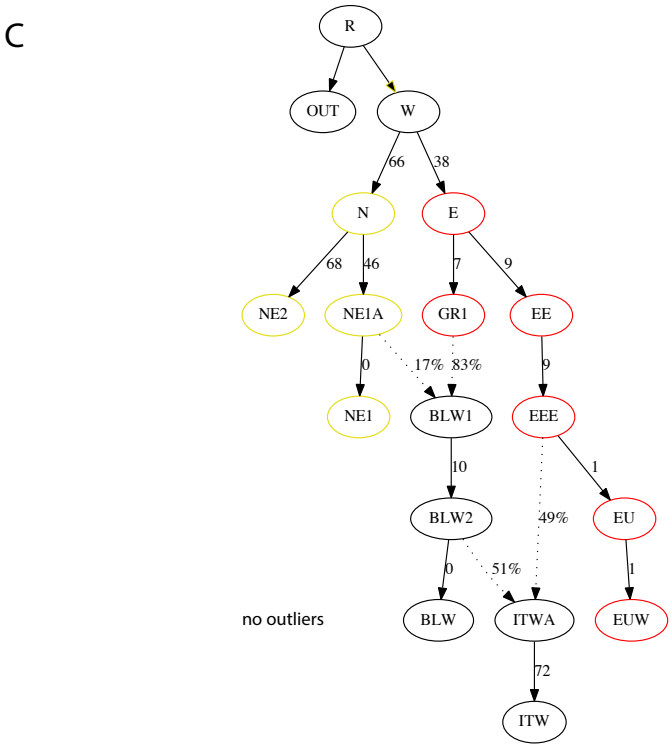


Fig. S12.
Results of *Qpgraph* analysis on wild boars genome. In this analysis population were split into NE2 (Iran and Armenia), NE1 (Turkey and Samos), BLW (Balkan), ITW (Italy and Switzerland), EUW (rest of Europe), OUT (*S. verrucosus*). **A.** Model with no gene-flow - this model left 11 outliers. **B.** Model involving gene-flow from Turkish (NE1) wild boars into Balkan wild boars (BLW) - this analysis left two outliers. **C.** Model involving gene-flow from Turkish (NE1) wild boars into Balkan wild boars (BLW) and gene-flow from Balkan into Italy (ITW) - this analysis left no outliers.

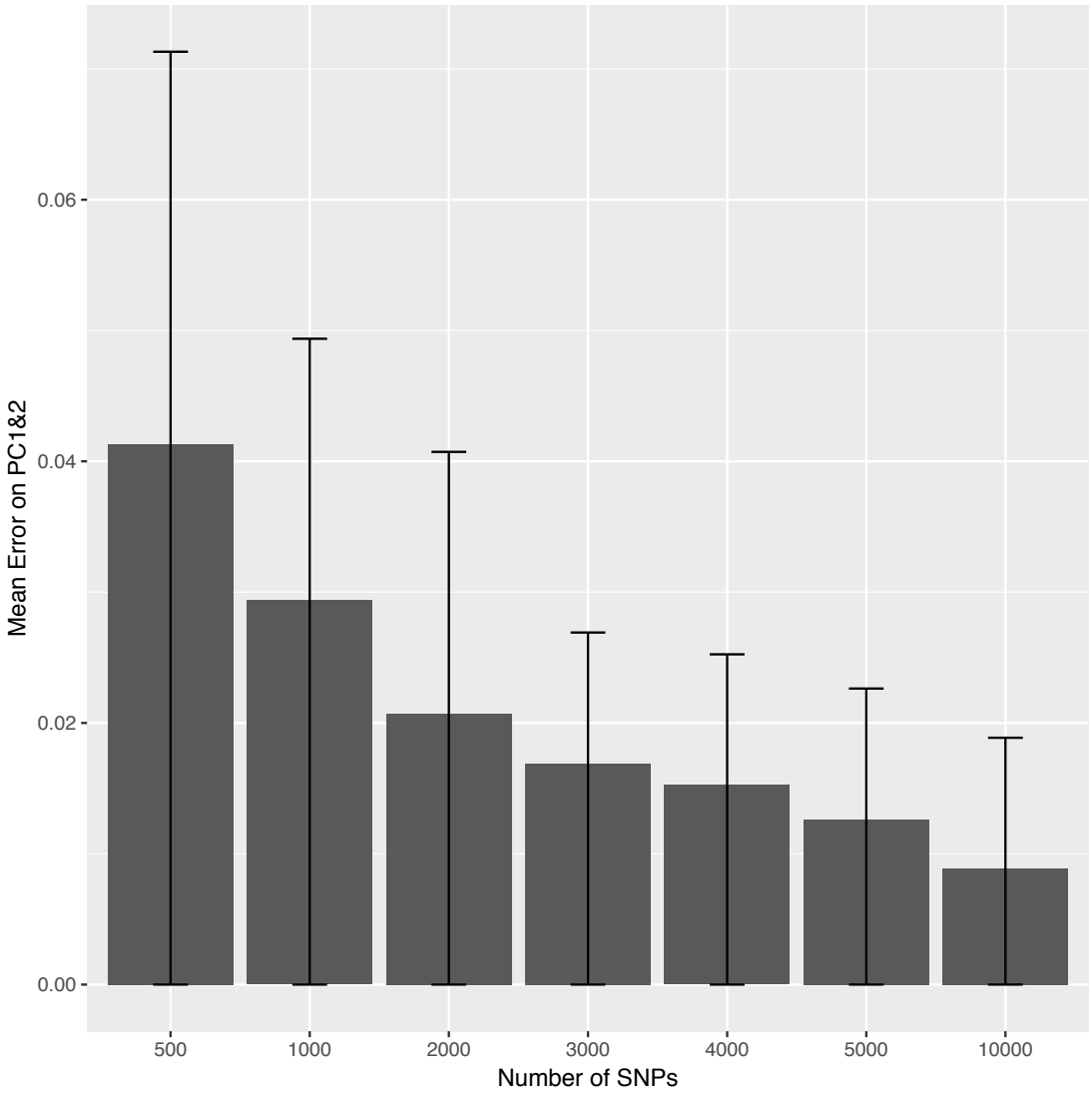


Fig. S13.
Mean error rate on PC1&2 for low coverage data with different levels of SNP coverage.

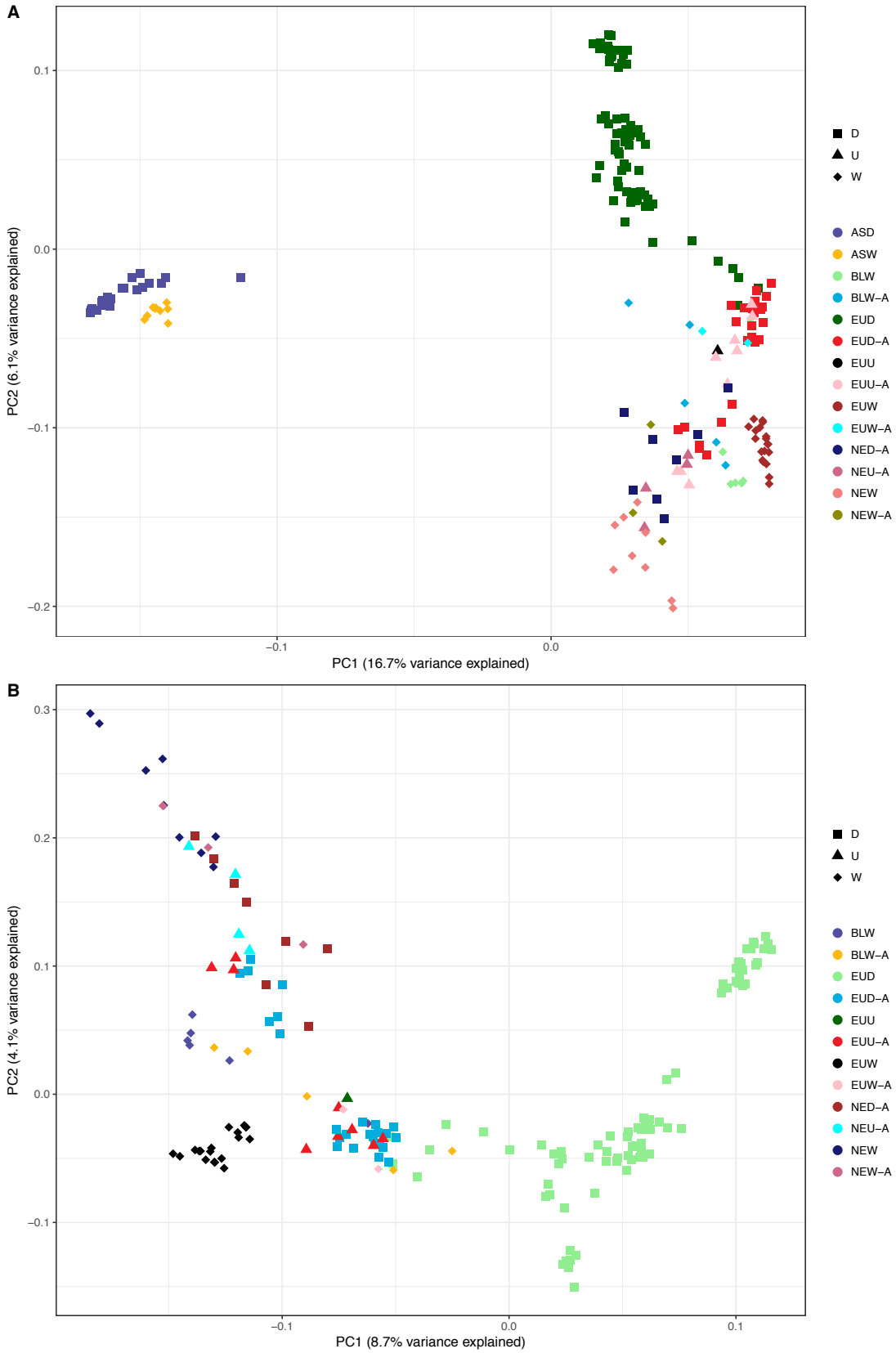


Fig. S14.
PCA including low coverage data (projected). **A.** PCA including Asian wild and domestic samples. **B.** PCA without Asian wild samples. Samples are split into multiple categories. Symbols: D=Domestic, U=Unknown status, W=Wild. Population code: ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, BLW-A=Balkan

wild ancient, EUD=European domestic, EUD-A= European domestic ancient, EUU-A=European unknown (either wild or domestic) ancient, EUW=European wild (including Italian), EUW-A=European wild (including Italian), NED-A= Near Eastern domestic ancient, NEU= Near Eastern unknown (either wild or domestic) ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient.

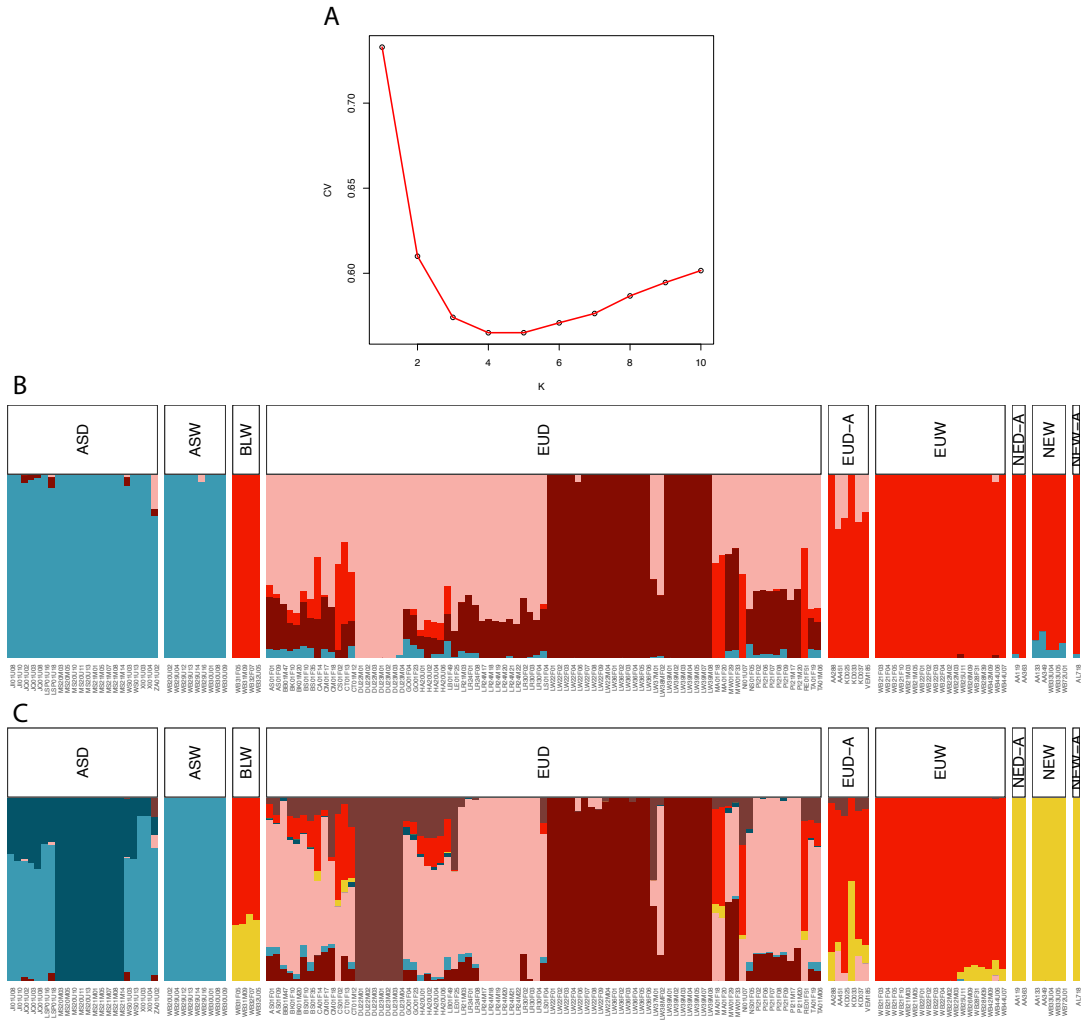


Fig. S15.
Results of ADMIXTURE analysis including all modern genomes and all high coverage ancient genome wild boar genomes. **A.** Cross validation for different K values. **B.** Results for K=4. **C.** Results for K=7. ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, BLW-A=Balkan wild ancient, EUD=European domestic, EUD-A= European domestic ancient, EUU-A=European unknown (either wild or domestic) ancient, EUW=European wild (including Italian), EUW-A=European wild (including Italian), NED-A= Near Eastern domestic ancient, NEU= Near Eastern unknown (either wild or domestic) ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient.

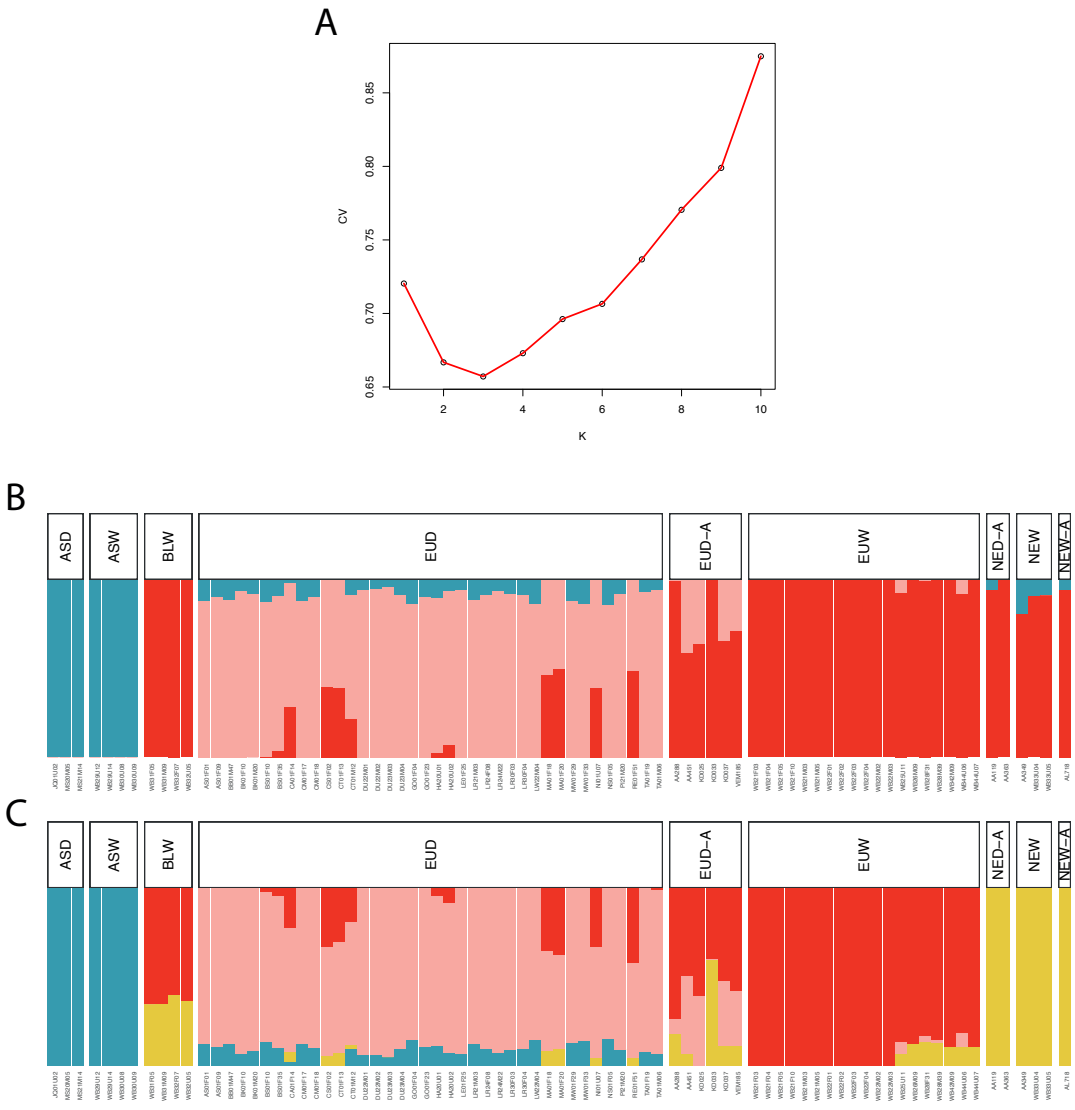


Fig. S16.
Results of ADMIXTURE analysis with only up to two genomes per modern breed. **A.** Cross validation for different K values. **B.** Results for K=3. **C.** Results for K=4. ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, BLW-A=Balkan wild ancient, EUD=European domestic, EUD-A= European domestic ancient, EUU-A=European unknown (either wild or domestic) ancient, EUW=European wild (including Italian), EUW-A=European wild (including Italian), NED-A= Near Eastern domestic ancient, NEU= Near Eastern unknown (either wild or domestic) ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient.

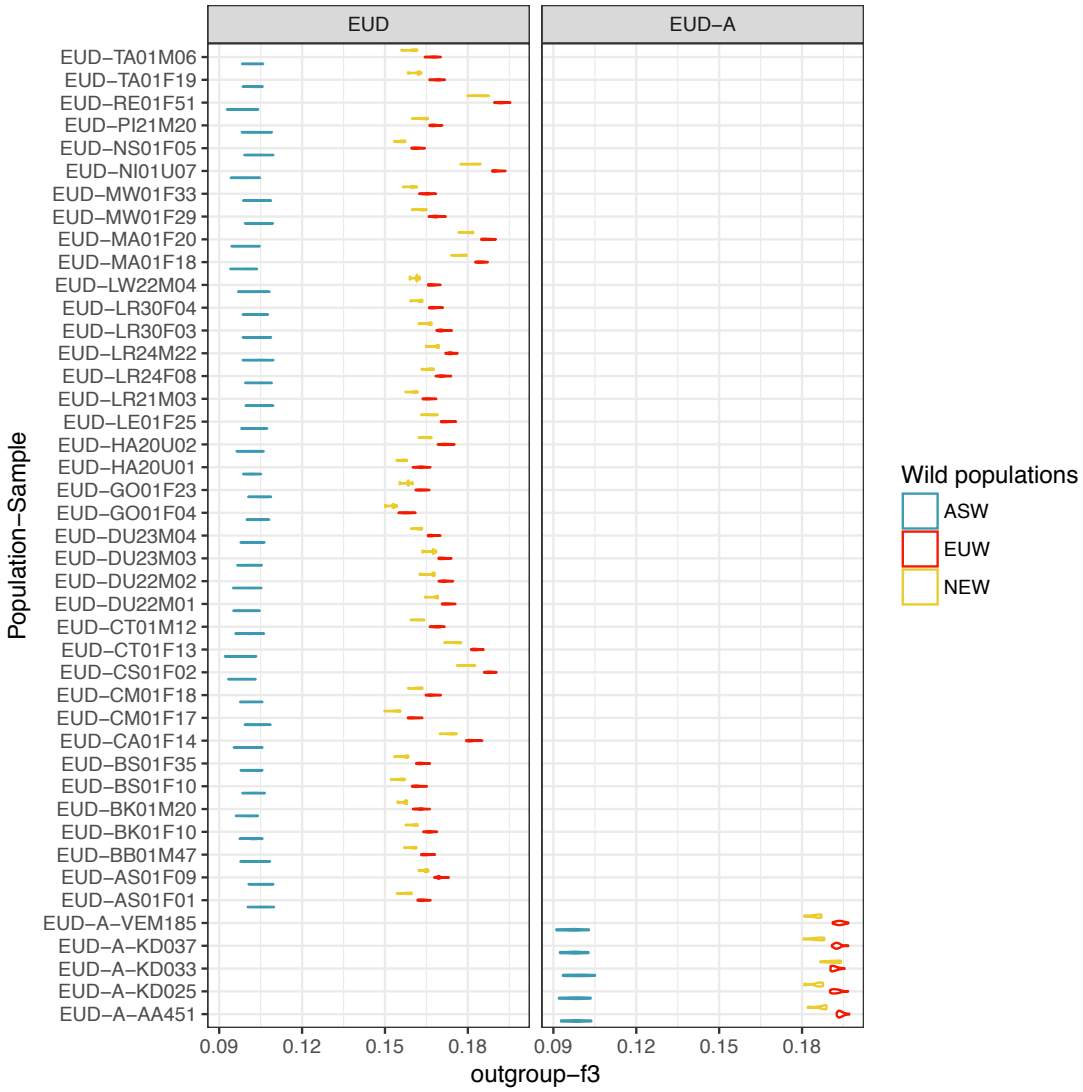


Fig. S17.
Violin plot representing the distribution of outgroup-f3 statistics of the form (*S. verrucosus*; [ASW/EUW/NEW], EUD) for each EUD genome.

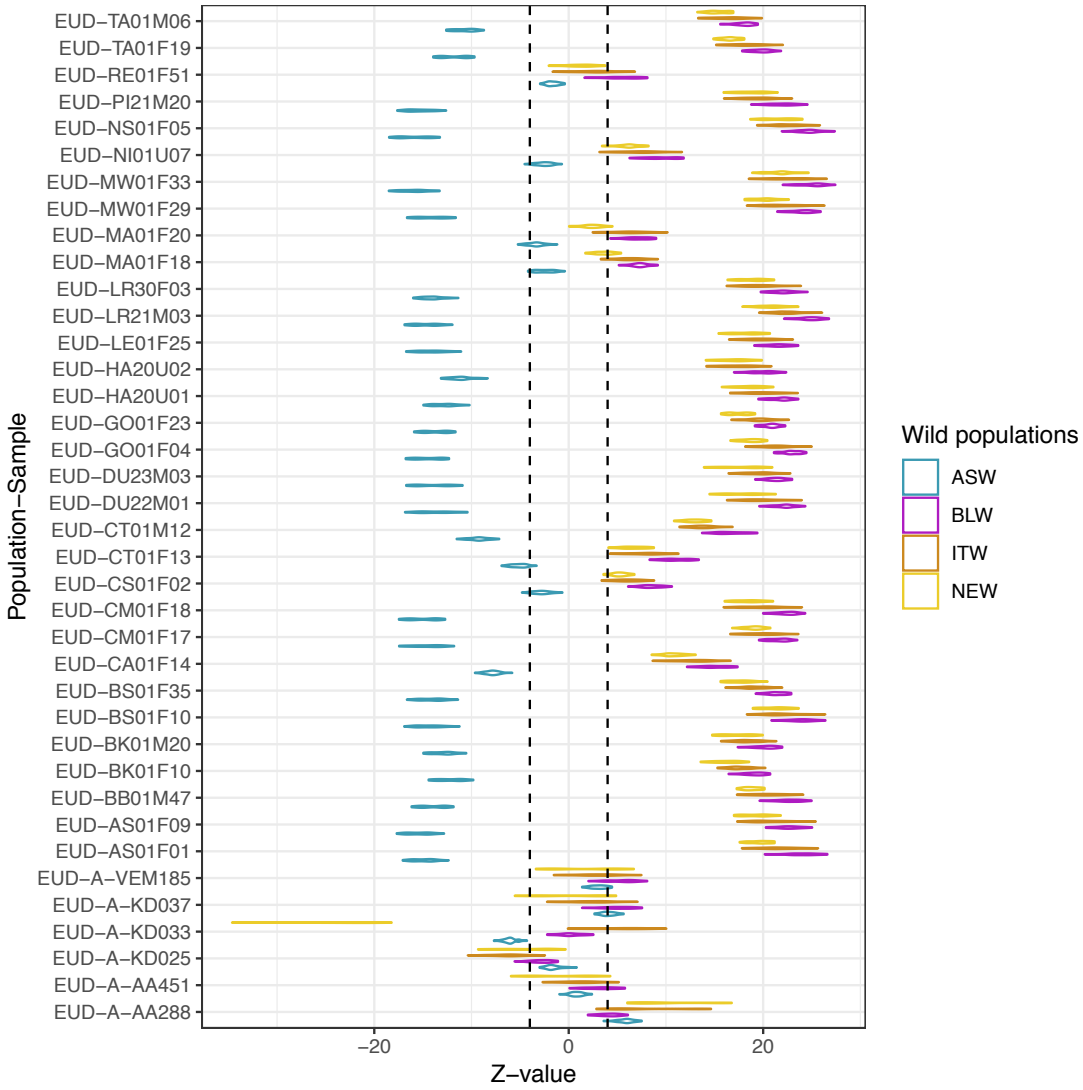


Fig. S18.
Violin plot representing the distribution of Dstatistics of the form D(S.verrucosus, [NEW/ITW/BLW/ASW]; EUD, EUW) for each EUD genome.

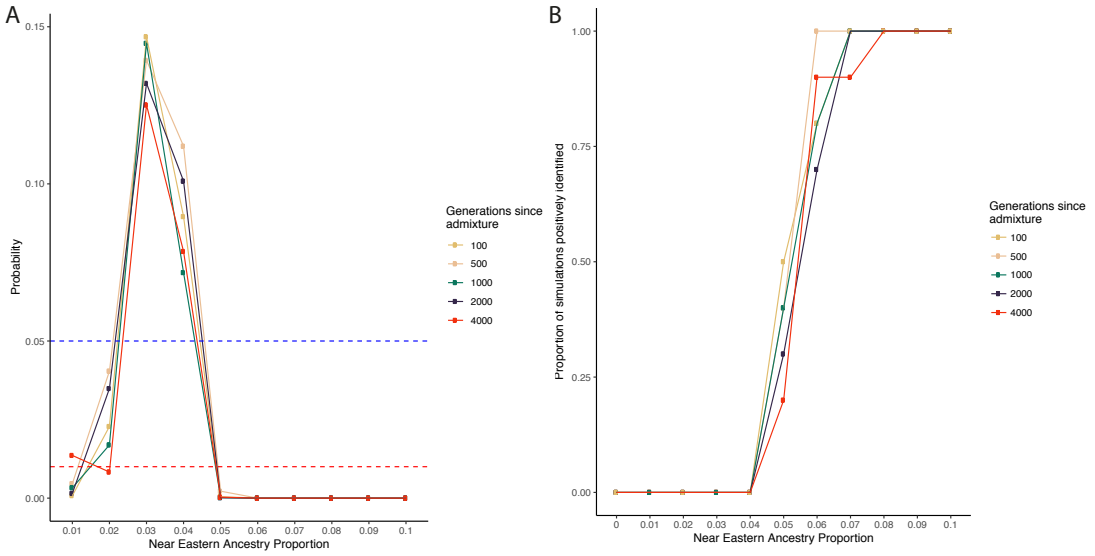


Figure S19.
A. Graph depicting the results of a simulation of European pig genomes using different timeframes of admixture with European wild boar, and different levels of Near Eastern ancestry (x-axis). These simulated genomes were analyzed using ADMIXTURE; the y-axis represents the probability (computed using a binomial distribution) of observing 8 out of 85 European domestic genomes with detectable Near Eastern ancestry (y-axis). **B.** Proportion of simulations that were (out of 10 replicates) for which *GLOBETROTTER* identified NEW ancestry in EUD for different levels of ancestry fraction and time since admixture.

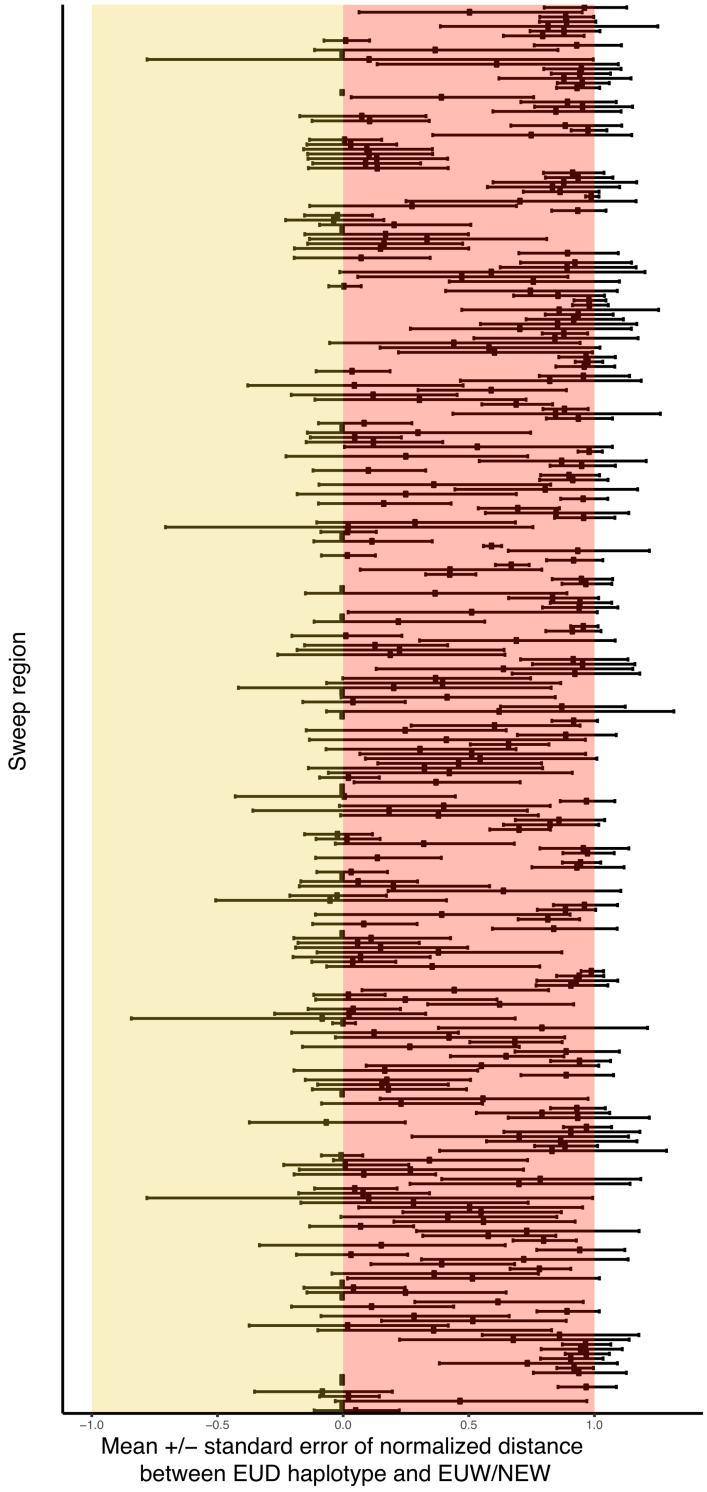


Figure S20.
A graph depicting the ancestry of haplotypes carried by domestic pigs in regions of the genome that underwent a selective sweep (338). Bars on the x-axis represent the mean and standard deviation of the normalized difference between the nucleotide distance of haplotypes found in domestic pigs to the closest European and Near Eastern wild boar haplotype. Positive and negative values indicate that, on average, haplotypes possessed by domestic pigs in a sweep region (y-axis) are closest to haplotypes found in European wild and Near East boars respectively.

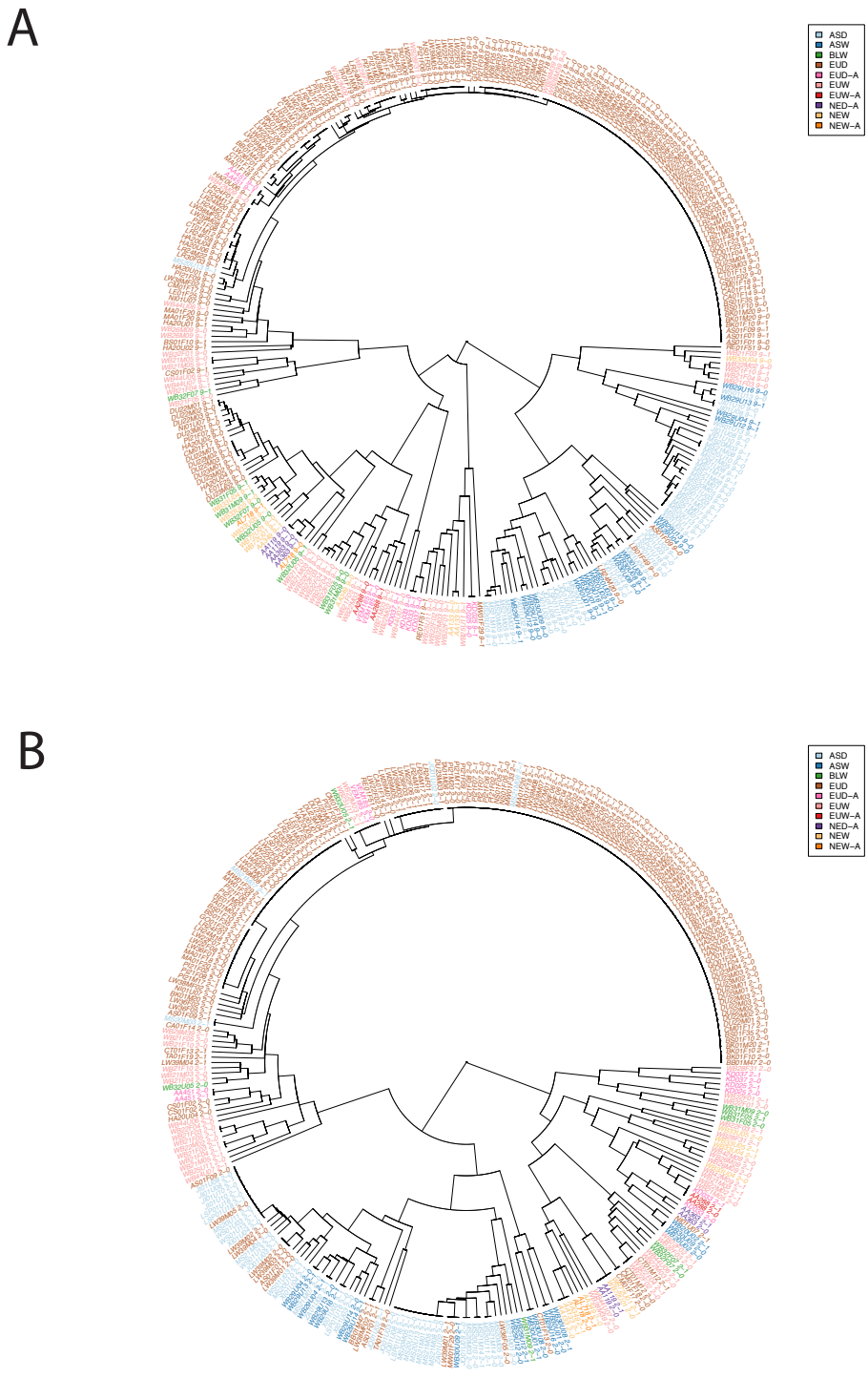


Fig. S21:
Ultrametric tree (UPGMA) in two regions detected as selective sweeps in European Domestic pigs (*335*). This figure shows a clear reduction of diversity in EUD in these specific regions - potentially due to a selective sweep. In both cases the selected haplotype in EUD appears to be much more closely related to EUW than NEW. **A** chromosome 2:73626576-74003521 **B** chromosome 9:109689664-109911665. Population codes: ASD=Asian domestic, ASW=Asian wild BLW=Balkan wild, BLW-A=Balkan wild ancient, EUD=European domestic, EUD-A= European

domestic ancient, EUU-A=European unknown (either wild or domestic) ancient, EUW=European wild (including Italian), EUW-A=European wild (including Italian), NED-A= Near Eastern domestic ancient, NEU= Near Eastern unknown (either wild or domestic) ancient, NEW=Near Eastern wild, NEW-A=Near Eastern wild ancient.

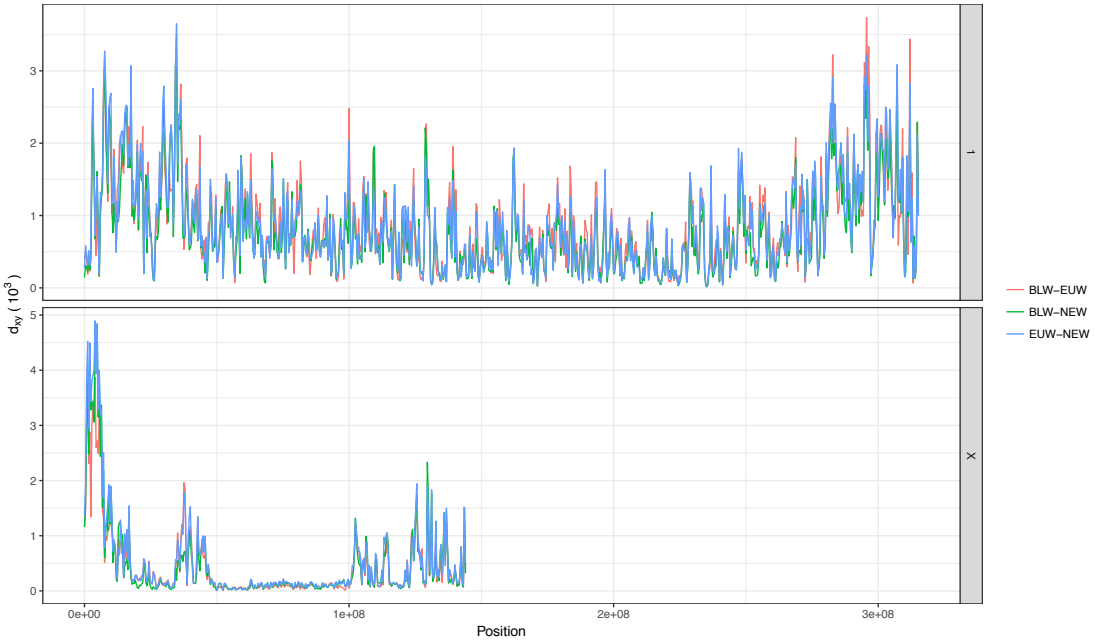


Fig. S22.
Dxy values plotted along chromosome 1 and X.

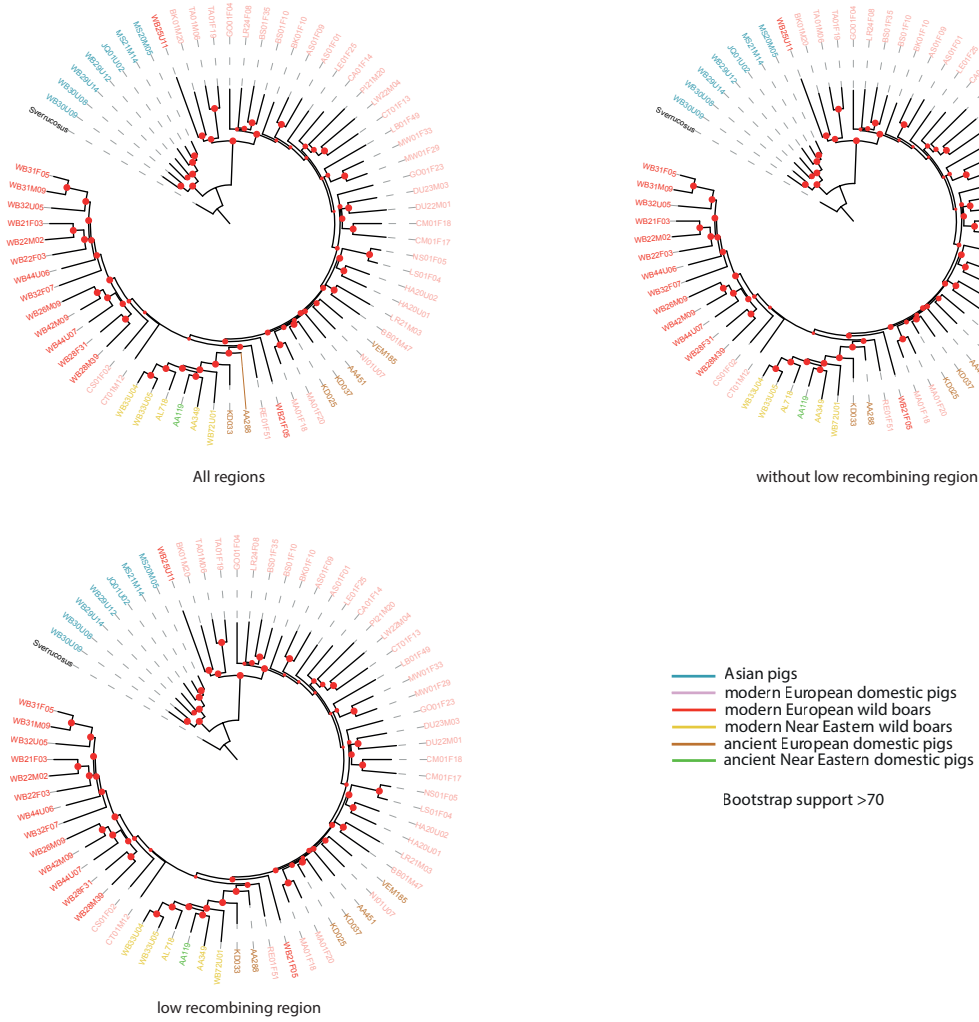


Fig. S23.
NJ tree for **A.** whole chromosome X **B.** only low recombining region of chromosome X **C.** without the low recombining region.

References

1. Benecke N, Bollongino R, Küßner M, Weber R (2006) Zur Datierung und Fauna des spätglazialen Schichtkomplexes am Abri Fuchskirche I bei Allendorf, Lkr. *Saalfeld-Rudolstadt Alt-Thüringen* 39:121–130.

2. Lechevallier M, et al. (1978) Abou Gosh et Beisamoun. Deux gisements du VII millénaire avant l'ère chrétienne en Israël. *Mémoires et Travaux du Centre de Recherches Préhistoriques Français de Jérusalem Jérusalem* 2.

3. Khalaily H, Marder O, ha-'atiḳot IR (2003) *The Neolithic site of Abu Ghosh: the 1995 excavations*.

4. Özgüç N (1966) EXCAVATIONS AT ACEMHOYtlK. *Anadolu (Anatolia)*. doi:10.1501/andl_0000000092.

5. Öztan A, Arbuckle BS (2001) Yılı Acemhöyük Kazıları. *Kazı Sonuçları Toplantısı* 24:1–39.

6. Arbuckle BS (2006) The Evolution of Sheep and Goat Pastoralism and Social Complexity in Central Anatolia. Dissertation (Harvard University).

7. Burenhult G (1997) *Ajvide och den moderna arkeologin* (Natur och Kultur).

8. Burenhult G, Brandt B (2002) The grave-field at Ajvide. *Remote Sensing. Applied Techniques for the Study of Cultural Resources and the Localization, Identification and Documentation of Sub-Surface Prehistoric Remains in Swedish Archaeology*, ed Burenhult G (Department of Archaeology, Stockholm University, Stockholm).

9. Burenhult G (1997) Säljägare och svinaherdar på Ajvide. *Ajvide Och Den Moderna Arkeologin. Natur Och Kultur*.

10. Gligor M (2009) *Aşezarea neolitică şi eneolitică de la Alba Iulia-Lumea Nouă în lumina noilor cercetări* (Ed. Mega).

11. Gligor M (2014) Începuturile eneoliticului timpuriu în Transilvania: o abordare Bayesiană [The beginning of Early Eneolithic in Transylvania: a Bayesian approach]. *Analele Banatului* 22:91–105.

12. Gligor M, McLeod K (2014) Disarticulation as a mortuary practice in Early Eneolithic Transylvania? A case study from Alba Iulia-Lumea Nouă. *Annales Universitatis Apulensis, Series Historica* 18(2).

13. Gligor M, Mcleod KKM (2015) Disposal of the dead. Uncommon mortuary practices from Alba Iulia-Lumea Nouă 2003 excavation. *Homines, Funera, Astra 2". Life Beyound Death in Ancient Times. Proceedings of the International Symposium on Funerary Anthropology. 23-26 September 2012, "1 Decembrie 1918" University (Alba Iulia, Romania)*, ed R. Kogălniceanu, M. Gligor, R. Curcă, S. Straton (Archaeopress, Oxford), pp 25–41.

14. Lundberg C, Gligor M PLACE OF DEATH AND PLACE OF REST: COMMINGLED HUMAN REMAINS FROM ALBA IULIA-LUMEA NOUĂ 2015 EARLY ENEOLITHIC FUNERARY DISCOVERY. *Annales Universitatis Apulensis, Series Historica* 19(2).

15. Szirmai K (2004) Feltárások és leletmentések az auxiliáris castellum vicusában Albertfalván. *Aquincumi Füzetek* 10:116–122.

16. Bökönyi S (1974) *History of domestic mammals in Central and Eastern Europe* (Akademiai Kiado).

17. Papathanasiou A (2005) Health status of the Neolithic population of Alepotrypa Cave, Greece. *Am J Phys Anthropol* 126(4):377–390.

18. Papathanasopoulos GA, Katsipanou-Margeli V, Kourtesi-Filippaki G (2011) To Neolithiko Diro. *Spilaio Alepotrypa* (Ekd. Oikos Melissa Athina).

19. Moser B (1986) Die Tierknochenfunde aus dem latènezeitlichen Oppidum von Altenburg-Rheinau: Charakterisierung des Fundgutes Pferd, Hund, Hausgeflügel und Wildtiere/von Barbara Moser. Dissertation (München).

20. Leuzinger U, und Kultur Thurgau DFE (2000) *Die jungsteinzeitliche Seeufersiedlung Arbon-Bleiche 3: Befunde* (Departement für Erziehung und Kultur des Kantons Thurgau Frauenfeld).

21. Marti-Grädel E, Deschler-Erb S (2004) Viehhaltung und Jagd : Ergebnisse der Untersuchung der handaufgelesenen Tierknochen. *Die Jungsteinzeitliche Seeufersiedlung Arbon, Bleiche 3 : Umwelt Und Wirtschaft*, Archäologie im Thurgau. (Kantonsarchäologie Thurgau, Frauenfeld), pp 158–252.

22. Özdoğan M (2007) Von Zentralanatolien nach Europa Die Ausbreitung der neolithischen Lebensweise. *Vor 12000 Jahren in Anatolien - Die ältesten Monumente Der Menschheit*, ed Lichter C (Badisches Landesmuseum, Karlsruhe), pp 150–160.

23. Benecke N (2001) Die Tierwelt Thrakiens im Mittelholozän (ca. 6000--2000 v. Chr.)—anthropogene und natürliche Komponenten. *Beiträge zur Archäozoologie und prähistorischen Anthropologie* 3:29–38.

24. Buitenhuis H (1997) Asikli Höyük: A "'protodomestication'"site. *Proceedings of the 7 '" ICAZ Conference (Konstanz, 26/09/94—01/10/94). Paris: Antrhopozoologica*, pp 25–26.

25. Özbaşaran M (2012) Aşıklı. *The Neolithic in Turkey. New Excavations and New Research, Vol. 3: Central Turkey*, eds Özdoğan M, Başgelen N, Kuniholm P (Archaeology and Art Publications, Istanbul), pp 135–158.

26. Stiner MC, et al. (2014) A forager–herder trade-off, from broad-spectrum hunting to sheep management at Aşıklı Höyük, Turkey. *Proc Natl Acad Sci U S A* 111(23):8404–8409.

27. Fortin M (1990) Rapport préliminaire sur la seconde campagne de fouilles à Tell 'Atij et la première à Tell Gudeda (automne 1987), sur le Moyen Khabour. *syria* 67(2):219–256.

28. Zeder MA (2003) Food Provisioning in Urban Societies: A View from Northern Mesopotamia. *The Social Construction of Ancient Cities*, ed Smith M (Smithsonian Press), pp 156–183.

29. Rufolo SJ (2011) Specialized pastoralism and urban process in third millennium BC northern Mesopotamia: A treatment of zooarchaeological data from the Khabur Basin, Syria. Dissertation (Johns Hopkins University).

30. Schibler J, Furger AR, Schmid E, Reich J, Kaufmann B (1988) *Die Tierknochenfunde aus Augusta Raurica:(Grabungen 1955-1974)* (Römermuseum Augst).

31. Furger A, Deschler-Erb S (1992) *Das Fundmaterial aus der Schichtenfolge beim Augster*

Theater : typologische und osteologische Untersuchungen zur Grabung Theater-Nordwestecke 1986/87 (Römermuseum Augst, Augst).

32. Deschler-Erb S, Plogmann HH (2012) Ein Kultmahl im privaten Kreis-zu den Tierknochen aus dem Vorratskeller der Publikumsgrabung (2008-2010.058) von Augusta Raurica (um/nach Mitte 3. Jahrhundert n. Chr.). *Jahresberichte aus Augst und Kaiseraugst* 33:237–268.

33. De Cupere B, Duru R, Umurtak G (2008) Animal husbandry at the Early Neolithic to Early Bronze Age site of Bademağacı (Antalya province, SW Turkey): evidence from the faunal remains. *Travaux de la Maison de l'Orient et de la Méditerranée* 49(1):367–405.

34. Gorbunov VS (1986) Abashev culture of Southern Ural region. *Ufa*, p 95.

35. Knipper C, et al. (2018) A knot in a network: Residential mobility at the Late Iron Age proto-urban centre of Basel-Gasfabrik (Switzerland) revealed by isotope analyses. *Journal of Archaeological Science: Reports* 17:735–753.

36. Tresset A (1996) Le rôle des relations homme/animal dans l'évolution économique et culturelle des sociétés des Ve-IVe millénaires en Bassin Parisien: approche ethno-zootechnique fondée sur les ossements animaux. Dissertation (Paris 1).

37. Bandi H-G (1963) *Birsmatten-Basisgrotte: eine mittelsteinzeitliche Fundstelle im unteren Birstal* (Stämpfli).

38. Bandi HG, Others (1963) Birsmatten-Basisgrotte. *Ein Mittelsteinzeitliche Fundstelle im unteren Birstal (Acta Bernensia I) Bern: Verlag Stämpfli and Cie Bern*.

39. Bollongino R, et al. (2013) 2000 years of parallel societies in Stone Age Central Europe. *Science* 342(6157):479–481.

40. Lipson M, et al. (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551:368.

41. Orschiedt J, Gehlen B, SCHöN W, GRöNING F (2012) The Neolithic and Mesolithic cave site Blätterhöhle at Hagen, Germany. *Notae Paehistoricae* 32:73–88.

42. Buko A, et al. (2013) A unique medieval cemetery from the 10th/11th century with chamber-like graves from Bodzia (central Poland): Preliminary result of the multidisciplinary research. *Archäologisches Korrespondenzblatt* 43(3):423–442.

43. Sobkowiak-Tabaka I, Kabaciński J (2012) Rescue excavation research conducted by the Team for Rescue Archaeological Research of the Centre for Prehistoric and Medieval Research of the Institute for Archaeology and Ethnology of the Polish Academy of Sciences (PAN) in motorway construction routes in years 2007-2008. *Raport 2007-2008* (Warszawa), pp 11–41.

44. Simonin D (1998) Boulancourt « le Châtelet », habitat fortifié du Néolithique moyen et du Bronze final. Opération de fouille programmée. *Rapport de synthèse, musée de Préhistoire d'Île-de-France, Nemours*:55.

45. Bălăşescu A, Simonin D, Vigne J-D (2008) La faune du Bronze final IIIb du site fortifié de Boulancourt «le Châtelet» (Seine-et-Marne). *Bulletin de la Société préhistorique française* 105(2):371–406.

46. Czerniak L (2002) Settlements of the Brześć Kujawski type on the Polish Lowlands.

Archeologické rozhledy 54:9–22.

47. Nielsen PO (2004) Causewayed camps, palisade enclosures and central settlements of the Middle Neolithic in Denmark. *Journal of Nordic Archaeological Science* 14:19–33.

48. Rowley-Conwy P, Dobney K (2007) Wild boar and domestic pigs in Mesolithic and Neolithic southern Scandinavia. *Pigs and Humans: 10,000 Years of Interaction*, eds Albarella U, Dobney K, Ervynck A (Oxford University Press, Oxford), pp 131–155.

49. Zeiler JT (2003) Faunaresten uit een nederzettingscomplex te Odijk, gemeente Bunnik (Midden-IJzertijd, Laat-Romeinse tijd en Vroege Middeleeuwen). *Leeuwarden: Archaeobone*:16.

50. Detry C, Cardoso JL (2010) On some remains of dog (Canis familiaris) from the Mesolithic shell-middens of Muge, Portugal. *J Archaeol Sci* 37(11):2762–2774.

51. Detry C (2007) Paleoecologia e paleoeconomia do baixo Tejo no Mesolítico Final: o contributo do estudo dos mamíferos de Muge. Universidade de Salamanca/Universidade Autónola de Lisboa. *Salamanca/Lisboa (unpublished PhD thesis)*.

52. Steadman SR, et al. (2015) The 2013 and 2014 Excavation Seasons at Çadir Höyük on the Anatolian North Central Plateau. *Anatolica* 41:87–123.

53. Steadman SR, Ross JC, McMahon G, Gorny RL (2008) Excavations on the north-central plateau: The Chalcolithic and Early Bronze Age occupation at Çadır Höyük. *Anatolian Studies* 58:47–86.

54. Arbuckle BS (2009) Chalcolithic caprines, Dark Age dairy, and Byzantine beef: a first look at animal exploitation at Middle and Late Holocene Çadır Höyük, North Central Turkey. *Anatolica* 35:179–224.

55. Schoop UD (2011) Çamlıbel Tarlası, ein metallverarbeitender Fundplatz des vierten Jahrtausends v. Chr. im nördlichen Zentralanatolien. *Anatolian Metal V Bochum: Deutsches Bergbaumuseum*:53–68.

56. Bartosiewicz L, et al. (2013) Chalcolithic pig remains from Çamlibel Tarlasi, Central Anatolia. *Archaeozoology of the Near East X: Proceedings of the Tenth International Symposium on the Archaeozoology of South-Western Asia and Adjacent Areas*, eds Cupere B, Linseele V, Hamilton-Dyer S (Editions Peeters), pp 101–120.

57. Motzoi-Chicideanu I (2016) *Cimitirul din epoca bronzului de la Cârlomăneşti (Jud. Buzău)* (Editura Academiei Române).

58. Dumitraşcu V, Bălăşescu A (2004) Fauna din complexele arheologice de la Cârlomăneşti-Armanu (jud. Buzău). *Mousaios* (9):51–56.

59. Ashley S, et al. (2007) The resources of an upland community in the fourth millennium BC. *Uplands of Ancient Sicily and Calabria. The Archaeology of Landscape Revisited.*, ed Fitzjohn M, pp 59–80.

60. Dumitrescu V (1986) Stratigrafia aşezării-tell de pe Ostrovelul de la Căscioarele. *Cultură şi Civilizaţie la Dunărea de Jos* 2:73–82.

61. Bălăşescu A, Radu V, Moise D (2005) *Omul si mediul animal intre mileniile VII-IV i.e.n. la Dunarea de Jos* (Editura Cetatea de Scaun, București).

62. Krauss R (2014) *Ovčarovo-Gorata: eine frühneolithische Siedlung in Nordostbulgarien*

(Habelt-Verlag).

63. Benecke N, Ninov L (2002) Zur Nahrungswirtschaft der neolithischen Bevölkerungen im Gebiet des heutigen Bulgariens nach archäozoologischen Befunden. *Eiträge Zu Jungsteinzeitlichen Forschungen in Bulgarien. Saarbrücker Beiträge Zur Altertumskunde*, eds Lichardus-Itten M, Lichardus J, Nikolov V (Habelt, Bonn), pp 555–573.

64. Facorellis Y (2011) Sequential radiocarbon dating and calculation of the marine reservoir effect. *The Cave of the Cyclops: Mesolithic and Neolithic Networks in the Northern Aegean, Greece Volume II: Bone Tool Industries, Dietary Resources and the Paleoenvironment, and Archeometrical Studies*:361–372.

65. Trantalidou K (2011) From Mesolithic Fishermen and Bird Hunters to Neolithic Goat Herders: The Transformation of an Island Economy in the Aegean. *The Cave of the Cyclops: Mesolithic and Neolithic Networks in the Northern Aegean, Greece, Volume II: Bone Tool Industries, Dietary Resources and the Paleoenvironment, and Archeometrical Studies*, ed Sampson A (Cambridge University Press, Cambridge), pp 53–150.

66. Trantalidou K (2014) Trends in faunal taxonomic representation during the 9th millennium in the Aegean and the ambiguous status of suids. *La Transition En Méditerranée. Comment Des Chasseurs Devinrent Agriculteurs. International Conference*, eds Manen C, Perrin T, Guilaine J (Muséum d'Histoire Naturelle de Toulouse), pp 141–163.

67. Moundrea-Agraphioti A (2011) The Mesolithic and Neolithic Bone Implements. *The Cave of Cyclops. Mesolithic and Neolithic Networks in the Northern Aegean, Greece*, ed Sampson A (INSTAP Academic Press, Philadelphia), pp 3–49.

68. Mylona D (2011) Fish vertebrae. *The Cave of Cyclops. Mesolithic and Neolithic Networks in the Northern Aegean, Greece*, ed Sampson A (INSTAP Academic Press, Philadelphia), pp 237–266.

69. Powell J (2011) Non-Vertebral Fish Bones. *The Cave of Cyclops. Mesolithic and Neolithic Networks in the Northern Aegean, Greece*, ed Sampson A (INSTAP Academic Press, Philadelphia), pp 151–235.

70. Bıçakçı E (1998) An essay on the chronology of the PrePottery Neolithic settlements of the Taurus Region (Turkey) with the building remains and 14C dates. *Light on Top of the Black Hill*, eds Arsebük G, Mellink MJ, Schirmer W (Ege Yayınları, Istanbul), pp 137–150.

71. Erim-Özdoğan A (2012) Çayönü. *The Neolithic in Turkey: New Excavations & New Research*, eds Özdoğan M, Başgelen N, Kuniholm P (Archaeology & Art Publications, Istanbul), pp 185–269.

72. Özdoğan M, Özdoğan A (1990) Çayönü: a conspectus of recent work. *Préhistoire de Levant II*, eds Aurenche O, Cauvin MC (Éditions du CNRS, Paris), pp 387–396.

73. Primas M (1985) *Cazis-Petrushügel in Graubünden* (Juris, Zürich).

74. Bougard F (1997) Fouilles de l'École française de Rome à Leopolis-Cencelle (information). *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 141(2):479–490.

75. Minniti C (2016) Meat consumption as identity of status and profession during the

Middle Ages: a review of the zooarchaeological evidence from Rome and Latium (Italy). *Objects, Environment, and Everyday Life in Medieval Europe*, eds Jervis B, Broderick LG, Grau Sologestoa I (Brepols Publisher, Turnhout), pp 137–156.

76. Esmonde Cleary AS (2013) *Chedworth: Life in a Roman Villa* (The History Press).

77. Czebreszuk J, Szmyt M (2011) Kurhan, rów i palisada w tradycji kultury ceramiki sznurowej. Próba rekonstrukcji scenariuszy rytualnych. *Kurhany I Obrządek Pogrzebowy W IV-II Tysiącleciu P.n.e.*, eds Kowalewska-Marszałek H, Włodarczak P (Instytut Archeologii i Etnologii Polskiej Akademii Nauk Instytut Archeologii Uniwersytetu Warszawskiego, Warsaw), pp 123–131.

78. Bonsall C, et al. (2012) Interrelationship of age and diet in Romania's oldest human burial. *Naturwissenschaften* 99(4):321–325.

79. Bonsall C, et al. (2016) The "Clisurean"finds from Climente II cave, Iron Gates, Romania. *Quat Int* 423:303–314.

80. Bonsall C (2008) The Mesolithic of the iron gates. *Mesolithic Europe*:238–279.

81. Bonsall C, et al. (2015) New AMS 14C Dates for Human Remains from Stone Age Sites in the Iron Gates Reach of the Danube, Southeast Europe. *Radiocarbon* 57(1):33–46.

82. Bond JM, O'Connor TP (1999) Bones from medieval deposits at 16-22 Coppergate and other sites in York. *The Archaeology of York*, 15. (Council for British Archaeology, York).

83. Domanska L (1991) Obozowisko kultury janislawickiej w Dębach, woj. *Wloclawskie* stanowisko(Inowroclaw):29.

84. Wąs M (2008) O dystrybucji „czekolady" w kulturze janisławickiej z perspektywy technologii krzemieniarstwa. *Krzemień Czekoladowy W Pradziejach", Studia Nad Gospodarką Surowcami Krzemiennymi W Pradziejach*, eds Borkowski W, Libera. J, Sałacińska B, Sałaciński S (Warszawa), pp 171–184.

85. Arbuckle BS, Erek CM (2012) Late Epipaleolithic hunters of the central Taurus: Faunal remains from Direkli Cave, Kahramanmaraş, Turkey. *Int J Osteoarchaeol* 22(6):694–707.

86. Erek CM (2010) A new Epi-paleolithic site in the Northeast Mediterranean region: Direkli Cave (Kahramanmaras, Turkey). *Adalya* 13:1–17.

87. Lichardus I, et al. (2003) Bericht über die bulgarisch-deutschen Ausgrabungen in Drama (1996-2002). *Bericht der Römisch-Germanischen Kommission* 84:157–221.

88. Damian P, Bâltâc A (2007) *The Civil Roman Settlement at Ostrov-Durostorum* (Muzeul Brăilei).

89. Bălăşescu A, Radu V (2013) Studiul faunei din epoca romană de la Ostrov, punctul Ferma 4, județul Constanța. *Cercetări Arheologice* 20:231–250.

90. Wainwright GJ, Longworth IH (1971) Durrington Walls, Excavations 1966--1968 (--Reports of the Research Committee of the Society of Antiquaries of London 29). London. *Society of Antiquaries*.

91. Pearson MP, Stonehenge Riverside Project (England) (2012) *Stonehenge: Exploring the Greatest Stone Age Mystery* (Simon and Schuster).

92. Craig OE, et al. (2015) Feeding Stonehenge: cuisine and consumption at the Late Neolithic site of Durrington Walls. *Antiquity* 89(347):1096–1109.

93. Viner S, Evans J, Albarella U, Parker Pearson M (2010) Cattle mobility in prehistoric Britain: strontium isotope analysis of cattle teeth from Durrington Walls (Wiltshire, Britain). *J Archaeol Sci* 37(11):2812–2820.

94. Wright E, Viner-Daniels S, Parker Pearson M, Albarella U (2014) Age and season of pig slaughter at Late Neolithic Durrington Walls (Wiltshire, UK) as detected through a new system for recording tooth wear. *J Archaeol Sci* 52:497–514.

95. Albarella U, Serjeantson D (2002) A passion for pork: meat consumption at the British Late Neolithic site of Durrington Walls. *Consuming passions and patterns of consumption*:33–49.

96. Breu W (1986) Tierknochenfunde aus einer germanischen Siedlung bei Eggolsheim in Oberfranken:(2.-5. Jh. n. Chr.). Dissertation (Verlag nicht ermittelbar).

97. Kaufmann D B14 Eilsleben. *Archäologie in Der Deutschen Demokratischen Republik. Denkmale Und Funde. Band 2 Fundorte Und Funde.*, Fundorte und Funde., ed Hermann J (Urania-Verlag, Leipzig, Jena, Berlin), pp 410–411.

98. Arsuaga JL, et al. (2014) Neandertal roots: Cranial and chronological evidence from Sima de los Huesos. *Science* 344(6190):1358–1363.

99. Carretero JM, et al. (2008) a late pleistocene-early Holocene archaeological sequence of portalón de Cueva mayor (Sierra de atapuerca, burgos, Spain). *Munibe* 59:67–80.

100. Galindo-Pellicena MA, Carretero JM, Arsuaga JL (2014) Primary or Secondary Products?: the Nature of Capra and Ovis Exploitation within the Chalcolithic and Bronze Age Levels at Portalón (Atapuerca Hill, Burgos, Spain). Greenfield H. Animal Secondary Products.

101. Alday A, et al. (2017) Pottery with ramiform-anthropomorphic decoration from El Portalón de Cueva Mayor site (Sierra de Atapuerca, Burgos) and the globalized symbolic world of the first Neolithic. *Quat Int*. doi:10.1016/j.quaint.2017.10.044.

102. Pérez-Romero A, et al. (2017) An unusual Pre-bell beaker copper age cave burial context from El Portalón de Cueva Mayor site (Sierra de Atapuerca, Burgos). *Quat Int* 433:142–155.

103. Sverrisdóttir OÓ, et al. (2014) Direct estimates of natural selection in Iberia indicate calcium absorption was not the only driver of lactase persistence in Europe. *Mol Biol Evol* 31(4):975–983.

104. Günther T, et al. (2015) Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A* 112(38):11917–11922.

105. Valdiosera C, et al. (2018) Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc Natl Acad Sci U S A* 115(13):3428–3433.

106. Anderung C, et al. (2005) Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proc Natl Acad Sci U S A* 102(24):8431–8435.

107. Lira J, et al. (2010) Ancient DNA reveals traces of Iberian Neolithic and Bronze Age

lineages in modern Iberian horses. *Mol Ecol* 19(1):64–78.

108. Bordaz J, Bordaz L (1976) Erbaba excavations, 1974. *Türk Arkeoloji Dergisi* 23(2):39–44.

109. Bordaz, Alper-Bordaz (1979) The prehistory of the Beyshehir-Suglu Basin Turkey: Excavations at Erbaba. *National Geographic Research Reports*:157–165.

110. Bordaz J, Bordaz LA (1982) Erbaba: the 1977 and 1978 seasons in perspective. *Türk Arkeoloji Dergisi* 26(1):85–93.

111. Bordaz J (1973) Current Research in the Neolithic of South Central Turkey: Suberde, Erbaba and Their Chronological Implications. *Am J Archaeol* 77(3):282–288.

112. Perkins D (1973) The Beginnings of Animal Domestication in the near East. *Am J Archaeol* 77(3):279–282.

113. Makarewicz CA (1999) The faunal remains of Neolithic Erbaba: processes of domestication and the herding economy. Dissertation (Brandeis University).

114. Arbuckle BS (2013) The late adoption of cattle and pig husbandry in Neolithic Central Turkey. *J Archaeol Sci* 40(4):1805–1815.

115. Arbuckle BS, et al. (2014) Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS One* 9(6):e99845.

116. Saguì L (2001) L'esedra della Crypta Balbi tra tardo antico e alto medioevo. *Roma dall'antichità al medioevo: archeologia e storia nel Museo Nazionale Romano Crypta Balbi*:593–595.

117. Saguì L (2013) L'esedra della Crypta Balbi: archeologia, storia, musealizzazione. *Forma Urbis* 18(2):14–21.

118. De Grossi Mazzorin J, Minniti C (2001) L'allevamento e l'approvvigionamento alimentare di una comunità urbana. L'utilizzazione degli animali a Roma tra il VII e il X secolo. *Roma Dall'antichità Al Medioevo*, eds Arena MS, et al. (Archeologia e Storia nel Museo Nazionale Romano, Rome), pp 69–78.

119. Eingartner J, Eschbaumer P, Weber G, Engelien-Schmidt I (1993) *Der römische Tempelbezirk in Faimingen-phoebiana* (Zabern).

120. Andersen SH (2000) "Køkkenmøddinger"(shell middens) in Denmark: a survey. *Proceedings of the Prehistoric Society* (Cambridge University Press), pp 361–384.

121. Luzi C, Courtin J (2001) La céramique des niveaux préchasséens de la baume Fontbrégoua (Salernes, Var). *Bulletin de la Société préhistorique française* 98(3):471–484.

122. Rowley-Conwy P, Gourichon L, Helmer D, Vigne J-D (2013) Early Domestic Animals in Italy, Istria, the Tyrrhenian Islands and Southern France. *The Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds Colledge S, Conolly J, Dobney K, Manning K, Shennan S (Left Coast Press Editions), pp 161–194.

123. Jacobsen TW, Farrand WR, Farrand WR, Cooper FA, Vitaliano CJ (1987) *Franchthi Cave and Paralia: Maps, Plans, and Sections* (Indiana University Press).

124. Gramsch B (1992) Friesack mesolithic wetlands. *The Wetland Revolution in*

Prehistory. Conference, pp 65–72.

125. Sarti L, Balducci C, Brilli P, Corridi C (2002) Frilli-area C: un sito della media età del Bronzo a Sesto Fiorentino. *Rivista di scienze preistoriche* (52):261–292.

126. Smith PEL (1974) Ganj Dareh Tepe. *Paléorient* 2(1):207–209.

127. Zeder MA, Hesse B (2000) The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* 287(5461):2254–2257.

128. BOLEnDER DJ, Steinberg JM, Damiata BN (2011) Farmstead relocation at the end of the Viking Age: results of the Skagafjörður archaeological settlement survey. *Archaeologia Islandica* 9:77–99.

129. Peters J, Schmidt K (2004) Animals in the symbolic world of Pre-Pottery Neolithic Göbekli Tepe, south-eastern Turkey: a preliminary assessment. *Anthropozoologica* 39(1):179–218.

130. Schmidt K (2006) *Sie bauten die ersten Tempel: das rätselhafte Heiligtum der Steinzeitjäger: die archäologische Entdeckung am Göbekli Tepe* (Beck).

131. Ervynck A, Dobney K, Hongo H, Meadow R (2001) Born Free? New Evidence for the Status of" Sus scrofa" at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient*:47–73.

132. Peters J, Arbuckle BS, Pöllath N (2014) Subsistence and beyond: animals in Neolithic Anatolia. *Neolithic in Turkey Istanbul: Arkeoloji ve Sanat Yayınları*.

133. Mahfroozi A (2007) The archaeology of eastern Mazandaran relying on excavations at Gohar Tepe. *Archaeological Reports* 7(2):347–368.

134. Soltysiak A, Mahfroozi A (2008) Short Fieldwork Report: Gohar Tepe and Goldar Tepe (Iran), seasons 2006–2007. *Bioarchaeology of the Near East* 2:71–77.

135. Brukner B (1988) *Die Siedlung der Vinča-Gruppe auf Gomolava (die Wohnschicht des Spätneolithikums und Frühäneolithikums-Gomolava Ia, Gomolava Ia-b und Gomolava Ib) und der Wohnhorizont des Äneolithischen Humus (Gomolava II)*.

136. Boric D (2009) Absolute dating of metallurgical innovations in the Vinča culture of the Balkans. *Metals and Societies : Studies in Honour of Barbara S. Ottaway*, eds Kienlin TL, Roberts BW (Habelt, Bonn).

137. Orton D (2012) Herding, Settlement, and Chronology in the Balkan Neolithic. *European Journal of Archaeology* 15(1):5–40.

138. Clason AT (1979) The Farmers of Gomolava in the Vinca and La Tene Period. *Palaeohistoria* 21:42–81.

139. Orton DC (2008) Beyond hunting and herding: humans, animals, and the political economy of the Vinča period. Dissertation (University of Cambridge).

140. Voigt MM (1994) Excavations at Gordion 1988-89: The Yassihoyuk Stratigraphic Sequence. *Anatolian Iron Ages 3: The Proceedings of the Third Anatolian Iron Ages Colloquium Held at Van, 6-12 August 1990= Anadolu Demir Caglari 3= III. Anadolu Demir Caglari Sempozyumu Bildirileri, Van, 6-12 Agustos 1990* (The British Institute of Archaeology at Ankara), pp 265–293.

141. Miller NF, Zeder MA, Arter SR (2009) From Food and Fuel to Farms and Flocks: The Integration of Plant and Animal Remains in the Study of the Agropastoral Economy at Gordion, Turkey. *Curr Anthropol* 50(6):915–924.

142. König E (1993) Tierknochenfunde aus einer Feuchtbodensiedlung der Chamer Gruppe in Griesstetten, Ldkr. Neumarkt. Dissertation (LMU-München).

143. Kaiser T, Forenbaher S (2016) Navigating the Neolithic Adriatic. *Fresh Fields and Pastures New*, eds Lillios KT, Chazan M (Sidestone Press, Leiden), pp 145–164.

144. Petrić N (2004) Kultura Gudnja i primjeri importa u neolitiku Dalmacije. *Prilozi Instituta za arheologiju u Zagrebu* 21(1):197–207.

145. Minichreiter K (2005) The first farmers of northern Croatia. *A Short Walk through the Balkans: The First Farmers of the Carpathian Basin and Adjacent Regions* (Institute of Archaeology, University College London, London).

146. Mortensen P (1975) Survey and soundings in the Holailan Valley 1974. *Proceedings of the Third Annual Symposium on Archaeological Research in Iran*, pp 1–2.

147. Zeder M (2008) Animal domestication in the Zagros: an update and directions for future research. *Travaux de la Maison de l'Orient et de la Méditerranée* 49(1):243–277.

148. Schmidt K (1995) Investigations in the Upper Mesopotamian Early Neolithic: Göbekli Tepe and Gürcütepe. *Neo-Lithics* 2(95):9–10.

149. Schmidt K (2000) Göbekli Tepe, Southeastern Turkey: A Preliminary Report on the 1995-1999 Excavations. *Paléorient* 26(1):45–54.

150. Mashkour M (2006) Boars and pigs: A view from the Iranian plateau. *De La Domestication Au Tabou: Le Cas Des Suidés Dans Le Proche-Orient Ancient. Travaux de La Maison d'Archéologie et d'Ethnologie de Nanterre 1*, eds Lion B, Michel C (De Boccard), pp 155–163.

151. Mohaseb Karimlu F (2012) Exploitation des animaux de l'Âge du Bronze au début de la période Islamique dans le Nord-ouest de l'Iran: l'étude archéozoologique de Haftavan Tepe. Dissertation (Paris 1).

152. Mohaseb AF, Mashkour M (2017) Animal Exploitation from the Bronze Age to the Early Islamic Period in Haftavan Tepe (Western Azerbaijan--Iran). *Archaeozoology of the Near East 9 Proceedings of the 2008 Al Ain-Abu Dhabi Conference*, eds Mashkour M, Beech MJ (Oxbow Books Limited), pp 146–170.

153. Rosenberg M, Redding RW (2002) Hallan Çemi and Early Village Organization in Eastern Anatolia. *Life in Neolithic Farming Communities: Social Organization, Identity, and Differentiation*, ed Kuijt I (Springer US, Boston, MA), pp 39–62.

154. Redding RW (2005) Breaking the mold: A consideration of variation in the evolution of animal domestication. *First Steps of Animal Domestication New Archaeozoological Approaches*, eds Vigne J-D, Peters J, Helmer D (Oxbow Books, Oxford), pp 41–48.

155. Popovici D, Haşotti P, Galbenu D, Nicolae C (1992) Cercetările arheologice din tell-ul de la Hârşova (1988). *Cercetări Arheologice* 9:8–18.

156. Popovici D, et al. (2000) Les recherches archéologiques du tell de Hârşova (dép. de Constanţa) 1997–1998. *Cercetari Arheologice* 11(1):13–123.

157. Balasescu A, Radu V, Moise D (2005) *Omul si mediul animal intre mileniile VII-IV ien la Dunarea de Jos*.

158. Kussinger S (1988) Tierknochenfunde von Lidar Höyük in Südostanatolien (Grabungen 1979-86).-229 S. Dissertation (Dissertation Universität München).

159. Stahl U (1989) Tierknochenfunde vom Hassek Höyük (Südostanatolien). Dissertation (Ludwig-Maximilians-Universität).

160. Boessneck J (1992) Besprechung der Tierknochen-und Molluskenreste von Hassek Höyük. *Hassek Höyük: naturwissenschaftliche Untersuchungen und lithische Industrie*:58–74.

161. Ottoni C, et al. (2013) Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol Biol Evol* 30(4):824–832.

162. Fewkes VJ (1932) Excavations in the Late Neolithic Fortress of Homolka in Bohemia A Preliminary Report. *Proc Am Philos Soc* 71(6):357–392.

163. Ehrich RW, Pleslová-Štiková E (1968) *Homolka: an eneolithic site in Bohemia* (Peabody Museum, Oxford).

164. Bogucki PI (1979) Mammal Remains from HutB at the Eneolithic Settlement of Homolka (Bohemia). *Archeologické Rozhledy Praha* 31(1):83–92.

165. van der Velde HM (2001) Houten, Hoogdijk terrein 89. *Onderzoek in Het Kader van de Vinexlocatie Loerik, Hofstad II* (Archeologisch Diensten Centrum, Bunschoten).

166. Grönvold K, et al. (1995) Ash layers from Iceland in the Greenland GRIP ice core correlated with oceanic and land sediments. *Earth Planet Sci Lett* 135(1):149–155.

167. Sigurgeirsson MÁ, Hauptfleisch U, Newton A, Einarsson Á (2013) Dating of the Viking Age Landnám Tephra Sequence in Lake Mývatn Sediment, North Iceland. *Journal of the North Atlantic* (21):1–11.

168. Ascough PL, et al. (2007) Reservoirs and radiocarbon: 14C dating problems in Mývatnssveit, Northern Iceland. *Radiocarbon* 49:947–961.

169. Boroneanț A, Bonsall C (2016) The Icoana burials in context. *Mesolithic burials--rites, symbols and social organisation of early Postglacial communities* 2:757–780.

170. Girdland-Flink L (2013) Investigating patterns of animal domestication using ancient DNA. Doctoral (Durham University).

171. Braidwood LS (1983) *Prehistoric archeology along the Zagros Flanks* (Oriental Inst Pubns Sales).

172. Flannery KV (1983) Early pig domestication in the Fertile Crescent: A retrospective look. *The Hilly Flanks and Beyond: Essays on the Prehistory of Southwestern Asia Studies in Ancient Oriental Civilization* 36:163–188.

173. Price MD, Arbuckle BS (2015) Early Pig Management in the Zagros Flanks: Reanalysis of the Fauna from Neolithic Jarmo, Northern Iraq. *Int J Osteoarchaeol* 25(4):441–453.

174. Stampfli HR (1983) The fauna of Jarmo with notes on animal bones from Matarrah,

the Amuq, and Karim Shahir. *Prehistoric Archaeology along the Zagros Flanks Oriental Institute Publications* 105:431–483.

175. Vagedes K (2001) Die Tierknochenfunde aus Karlburg: ein Vergleich zwischen Burg und Talsiedlung. *Karlburg--Rosstal--Oberammerthal: Studien zum Frühmittelalterlichen Burgenbau in Nord Bayern, Frühgeschichtliche und ProvinzialRömische Archäologie Materialen und Forschungen* 5:305–315.

176. Manaseryan N (2013) Armenia: wild boar in all issues. *Studii de Preistorie* 10:245–248.

177. Vigne J-D, et al. (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proceedings of the National Academy of Sciences* 109(22):8445–8449.

178. Vigne J-D, et al. (2017) Klimonas, a late PPNA hunter-cultivator village in Cyprus: new results. *Nouvelles Données Sur Les Débuts Du Néolithique à Chypre / New Data on the Beginnings of the Neolithic in Cyprus*, séances en ligne., eds Vigne J-D, Briois F, Tengberg M (Société Préhistorique Française, Paris), pp 21–46.

179. Vigne J-D, et al. (2009) Pre-Neolithic wild boar management and introduction to Cyprus more than 11,400 years ago. *Proc Natl Acad Sci U S A* 106(38):16135–16138.

180. Zazzo A, Lebon M, Quiles A, Reiche I, Vigne J-D (2015) Direct Dating and Physico-Chemical Analyses Cast Doubts on the Coexistence of Humans and Dwarf Hippos in Cyprus. *PLoS One* 10(8):e0134429.

181. Vigne J-D, Carrère I, Briois F, Guilaine J (2011) The Early Process of Mammal Domestication in the Near East: New Evidence from the Pre-Neolithic and Pre-Pottery Neolithic in Cyprus. *Curr Anthropol* 52(S4):S255–S271.

182. Seidel U (2008) *Michelsberger Erdwerke im Raum Heilbronn: Neckarsulm-Obereisesheim "Hetzenberg" und Ilsfeld "Ebene", Lkr. Heilbronn, Heilbronn-Klingenberg "Schloßberg", Stadtkreis Heilbronn. Text, Literatur und Anhänge* (K. Theiss, Stuttgart).

183. Seidel U, et al. (2016) Die Zeit der großen Gräben: Modelle zur Chronologie des Michelsberger Fundplatzes von Heilbronn-Klingenberg „Schlossberg", Stadtkreis Heilbronn, Baden-Württemberg. *Praehistorische Zeitschrift* 91(2). doi:10.1515/pz-2016-0022.

184. Maziar S (2015) Settlement dynamics of the Kura-Araxes culture: An overview of the Late Chalcolithic and Early Bronze Age in the Khoda Afarin Plain, North-Western Iran. *Paléorient* 41(1):25–36.

185. Füleky G, Márity E (1997) Environmental changes in Budaújlak (Pannonia Province, Hungary) in the Roman period. *Landscapes in Flux: Central and Eastern Europe in Antiquity*, Colloquia Pontica 3., eds Chapman J, Dolukhanov PM (Oxbow, Oxford), pp 231–239.

186. Márity E (1996) Kolosy tér 6.: római kor, középkor, őskor. *Régészeti Füzetek* 1(47):35.

187. Krauss R (2006) *Die prähistorische Besiedlung am Unterlauf der Jantra vor dem Hintergrund der Kulturgeschichte Nordbulgariens: Praistoriiata po dolnoto techenie na r. Iantra na fona na kulturnata istoriia na severna Bulgariia* (VML, Verlag Marie Leidorf).

188. Lichardus-Itten M, Demoule JP, Pernicheva L, Grębska-Kulova M, Kulov IK (2006) an Early Neolithic site in South-West Bulgaria and its importance for European Neolithization. *Aegean--Marmara--Black Sea: the Present State of Research on the Early Neolithic Langenweissbach* 94.

189. Honkisz B (2014) Gospodarcze i społeczne znaczenie nieudomowionych zwierząt dla społeczności wczesnoneolitycznych w wybranych rejonach Europie Środkowej. *Folia Praehistorica Posnaniensia* 19.

190. Kołodziej B (2011) Pochówki zwierzęce w neolicie na terenie ziem Polski. *Materiały i Sprawozdania* 32:55–106.

191. Ott-Luy S (1988) *Die Tierknochenfunde aus der mittelneolithischen Station von Künzing-Unternberg, Ldkr. Deggendorf*.

192. Boessneck J (1958) *Tierknocken aus spätneolithischen Siedlungen Bayerns* (Munchen).

193. Kennet D (1997) Kush: a Sasanian and Islamic-period archaeological tell in Ras al-Khaimah (UAE). *Arabian archaeology and epigraphy* 8(2):284–302.

194. L. B, F. D, Rousselet O, Manen C (2018) Early farming economy in Mediterranean France: fruit and seed remains from the Early to Late Neolithic levels of the site of Taï (ca 5300–3500 cal bc). *Vegetation History and Archaeobotany* 28(1):17–34.

195. Rova E, Weiss H (2002) The origins of North Mesopotamian Civilization, Ninevite 5 Chronology, Economy, Society (Subartu IX).

196. Ervynck A (1994) "De Burcht" te Londer-zeel. Bewonings-ge-schiedenis van een motte en een bakstenen kasteel. *Archeologie in Vlaanderen Monografie I*, ed Ervynck A (Instituut voor het Archeologisch Patrimonium, Zelik).

197. Ervynck A, Van Neer W, Van der Plaetsen P (1994) Dierlijke resten. *"De Burcht" Te Londer-zeel. Bewonings-ge-schiedenis van Een Motte En Een Bakstenen Kasteel (Archeologie in Vlaanderen Monografie I)*, ed Ervynck A (Instituut voor het Archeologisch Patrimonium, Zellik), pp 99–170.

198. Pyzel J (2012) Preliminary Results of Large Scale Emergency Excavations in Ludwinowo 7, Comm. Włocławek. *Siedlungsstruktur Und Kulturwandel in Der Bandkeramik: Beiträge Der Internationalen Tagung "Neue Fragen Zur Bandkeramik Oder Alles Beim Alten*:160–166.

199. Micu C, Maille M (2002) Recherches archéologiques dans le cadre de l'établissement-tell de Luncaviţa (dép. de Tulcea). *Studii de preistorie* 1:115–129.

200. Comşa E (1952) Săpături arheologice la Luncaviţa. *Materiale* 5:221–225.

201. Bălăşescu A (2003) L'étude de la faune des mammifères découverts a Luncaviţa. *Peuce* 1(14):453–468.

202. Commenge C (2009) Neolithic settlement patterns in the alluvial plains of Macedonia. *Ol'Man River. Geo-Archaeological Aspects of Rivers and River Plains*, Archaeological Reports Ghent University., eds De Dapper M, Vermeulen F, Deprez S, Taelman D (Ghent Academia Press, Ghent), pp 229–241.

203. Bălăşescu A (2014) Arheozoologia neo-eneoliticului de pe Valea Teleormanului. *Muzeul Naţional de Istorie a României, Colecţia Muzeului Naţional, Seria Cercetări*

Puridisciplinare (Editura Mega, Cluj Napoca), p 216.

204. Evin A, et al. (2015) Unravelling the complexity of domestication: a case study using morphometrics and ancient DNA analyses of archaeological pigs from Romania. *Philos Trans R Soc Lond B Biol Sci* 370(1660):20130616.

205. Çakırlar C (2013) Rethinking Neolithic subsistence at the gateway to Europe with new archaeozoological evidence from Istanbul. *Barely Surviving or More than Enough*, eds Groot M, Lentjes D, Zeiler J (Sidestone Press, Leiden), pp 59–79.

206. Thuesen I (1994) Mashnaqa. *Am J Archaeol* 98(1):111–112.

207. Hole FA (2007) Khabur Basin Project - 1986-2001. *Ann Archeol Syrie* (45-46):11–20.

208. Braidwood RJ, Braidwood L, Smith JG, Leslie C (1952) Matarrah: A Southern Variant of the Hassunan Assemblage, Excavated in 1948. *J Near East Stud* 11(1):1–75.

209. Mignot PH (1994) Le Château de Montaigle. *Fiche Patrimoine* 94:12.

210. Gautier A, Fiers E (2009) Restes animaux des occupations de l'âge de Fer, gallo-romaine et médiévale du site castral de Montaigle (Namur). *VA, VIE ARCHEOLOGIQUE* (68):85–100.

211. Gourichon L, Helmer D (2008) Étude archéozoologique de Mureybet. *Tell Mureybet, Un Site Néolithique Dans Le Moyen Euphrate Syrien*, ed Ibañez JJ (Archaeopress, Oxford), pp 115–227.

212. Willcox G, Fornite S, Herveux L (2008) Early Holocene cultivation before domestication in northern Syria. *Veg Hist Archaeobot* 17(3):313–325.

213. Gurina NN (1955) New Neolithic sites in eastern Estonia. *Drevnie Poseleniya I Gorodishcha* (Tallinn), pp 153–175.

214. Paaver K (1965) Theriofauna genesis and variability of mammals in Baltic Republic in Holocene. *Institut Zoologii I Botaniki* (Tartu), p 494 pp. + XII tabl.

215. Mol D, Post K, Reumer JWF, de Vos J, Laban C (2003) Het Gat: preliminary note on a Bavelian fauna from the North Sea with possibly two mammoth species. *Deinsea* 9(1):253–266.

216. Mol D, et al. (2004) The Eurogeul - first report of paleontological, palynological and archeological investigations of this part of the North Sea. *Quat Int* 142-143(2006):178–185.

217. Mol D, Reumer JWF (2010) The North Sea--Maasvlakte 2 project. *Quaternaire, Hors Serie, 3. Volume of Abstracts of the Vth International Conference on Mammoths and Their Relatives, Le Puyen-Velay, August 30--September 5*.

218. Kolfschoten T van, Laban C, Others (1995) Pleistocene terrestrial mammal faunas from the north Sea area. *Mededelingen Rijks Geologische Dienst* 52:135.

219. Reumer JWF, et al. (2003) Late Pleistocene survival of the saber-toothed cat Homotherium in northwestern Europe. *J Vert Paleontol* 23(1):260–262.

220. Mol D, et al. (2008) *Kleine encyclopedie van het leven in het Pleistoceen: mammoeten, neushoorns en andere dieren van de Noordzeebodem* (Veen Magazines,

Diemen).

221. Kortenbout van der Sluijs G (1971) Bones of mammals from the Brown Bank area (North Sea). *Mesolithic bone and antler implements from the North Sea and from the Netherlands Berichten van de Rijksdienst voor het Oudheidkundig Bodemonderzoek*:20–21.

222. Van Der Sluijs G. K (1971) Fossil bones from the North Sea.- in: Proceedings Nederlandse Dierkundige Vereniging (Netherlands Zoological Society). *Neth J Zool* 21(2):214.

223. Mordant D, Mordant C (1992) Noyen-sur-Seine: a Mesolithic waterside settlement. *The Wetland Revolution in Prehistory. Conference*, pp 55–64.

224. Mordant D, Valentin B, Vigne J-D (2013) Noyen-sur-Seine, twenty-five years on. *Mesolithic Palethnography Research on Open-Air Sites between Loire and Neckar (Proceedings from the International Round-Table Meeting, Paris, November 26 – 27, 2010)*, Séances de la Société préhistorique française, 2012., eds Valentin B, et al. (Société préhistorique française, Paris), pp 37–49.

225. Marinval-Vigne M-C, et al. (1993) Archéologie et paléo-environnement: Noyen-sur-Seine (Seine-et-Marne). *Mémoires du Groupement archéologique de Seine-et-Marne* (1):21–31.

226. Bridault A, et al. (2000) Wild boar - age at death estimates: the relevance of new modern data for archaeological skeletal material. 1. Presentation on the corpus. Dental and epiphyseal fusion ages. *Anthropozoologica* 31:11–38.

227. Hofmann R (2007) *Prospektionen und Ausgrabungen in Okoliste (Bosnien-Herzegowina): Siedlungsarchäologische Studien zum zentralbosnischen Spätneolithikum (5300-4500 v. Chr.)* (Bericht der Römisch-Germanischen Kommission).

228. Vanhoutte S (2015) The late Roman coastal fort of Oudenburg (Belgium): spatial and functional transformations within the fort walls. *Roman Military Architecture on the Frontiers. Armies and Their Architecture in Late Antiquity*, eds Collins R, Symonds M, Weber M (Oxbow Books, Oxford & Philadelphia), pp 62–75.

229. Ervynck A, et al. (2017) Het castellum aan het eind van de wereld. Vlees en vis voor de soldaten te Oudenburg. *Signa* 6:41–45.

230. Roudil J-L (1987) Le gisement Néolithique de la Baume d'Oullins le Garn-Gard. *Premières Communautés Paysannes En Méditerranée Occidentale*, eds Guilaine J, Courtin J, Roudil J-L, Vernet J-L (CNRS Éditions), pp 523–529.

231. Roudil JL, Van Willigen S (2002) La céramique du Néolithique ancien de la Baume d'Oullins (Le Garn, Gard). *Ardèche Archéologie* 19:1–26.

232. Helmer D, Gourichon L, Sidi Maamar H, Vigne J-D (2005) L'élevage des caprinés néolithiques dans le sud-est de la France: saisonnalité des abattages, relations entre grottes-bergeries et sites de plein air. *Anthropozoologica* 40(1):167–189.

233. Turnbull PF, Reed CA (1974) The Fauna from the Terminal Pleistocene of Palegawra Cave, A Zarzian Occupation Site in Northeastern Iraq. *Fieldiana* 63(3):81–146.

234. Radi G, Tozzi C, Arobba D, Battentier J (2017) Modelling the earliest north-western dispersal of Mediterranean Impressed Wares: new dates and Bayesian chronological model. *Documenta Praehistorica* 44:54–77.

235. Benecke N, et al. (2013) Pietrele in the Lower Danube region: integrating archaeological, faunal and environmental investigations. *Documenta Praehistorica* 40:175.

236. Garašanin MV (1951) *Hronologija vinčanske grupe* (Univerza, Arheološki seminar, Ljubljana).

237. Radivojević M, Kuzmanović-Cvetković J (2014) Copper minerals and archaeometallurgical materials from the Vinča culture sites of Belovode and Pločnik: overview of the evidence and new data. *Starinar* (64):7–30.

238. Stalio B (1960) Plocnik-Prokuplje-naselje. *Arheološki pregled* 2:33–36.

239. Palincaş N (1996) Valorificarea arheologică a probelor 14C din fortificaţia aparţinând Bronzului târziu de la Popeşti (jud. Giurgiu). *SCIVA* 47(3):239–288.

240. Dumitrascu V (2004) THE 2001 CAMPAIGN IN THE EARLY IRON AGE SETTLEMENT AT POPE: A ZOOARCHAEOLOGICAL ANALYSIS. *Dacia* 48:71–75.

241. Miracle P (1997) Early Holocene foragers in the karst of northern Istria. *Poročilo O Raziskovanju Paleolita, Neolita in Eneolita v Sloveniji*, XXIV., pp 43–61.

242. Czerniak L, et al. (2016) House time: Neolithic settlement development at Racot during the 5th millennium CAL BC in the Polish lowlands. *J Field Archaeol* 41(5):618–640.

243. Rieckhoff S (1987) *Archäologisches Museum im BMW-Werk Regensburg* (Bayer. Motoren-Werke Aktienges.).

244. Vagedes K, Peters J (2001) Die Faunenreste aus der karolingisch-ottonischen Reichsburg in Rosstal, Landkreis Fürth. *Studien Zum Frühmittelalterlichen Burgbau in Nordbayern*, eds Ettel P, Karlburg-Rosstal-Oberammertal. (Rahden/Westf., Leidorf), pp 317–339.

245. Lesur J, Gasco J, Tresset A, Vigne J-D (2001) Un approvisionnement chasséen caussenard exclusivement fondé sur la chasse? La faune de Roucadour (Lot). *Préhistoire du Sud-Ouest* (8):71–90.

246. De Cupere B (2001) *Animals at ancient Sagalassos evidence of the faunal remains* (Brepols, Turnhout).

247. Malez M (1964) Šandalja bei Pula-ein neuer und wichtiger paläolithischer fundort in Istrien. *Bulletin Scientifique du Conseil des Académies des Sciences et des Arts de la RSF Yougoslavie* 9(6):154–155.

248. Miracle PT (1995) Broad-spectrum adaptations re-examined: Hunter-gatherer responses to Late Glacial environmental changes in the eastern Adriatic. Dissertation (The University of Michigan).

249. Oros Sršen A, Brajković D, Radović S, Mauch Lenardić J, Miracle PT (2014) The Avifauna of Southern Istria (Croatia) During the Late Pleistocene: Implications for the Palaeoecology and Biodiversity of the Northern Adriatic Region. *Int J Osteoarchaeol* 24(3):289–299.

250. Freguglia M, Lo Vetro D, Volante N (2005) Santa Maria in Selva di Treia, Macerata: l'area 1. *Atti della XXXVIII Riunione scientifica*:856–860.

251. Dinu A, Soficaru A, Miritoiu D (2007) The Mesolithic at the Danube's Iron Gates: new radiocarbon dates and old stratigraphies. *Documenta Praehistorica* 34(31):e52.

252. Beregovaya NF (1960) Paleolithic sites of USSR. *Materialy i Issledovaniya po Archeologii SSSR* 81:1–218.

253. Benecke N (2006) Zur Datierung der Faunensequenz am Abri Šan-Koba (Krim, Ukraine). *Beiträge zur Archäozoologie und Prähistorischen Anthropologie* 12:15.

254. Bibikov SN, Stanko VN, Koen VY (1994) Final Paleolithic and Mesolithic of Crimean mountain. *Vest, Odessa*:238.

255. Kośko A, Kirkowski R (1997) Siniarzewo, st. 1, gm. Zakrzewo, woj. włocławskie, AZP 46-44/115. *Informator Archeologiczny: badania* 31.

256. Vogler U (1997) Faunenhistorische Untersuchungen am Sirkeli Höyük/Adana, Türkei. Diss. med. vet. (LMU-München).

257. Larina OV, Wechler KP, Dergačev VA, Kovalenko SI, Bikbaev VM (1997) New field surveys on the Mesolithic and Neolithic of Moldova. *Vestigii Arheologice din Moldova*:62–110.

258. Yanevich OO (1993) Shpan Mesolithic culture. *Archeologia* `:3–15.

259. Benecke N (2000) Mesolithic hunters of the Crimean mountains: The fauna from the rock shelter of Shpan'-Koba. *Archaeozoology of the Near East*, eds Mashkour M, Choyke AM, Buitenhuis H, & Poplin F (Groningen), pp 107–120.

260. Martini F, Poggiani Keller R, Sarti L Bell Beaker in everyday life. *Proceedings of the 10th Meeting Archéologie et Gobelets* (Museo Fiorentino di Preistoria , Firenzevia).

261. Sarti L (2006) Presenze di vasi a bocca quadrata in Italia centrale: problemi e prospettive. *Preistoria dell'Italia Settentrionale*, eds A. P, Visentini P (Mueo Friulano di Storia naturale, Udine), pp 193–210.

262. Volante N (2014) La cultura dei vasi a Bocca quadrata e il pieno Neolitico. *A Cura Di, Passaggi a Nord Ovest*, eds Poggesi G, Sarti L (Millenni 10, Firenze), pp 46–50.

263. Kuz'min SL (2008) Ladoga in early Middle Ages (middle VIII – beginning XII century). *Issledovanie Archeologicheskikh Pamyatnikov Epokhi Srednevekoviya*, ed Vinogradov AV (St. Petersburg), pp 69–94.

264. Bordaz J (1970) *The Suberde excavations, Southwestern Turkey: an interim report* (Turk Tarih Kurumu Basimevi).

265. Solecki RS (1964) An archaeological reconnaissance in the Beysehir-Sugla area of south western Turkey'. *Turk Arkeoloji Dergisi* 13(1):129–125.

266. Arbuckle BS (2008) Revisiting Neolithic Caprine Exploitation at Suberde, Turkey. *J Field Archaeol* 33(2):219–236.

267. Perkins D Jr, Daly P (1968) A hunters' village in neolithic Turkey. *Sci Am* 219(5):97–106.

268. King T (2017) Analysis of the suid (Sus scrofa) faunal assemblage from Neolithic Suberde Höyük, Southwest Turkey. Curriculum in Archaeology (UNC-CH).

269. McComish JM (2015) Archaeological excavations at 12-18 Swinegate, 14 Little Stonegate and 18 Back Swinegate.

270. Mazar A, Amitai-Preiss N (2006) *Excavations at Tel Beth-Shean 1989-1996* (Israel Exploration Society, Jerusalem).

271. Horwitz LK Mammalian remains from Areas H, L, P and Q. *Excavations at Tel Beth-Shean 1989-1996 Vol. I. From the Late Bronze Age IIB to the Medieval Period.*, ed Mazar A (Israel Exploration Society, Jerusalem), pp 689–710.

272. Maeir AM (2012) *Tell es-Safi/Gath I: Report on the 1996--2005 Seasons. Ägypten und Altes Testament 69* (MANFRED GÖRG, MÜNCHEN).

273. Meiri M, et al. (2013) Ancient DNA and population turnover in southern levantine pigs--signature of the sea peoples migration? *Sci Rep* 3:3035.

274. Gitin S (2012) Tel Miqne-Ekron. Summary of Fourteen Seasons of Excavation 1981–1996 and Bibliography 1982–2012. 2012. *The Tel Miqne-Ekron Excavation and Publications Project* (W.F. Albright Institute of Archaeological Research and Hebrew University, Jerusalem).

275. Dothan T, Garfinkel Y, Gitin S (2016) *Tel Miqne-Ekron Excavations, 1985–1988, 1990, 1992–1995: Field IV Lower—The Elite Zone, Part 9/3A: The Iron Age I and IIC Early and Late Philistine Cities Database* (Eisenbrauns, Winona Lake).

276. Hesse B (1986) Animal Use at Tel Miqne-Ekron in the Bronze Age and Iron Age. *Bull Am Schools Orient Res* (264):17–27.

277. Lev-Tov JSE (2000) Pigs, Philistines, and the Ancient Animal Economy of Ekron from the Late Bronze Age to the Iron Age II. Dissertation (University of Tennessee, Knoxville).

278. Maher EF (2004) Food for the Gods: The Identification of Philistine Rites of Animal Sacrifice. PhD (University of Illinois at Chicago).

279. Maher EF, and Hesse B (2016) The Middle Bronze Age II and Iron Age I faunal remains. Tel Miqne-Ekron Excavations 1985-1988, 1990, 1992-1995. Field IV Lower – The Elite Zone. Part 1. The Iron Age I Early Philistine City. Winona Lake. *Eisenbrauns*, eds Dothan T, Garfinkel Y, and Gitin S, pp 515–542.

280. Bartosiewicz L (2016) Halaf Period Animal Remains from Tell Aqab, Northeastern Syria. *Bones and Identity: Zooarchaeological Approaches to Reconstructing Social and Cultural Landscapes in Southwest Asia*, eds Marom N, Yeshurun R, Weissbrod L, and Bar-Oz G (Oxbow Books, Oxford & Philadelphia), pp 125–155.

281. Hilditch J, Kiriatzi E, Psaraki K, Aravantinos V (2008) Early Helladic II pottery from Thebes: An integrated typological, technological and provenance study. *BAR INTERNATIONAL SERIES* 1746:263.

282. Vika E (2011) Diachronic dietary reconstructions in ancient Thebes, Greece: results from stable isotope analyses. *J Archaeol Sci* 38(5):1157–1163.

283. Aravantinos V (1997) Θήβα. *ArchDeltio* 52:354–359.

284. Mashkour M, Bailon S (2010) Animal remains from Tol-e Bashi (Ramjerd plain-Fars), a Late Neolithic / Chalcolithic settlement in South-West Iran. Social Life in a Neolithic Village. *The 2003 Excavations at Tol-E-Baši, Iran*, eds Pollock S, Bernbeck R,

Abdi K (Archäologie Mitteilungen, Turan), pp 215–229.

285. Andersen NH (1982) A Neolithic Causewayed Camp at Trelleborg near Slagelse, West Zealand. *Journal of Danish Archaeology* 1(1):31–33.

286. Davidsen K (1978) *The final TRB culture in Denmark: a settlement study* (Akademisk Forlag, Copenhagen).

287. Miller R, Zwyns N, Otte M, Stevens C, Stewart J (2012) La séquence mésolithique et néolithique du Trou Al'Wesse (Belgique): résultats pluridisciplinaires. *Anthropologie* 116(2):99–126.

288. Grundbacher HR, Stampfli B (1977) Tierknochenfunde. Erster Bericht. *Die Neolithischen Ufersiedlungen von Twann* (Band, Bern).

289. Becker C (1981) Tierknochenfunde. Dritter Bericht. *Die Neolithischen Ufersiedlungen von Twann* (Band, Bern).

290. Becker C, Johansson F (1981) Tierknochenfunde. Zweiter Bericht. *Die Neolithischen Ufersiedlungen von Twann* (Band, Bern).

291. Stampfli HR (1980) Tierknochenfunde: Dokumentation. *Die Siedlungen Der Horgener Kultur. Die Neolithischen Ufersiedlungen von Twann*, ed Furger AR (Band, Bern), pp 141–160.

292. Schier W, Draşovean F (2004) Der spätneolithisch-frühkupferzeitliche Tell von Uivar, jud. Timiş, Rumänien. Vorbericht über die Prospektionen und Ausgrabungen 1998–2002. *Praehist Zeitschr* 79:145–230.

293. El Susi G (2017) Animal Bones from the Neolithic (Szakálhát) Levels at Uivar (Timiş County). *Georgeta El Susi* 31:29–54.

294. Çilingiroğlu Ç (2012) *The Neolithic pottery of Ulucak in Aegean Turkey: organization of production, interregional comparisons and relative chronology* (BAR International Series, Oxford).

295. Çilingiroğlu Ç, Çakırlar C (2013) Towards configuring the neolithisation of Aegean Turkey. *Documenta Praehistorica* 40:21–29.

296. Çakırlar C (2012) The evolution of animal husbandry in Neolithic central-west Anatolia: the zooarchaeological record from Ulucak Höyük (c. 7040–5660 cal. BC, Izmir, Turkey). *Anatolian Studies* 62:1–33.

297. Zeder MA (1994) After the revolution: post-Neolithic subsistence in northern Mesopotamia. *Am Anthropol* 96(1):97–126.

298. Church MJ, et al. (2005) Puffins, Pigs, Cod and Barley: Palaeoeconomy at Undir Junkarinsfløtti, Sandoy, Faroe Islands. *Environ Archaeol* 10(2):179–197.

299. Arge SV, Church MJ, Brewington SD (2009) Pigs in the Faroe Islands: An Ancient Facet of the Islands' Paleoeconomy. *Journal of the North Atlantic*:19–32.

300. Brewington S, et al. (2015) Islands of change vs. islands of disaster: Managing pigs and birds in the Anthropocene of the North Atlantic. *Holocene* 25(10):1676–1684.

301. Caduff B (2005) Die ur-und frühgeschichtlichen Fundstellen auf dem Haselboden in Untervaz (GR). *Archäologie der Schweiz* 28(3):16–23.

302. Comşa E (1961) Săpăturile arheologice de la Boian. *Materiale şi cercetări arheologice* 7:63–70.

303. Bolomey A (1966) Fauna neolitica din asezarea Boian A de la Varasti. *Studii si cercetari de antropologie* 3(1):27–34.

304. Volante N (2003) Neto-Via Verga (Sesto Fiorentino): la produzione vascolare dell'area 1. *Rivista di scienze preistoriche* (53):375–504.

305. Sarti L, Volante N (2011) Il pieno Neolitico in Toscana: variabilità delle produzioni ceramiche e litiche nel contesto dell'Italia centrale. *Rivista di studi liguri* (77):497–504.

306. Coradeschi G (2013) O sítio da Idade do Bronze de Via Neruda em Sesto Fiorentino (Florença, Itália): exploração dos recursos arbóreos. *Techne* 1(1).

307. Tasić NN, Ignjatović M (1908) Od tradicionalne do moderne metodologije. Istraživanja u Vinči 1978--2008. *Vinča—praistorijska metropola, istraživanja* 2008:87–119.

308. Dimitrijević V (2006) Vertebrate fauna of Vinča-Belo Brdo: Excavation campaigns 1998-2003. *Starinar* (56):245–269.

309. Andreescu R, Mirea P, Apope Ş (2001) Cultura Gumelniţa în vestul Munteniei. Aşezarea de la Vităneşti, jud Teleorman. O Civilizaţie Necunoscută. Gumelniţa.

310. Balasse M, et al. (2016) Wild, domestic and feral? Investigating the status of suids in the Romanian Gumelniţa (5th mil. cal BC) with biogeochemistry and geometric morphometrics. *Journal of Anthropological Archaeology* 42:27–36.

311. Bréhard S, Bălăşescu A (2012) What's behind the tell phenomenon? An archaeozoological approach of Eneolithic sites in Romania. *J Archaeol Sci* 39(10):3167–3183.

312. Srejović D, Letica Z (1978) *Vlasac: A mesolithic settlement in the iron gates* (Serbian Academy of Science and Art).

313. Borić D, et al. (2014) Late Mesolithic lifeways and deathways at Vlasac (Serbia). *J Field Archaeol* 39(1):4–31.

314. Berglund J (1982) Kirkebjerget—A Late Bronze Age Settlement at Voldtofte, South-West Funen: An Interim Report on the Excavations of 1976 and 1977. *Journal of Danish Archaeology* 1(1):51–63.

315. Barker P, White R, Pretty K, Bird H, and Corbishley M (1997) *The Baths Basilica Wroxeter: Excavations 1966–90* (English Heritage, London).

316. Hammon A (2011) Understanding the Romano-British–Early Medieval Transition: A Zooarchaeological Perspective from Wroxeter (Viroconium Cornoviorum). *Britannia* 42:275–305.

317. Albarella U, Davis S, Detry C, Rowley-Conwy P (2005) Pigs of the "Far West": the biometry of Sus from archaeological sites in Portugal. *Anthropozoologica* 40:27–54.

318. Pyzel J (2001) *Naczynie mieszkowate kultury ceramiki wstęgowej rytej z Żegotek, gm. Strzelno, woj. kujawsko-pomorskie, stan. 2* (Instytut Archeologii i Etnologii Polskiej Akademii Nauk, Kraków).

319. Schibler J (1997) *Ökonomie und Ökologie neolithischer und bronzezeitlicher Ufersiedlungen am Zürichsee: Ergebnisse der Ausgrabungen Mozartstrasse, Kanalisationssanierung Seefeld, AKAD/Pressehaus und Mythenschloss in Zürich* (Direktion der Öffentlichen Bauten des Kantons Zürich, Hochbauamt, Abt. Kantonsarchäologie).

320. Ebersbach R, Ruckstuhl B, Bleicher N (2015) Zürich "Mozartstrasse". Neolithische und bronzezeitliche Ufersiedlungen. Band 5: Die neolithischen Befunde und die Dendroarchäologie. *Monographien Der Kantonsarchäologie Zürich 47* (Zürich und Egg).

321. Larson G, et al. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci U S A* 104(39):15276–15281.

322. Binladen J, et al. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2(2):e197.

323. Dabney J, et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* 110(39):15758–15763.

324. Damgaard PB, et al. (2015) Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* 5:11184.

325. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):db.prot5448.

326. Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D (2015) Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci* 370(1660):20130624.

327. Korlević P, et al. (2015) Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* 59(2):87–93.

328. Svensson EM, et al. (2007) Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Anim Genet* 38(4):378–383.

329. Briggs AW, Heyn P (2012) Preparation of Next-Generation Sequencing Libraries from Damaged DNA. *Ancient DNA: Methods and Protocols*, eds Shapiro B, Hofreiter M (Humana Press, Totowa, NJ), pp 143–154.

330. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

331. Schubert M, et al. (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13(1):178.

332. Li H, et al. (2009) The Sequence Alignment / Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. *Data Processing*:1–2.

333. Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods Mol Biol* 840:197–228.

334. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

335. Skoglund P, Ersmark E, Palkopoulou E, Dalén L (2015) Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr Biol* 25(11):1515–1519.

336. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.

337. Frantz LAF, et al. (2015) Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet* 47(10):1141–1148.

338. Bosse M, et al. (2014) Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat Commun* 5:4392.

339. Alexandri P, et al. (2012) The Balkans and the colonization of Europe: the post-glacial range expansion of the wild boar, Sus scrofa : Post-glacial range expansion of wild boar. *J Biogeogr* 39(4):713–723.

340. Frantz LAF, et al. (2013) Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol* 14(9):R107.

341. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

342. Larson G, et al. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307(5715):1618–1621.

343. Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26(12):1569–1571.

344. Larson G, et al. (2007) Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proc Natl Acad Sci U S A* 104(12):4834–4839.

345. Frantz L, et al. (2018) Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events. *bioRxiv*:241448.

346. Caye BK, Jay F, Michel O, François O (2016) Fast Inference of Individual Admixture Coefficients Using Geographic Data. *bioRxiv*:080291.

347. Martins H, Caye K, Luu K, Blum MGB, François O (2016) Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *bioRxiv*:054585.

348. Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour* 16(2):540–548.

349. Khalilzadeh P, et al. (2016) Contact Zone of Asian and European Wild Boar at North West of Iran. *PLoS One* 11(7):e0159499.

350. Krause-Kyora B, et al. (2013) Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nat Commun* 4:2348.

351. Caliebe A, Nebel A, Makarewicz C, Krawczak M, Krause-Kyora B (2017) Insights into early pig domestication provided by ancient DNA analysis. *Sci Rep* 7:44550.

352. Fang M, Berg F, Ducos A, Andersson L (2006) Mitochondrial haplotypes of European wild boars with 2n = 36 are closely related to those of European domestic pigs with 2n = 38. *Anim Genet* 37(5):459–464.

353. Fang M, Larson G, Ribeiro HS, Li N, Andersson L (2009) Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet* 5(1):e1000341.

354. Koutsogiannouli EA, Moutou KA, Sarafidou T, Stamatis C, Mamuris Z (2010) Detection of hybrids between wild boars (Sus scrofa scrofa) and domestic pigs (Sus scrofa f. domestica) in Greece, using the PCR-RFLP method on melanocortin-1 receptor (MC1R) mutations. *Mamm Biol* 75(1):69–73.

355. Frantz AC, Zachos FE, Kirschning J (2013) Genetic evidence for introgression between domestic pigs and wild boars (Sus scrofa) in Belgium and Luxembourg: a comparative approach with multiple marker …. *Biol J Linn Soc Lond* 110(1):104–11.

356. Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592–593.

357. Goedbloed DJ, et al. (2013) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol Ecol* 22(3):856–866.

358. Purcell S, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.

359. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.

360. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.

361. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.

362. Groenen MAM, et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491(7424):393–398.

363. White S (2011) From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History. *Environ Hist Durh N C* 16(1):94–120.

364. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.

365. Falush D, van Dorp L, Lawson D (2016) A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. *bioRxiv*:066431.

366. Hellenthal G, et al. (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751.

367. O'Connell J, et al. (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* 10(4):e1004234.

368. Tortereau F, et al. (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* 13(1):586.

369. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.

370. Leslie S, et al. (2015) The fine-scale genetic structure of the British population. *Nature* 519(7543):309–314.

371. Bosse M, et al. (2014) Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. *Mol Ecol* 23(16):4089–4102.

372. Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proc Biol Sci* 273(1595):1803–1810.

373. Price AL, et al. (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 5(6):e1000519.

374. Ai H, et al. (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet* 47(3):217–225.

375. Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23(13):3133–3157.

Chapter 8, “Discussion and Conclusions concerning Part II”

To reiterate the chapters of Part II, I describe in Chapter 5 different evolution-based hypotheses that ancient DNA can be used to address, of which Chapters 6 and 7 can be viewed as case studies. In Chapter 6, I use ancient mitochondrial genomes and different statistical methods to answer broader questions about evolution, specifically to find evidence of time-dependency in molecular rate estimates of several animal taxa. In Chapter 7, I and my co-authors use ancient pig mitochondrial genomes to answer more specific questions about the evolutionary history of domestic pigs. I demonstrate how most Near Eastern ancestry in the genomes of European domestic pigs have disappeared over 3,000 years as a result of interbreeding with local wild boars.

The format of my final discussion chapter is a bit different: Because the research presented in Chapter 6 is unpublished, I first briefly discuss further work that will need to be completed before the manuscript can be submitted to a journal. What follows in the rest of this chapter is a broader discussion of exciting new developments in the field of ancient DNA/paleogenomics, and a few hypotheses generally pertaining to my interest in infection and immunity which ancient DNA (or even RNA) can be utilised to address.

1. Next steps for Chapter 6

In Chapter 6, I focus on using ancient mitochondrial genomes and different statistical methods to answer broader questions about evolution, specifically

to find evidence of time-dependency in molecular rate estimates of several animal taxa. Time-calibrated ancient DNA datasets provide an excellent opportunity to clarify the discrepancy between short- and long-term rate estimates of pedigree and fossil-calibrated phylogenetic studies. I analyse mitochondrial genomes from modern, ancient and historical archaeological specimens of seven taxa: bison, chickens, dogs, goats, horses, pigs, and sheep. I use three different rapid dating methods (based on maximum-likelihood inference) as well as Bayesian phylogenetic inference to estimate mitochondrial substitution rates in these animal taxa across different time scales. I find some evidence of time-dependence in evolutionary rate estimates in all datasets, but the strength of the effect is dependent on the part of the mitogenome and the method used. Only the bison dataset shows a consistent time-dependent pattern in rate estimates across all methods and mitogenome partitions. Most of the other datasets show discrepancies between rate estimates from the narrower and wider sampling time windows, but a consistent time-dependent effect is seen only in the faster-evolving D-loop and 3rd codon positions of all genes. This study shows the importance of taking into consideration that different parts of the mitogenome evolve at different rates when measuring the time-dependence of rate estimates using datasets from slower-evolving species over short timescales.

I use the date-randomisation test [382][394] to assess the temporal signal in the datasets and validate rate estimates. In this test, the ages of samples are shuffled and the analysis is repeated to generate a set of rate estimates from the date-randomised datasets. The random assignment of sampling dates to sequences breaks the association between the nucleotide substitution and time (the molecular clock) [351]. The expectation is that if evolution is clock-like, the random assignment of dates to sequences will lead to rate estimates that are significantly different from those estimated using the correct sampling dates [351]. Because of time constraints, our results are based on

only five date-randomised replicates of the original mitogenome datasets for each taxon and each method. Before submitting this paper for publication, I will perform 20 date-randomisation replicates for each partition, as previously recommended [351]. With 20 replicates, an allowance of one failure corresponds to an error rate of 0.05.

The results of the chicken dataset analyses are likely to be biased due to considerable population structure and a strong sampling bias. The ancient samples are all from European haplogroup E1, and only a handful of modern samples have the same E1 haplogroup. The majority of the modern samples belong to a different haplogroup with deeper lineages and ancestry from China [426]. To perform appropriate molecular clock analyses it would be necessary to use only modern E1 samples which may necessitate sequencing more modern samples to attain the same statistical power. Alternatively, more ancient samples from different haplogroups would need to be added to the dataset, but this option is unlikely given the lack of chicken archaeological material from outside of western Europe.

Finally, I write in the Discussion of Chapter 6 that the variability in time-dependency and temporal signal in these different datasets may be affected in part by postmortem DNA damage. It may be prudent to perform separate Bayesian inference analyses using the age-independent [410] and age-dependent [411] models of postmortem DNA damage, and compare these results to analyses without these models. The age-independent model of postmortem damage adds a single parameter, δ, to the model, which represents the average amount of damage per nucleotide. The addition of δ additional substitutions per site to each external branch represents substitutions due to DNA damage [410]. In contrast, the age-dependent model [411] assumes that because DNA damage accumulates through time, it is more likely for sequences derived from older specimens to have more

miscoding lesions. In this model the probability that any given nucleotide remains undamaged is assumed to decay exponentially with the age of the sample [411]. These additional analyses would have to be done in BEAST, as the postmortem DNA damage models are not currently implemented in BEAST2, which was used for all of the Bayesian analyses presented in this chapter. For consistency it would then make sense to repeat all analyses in BEAST, if possible.

2. Discussion: Retrieval of ancient DNA

Museums and their ethnographic, archaeobotanical, and zooarchaeological collections represent a vast resource for ancient DNA, and a source of unique insights into the evolutionary history of many species. However, preservation methods commonly used in modern museum collections also cause difficulties in retrieving ancient DNA. Such conservation practices include the use of arsenic, glue, shellac, pesticides, heat, and chemical maceration with boiling water and benzene vapours [427]. Consequently, assumptions regarding ancient DNA preservation should not be made solely on the specimen age or their provenience, but also their postmortem treatment. For example, it has been reported that DNA yields from 19th and 20th-century historic dental calculus specimens are consistently lower in comparison to DNA recovered from ancient archaeological remains dating from 3,000 BCE to the 19th century CE [427].

The retrieval of ancient DNA from archaeological remains is limited by preservation conditions, where specimens that come from cold and dry permafrost conditions fare better than those derived from hot and wet environments [428][429]. As a result, ancient DNA datasets are biased heavily toward megafaunal species that lived on the so-called mammoth steppe, the dry, cold biome that stretched from the Iberian peninsula across

Eurasia and Beringia to Alaska during the Pleistocene until the end of the last Ice Age [430][431]. Similarly, ancient DNA is more likely to be recovered from historical samples found in dry desert conditions [347][348]. Recent progress in extraction and sequencing methods has however led to improved protocols that make it possible to recover ancient DNA from younger archaeological remains that are also from warmer geographical regions (e.g. Neolithic sites in Western Asia and Europe [384][217][385]), and even tropical environments [432]. Additionally, it can be possible to retrieve short (≥35 bp) or even ultrashort (≥25 bp) DNA fragments [433].

3. Discussion: From aDNA to aRNA

Although RNA is generally not expected to survive in the archaeological record [434], several studies in recent years have explored the recovery of RNA from ancient and historical samples, particularly in plant materials. Double-stranded viral RNA genomes have been found in ~750 year old barley grain [435], 700-year old caribou faeces [436], and RNA transcripts have been recovered from maize kernels that date to 1915 [437]. Environmentally-induced regulation patterns of microRNA [438] and RNA-induced demethylation patterns in barley grain [436] has also been uncovered. In human soft tissue remains, microRNAs have been successfully retrieved from the 5,300 year old Tyrolian "Iceman" using a target quantitative PCR (qPCR approach) [439]. Most recently, ancient RNA has been reportedly generated from mammalian soft tissue using NGS technology [440]. The ancient RNA was sequenced from permafrost-preserved liver tissue of a 14,300 year old Pleistocene canid, and two historical wolf skins. The ancient RNA isolated has evidence of biologically relevant tissue specificity and exhibit similarity to equivalent data from modern-day control tissue. In addition, the authenticity of the data is confirmed by the presence of exon-exon junctions and high endogenous ribosomal RNA (rRNA) [440].

The authors emphasise that the results should not be considered an "average" representation of nucleic acid preservation in ancient soft tissues because of the exceptional preservation conditions of the specimens. The degradation of RNA in a deceased individual occurs predominantly due to the enzymatic activity of cellular RNAses [441]. The susceptibility of modern RNA to degradation in a laboratory setting could be attributable to contamination from RNAses that could be active even when frozen. The authors point out that the complexities of purified RNA survival within a long-term laboratory storage setting is well-documented [442][443], and that the thermodynamics of RNA stability and its enzymatic reactions within long-term diagenetic postmortem contexts [441] is poorly understood [440]. However, under favourable circumstances such as desiccation and permafrost conditions, the enzymatic and autolytic processes of decomposition, including that of RNAse activity, can be sufficiently arrested in animal remains, so that ancient RNA can be recovered [440].

4. Discussion: Reconstructing patterns of past population movement

Ancient DNA has the potential to help reconstruct patterns of past population movement and interaction during important periods of economic and cultural transition. This is clearly demonstrated in Chapter 7, where we use pig mitochondrial and nuclear genomes to show that Anatolian wild boars domesticated ~10,500 years ago were the ancestors of domestic pigs that were transported into Europe by Neolithic farmers two thousand years later. Over the course of 3 millennia, hybridisation between Near Eastern domestic pigs and European wild boars in continental Europe resulted in a near-complete genomic replacement and the gradual disappearance of almost all Near Eastern ancestry.

Another example where ancient DNA is utilised in reconstructing past population dynamics is seen in a recent study examining the spread of pastoralism starting around 5,000 years ago in East Africa. It had been difficult to identify the impact of the local foraging communities on the migrating food producers who brought livestock such as cattle, sheep, and goats with them, and vice-versa. Genomes from 41 individuals that lived in East Africa ~100 to 4,000 years ago were analysed, and the authors were surprised to find relatively little genetic admixture between the migrating food producers and local hunter-gatherers [444]. A further example can be found in a genome-wide analysis of 6 Mesolithic and 67 Neolithic individuals found in Britain, dating from 10,500 to 4,500 years ago, which supports a scenario where agriculture was introduced by incoming farmers who migrated from the Aegean, with small, geographically-structured levels of hunter-gatherer ancestry. These results support the slow and steady assimilation of Neolithic farmers with hunter-gatherer communities, as opposed to a complete population replacement, which would be expected if there was limited interaction between populations, as in the east African example. Furthermore, these results also contrast to the expectation seen in other European Neolithic populations where there was usually a resurgence of hunter-gatherer ancestry during the Neolithic [445].

5. Discussion: Reconstruction of speciation history

Successful extraction and sequencing of DNA from extinct animals can enable us to place these species in a phylogenetic context that can challenge previously-held assumptions of the species' biogeography, diversification history, and morphological evolution. We see in Chapter 7 that European domestic pigs did not originate from an independent domestication process,

but rather from the continuous management of herds that were interbred (however intentionally) with local wild boar.

A different example regarding the dichotomy of morphological and genetic evolution of species can be seen in the case of dire wolves, reported in Appendix Chapter 10. We report that despite being morphologically highly similar, and being classified as closely related species, dire wolves and gray wolves are highly differentiated species that diverged around 5 million years ago, with the gray wolf more closely related to the dhole (genus *Cuon*) and African wild dog (genus *Lycaon*) than to either the dire wolf or African jackals (both genus *Canis*). Moreover, with ancient DNA we can also address questions as to why these species became extinct (e.g. dire wolves may not have been as well-adapted to the warming climate of North America as the gray wolves were).

Within the past two years, ancient DNA studies on extinct species include the giant sloth [446], the Tasmanian devil [447], the Jamaican monkey *Xenothrix* [448], the giant scops owl from the Mauritius and Rodrigues Islands in southwestern Indian Ocean [449], and a 2,500 year old extinct bird called the *Caracara creightoni*. Ancient DNA in birds is generally more difficult to obtain, but even more so in volant birds with light, hollow bones. Remarkably, the bone of the *Caracara creightoni* was recovered from the anoxic saltwater environment of a tropical blue hole in the Bahamas [450].

6. Discussion: Historical virus epidemiology

Epidemics and disease outbreaks have always been a major part of human history. When infectious diseases decimate large proportions of the human population, economies are devastated, and civilisations and societies irrevocably altered. To name a few examples, the Justinian plague has been

linked to the fall of the Roman empire, measles and smallpox epidemics that swept through the Americas after European contact likely facilitated colonisation, and the Black Death resulted in large changes to society, diet and land-use across Europe. The cyclic characteristics of some infectious diseases that may appear new are often the result of fluctuating disease patterns throughout our history [434]. Understanding the genetic basis of virulence in historical diseases can potentially offer important insights into the dynamics of emerging and re-emerging infectious diseases that impact the world today. While it is rare to observe adaptive mutations over the course of an infectious disease outbreak, studying ancient pathogen genomes can allow us to observe "evolution in action", by reconstructing the evolutionary history of how and why a pathogen's virulence may change over time. Interestingly, it has been predicted in the past through mathematical modelling that HIV-1 will evolve to become less virulent over time [451], and recent evidence with long-term cohorts from Uganda suggest validation of this prediction [452].

Overall, from ancient genomes, we can identify potentially important mutations that result in greater virulence and pathogenicity, determine the timing of the pathogen's geographic spread, and understand the pathogen's relationship to historical events and environmental changes that result in infectious disease outbreaks. Microbial genomics have provided novel insights into the aetiology of historically important infectious disease epidemics, such as plague, leprosy, syphilis, and tuberculosis (reviewed in [434]). For example, isolation of Hepatitis B virus (HBV) from ancient bones and teeth [340] show that despite the fact that the virus primarily replicates in the liver, it can also enter the bloodstream and disseminate throughout the body. It is therefore probable that this behaviour can be seen in other types of clinically important DNA viruses that have yet to be investigated in human skeletal remains, such as papillomaviruses and herpesviruses. Ancient DNA

from a papillomavirus has been isolated from a 27,000 year old bushy-tailed packrat, and concordance in the phylogenetic trees of both host and virus suggests a shared evolutionary history [453].

Analysis of human parvovirus B19 and HBV genomes recovered from human skeletal remains show that these viruses have been circulating in the human population for more than 7,000 years [340][454][341]. Additionally, the ancient HBV isolates are distinct lineages that have possibly gone extinct, since they have no close modern relatives. Increased sampling of ancient HBV genomes from archaeological remains can help us better understand the origins and the evolutionary history of the virus, specifically the strains that are endemic in the world today. The increased sampling of ancient HBV genomes has also revealed a time-dependent effect [455], the phenomenon that is addressed at length in Chapter 6.

Ancient and historic mammalian RNA viruses have not been isolated yet, with the notable exceptions of the 1918 Spanish influenza virus [456], the oldest known HIV-1 infection from 1959 [457], and a near-full length HIV-1 virus genome from 1966 [458], all preserved in formalin-fixed paraffin-embedded tissue samples. These formalin-fixed, paraffin-embedded tissue samples from many diseases that are stored in various public health archives across the world can provide an excellent resource for further study of significant infectious diseases from recent history [459].

Ancient pathogen nucleic acids from animals may allow us to address questions involving the evolutionary history of historically and medically important viruses, particularly those at high risk for zoonotic spillover, such as influenza virus. Influenza has been remarked upon in the historical record for millennia, beginning in early Greek writings from 412 BC [460]. Influenza epidemics occurred throughout Europe in the following centuries, but the first probable influenza pandemic occurred in 1510, spreading from Africa to

Europe. There was another possible outbreak in 1557, but a more likely pandemic attributed to influenza was in 1580, originating during the summer in Asia, which then spread to Africa, and then to Europe along two corridors from Asia Minor and Northwest Africa [461]. Within 6 months, the disease disseminated throughout Europe from the south to the north, and then spread to America [461][462]. 8,000 deaths were reported in Rome, and populations of several Spanish cities were reportedly decimated [462][463].

Influenza viruses circulate endemically in wild bird populations, typically infecting wild waterfowl and regularly entering commercial poultry flocks [464]. Avian influenza is generally poorly transmissible to humans [465][466][467], and onward human-human transmissions after a spillover event are rare. However, pigs often serve as an intermediate host (or a "mixing vessel") for influenza viruses where genetic reassortment occurs which may increase the likelihood of further cross-species transmission [468][469] and result in human epidemics [470]. In addition, influenza viruses are also known to "spill back" from humans to pigs [471][472][473][474]. Avian influenza was first described in the historical record in 1878 northern Italy in 1878 as "fowl plague" [475][476]. Intriguingly, there is some evidence suggesting that the virus that caused the 1918 influenza pandemic did not acquire its haemagluttinin (HA) subtype from a bird [477]. Archaeological collections of pig and bird remains provide an enticing resource to test for the presence of influenza virus.

As the methodology in recovering ancient pathogen genetic material continues to develop, new insights into these ancient epidemics will continue to emerge, as well as our understanding of how medicine, cultural and social practices with respect to infectious diseases have changed over time [434][478].

7. Discussion: Ancient DNA to address co-evolution of human immunity and viruses

Humans are in a constant co-evolutionary relationship with viruses, bacteria, and other parasites, and some of the most important selective forces experienced by humans during our evolutionary history are due to pathogens and the diseases they cause. Throughout our evolutionary history, a number of environmental and/or cultural constraints have been created that can have consequences, which is seen in some populations having different degrees of introgression of alleles from other hominin species.

There are several known examples of adaptive introgression of archaic alleles that have significantly shaped modern human immunity. Archaic humans that have inhabited Western Asia and Europe for more than 200,000 years were likely to be adapted to their local environments and its pathogens. Through admixture with Denisovans inhabiting Western Asia, modern humans acquired a polymorphic human leukocyte antigen (HLA) class I allele, which then introgressed into modern Eurasian and Oceanian populations, and now represent more than half the HLA alleles of modern Eurasians [479]. As another example, 3 Toll-like receptors, key components of the innate immune system and the first line of host defence against pathogens, also entered the modern human genome through repeated introgression from archaic humans, resulting in differences in disease phenotypes in modern Europeans [480][481][482]. Finally, a haplotype of the innate immune gene *STAT2* that closely resembles the Neanderthal *STAT2* is carried by Melanesian populations at a 10-fold higher frequency than in Eurasians, and the haplotype is not seen in sub-Saharan Africans [483].

It was reported recently that long, frequent, and likely adaptive segments of Neanderthal ancestry in modern humans are enriched for virus-interacting immune genes [484]. These introgressed segments are more likely to be seen

in modern Europeans compared to those of East Asian ancestry. Interactions between different archaic populations such as Neanderthals and modern humans would have resulted in viral pathogen exchange either by direct contact or from their shared environment. These introgressed sequences may have provided a measure of protection against such viral infections, leading to positive selection for these sequences during admixture between populations [484]. Most adaptations targeting coding variation in innate immune genes have occurred in the last 6,000-13,000 years, when human populations shifted from being hunter-gatherers to farmers [481]. Taken altogether, the evidence indicates that the geographic distribution of human susceptibility to different diseases has a deep evolutionary history.

In order to improve our understanding of why one person develops disease while another does not, it is important to consider the evolutionary pathways to variation in disease risk. A potentially promising subject to test in a case study is *IFNL4*, a polymorphic pseudogene that encodes for interferon (IFN) lambda 4 protein, an inducer of antiviral gene expression. Interferons are produced in response to viral infection and block viral replication by expressing other antiviral genes. Humans are polymorphic for the dinucleotide TT/ΔG allele, where compared to the ancestral state in non-human primates, the TT allele produces a frameshift in the coding region, rendering the gene nonfunctional. This impaired allele has increased rapidly in frequency and appears to be under positive selection. There are population-specific differences in treatment outcomes of HCV-infected individuals who have the intact ΔG allele [485], which has been selected against in non-African human populations [486]. People who have the TT allele (nonfunctional *IFNL4*) exhibit better clinical outcomes than patients who have the intact and fully active ΔG allele [487]. Interestingly, HCV is thought to have originated in West Africa, where the TT allele is prevalent. The most divergent HCV strains in rural populations indicate centuries of

stable human-to-human transmission, in the absence of any consistent and widespread transmission route. This pattern would be most explained if the virus diversity originated from a (yet unknown) animal reservoir [488], and that each HCV genotype arose from cross-species transmission from separate zoonotic sources in different locations [489].

It is possible that intact *IFNL4* had emerged as a host defense factor in the immune response against ancient viral pathogens, but was later negatively selected in shifting host–pathogen relationships [490]. Many questions still remain regarding the evolutionary history of the *IFNL4* variants, including the date of its divergence, and which non-African ancestral populations show more adaptive selection for this allele. Sub-Saharan Africa has a substantial HCV disease burden where 10-15 million individuals are viraemic [491], although it is unclear how much current sub-Saharan African HCV prevalence is due to iatrogenic transmission from colonial public health campaigns and how much is due to endemic infection [491]. In addition, it is unknown what types of viral pathogens the ΔG variant protects against, and whether African populations have increased susceptibility to HCV infection or progression of HCV-associated disease, such as liver cirrhosis and liver cancer. With so many emerging and reemerging viruses in the 21st century, it is important to gain a better understanding of the genetic underpinnings of human disease susceptibility from pathogens.

8. Conclusions

In 1984, Russell Higuchi and colleagues for the first time successfully sequenced the DNA of an extinct animal. The quagga, a relative of a zebra, had been extinct since 1883 [492]. The authors had used preserved soft tissue from a museum specimen, and the extraction and sequencing process was laborious, requiring the use of bacterial vectors to directly clone a 229 bp

fragment of the quagga mitochondrial genome. Thirty years after the fragment of the quagga mitochondrial genome was sequenced, the complete nuclear genome of the quagga was sequenced to an ~8× average depth of coverage [493]. In combining high-throughput sequencing with improved extraction methods, the time window in which ancient DNA molecules have been recovered range from at least 1 million years in permafrost regions [331] to 500,000 years in temperate caves [240][494][495]. The oldest complete ancient genome to date originates from a ~700,000 year old horse bone preserved in permafrost [331].

Bolstered by technological advances in recent years, ancient DNA is playing an increasingly important role in studies of phylogenetics, conservation genetics, and population genetics. The promise of recovering ancient RNA from archaeological material would allow for the metagenomic palaeopathology of RNA viruses, tissue identification, and the identification of transcriptomic processes *in vivo* that can be compared to individual ancient genomes within the context of influence from climate, diet, trauma, and disease [440]. The palaeo-epigenome is currently being characterised, with the first genome-wide maps of methylation and nucleosome patterns reconstructed from postmortem patterns of DNA degradation [496][497], and from direct enrichment of methylated CpG sites [496][497][498]. Such studies, in concert with the inclusion of ancient RNA, will hopefully yield important insights into ancient gene expression levels and changes in gene regulation during the evolutionary processes of species divergence, adaptation, and functional responses to known selection pressures [499][195].

Some of the most important selective forces on human evolutionary history are pathogens, and a major feature of the human genome are mechanisms to cope with the constant impact of infectious disease. Human immunity has been shaped by positive selection on immune genes, including introgression

from other ancient populations that were adapted to their environment and its local pathogens. Ancient pathogen DNA and RNA can facilitate the discovery of previously unknown animal reservoirs, the reconstruction of ancient pathogen migration and transmission, and by comparing historical and contemporary lineages, track genomic changes through time [434].

To summarise, I show in the previous chapters how ancient DNA can be used to ask fundamental evolutionary biology questions as well as reconstruct early animal domestication processes. As shown in Chapter 6, time-series ancient DNA datasets provide a useful resource to study the strength of time-dependency in molecular rate estimates and bridge the gap between between short- and long-term rate estimates of pedigree and fossil-calibrated phylogenetic studies [406]. Not taking the time-dependent effect into account has important implications for studies across recent evolutionary timescales, such as overestimating the times of coalescent events if the phylogeny is calibrated using fossil data. Ancient DNA can help reconstruct patterns of past population movement, interaction, and dispersal of humans and domesticated animals during important human cultural periods either during phases of cultural transition or over long spans of time. I and my co-authors demonstrate in my previous chapters (in particular Chapter 7 and Appendix Chapter 11) that ancient DNA allows us to better understand the timing, location, mechanisms, and evolutionary dynamics of domestication. Ancient DNA allows for the incorporation of extinct species into phylogenies (as seen in Appendix Chapter 10) and adds time depth into population genetics studies [500]. As methodologies to recover ancient DNA and RNA continues to advance, I expect that as quickly as we find new insights into long-lasting questions involving evolutionary processes, population dynamics, domestication, speciation, and extinction, new questions and corresponding challenges are sure to arise. There is seemingly no end to the excitement that this field brings.

Appendix

The appendix of my thesis consists of 4 additional chapters that are projects I contributed to during my DPhil that were not at the heart of my main objectives. Chapter 9 consists of the paper published in PNAS, titled "Human Endogenous Retrovirus-K HML-2 integration within RASGRF2 is associated with intravenous drug abuse and modulates transcription in a cell-line model". Chapters 10 and 11 consists of work that I have done with collaborators. The manuscript of Chapter 10, "Dire wolves were the last of an ancient New World canid lineage," is currently undergoing the second round of reviews in Science. The manuscript of Chapter 11, "Specialised sledge dogs aided Inuit dispersal across the North American Arctic," has been published in Proceedings of Royal Society B.

Chapter 9, "Human Endogenous Retrovirus-K HML-2 integration within RASGRF2 is associated with intravenous drug abuse and modulates transcription in a cell-line model"

Summary

HERV-K (HML-2) may still be proliferating in the human genome, and is also insertionally polymorphic in the human population. This paper focuses primarily on one specific HERV-K (HML-2) provirus is found within an intron of a gene called RASGRF2. This gene is involved modulating dopaminergic activity in the brain, is more frequently found in drug addicts, and is significantly associated with addiction. We experimentally show that HERV-K (HML-2) can manipulate nearby genes, and that the HERV-K (HML-2) insertion in RASGRF2 is significantly more present in the genomes of people who inject drugs (PWIDs) compared to the non-PWID controls. This

paper gives evidence that polymorphic HERV-K (HML-2) can contribute to the burden of pathogenesis of genetic disease in humans.

Authors

Timokratis Karamitros, Tara Hurst, Emanuele Marchi, Eirini Karamichalie, Urania Georgopouloue, Andreas Mentis, Joey Riepsaame, ***Audrey Lin****, Dimitrios Paraskevis, Angelos Hatzakis, John McLauchlan, Aris Katzourakis, and Gkikas Magiorkinis*

Status

Published (2018) Proceedings of the National Academy of Sciences of the United States of America. doi:10.1073/pnas.1811940115

My contribution

I designed a set of primers targeting a exon-exon junction in the human RASGRF2 gene to be used in PCR tests. In addition, I also performed the PCR screening tests on the Greek cohort (*n*=202) of HIV-infected people who inject drugs (PWID) and non-PWID samples.

Human Endogenous Retrovirus-K HML-2 integration within *RASGRF2* is associated with intravenous drug abuse and modulates transcription in a cell-line model

Timokratis Karamitros[a,b], Tara Hurst[a,c], Emanuele Marchi[d], Eirini Karamichali[e], Urania Georgopoulou[e], Andreas Mentis[b], Joey Riepsaame[f], Audrey Lin[a], Dimitrios Paraskevis[g], Angelos Hatzakis[g], John McLauchlan[h], Aris Katzourakis[a,1], and Gkikas Magiorkinis[a,g,1]

[a]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; [b]Public Health Laboratories, Department of Microbiology, Hellenic Pasteur Institute, 11527 Athens, Greece; [c]Division of Virology, National Institute for Biological Standards and Control, Potters Bar EN6 3QG, United Kingdom; [d]Nuffield Department of Medicine, University of Oxford, Oxford OX1 3SY, United Kingdom; [e]Molecular Virology Laboratory, Department of Microbiology, Hellenic Pasteur Institute, 11527 Athens, Greece; [f]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom; [g]Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, 11527 Athens, Greece; and [h]Medical Research Council–University of Glasgow Centre for Virus Research, Glasgow G61 1QH, United Kingdom

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved August 23, 2018 (received for review July 20, 2018)

HERV-K HML-2 (HK2) has been proliferating in the germ line of humans at least as recently as 250,000 years ago, with some integrations that remain polymorphic in the modern human population. One of the solitary HK2 LTR polymorphic integrations lies between exons 17 and 18 of *RASGRF2*, a gene that affects dopaminergic activity and is thus related to addiction. Here we show that this antisense HK2 integration (namely RASGRF2-int) is found more frequently in persons who inject drugs compared with the general population. In a Greek HIV-1–positive population (n = 202), we found RASGRF2-int 2.5 times (14 versus 6%) more frequently in patients infected through i.v. drug use compared with other transmission route controls (P = 0.03). Independently, in a United Kingdom-based hepatitis C virus-positive population (n = 184), we found RASGRF2-int 3.6 times (34 versus 9.5%) more frequently in patients infected during chronic drug abuse compared with controls (P < 0.001). We then tested whether RASGRF2-int could be mechanistically responsible for this association by modulating transcription of *RASGRF2*. We show that the CRISPR/Cas9-mediated insertion of HK2 in HEK293 cells in the exact *RASGRF2* intronic position found in the population resulted in significant transcriptional and phenotypic changes. We also explored mechanistic features of other intronic HK2 integrations and show that HK2 LTRs can be responsible for generation of *cis*-natural antisense transcripts, which could interfere with the transcription of nearby genes. Our findings suggest that RASGRF2-int is a strong candidate for dopaminergic manipulation, and emphasize the importance of accurate mapping of neglected HERV polymorphisms in human genomic studies.

HERV-K HML-2 | endogenous retrovirus | RASGRF2 | persons who inject drugs | addiction

The human genome is littered with retroviral elements as a result of ancient retroviral infections in the germ line of our primate ancestors. Some of these retroviral invasions proliferated successfully by continuously reintegrating in their host genomes (1). Proliferation success of endogenous retroviruses (ERVs), measured by the relative abundance within different hosts, depends on a complex interplay of factors, including viral life cycle (2) and host life history (3). Furthermore, hosts have multiple diverse control mechanisms that inactivate ERVs in place. Some of these mechanisms that reduce potentially deleterious retroviral activity include the accumulation of inactivating mutations, internal recombination events between their long terminal repeats (LTRs), host-derived restriction factors, and transcriptional silencing (4).

While pathogenic ERVs have been described throughout the animal kingdom, none have been definitively linked with harmful effects in humans. Furthermore, ERV germ-line proliferation in humans and other great apes is lower compared with other primates (5), with the majority of human endogenous retroviruses (HERVs) being defective and fixed within the population. Perhaps retroviral proliferation poses a disproportionate burden for humans compared with other mammals. Although most HERVs ceased proliferating millions of years ago, HERV-K (HML-2) (referred to in the text as HK2) continued proliferating in the germ line of our ancestors after the human–chimpanzee divergence ~5 to 6 Mya (1, 6), with some integrations still being polymorphic in the population (7, 8). As a consequence, some HK2 proviruses are exceptionally well preserved, express viral proteins, and produce viral particles (9). HK2 integrations tend to be found near genes, with an antisense bias (10), suggesting that antisense integrations are less likely to impose a pathogenic burden (11, 12). Some evidence suggests that intronic HK2 integrations modulate transcription of nearby genes (13), but their phenotypic role remains largely unknown. Intronic HK2, as well as other intronic transposable elements, have been proposed to have contributed to host evolution by modulating

Significance

The human genome is "littered" with remnants of ancient retrovirus infections that invaded the germ line of our ancestors. Only one of these may still be proliferating, named HERV-K HML-2 (HK2). Not all humans have the same HK2 viruses in their genomes. Here we show that one specific uncommon HK2, which lies close to a gene involved in dopaminergic activity in the brain, is more frequently found in drug addicts and thus is significantly associated with addiction. We experimentally show that HK2 can manipulate nearby genes. Our study provides strong evidence that uncommon HK2 can be responsible for unappreciated pathogenic burden, and thus underlines the health importance of exploring the phenotypic roles of young, insertionally polymorphic HK2 integrations in human populations.

Author contributions: T.K., J.R., A.K., and G.M. designed research; T.K., T.H., E.K., U.G., A.M., J.R., A.L., A.K., and G.M. performed research; D.P., A.H., and J.M. contributed new reagents/analytic tools; T.K., T.H., E.K., U.G., A.M., A.K. and G.M. analyzed data; and T.K., T.H., E.M., E.K., U.G., A.M., J.R., A.L., D.P., A.H., J.M., A.K., and G.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: The sequences reported in this paper have been deposited in Genbank, https://www.ncbi.nlm.nih.gov/genbank/ (accession nos. MH626561–MH626576).

[1]To whom correspondence may be addressed. Email: aris.katzourakis@zoo.ox.ac.uk or gmagi@med.uoa.gr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811940115/-/DCSupplemental.

Published online September 24, 2018.

transcription of the surrounding genes (14, 15). Moreover, some evidence suggests that intronic integrations of transposable elements could reshape the transcriptome upon reactivation during cancer development (16), although there is no evidence that HK2 reactivation could result in gene modulation in cancer development (17).

One of the polymorphic HK2 integrations is found between exons 17 and 18 of *RASGRF2* (hereafter RASGRF2-int), a gene involved in signaling pathways. RASGRF2 has at least two domains and is expressed in T cells, the heart, and the brain. The deregulation of the gene could thus be involved in complex phenotypes (or disease syndromes) involving more than one system (the brain and immune system). Double knockout mice mutants of $Rasgrf2^{tm1Esn}$ displayed no significant pathological phenotype with respect to growth and development (18), but experiments into its role in T-cell signaling responses showed diminished immune responses (19); the importance of the deficient immune response has not been explored in humans.

Stronger evidence exists for the brain-related phenotype of RASGRF2 in humans; a genome-wide association study on alcohol addiction provided a potential hit for the SNP rs26907 that is located between exons 3 and 4 (20). This SNP was subsequently shown to modulate addictive behavior (21) due to modifications of noradrenergic and serotonergic responses (22). Young carriers of rs26907 were more likely to have alcohol-induced reinforcement and show enhanced reward-related dopaminergic activity in functional MRI experiments (20). Crucially, the rodent model of addiction is simple, reproducible, and predictive of addiction in humans (23), with double knockout $RASGRF2^{(-/-)}$ mice being remarkably resistant to addiction (21).

Here we show that RASGRF2-int is significantly associated with drug addiction in two independent, genetically distinct human populations. We also provide evidence in support of a mechanistic interaction between HK2 integration and *RASGRF2* transcription, suggesting that the observed association is likely to be causal.

Results

RASGRF2-Int Is Associated with Drug Addiction. The most well-established phenotype of RASGRF2 is linked with addiction. *RASGRF2* is expressed most intensively in the brain, but also in the heart, lungs, and T cells (24, 25). Thus, we initially hypothesized that if the proviral HK2 modulated *RASGRF2* in humans, it would be found at higher frequency among individuals with well-defined strong addictive behavior such as persons who inject drugs (PWIDs). We tested this by using PCR to screen the genomic DNA of 202 fully anonymized Greece-based HIV-positive individuals with blinded samples for the presence or absence of this integration (Fig. 1 *A–D*). The cohort consisted of 102 PWIDs (reported i.v. drug use within the past 6 mo as the likely transmission route) and 100 controls (infected by other transmission routes). In this population, we did not test for potential confounding variables including sex and age among groups. The power of our approach is 80% for recovering fourfold higher frequency in the PWIDs versus the expected 4% (7) in the general population. We found that the integration was present in 14 PWIDs vs. 6 non-PWIDs ($P = 0.03$, χ^2, one-sided test) in the form of a solo-LTR, suggesting a more than twofold higher frequency in populations with long-term addictive behavior. Individuals who carried the integration were heterozygotic as determined by another PCR, specific for the preintegration site, which also served as an internal control to confirm the absence of the integration in other individuals.

We further examined whether the association would be observed in an independent, genetically distinct population. Thus, we used a United Kingdom-based population of individuals with chronic hepatitis C virus (HCV) infection and tested the frequency of RASGRF2-int with respect to the route of transmission. Here, we posed a stricter criterion on the history of addictive behavior, as we selected PWIDs who had injected within the previous 6 mo from the sampling date and reported having their first injection at least 2 y before sampling (thus establishing long-term addictive behavior). The control population consisted of subjects who had been infected with HCV through bleeding disorders and matched the PWID population by age (±10 y), gender, and ethnicity. We found RASGRF2-int in 34 out of 100 PWIDs compared with 8 from the 84 controls, revealing a 3.5-fold higher frequency in populations with long-term addictive behavior ($P < 0.001$, χ^2 test, two-sided test) compared with matched controls. We further tested potential confounding of gender, age, alcohol use, and smoking in multivariate models and found that the association of RASGRF2-int with drug addiction remained significant ($P < 0.001$) while no other parameters were found to be significantly associated with RASGRF2-int. A pooled analysis of the Greece- and United Kingdom-based cohorts further supported the strong significance for the association of RASGRF2-int with PWID (23.8% in 202 PWIDs versus 7.6% in 184 control patients; $P < 0.0001$, χ^2 test, two-sided test).

Artificial Insertion of the HK2 LTR Can Modulate *RASGRF2* Transcription. We then explored whether the above-found associations were due to a causal relationship between RASGRF2-int and addiction. Our primary hypothesis is that RASGRF2-int is modulating transcription of *RASGRF2* (Fig. 1), as some intronic ERVs have been shown to modulate transcription in mice (12, 26, 27). RASGRF2-int, like the majority of HK2 intronic integrations of the human genome, is antisense compared with *RASGRF2* (10). Intronic ERVs in mice are mostly antisense and the majority of them do not disrupt normal gene transcription (26), while a minority of antisense intronic mouse ERVs have been shown to disrupt transcription (26, 28). The alternative explanation for the above-found associations is that RASGRF2-int is a proxy of a genetically linked (yet unknown to us) polymorphism of *RASGRF2*, which bears the true causal effect.

To test our primary hypothesis, we used the CRISPR/Cas9 approach (Fig. 1*E*) to introduce the LTR in the same position observed in the human population within the HEK293 cell line, derived from kidney cells but known to have a neuronal transcriptome (29). We performed PCR and Sanger sequencing of the integration and preintegration sites to show that the integration was heterozygous and that there was no off-target editing of the preintegration site. Eighteen days after clonal selection, we evaluated the transcription levels of *RASGRF2* exons. We detected a significant modification of the normal transcription of *RASGRF2* (Fig. 1*G*). More specifically, transcription and splicing of the early exons were significantly increased by more than five times while transcription and splicing of the surrounding exons were significantly diminished by ∼70% compared with the wild type (see Fig. 1*G* legend for detailed statistics). In a previous study, down-regulation of *RASGRF2* by 70% with RNAi resulted in a marked decrease of the RASGRF2 protein (30), suggesting that our observed down-regulation can be also significant at the translational level. Crucially, according to ENSEMBL, two protein-coding transcripts are produced from *RASGRF2* (RASGRF2-201 and RASGRF2-206; Fig. 1*H*), the presence of which in HEK293 cells we confirmed by analyzing publicly available RNA-sequencing (RNA-seq) datasets. RASGRF2-201 includes all of the exons, while RASGRF2-206 includes only the first 10 exons, suggesting that our findings could potentially be explained through down-regulation of RASGRF2-201 and up-regulation of RASGRF2-206. This hypothesis, however, needs to be explored with a stabilized cell-line model which will allow in-depth study of the potential underlying mechanism.

RASGRF2 has at least two independent domains responsible for signaling activities through the guanine exchange factor (GEF), one for Ras and one for Rac1 (Fig. 1*H*). The exons with "diminished" expression are proxies for Ras-GEF activity, and their disruption should result in a decelerated rate of cell division. We therefore expected a fitness cost for the edited cells compared with wild type. Indeed, genome editing produced a slightly diminished survival-under-stress phenotype for the population of the edited cells; 3 out of 48 wells seeded with the clonally expanded edited cell line survived de novo clonal

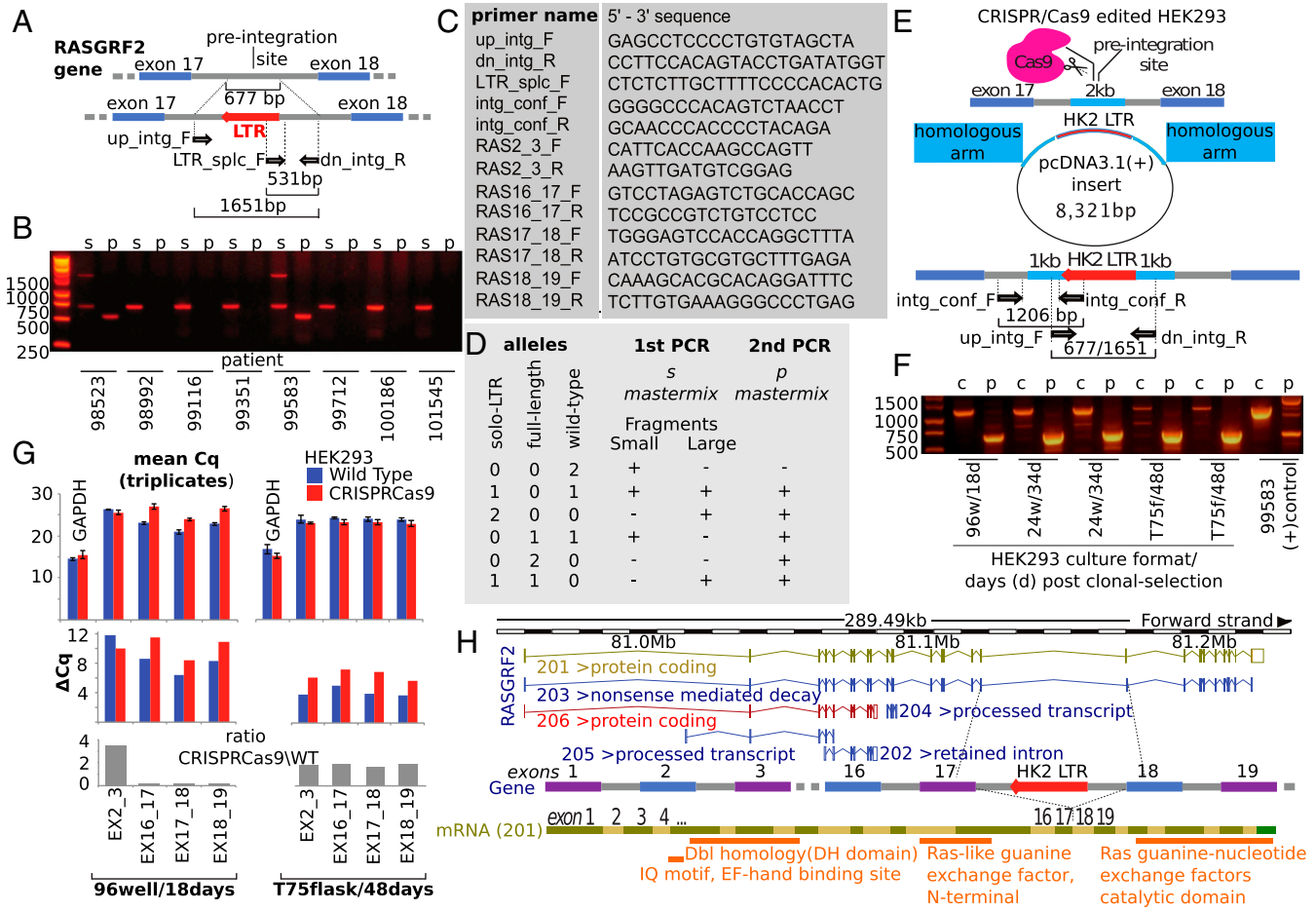


Fig. 1. (*A*) LTR integration screening—PCR design. Primer mapping and product sizes relative to the wild type (*Top*) and the HK2-LTR (red) integrated allele (*Bottom*). Exonic and intronic regions are in blue and gray, respectively. (*B*) Integration screening results of eight (p1 to p8) random patients: Primers LTR_splc_F and dn_intg_R were used in "p," while primers up_intg_F and dn_intg_R were used in "s" mastermix. Patients 98523 and 99583 are positive (heterozygous) for the integration. (*C*) Primer sequences used for the LTR integration screening and the confirmation of the editing in HEK293 cells and for the expression assessment of the RASGRF2 exonic junctions. (*D*) Tabular index for the genotypic interpretation of the PCR products in *B*. (*E*) CRISPR/Cas9 editing of the HEK293 cell line to incorporate RASGRF2-int. pcDNA3.1(+) plasmid containing the LTR sequence flanked by 1-kb preintegration site homologous arms (light blue) was used for the homology directed repair (HDR) insertion mechanism. Primers intg_conf_F (mastermix "c"), which avoid the false-positive amplification of the HDR plasmid, and ingr_conf_R (mapping the 5′ LTR splice site) were used in conjunction with primers up_intg_F and dn_intg_R to confirm integration/genotyping of the cell line. (*F*) Post CRISPR/Cas9 clonal selection and screening/genotyping for the RASGRF2-int HEK293 cells. The gel is annotated according to the scale of the cells in culture over the days posttransfection: 96w/18d (96-well plate, 18 d), 24w/34d (24-well plate, 34 d), and T75f/48d (T75 flask, 48 d). The aneuploidy of the cell line selects for the wild-type alleles versus the edited ones. (*G*) Differential expression of exons (triplicates of the same clone) upon editing of HEK293 cells: SYBR Green qPCR was used to evaluate the expression levels of exon–exon junctions 2–3, 16–17, 17–18, and 18–19. Error bars represent 95% confidence intervals. The expression of the exons around the integration is reduced by more than >70% ($P = 0.01$, $P = 0.047$, and $P = 0.002$ for 16–17, 17–18, and 18–19, respectively, *t* test), while the expression of exons 2–3 is increased by more than threefold ($P = 0.01$, *t* test) compared with the wild-type HEK293 cells, 18 d posttransfection. The expression levels of the exons revert to normal after 48 d posttransfection. (*H*) RASGRF2 gene and alternative transcripts. Positioning of HK2 solo-LTR integration, between exons 17 and 18 of the main, 201, transcript.

selection in nonenriched media compared with 13 out of 48 of the wild type, suggesting a selective disadvantage of the cells harboring the integration ($P < 0.002$, binomial test). Furthermore, chromosome 5 of HEK293 cells (where RASGRF2 lies) is aneuploid, the copy number of which fluctuates during passages (31). We found that within 30 d during serial passaging, the cells were losing the edited allele (Fig. 1), suggesting a selective disadvantage of the allele carrying the integration and a recovery of the normal phenotype (Fig. 1*H*). Remarkably, the modulated transcription profile of *RASGRF2* exons was restored in the cells, which eventually lost the RASGRF2-int allele. The concurrent transcriptional recovery following the loss of the integration also supported that the observed fitness cost was indeed due to the *RASGRF2* editing by the knocking down of the Ras-GEF activity and not due to potential off-target effects.

Unfortunately, the fitness disadvantage resulted in destabilizing the HEK293 clone of RASGRF2-int; we attempted to stabilize the clone twice, but on both occasions loss of the clone as described above was observed (suggesting also that the fitness cost experiment is reproducible). We then attempted to edit eHAP1 cells, a haploid cell line which in theory should allow us to stabilize edited clones if the disruption caused by the integration is not deleterious when in the homozygotic state. On the other hand, if homozygosity has a significant deleterious cost, this would result in failure to expand the edited clone. After screening 92 potential clones, we obtained nine candidates, none of which survived upon expansion (see also *SI Appendix*). This prevented us from obtaining sufficient quantity of the edited clone that would allow us to perform direct detailed in-depth transcriptomic and proteomic analyses of RASGRF2-int. Although multiple mechanisms have been described for disruption of nearby genes by intronic integrations for

other classes of human transposable elements (32), only a few studies have suggested a potential importance of intronic antisense HERVs (13), while the role of polymorphic intronic HERVs is largely unknown. To indirectly explore potential mechanisms of transcriptional modulation that could be involved in the observed in vivo and in vitro phenotypic changes, we then studied intronic transcriptional dynamics of HK2 integrations in a cell line known to highly express HK2.

Intronic HK2 Integrations Can Initiate Transcription. Previous studies showed that at least 50% of HK2 LTRs can initiate transcription (33). To provide in-depth analysis of HK2 intronic transcription, we examined RNA-seq data from teratocarcinoma NCCIT cells known for their embryonic stem-cell identity and elevated HK2 transcription (34). To avoid potential bias as a consequence of the presence/absence of polymorphic integrations, we characterized the HK2 integration sites of NCCIT cells using a target enrichment sequencing approach (*SI Appendix*). We marked 10 differences compared with the hg19 reference genome, 4 integrations existing in hg19 that were missing and 6 integrations absent from hg19 integrations that were present. Based on the HK2 integration sites in NCCIT cells, we analyzed the HK2-related transcriptome in previously published RNA-seq datasets derived from this cell line (35) (Fig. 2). HK2 LTRs have enhancer and promoter signals, and therefore we focused on RNA reads spanning the integration sites (LTR–host junction) of HK2 to evaluate the potential of transcriptional modulatory activity of these integrations. The transcriptional activity of retroviral LTRs is known to be bidirectional (36), with promoter activity being mainly unidirectional (whether the promoter is on the sense or antisense strand), while enhancer activity can be bidirectional or 2D (i.e., acting on promoters found at both sense and antisense DNA strands simultaneously). We categorized the HK2–host junctions as LTR-start and LTR-end junctions and considered that LTR-start RNAs reflect antisense-5′ (A5) transcriptional activity while LTR-end RNAs reflect sense-3′ (S3) transcriptional activity of the LTR (Fig. 2*A*). We note that we only measured reads spanning the LTR–host junctions, that is, we only count "chimeric" reads that are partially HK2 LTR (edge), partially host sequence. These reads, unless HK2 integrations were produced as a result of duplication events, should be unique and, thus, less likely to suffer from mapping issues (e.g., due to the highly repetitive nature of HK2 LTRs or when the host flanking sequence is a highly repetitive region itself).

There was a clear distinction in the abundance of the two classes of RNAs (i.e., those spanning the 5′ or 3′ ends of the LTR–host junction), suggesting that strong transcription is initiated within the LTRs and is due to the enhancer/promoter roles (Fig. 2*B*). The LTR-end RNA abundance contrast observed in the highly transcribed integrations suggests a clear dichotomy of the S3 and A5 activities (Fig. 2*B*), and that strong bidirectional transcriptional activity is not commonly observed in HK2 LTRs. The only exception is a polymorphic integration at chromosome 8 (position 91,696,259), which either suggests bidirectional transcriptional activity or, most likely, is part of the long noncoding RNA NONHSAT127661 that runs through this sequence. It is, consequently, not initiated within and by the LTR. Based on the biology of LTRs, we expected that the S3 activity would be more potent and, indeed, we find that the 3′ RNAs are more abundant than 5′ RNAs ($P = 0.023$).

We then looked at the evolutionary dynamics of the transcriptional activity of the HK2 integration sites. If the ability of the HK2 integration sites to induce RNA transcription were random, then transcription should be eventually diminished in older integrations as a result of random deleterious mutations or local accumulation of repressive epigenetic marks. In effect, we would see the reduction of transcriptionally active integrations as we move back in time (i.e., more recently integrated proviruses should be more transcriptionally active than older proviruses). Based on this model and due to the more potent S3 activity, we would normally expect that the A5 activity would be more rapidly reduced. In striking contrast, we see that the S3 activities have been significantly reduced through time ($P = 0.017$, one-sided, Z test; Fig. 2*F*) (i.e., older integrations have less activity compared with younger integrations), while the A5 has remained resistant to silencing with ∼8% of the LTRs remaining active ($P \sim 0.3$, one-sided, Z test; Fig. 2*F*) even after 30 My. We also observe, for both the A5 and S3 activities, that, within nonsilenced elements, activity is higher for younger (integrated within the last 5 My) rather than older elements (Fig. 2*G*).

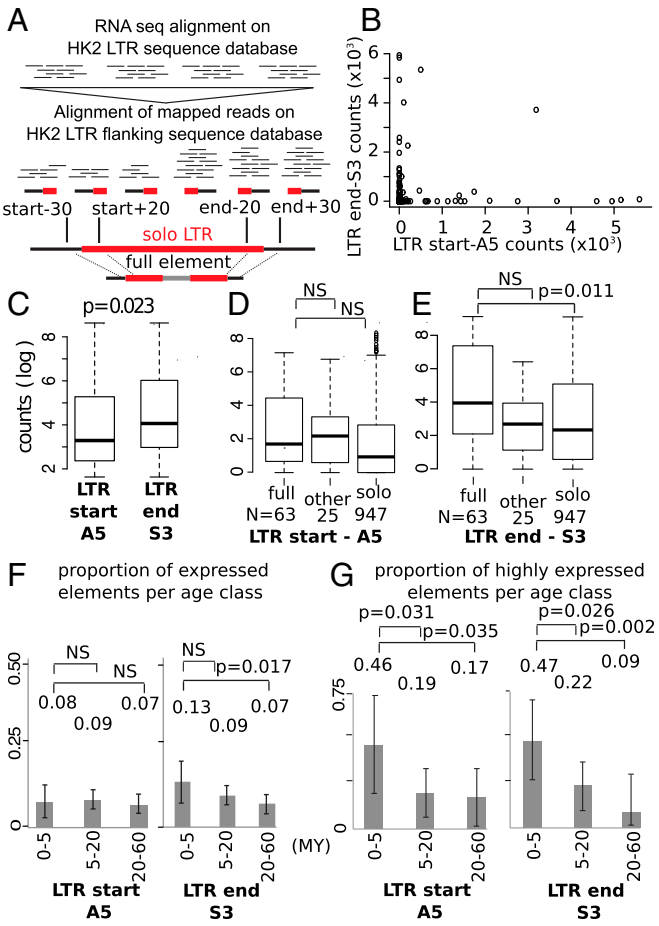


Fig. 2. (*A*) Schematic representation of the pipeline used for the estimation of the expression at the edges of the HK2 integrations in NCCIT cells. LTR (in red) and full-length HK2 elements were used to build a sequence database. RNA-seq HK2-filtered reads were locally realigned against an HK2–host junction database. (*B*) A scatterplot of counts spanning LTR-start vs. LTR-end edges. (*C*–*E*) (Log) Boxplots of mean counts vs. type of element, over LTR-start and LTR-end edges. (*F*) Active (>5 reads) vs. not active (0 to 5 reads) elements, in combination with their age. (*G*) The magnitude of activity (more/less than 500 reads) in combination with the age of the elements, over LTR-start and LTR-end edges. NS, not statistically significant. Error bars indicate 95% confidence intervals.

We then explored whether such a transcriptional dichotomy occurs within introns, which could disrupt transcription through multiple mechanisms including transcriptional collision and RNA interference by *cis*-naturally occurring antisense transcripts (*cis*-NATs) (13, 37–39). HK2 is integrated intronically or within 2 kb of 359 genes (Dataset S1), some of which are important for immunity (*CD4*, *NFKB2*, *APOBEC3C*) and signaling (*DEK*, *RASGRF2*). We found a clear transcriptional dichotomy in 23 genes, most notably *CD4*, *NFKB2*, and *DEK*, suggesting that HK2 LTRs can be triggered to initiate strong transcription within introns, which could in effect modify the functional expression of the respective genes.

Finally, to reproduce the potential of the HK2 LTR for initiating transcription, we used a plasmid reporter gene system to examine whether the HK2 LTR can act as a promoter in both the sense and

antisense directions. After successfully transfecting two cell lines (HeLa and Huh7), we were able to see that both directions initiated transcription, although the antisense direction was more potent than the sense in both cell lines. The finding agrees with the previously reported bidirectional promoter activity of HERV-K LTRs (40) and further supports our hypothesis for a disruption and modulation mechanism of RASGRF2-int, either through *cis*-NAT interference or transcriptional collision.

Discussion

The Potential for Transcriptional Modulation by Intronic HK2. Thirty-three percent of polymorphic HK2 integrations are within or near genes (within 2 kb), of which 60% are intronic (8). Retroviral integrations preferentially occur near genes, because euchromatic (41) areas are more easily accessible to retroviral integrase. Perhaps intronic integrations are more likely to survive within the hypomethylated introns of genes (42) where retroviral transcription is allowed, rather than within heavily methylated areas of the genome. These intronic integrations exhibit an antisense bias with respect to their directionality compared with the nearby host genes (43), suggesting stronger negative selection on sense-oriented insertions, possibly due to their interference with transcription signals (44). Indeed, at the time of integration, the reconstituted HK2 (HERVKcon) ancestral virus does not exhibit directional bias in vitro with respect to genes (45), and younger integrations seem to have less bias than older elements, although this is not statistically significant (Dataset S1). It has thus been hypothesized that antisense intronic solo-LTRs have no or minimal pathogenicity (10).

Based on the RNA-seq analysis and our functional experiment with the *RASGRF2* integration, we suggest that HK2 antisense integrations can modify the human transcriptome. We thus suggest a model for transcriptional activity of HK2 elements where, at the time of integration or soon after, novel proviruses are either silenced or remain active but an adjustment in their regulatory activity results in intermediate transcription over time.

Genetics and in Vitro and in Vivo Phenotypes of RASGRF2-Int. In accordance with the well-established role of RASGRF2 with addiction, we observed a statistically significant association of RASGRF2-int with i.v. drug use in two independent genetically distinct populations. PWIDs are three times more likely to carry the RASGRF2-int allele than non-PWID controls. Larger studies are required to solidify our findings in PWIDs and potentially with other addiction phenotypes. A reward-related phenotype such as addictive behavior is conditionally pathogenic and, in theory, might even promote beneficial behavior under different circumstances (46). We find that RASGRF2-int had a slightly higher frequency in the United Kingdom-based population compared with the Greece-based population, although the study design does not allow for a safe comparison between populations. There are, however, multiple sources of evidence that the frequency of RASGRF2-int in the human population varies in time and space.

RASGRF2-int has been found in both full genomes of Neanderthal and Denisovan, suggesting that the time of integration was older than diversification from modern humans and most importantly at a much higher frequency in the archaic hominin lineage (47). On the other hand it is rare (~5%) in modern human populations, with no homozygotic individual recovered in our study, since RASGRF2-int was heterozygous in every person tested. This could be due to either our sample size not being sufficiently large or because the homozygous state is deleterious for the host and thus under strong negative selection. If the homozygous state is indeed deleterious while the heterozygous state provides beneficial traits to the host, the prolonged existence of RASGRF2-int in human populations could be explained by balancing selection. In this case, the reward trait of RASGRF2 could have had a much more significant effect in archaic hominin populations compared with modern humans. Interestingly, in line with our PCR screening of controls, RASGRF2-int is found in ~5% of the human population of non-African origin, but has recently been reported to be absent among East Asian populations (8). Larger studies will, thus, also clarify the possible existence of homozygotes of RASGRF2-int, determine their potential phenotype, and allow us to test for the hypothesis of balancing selection.

Accordingly, our CRISPR/Cas9-edited cells showed a transcriptional modification of *RASGRF2* even though HK2 integration was only in one of the alleles. We would typically expect the wild-type allele to sufficiently express the gene product, but the intensity of the effect suggests that the "edited" allele has a *trans* effect on "unedited" wild-type transcripts. This can be explained through RNA interference of the overhanging non-LTR homologous segment of the resulting NAT, but we cannot exclude that transcription collision also operates in the edited allele. We then showed that RASGRF2-int had a fitness cost in cell-culture conditions. We should, however, note that RASGRF2-int fitness cost was nonlethal in heterozygotic cells, as the cells were able to survive and proliferate though at a slower rate. Based on our experiments, we cannot, however, exclude that the fitness cost of RASGRF2-int in the homozygotic state is more deleterious; thus, disentangling the in vivo phenotype of RASGRF2-int is challenging.

The function of *RASGRF2* has been well-studied in vivo and there were clear predictions to be tested with an epidemiological association study. Based on our cell-line experiments, we suggest that RASGRF2-int leads to enhancement of dopaminergic activity through higher expression of the first exons of *RASGRF2*, which then results in increased potential for addiction. Our study provides the rationale to explore a wider effect of RASGRF2-int in the dopaminergic activity through independent studies on the frequency of RASGRF2-int in persons with drug use and other addictive behaviors, as well as functional MRI experiments of the RASGRF2-int carriers.

Mechanism and Pathogenicity of HERV-K HML-2 Transcriptional Rewiring. We suggest that intronic HERVs can trigger transcriptional collision and/or RNA interference through *cis*-NATs (48). Under this mechanistic model, the implicated genes, even if disrupted, still have ongoing transcription which can be sufficiently functional. When intronic HK2 transcription becomes upregulated, the interrupted genes will have a modified transcriptional profile; whether the resulting phenotypes are beneficial or pathogenic will depend on the protein domains upstream and downstream of the integration, the directionality of the integration, and the environmental exposure. It is possible that other unknown potentially pathogenic mechanisms can operate as a result of HK2 activity; an intriguing hypothesis would be that HK2 integrations can modify transcription through remodeling of chromatin structure, as has been recently demonstrated for human T-lymphotropic virus (49). It needs, however, to be noted that although there is a well-established link between mRNA levels and protein translation in the case of *RASGRF2* (30), this cannot be generalized to other genes; depending on the characteristics of the translated protein (e.g., half-life), a disruption at the mRNA level might have less dramatic effects (50). Our findings thus point towards a conditionally pathogenic role for HK2 integrations through rewiring of the transcriptome; although disease associations and HK2 integrations have been reported, causality has not been sufficiently supported.

Our study demonstrates that at least a percentage of genomic contributions to disease and phenotypes can be missed as a result of inaccurate mapping of polymorphisms related to repetitive elements such as HERVs and other retrotransposons. Developing algorithms and technologies that will accurately recover this difficult part of the human genome and transcriptome will be important to further establish links between HERVs and human phenotypes.

Materials and Methods

Patient Samples. Ethical approval for the use of Greek-population DNA samples (PWIDs and controls) was obtained from the Research Ethics Committee of the University of Athens. The DNA samples of the United Kingdom-based population were obtained from HCV Research UK. Ethics

approval for HCV Research UK was given by National Health Services Research Ethics Committee East Midlands-Derby 1 (Research Ethics Committee reference 11/EM/0314). Informed consent was obtained at the time of collection.

Wet Laboratory. The wet laboratory procedures are described in detail in *SI Appendix* and include the design of target enrichment and high-throughput sequencing, procedures used to culture cells, design and protocols used for genome engineering (CRISPR/Cas9 design and execution), PCR screening for RASGRF2-int patient samples and qPCR on cell lines, and assessment of the HK2 LTR promoter in plasmids.

Bioinformatics.

Characterization of the HK2 integrome of NCCIT cells. NCCIT cells were sequenced using a customized target-enrichment method (*Wet Laboratory*) to capture HK2 LTRs with flanking host sequences. These genomic fragments prepared in specially optimized large insert sizes (800 to 1,200 bp) and 2 × 300-bp paired-end sequencing resulted in reads that were mapped using Novoalign (www.novocraft.com/products/novoalign/) allowing for soft clipping. Reads were then overlapped with a previous list of known HK2 integration sites to mark presence/absence (8).

RNA-Seq. Previously published RNA-seq data (50-bp-long reads) from NCCIT cells (51) were filtered after local mapping alignment [Bowtie2 default settings (52)] on an HK2 LTR sequence database. The partially mapped reads were then locally mapped against the HK2 host–junction sequence database after extracting 50-bp fragments at the edges of each element (30 nt before and 20 nt after the start coordinate or 20 nt before and 30 nt after the end coordinate of each element). We made a comprehensive list of NCCIT-integrated elements in the datasets derived from our HK2 target-enrichment NCCIT libraries. We estimated the mean read counts by calculation of the mean read depth across each 50-bp fragment. The mean counts were then adjusted to match the orientation of each element; for example, the 3′ counts were considered to be "LTR-start" counts for negative-oriented elements.

ACKNOWLEDGMENTS. We thank Paul Klenerman for critical reading of the manuscript. We wish to acknowledge the role of the HCV Research UK Biobank (Award C0365) in collecting and making available samples and data for this publication. The study has been supported by Medical Research Council UK (Project MR/K010565/1). J.M. was supported by an MRC Award (MC_UU_12014/1). A.K. was funded by the Royal Society.

1. Bannert N, Kurth R (2006) The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet* 7:149–173.
2. Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R (2012) Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci USA* 109: 7385–7390.
3. Katzourakis A, et al. (2014) Larger mammalian body size leads to lower retroviral activity. *PLoS Pathog* 10:e1004214.
4. Stoye JP (2012) Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol* 10:395–406.
5. Magiorkinis G, Blanco-Melo D, Belshaw R (2015) The decline of human endogenous retroviruses: Extinction and survival. *Retrovirology* 12:8.
6. Medstrand P, Mager DL (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol* 72:9782–9787.
7. Marchi E, Kanapin A, Magiorkinis G, Belshaw R (2014) Unfixed endogenous retroviral insertions in the human population. *J Virol* 88:9529–9537.
8. Wildschutte JH, et al. (2016) Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci USA* 113:E2326–E2334.
9. Löwer R, et al. (1993) Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci USA* 90:4480–4484.
10. Brady T, et al. (2009) Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev* 23:633–642.
11. Wu X, Burgess SM (2004) Integration target site selection for retroviruses and transposable elements. *Cell Mol Life Sci* 61:2588–2596.
12. Zhang Y, Romanish MT, Mager DL (2011) Distributions of transposable elements reveal hazardous zones in mammalian introns. *PLoS Comput Biol* 7:e1002046.
13. Gogvadze E, Stukacheva E, Buzdin A, Sverdlov E (2009) Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions. *J Virol* 83: 6098–6105.
14. Sverdlov ED (2005) *Retroviruses and Primate Genome Evolution* (Landes Bioscience, Georgetown, TX).
15. Kazazian HH, Jr (2004) Mobile elements: Drivers of genome evolution. *Science* 303: 1626–1632.
16. Wang T, et al. (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci USA* 104:18613–18618.
17. Montesion M, Bhardwaj N, Williams ZH, Kuperwasser C, Coffin JM (2017) Mechanisms of HERV-K (HML-2) transcription during human mammary epithelial cell transformation. *J Virol* 92:e01258-17.
18. Fernández-Medarde A, et al. (2002) Targeted disruption of Ras-Grf2 shows its dispensability for mouse growth and development. *Mol Cell Biol* 22:2498–2504.
19. Ruiz S, Santos E, Bustelo XR (2007) RasGRF2, a guanosine nucleotide exchange factor for Ras GTPases, participates in T-cell signaling responses. *Mol Cell Biol* 27:8127–8142.
20. Schumann G, et al. (2011) Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci USA* 108:7119–7124, and erratum (2011) 108:9316.
21. Stacey D, et al.; IMAGEN Consortium (2012) RASGRF2 regulates alcohol-induced reinforcement by influencing mesolimbic dopamine neuron activity and dopamine release. *Proc Natl Acad Sci USA* 109:21128–21133.
22. Fasano S, et al. (2009) Ras-guanine nucleotide-releasing factor 1 (Ras-GRF1) controls activation of extracellular signal-regulated kinase (ERK) signaling in the striatum and long-term behavioral responses to cocaine. *Biol Psychiatry* 66:758–768.
23. Barak S, Carnicella S, Yowell QV, Ron D (2011) Glial cell line-derived neurotrophic factor reverses alcohol-induced allostasis of the mesolimbic dopaminergic system: Implications for alcohol reward and seeking. *J Neurosci* 31:9885–9894.
24. Anborgh PH, et al. (1999) Ras-specific exchange factor GRF: Oligomerization through its Dbl homology domain and calcium-dependent activation of Raf. *Mol Cell Biol* 19: 4611–4622.
25. Lutchman M, et al. (2002) Dematin interacts with the Ras-guanine nucleotide exchange factor Ras-GRF2 and modulates mitogen-activated protein kinase pathways. *Eur J Biochem* 269:638–649.
26. Zhang Y, Babaian A, Gagnier L, Mager DL (2013) Visualized computational predictions of transcriptional effects by intronic endogenous retroviruses. *PLoS One* 8: e71971.
27. Maksakova IA, et al. (2006) Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genet* 2:e2.
28. Babaian A, Mager DL (2016) Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 7:24.
29. Shaw G, Morse S, Ararat M, Graham FL (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J* 16: 869–871.
30. Ma X, et al. (2014) βArrestin1 regulates the guanine nucleotide exchange factor RasGRF2 expression and the small GTPase Rac-mediated formation of membrane protrusion and cell motility. *J Biol Chem* 289:13638–13650.
31. Lin YC, et al. (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* 5:4767.
32. Kaer K, Speek M (2013) Retroelements in human disease. *Gene* 518:231–241.
33. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006) At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. *J Virol* 80:10752–10762.
34. Boller K, et al. (1993) Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* 196:349–353.
35. Grow EJ, et al. (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522:221–225.
36. Arpin-André C, Laverdure S, Barbeau B, Gross A, Mesnard JM (2014) Construction of a reporter vector for analysis of bidirectional transcriptional activity of retrovirus LTR. *Plasmid* 74:45–51.
37. Osato N, Suzuki Y, Ikeo K, Gojobori T (2007) Transcriptional interferences in *cis* natural antisense transcripts of humans and mice. *Genetics* 176:1299–1306.
38. Carmichael GG (2003) Antisense starts making more sense. *Nat Biotechnol* 21: 371–372.
39. Zinad HS, Natasya I, Werner A (2017) Natural antisense transcripts at the interface between host genome and mobile genetic elements. *Front Microbiol* 8:2292.
40. Domansky AN, et al. (2000) Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett* 472:191–195.
41. Schröder ARW, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521–529.
42. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* 27:404–408.
43. Medstrand P, van de Lagemaat LN, Mager DL (2002) Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res* 12:1483–1495.
44. van de Lagemaat LN, Medstrand P, Mager DL (2006) Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biol* 7:R86.
45. Bushman F, et al. (2005) Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* 3:848–858.
46. Spanagel R, Weiss F (1999) The dopamine hypothesis of reward: Past and current status. *Trends Neurosci* 22:521–527.
47. Agoni L, Golden A, Guha C, Lenz J (2012) Neandertal and Denisovan retroviruses. *Curr Biol* 22:R437–R438.
48. Pelechano V, Steinmetz LM (2013) Gene regulation by antisense transcription. *Nat Rev Genet* 14:880–893.
49. Satou Y, et al. (2016) The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. *Proc Natl Acad Sci USA* 113:3054–3059.
50. Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583:3966–3973.
51. Jung I, et al. (2012) H2B monoubiquitylation is a 5′-enriched active transcription mark and correlates with exon-intron structure in human cells. *Genome Res* 22:1026–1035.
52. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

Chapter 10, "Dire wolves were the last of an ancient New World canid lineage"

Summary

This study reports the first ancient nuclear genomes derived from five dire wolves, an iconic extinct mega-carnivore species that lived in North America during the Pleistocene, as well as the genome sequence of the two African jackal species (the last species of the genus that were not yet sequenced), and the first collagen sequence obtained from a La Brea Tar Pits dire wolf. Despite being morphologically highly similar, and being classified as closely related species, dire wolves and gray wolves are highly differentiated species that diverged around 5 million years ago. This paper uses biogeographic and paleontological evidence to show that the dire wolf was in fact the last species of a canid lineage that thrived in North America for over two million years. Also, even though dire wolves and gray wolves overlapped in North America for thousands of years, our results suggest that they did not interbreed.

Authors

Angela R. Perri, Kieren J. Mitchell[2], Alice Mouton, Sandra Alvarez Carretero, Ardern Hulme-Beaman, James Haile, Alexandra Jamieson, Julie Meachen, ***Audrey T. Lin****, Blaine W. Schubert, Carly Ameen, Pere Bover, Selina Brace, Alberto Carmagnini, Christian Carøe, Jose A. Samaniego Castruita, James C. Chatters, Keith Dobney, Mario Dos Reis, Allowen Evin, Philippe Gaubert, Shyam Gopalakrishnan, Graham Gower, Holly Heinger, Kristofer Helgen, Josh Kapp, Anna Linderholm, Andrew T. Ozga, Samantha Presslee, Alexander Salis, Nedda F. Saremi, Colin Shew, Katherine Skerry, Mary Thompson, Matthew J. Collins, Mikkel-Holger S. Sinding, M. Thomas P. Gilbert, Anne C. Stone, Beth Shapiro, Blaue Van Valkenburg, Robert K. Wayne, Greger Larson, Laurent A. F. Frantz, and Alan Cooper.*

Status

Under review in Science.

My contribution

In this paper, I infer the evolutionary history of dire wolves and extant canids by constructing a phylogeny using protein sequences. I built a phylogenetic tree using the COL1A1 and COL1A2 amino acid sequence generated by MS/MS analysis from a dire wolf and other Carnivora species, whose sequences were publically available. These Carnivora species consist of: redfox (*Vulpes vulpes*; XP_025851655, XP_025859557), dog (*Canis familiaris*; NP_001003090, NP_001003187), dingo (*Canis familiaris dingo*; XP_025295726, XP_025327115), grizzly bear (*Ursus arctos*; XP_026368913, 800 XP_026361636), northern seal (*Callorhinus ursinus*; XP_025715155, XP_025728689) and a cat (*Felis catus*; XP_003996748, XP_003982813). The COL1A1 and COL1A2 gene sequences from all other modern canid were also downloaded from Genbank: gray fox, andean fox, black backed jackal, side striped jackal, african wild dog, dhole, ethiopian wolf, golden wolf, wolf, and yellowstone wolf. To extract the amino acid sequences from the DNA sequences, the GFF file from Ensembl (v95) for canFam3.1 was downloaded. The coding sequences of COL1A1 and COL1A2 were then extracted and translated using gffread v0.10.5. The amino acid sequences for COL1A1 and COL1A2 were then concatenated for each species and then aligned using MAFFT v7.123b (97). All leucine (L) and isoleucine (I) from the alignment were masked because these are isobaric, and low-energy MS/MS sequencing is not capable of discriminating between them. A phylogenetic tree was generated using MrBayes version 3.2.1 (65) with the amino-acid model estimated from the data (prset aamodellpr = mixed). Two runs of four chains were run each with 1,000,000 generations. Convergence was assessed by ensuring all ESS for all parameters where higher than 100. The result of this work can be seen in Figure S12.

Title: Dire wolves were the last of an ancient New World canid lineage

Authors: Angela R. Perri[1,*§], Kieren J. Mitchell[2,*§], Alice Mouton[3,*], Sandra Álvarez-Carretero[4,*], Ardern Hulme-Beaman[5,6], James Haile [7], Alexandra Jamieson[7], Julie Meachen[8], Audrey T. Lin[7,9], Blaine W. Schubert[10], Carly Ameen[11], Ekaterina E. Antipina[12], Pere Bover[13], Selina Brace[14], Alberto Carmagnini[4], Christian Carøe[15], Jose A. Samaniego Castruita[15], James C. Chatters[16], Keith Dobney[5], Mario dos Reis[4], Allowen Evin[17], Philippe Gaubert[18], Shyam Gopalakrishnan[15], Graham Gower[2], Holly Heiniger[2], Kristofer M. Helgen[19], Josh Kapp[20], Pavel A. Kosintsev[21], Anna Linderholm[7, 22], Andrew T. Ozga[23, 24], Samantha Presslee[25], Alexander T. Salis[2], Nedda F. Saremi[20], Colin Shew[3], Katherine Skerry[24], Dmitry E. Taranenko[26], Mary Thompson[27], Mikhail V. Sablin[28],Yaroslav V. Kuzmin[29, 30], Matthew J. Collins[31, 32], Mikkel-Holger S. Sinding[15, 33], M. Thomas P. Gilbert[15, 34], Anne C. Stone[23 ,24], Beth Shapiro[20, 35], Blaire Van Valkenburgh[3], Robert K. Wayne[3], Greger Larson[7], Laurent Frantz[4, 7#§], and Alan Cooper[2#§]

Affiliations:

[1]*Department of Archaeology, Durham University, Durham, UK*
[2]*Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Australia*
[3]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA*
[4]*School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*
[5]*Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK*
[6]*School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK*
[7]*The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK*
[8]*Department of Anatomy, Des Moines University, Des Moines, IA, USA*
[9]*Department of Zoology, University of Oxford, Oxford, UK*
[10]*Center of Excellence in Paleontology & Department of Geosciences, East Tennessee State University, Johnson City, TN, USA*
[11]*Department of Archaeology, University of Exeter, Exeter UK*
[12]*Institute of Archaeology, Russian Academy of Sciences, Moscow, Russia*
[13]*ARAID Foundation, Instituto Universitario de Investigación en Ciencias Ambientales (IUCA) - Aragosaurus Group, Universidad de Zaragoza, Spain*
[14]*Department of Earth Sciences, Natural History Museum, London, UK*
[15]*Section for Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen, Denmark*
[16]*Applied Paleoscience, Bothell, WA, USA*
[17]*Institut des Sciences de l'Evolution - Montpellier, CNRS, Université de Montpellier, IRD, EPHE, France*
[18]*Laboratoire Evolution & Diversité Biologique, UPS/CNRS/IRD, Université Paul Sabatier, Toulouse, France*
[19]*School of Biological Sciences and Environment Institute, University of Adelaide, Australia*
[20]*Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA,USA*

[21]*Institute of Plant and Animal Ecology, Urals Branch of the Russian Academy of Sciences, Yekaterinburg, Russia*
[22]*Department of Anthropology, Texas A&M University, College Station, TX, USA*
[23]*Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA*
[24]*School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA*
[25]*Department of Archaeology, University of York, York, UK*
[26]*Institute of Systematics and Ecology of Animals, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*
[27]*Idaho Museum of Natural History, Idaho State University, Pocatello, ID, USA*
[28]*Zoological Institute of the Russian Academy of Sciences, Universitetskaya nab. 1, 199034, St. Petersburg, Russia Federation*
[29]*Sobolev Institute of Geology and Mineralogy, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*
[30]*Tomsk State University, Tomsk, Russia*
[31]*Department of Biology, University of Copenhagen, Copenhagen, Denmark*
[32]*McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK*
[33]*Greenland Institute of Natural Resources, Nuuk, Greenland*
[34]*NTNU University Museum, Trondheim, Norway*
[35]*Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA, USA*

§ Corresponding author. Email: angela.r.perri@durham.ac.uk (A.R.P.); kieren.mitchell@adelaide.edu.au (K.J.M.); laurent.frantz@qmul.ac.uk (L.A.F.F.); alanjcooper42@gmail.com (A.C.)
[#] These authors co-supervised this work.
*These authors contributed equally to this work.

Abstract: The dire wolf is considered one of the most common and widespread large carnivores in Pleistocene America, yet relatively little is known about its evolution or extinction. To reconstruct the evolutionary history of dire wolves, we sequenced five genomes from sub-fossil bones dating from 13,000 to over 50,000 years ago. Our results indicate that though they were similar morphologically to the extant gray wolf, dire wolves were a highly divergent lineage that split from living canids ~5.5 million years ago. In contrast to numerous examples of hybridization across Canidae, there is no evidence for gene flow between dire wolves and either North American gray wolves or coyotes. This suggests that the dire wolf evolved in isolation from the Pleistocene ancestors of these species. Our results also support an early New World origin of the dire wolf, while the ancestors of the gray wolf, coyote, and dhole evolved in Eurasia and only colonized North America relatively recently.

Main Text: The dire wolf (*Canis dirus*) was a large (~68 kg) wolf-like canid and among the most common extinct large carnivores of the American Late Pleistocene megafauna (*1*). Dire wolf remains are present in the North American paleontological record from at least ~250,000 to ~13,000 years ago, at the end of the Pleistocene, particularly in the lower latitudes (*2*) (Figure 1A). Other canid species present in Late Pleistocene North America include the slightly smaller gray wolf (*C. lupus*), the much smaller coyote (*C. latrans*) and the dhole (or Asiatic wild dog; *Cuon alpinus*), though dire wolves appear to be more common overall (*1*). For example, >4,000

individuals have been excavated in California's fossil-rich Rancho La Brea tar seeps alone, where they outnumber gray wolves more than 100-fold (*3*).

Despite the abundance of dire wolf fossils, the origin, taxonomic relationship, and ultimate driver of this species' extinction remain unclear. The leading hypothesis is that dire wolves, unlike gray wolves and coyotes, were highly specialized, and were unable to survive when their megafaunal prey went extinct (e.g. (*4–6*)). Dire wolves, however, are generally described as a sister species to (*7–9*), or even conspecific with (*10*) the gray wolf, and although the two species can be differentiated using geometric morphometric techniques (*11*), their overall morphological similarity (Figure 1B; Figure S1-6) supports a close evolutionary relationship. Given these apparent similarities, the disappearance of the dire wolf during the end-Pleistocene extinction is perplexing. One hypothesis is that these morphological similarities are the result of convergence and that the dire wolf instead represents a species belonging to a separate taxonomic lineage (sometimes classified in a separate genus *Aenocyon*; "terrible or dreadful wolf" (*12*)). Conceivably, this unique species was more susceptible to end-Pleistocene events, such as the extinction of the megafauna, relative to the gray wolf.

To resolve the evolutionary history of the dire wolf, we screened 46 sub-fossil specimens for the presence of preserved genomic DNA (Table S1). We identified five samples from Idaho (DireAFR & DireGB), Ohio (DireSP), Tennessee (DireGWC), and Wyoming (DireNTC) ranging in age from 12,900 to >50,000 years before present that possessed sufficient endogenous DNA to obtain both mitochondrial genomes (between ~1x and 31x coverage) and low-coverage nuclear genome sequences (~0.01x to 0.23x coverage) using hybridization capture or shotgun sequencing methods (*11*). Although we did not successfully sequence DNA from the La Brea tar seeps dire wolf specimens, one specimen did contain type I collagen (COL 1) suitable for sequencing using paleoproteomic methods (*11*)(Table S1).

Analyses of the dire wolf COL 1 sequence suggested that they were highly divergent from gray wolves, coyotes, African wolves (*C. anthus*), and dogs (*C. familiaris*) (Figure S7), but these data could not confidently resolve the relationships between more distantly related canids due to a lack of lineage-specific amino acid changes among these species (*13*). Phylogenetic analyses of the mitochondrial genomes indicated that dire wolves form a well-supported monophyletic group that is highly divergent from gray wolves and coyotes (Figure S10; see Table S14-15 for a list of the thirteen species used in this analysis), contradicting recent paleontological analyses (*7–9*). Canid mitochondrial phylogenies, however, are known to be impacted by both admixture and incomplete lineage sorting (*14*).

In order to fully resolve the phylogenetic relationships of dire wolves, we analyzed our novel dire wolf nuclear genomic data with previously published genomic data from eight extant canids: gray wolf, coyote, African wolf, dhole, Ethiopian wolf (*C. simensis*), African wild dog (*Lycaon pictus*), Andean fox (*Lycalopex culpaeus*), and gray fox (*Urocyon cinereoargenteus*; an outgroup). Of these species, the geographical ranges of the gray wolf, coyote, dhole, and gray fox formerly overlapped with that of the dire wolf during the Pleistocene (Figure 1A). We also generated new nuclear genome sequences from the two endemic African jackals, the black-backed and side-striped jackal (*C. mesomelas* and *C. adustus*, respectively), in order to ensure all extant members of the "wolf-like canid" clade (comprising *Canis*, *Lycaon*, *Cuon*, and their extinct relatives) were represented (Table S14). Although the maximum likelihood phylogenetic analyses of the nuclear genome sequences confirmed a distant relationship between the dire wolf and the other wolf-like canids, we could not definitively resolve whether the dire wolf was the

basal member of this clade or the second lineage to diverge after the ancestor of the two African jackals (Figure 2A; Figure S11; Figure S15).

We investigated canid phylogenetic relationships in greater detail using a range of species tree analyses (*11*). Each approach produced highly concordant trees that support the monophyly of three primary lineages: dire wolves, African jackals, and a clade comprising all other extant wolf-like canids (Figure 2A; Figures S11-S15). Although our species tree analyses provided equivocal results regarding the relationships among these lineages, the gray wolf (genus *Canis*) is clearly more closely related to the dhole (genus *Cuon*) and African wild dog (genus *Lycaon*) than to either the dire wolf or African jackals (both genus *Canis*). This finding is consistent with previously proposed designations of genus *Lupulella (15)* for the African jackals and *Aenocyon* (*12*) for the dire wolf. Molecular dating of the genomic data confirmed that the initial divergences of the three primary wolf-like canid lineages occurred rapidly, which likely contributed to the poor resolution in the tree as a result of incomplete lineage sorting and ancestral admixture (Figure 2A). The dire wolf lineage last shared a common ancestor with extant wolf-like canids ~5.5 million years ago (95% HPD=3.9-8.2 million years ago; Figure 2A), followed by the divergence of the African jackals ~4.9 million years ago (95% HPD=3.5-7.4 million years ago; Figure 2B).

Given the propensity for sympatric canid species to interbreed (*14*), we tested for genomic signals of admixture between extant North American canids and the dire wolf using D statistic (*11*). Specifically, we computed statistics of the form D (outgroup [gray fox]; dire wolf; North American canid [gray wolf or coyote], African wolf/Eurasian wolf) and found no significant excess of shared derived alleles between the dire wolf and the extant North American canids. This suggests that dire wolves did not hybridize with gray wolves, coyotes, or their recent ancestors in North America (Figure 2B; Figure S18). We did, however, find that the African wild dog shares significantly fewer derived alleles with the dire wolf than the gray wolf, coyote, African wolf, dhole or the Ethiopian wolf (Figure 2C; Figure S19). This indicates an episode of ancient admixture at least ~4.0-5.0 million years ago (Figure 2A) between the ancestor of the dire wolf and the ancestor of the wolves, coyote and dhole, which may have contributed to the difficulty in definitively resolving the branching order of the basal wolf-like canid lineages (Figure 2A).

Hybridization frequently occurs between divergent wolf-like canid lineages where population ranges overlap. For example, modern gray wolves and coyotes readily hybridize in North America (e.g., (*16*)). In addition, genomic data suggest gene flow occurred between the dhole and African wild dogs during the Pleistocene (*14*), millions of years after their divergence. Consequently, the lack of evidence for gene flow between the dire wolf and gray wolves, coyotes or their common ancestor, despite substantial range overlap with dire wolves during the Late Pleistocene, suggests that these lineages diverged in geographical isolation. This result is consistent with the hypothesis that dire wolves originated in the Americas (*1*, *17–19*), likely from the extinct Armbruster's wolf (*C. armbrusteri (7)*).

Long term isolation of the dire wolf lineage in the Americas implies that other American fossils, such as the Pliocene *C. edwardii*, a proposed relative of the coyote (*7*), could also belong to the dire wolf lineage. A corollary of this suggestion is that diversification of the extant wolf-like canids occurred in parallel outside of the Americas, and perhaps began earlier than previously supposed, with the living *Canis* species possibly descending from Old World members of the extinct genus *Eucyon*, which first appear in the fossil record of Africa and Eurasia at the end of

the Miocene (see (*20*)). Geographic isolation since the late Miocene is consistent with our molecular estimates for the age of the dire wolf lineage, and may have allowed the dire wolf to evolve some degree of reproductive isolation prior to the Late Pleistocene North American arrival of other taxa such as the gray wolf, coyote, dhole, and *Xenocyon* (another extinct wolf-like canid).

Despite their overall phenotypic similarities, gray wolves and coyotes survived the Late Pleistocene megafaunal extinctions while dire wolves did not. One possible reason may be that both gray wolves and coyotes possessed greater morphological plasticity (*21*) and dietary flexibility, thus allowing them to avoid extinction and become the dominant terrestrial predators in North America (*6*, *22*). This possibility is supported by the late date we obtained from the DireGWC specimen (12,820-12,720 calBP), which suggests that the dire wolf survived until at least the Younger Dryas cold reversal, a period that also witnessed the latest known dates for other specialized North American mega-carnivores such as the American lion (*Panthera atrox*) and giant short-faced bear (*Arctodus simus*)(*23*). Alternatively, gray wolves and coyotes may have survived as a result of their ability to hybridize with other canids. Through adaptive introgression, North American gray wolves are known to have acquired traits related to coat color, hypoxia, and immune response (*24*, *25*). Specifically, enhanced immunity may have allowed them to resist novel diseases carried by newly arriving taxa. Our results suggest that dire wolves were unable to interbreed with other wolf-like canids, and this lack of genomic flexibility may have prevented them from acquiring the traits necessary to survive into the Holocene.

general, this should include a brief (1-2 paragraph) introduction, followed by a statement of the specific scope of the study, followed by results and then interpretations. Please avoid statements of future work or claims of priority, and avoid repeating the conclusions at the end.

References and Notes:

1. R. G. Dundas, *Boreas*. **28**, 375–385 (1999).
2. J. A. Meachen, A. L. Brannick, T. J. Fry, *Ecol. Evol.* **6**, 3430–3438 (2016).
3. J. A. Leonard *et al.*, *Curr. Biol.* **17**, 1146–1150 (2007).
4. R. W. Graham, J. I. Mead, *North America and adjacent oceans during the last deglaciation*, 371–402 (1987).
5. A. D. Barnosky, *Mass extinctions: processes and evidence*, 235–254 (1989).
6. L. R. G. DeSantis *et al.*, *Curr. Biol.* **29**, 2488–2495.e2 (2019).
7. R. H. Tedford, X. Wang, B. E. Taylor, *Bull. Am. Mus. Nat. Hist.* **325**, 1–218 (2009).
8. F. J. Prevosti, *Cladistics*. **26**, 456–481 (2010).
9. J. Zrzavý, P. Duda, J. Robovský, I. Okřinová, V. Pavelková Řičánková, *Zool. Scr.* **47**, 373–389 (2018).
10. G. D. Goulet, Comparison of Temporal and Geographical Skull Variation Among Nearctic

Modern, Holocene and Late Pleistocene Gray Wolves (Canis Lupus) (and Selected Canis), University of Manitoba (1993).

11. Supplementary Material.

12. J. C. Merriam, *Bulletin of the Department of Geology of the University of California*. **10**, 531–533 (1918).

13. M. Buckley, V. L. Harvey, A. T. Chamberlain, *Boreas*. **46**, 402–411 (2017).

14. S. Gopalakrishnan *et al.*, *Curr. Biol.* **28**, 3441–3449.e5 (2018).

15. D. Geraads, *Palaeontology*. **54**, 429–446 (2011).

16. B. M. vonHoldt *et al.*, *Sci Adv*. **2**, e1501714 (2016).

17. R. M. Nowak, North American Quaternary Canis, Monograph of the Museum of Natural History, The University of Kansas (1979).

18. B. Kurtén, E. Anderson, *Pleistocene mammals of North America* (Columbia University Press, 1980).

19. R. M. Nowak, *Wolves: Behavior, ecology, and conservation*, 239–258 (2003).

20. M. Sotnikova, L. Rook, *Quat. Int.* **212**, 86–97 (2010).

21. R. Driggers, *Appl. Opt.* **54**, ED12 (2015).

22. B. Van Valkenburgh, F. Hertel, *The decline of North American predators during the late Pleistocene* (4320).

23. A. Cooper *et al.*, *Science*. **349**, 602–606 (2015).

24. R. M. Schweizer *et al.*, *Mol. Biol. Evol.* **35**, 1190–1209 (2018).

25. T. M. Anderson *et al.*, *Science*. **323**, 1339–1343 (2009).

26. L. D. Martin, B. Miles Gilbert, (1978).

27. M. J. Kohn, M. P. McKay, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **326-328**, 42–53 (2012).

28. W. E. Scott, K. L. Pierce, J. P. Bradbury, R. M. Forester, *Cenozoic Geology of Idaho: Idaho Bureau of Mines and Geology Bulletin*. **26**, 581–595 (1982).

29. K. B. Tankersley, *Geoarchaeology*. **12**, 713–724 (1997).

30. K. M. Ford, A. R. Bair, J. A. Holman, *Mich. Acad.* **28**, 135–145 (1996).

31. H. G. McDonald, *The First Discovery of America: Archaeological Evidence of the Early Inhabitants of the Ohio Area, The Ohio Archaeological Council, Columbus*, 23–39 (1994).

246 32. J. A. Holman, *Mich. Acad.* **29**, 1–20 (1997).

247 33. T. M. Bills, H. G. McDonald, *Current Research in the Pleistocene*. **15**, 101–103 (1998).

248 34. A. S. Nye, thesis, East Tennessee State University (2007).

249 35. C. Stock, J. M. Harris, *Natural History Museum of Los Angeles County, Los Angeles, CA*
250 (1992).

251 36. J. Rohlf, *Book PSDIG, version*. **1** (2001).

252 37. R. Core Team, Others, *R Foundation for statistical computing, Vienna* (2013).

253 38. I. L. Dryden, K. V. Mardia, *Statistical Shape Analysis* (Wiley, 1998).

254 39. C. Goodall, *J. R. Stat. Soc. Series B Stat. Methodol.* **53**, 285–321 (1991).

255 40. BOOKSTEIN, F. L, *Image Fusion and Shape Variability Techniques* (1996).

256 41. S. Schlager, in *Statistical shape and deformation analysis* (Elsevier, 2017), pp. 217–256.

257 42. I. L. Dryden, M. I. Dryden, *Vienna, Austria: R Foundation for Statistical Computing* (2012).

258 43. M. Foote, *Paleobiology*. **19**, 403–419 (1993).

259 44. M. J. Anderson, *Biometrics*. **62**, 245–253 (2006).

260 45. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful
261 Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
262 *(Methodological)*. **57** (1995), pp. 289–300.

263 46. D. C. Adams, M. L. Collyer, A. Kaliontzopoulou, E. Sherratt, Geomorph: software for
264 geometric morphometric analyses. R package version 3.0. 6 (2016).

265 47. A. Evin *et al.*, *J. Archaeol. Sci.* **40**, 735–743 (2013).

266 48. M. Baylac, M. Frieß, in *Modern Morphometrics in Physical Anthropology*, D. E. Slice, Ed.
267 (Springer US, Boston, MA, 2005), pp. 145–165.

268 49. C. Meloro, P. Raia, *Evol. Biol.* **37**, 177–186 (2010).

269 50. P. Piras *et al.*, *Zoologischer Anzeiger - A Journal of Comparative Zoology*. **251**, 335–343
270 (2012).

271 51. J. Felsenstein, *Am. J. Hum. Genet.* **25**, 471–492 (1973).

272 52. M. A. Butler, A. A. King, *Am. Nat.* **164**, 683–695 (2004).

273 53. L. J. Harmon *et al.*, *Evolution* (2010).

274 54. M. Pagel, *Nature*. **401**, 877–884 (1999).

55. L. J. Harmon, J. T. Weir, C. D. Brock, R. E. Glor, W. Challenger, *Bioinformatics*. **24**, 129–131 (2008).

56. S. P. Blomberg, T. Garland Jr, A. R. Ives, *Evolution*. **57**, 717–745 (2003).

57. L. J. Revell, *Methods Ecol. Evol.* (2012) (available at https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2011.00169.x).

58. D. C. Adams, *Syst. Biol.* **63**, 685–697 (2014).

59. C. P. Klingenberg, N. A. Gidaszewski, *Syst. Biol.* **59**, 245–261 (2010).

60. E. K. Rueness *et al.*, *PLoS One*. **6**, e16385 (2011).

61. B. T. Fuller, J. M. Harris, A. B. Farrell, G. Takeuchi, J. Southon, *La Brea and beyond: The paleontology of asphalt-preserved biotas, ed. JM Harris. Natural History Museum of Los Angeles County, Science Series*, 151–167 (2015).

62. F. Welker, M. Soressi, W. Rendu, J.-J. Hublin, M. Collins, *J. Archaeol. Sci.* **54**, 279–286 (2015).

63. F. Welker *et al.*, *Nature*. **522**, 81–84 (2015).

64. B. Demarchi *et al.*, *Elife*. **5** (2016), doi:10.7554/eLife.17092.

65. N. L. van Doorn, H. Hollund, M. J. Collins, *Archaeol. Anthropol. Sci.* **3**, 281–289 (2011).

66. M. C. Chambers *et al.*, *Nat. Biotechnol.* **30**, 918–920 (2012).

67. B. Ma *et al.*, *Rapid Commun. Mass Spectrom.* **17**, 2337–2342 (2003).

68. J. Zhang *et al.*, *Mol. Cell. Proteomics*. **11**, M111.010587 (2012).

69. K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **30**, 772–780 (2013).

70. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics*. **19**, 1572–1574 (2003).

71. J. Dabney *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).

72. P. B. Damgaard *et al.*, *Sci. Rep.* **5**, 11184 (2015).

73. M. Meyer, M. Kircher, *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

74. C. Gamba *et al.*, *Nat. Commun.* **5**, 5257 (2014).

75. P. Brotherton *et al.*, *Nat. Commun.* **4**, 1764 (2013).

76. M.-T. Gansauge *et al.*, *Nucleic Acids Res.* **45**, e79 (2017).

77. C. Gamba *et al.*, *Mol. Ecol. Resour.* **16**, 459–469 (2016).

78. M.-T. Gansauge, M. Meyer, *Nat. Protoc.* **8**, 737–748 (2013).

79. E. Ersmark, L. Orlando, E. Sandoval-Castellanos, *Open Quaternary* (2015).

80. M. E. Allentoft *et al.*, *Nature*. **522**, 167–172 (2015).

81. C. Carøe *et al.*, *Methods Ecol. Evol.* (2017).

82. V. Slon *et al.*, *Quat. Int.* **398**, 210–218 (2016).

83. M. A. Nieves-Colón *et al.*, *Am. J. Phys. Anthropol.* (2018), doi:10.1002/ajpa.23472.

84. J. Sambrook, D. W. Russell, *CSH Protoc.* **2006** (2006), doi:10.1101/pdb.prot4455.

85. S. Verscheure, T. Backeljau, S. Desmyter, *Forensic Sci. Int. Genet.* **12**, 60–68 (2014).

86. T. Maricic, M. Whitten, S. Pääbo, *PLoS One*. **5**, e14004 (2010).

87. T. Maricic, M. Whitten, S. Pääbo, *PLoS One*. **5**, e14004 (2010).

88. P. Gaubert *et al.*, *PLoS One*. **7**, e42740 (2012).

89. S. S. T. Mak *et al.*, *Gigascience*. **7** (2018), doi:10.1093/gigascience/giy151.

90. H. Li, R. Durbin, *Bioinformatics*. **25**, 1754–1760 (2009).

91. A. McKenna *et al.*, *Genome Res.* **20**, 1297–1303 (2010).

92. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, *BMC Bioinformatics*. **15**, 356 (2014).

93. R. E. Green *et al.*, *Science*. **328**, 710–722 (2010).

94. P. Skoglund *et al.*, *Nature*. **538**, 510–513 (2016).

95. W. Haak *et al.*, *Nature*. **522**, 207–211 (2015).

96. S. Lindgreen, *BMC Res. Notes*. **5**, 337 (2012).

97. M. Schubert *et al.*, *BMC Genomics*. **13**, 178 (2012).

98. M. Kircher, *Methods Mol. Biol.* **840**, 197–228 (2012).

99. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, *Bioinformatics*. **29**, 1682–1684 (2013).

100. M. L. Borowiec, *PeerJ*. **4**, e1660 (2016).

101. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, *Mol. Biol. Evol.* **34**, 772–773 (2017).

102. M. A. Miller, W. Pfeiffer, T. Schwartz, in *2010 Gateway Computing Environments*

Workshop (GCE) (2010), pp. 1–8.

103. S. Kvist, M. E. Siddall, *Cladistics*. **29**, 435–448 (2013).

104. Z. Xi, L. Liu, C. C. Davis, *Mol. Biol. Evol.* **33**, 838–860 (2016).

105. K.-P. Koepfli *et al.*, *Curr. Biol.* **25**, 2158–2165 (2015).

106. B. M. vonHoldt *et al.*, *Sci Adv*. **2**, e1501714 (2016).

107. M.-H. S. Sinding *et al.*, *PLoS Genet.* **14**, e1007745 (2018).

108. A. R. Quinlan, I. M. Hall, *Bioinformatics*. **26**, 841–842 (2010).

109. A. Stamatakis, *Bioinformatics*. **22**, 2688–2690 (2006).

110. D. Bryant, R. Bouckaert, J. Felsenstein, N. A. Rosenberg, A. RoyChoudhury, *Mol. Biol. Evol.* **29**, 1917–1932 (2012).

111. Z. Yang, *Current Zoology*. **61** (2015), pp. 854–865.

112. S. Mirarab, T. Warnow, *Bioinformatics*. **31**, i44–52 (2015).

113. E. Sayyari, J. B. Whitfield, S. Mirarab, *Mol. Phylogenet. Evol.* **122**, 110–115 (2018).

114. H. Li *et al.*, *Data Processing*, 1–2 (2009).

115. O. Tange, *GNU Parallel 2018* (Lulu.com, 2018).

116. A. Stamatakis, *Bioinformatics*. **30**, 1312–1313 (2014).

117. Z. Yang, *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

118. Z. Yang, B. Rannala, *Mol. Biol. Evol.* **23**, 212–226 (2006).

119. J. Alroy, *Paleobiology*. **26**, 707–733 (2000).

120. A. Rambaut, N. C. Grassly, *Comput. Appl. Biosci.* **13**, 235–238 (1997).

121. W.-C. Chen, thesis, Iowa State University (2011).

122. J. L. Thorne, H. Kishino, I. S. Painter, *Mol. Biol. Evol.* **15**, 1647–1657 (1998).

123. B. Rannala, Z. Yang, *Syst. Biol.* **56**, 453–466 (2007).

124. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, *Mol. Biol. Evol.* **27**, 1877–1885 (2010).

125. M. D. Reis *et al.*, *Syst. Biol.* **67**, 594–615 (2018).

126. M. Dos Reis, T. Zhu, Z. Yang, *Syst. Biol.* **63**, 555–565 (2014).

127. N. Patterson *et al.*, *Genetics*. **192**, 1065–1093 (2012).

128. V. M. Bravo-Cuevas, J. Priego-Vargas, M. Á. Cabral-Perdomo, M. A. Pineda Maldonado, *Fossil Record*. **19**, 131–141 (2016).

129. N. J. Czaplewski, K. J. Rogers, C. Russell, *J. Cave Karst Stud.* **80** (2018)).

130. F. J. Prevosti, A. D. Rincón, *J. Paleontol.* **81**, 1053–1065 (2007).

131. S. R. Holen *et al.*, *Nature*. **544**, 479–483 (2017).

132. E. Scott, K. B. Springer, *PeerJ*. **4**, e2151 (2016).

133. M. R. Smith, P. D. Polly, *J. Vert. Paleontol.* **33**, 410–420 (2013).

134. B. W. Schubert, S. C. Wallace, *Boreas*. **38**, 482–492 (2009).

Acknowledgments: We thank the staff at the Kansas Museum of Natural History, the McClung Museum, the Cincinnati Museum Center, the Idaho Museum of Natural History, the United States Bureau of Reclamation, the La Brea Tar Pits and Museum, University of Texas at El Paso, University of California Museum of Paleontology, the University of Washington Burke Museum, the Carnegie Museum of Natural History, the British Natural History Museum, the Harrison Zoological Museum, National Museums Scotland, Museum National d'Histoire Naturelle, Ludwig Maximillian University, Naturalis Biodiversity Center, Danish Zoological Museum, Swedish Naturhistoriska Riksmuseet, Institute of Zoology (Chinese Academy of Sciences), Smithsonian National Museum of Natural History, Naturhistorisches Museum Bern, Harvard Museum of Comparative Zoology, Instituto de Conservacao da Natureza e das Florestas, Zoological Institute (Russian Academy of Sciences; state assignment № AAAA-A17-117022810195-3), Institute of Archaeology (Russian Academy of Sciences), Institute of Plant and Animal Ecology (Russian Academy of Sciences), Institute of Systematics and Animal Ecology (Russian Academy of Sciences) for access to specimens in their care. We also thank T. Barnosky, S. Bray, A. Farrell, R. Fischer, A. Harris, J. Harris, A. Henrici, P. Holroyd, R. MacPhee, A. Philpot, J. Saunders, J. Southon, G. Storrs, G. Takeuchi, X. Wang and C. Widga for their assistance. A.M. used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group. Further, the authors would like to acknowledge the assistance of the Danish National High-Throughput Sequencing Centre, BGI-Europe, the Garvan Institute of Medical Research and the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility for assistance in Illumina and BGIseq500 data generation. **Funding:** A.R.P. was supported by a Marie Curie COFUND Junior Research Fellowship (Durham University). A.M. was supported by an NSF grant (award number: 1457106) and the QCB Collaboratory Postdoctoral Fellowship (UCLA). DireGWC was sequenced using the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. L.A.F.F.,

J.H., A.H-B. and G.L. were supported either by a European Research Council grant (ERC-2013-StG-337574-UNDEAD) or Natural Environmental Research Council grants (NE/K005243/1 and NE/K003259/1), or both. K.S. was supported by a grant from Barrett, the Honors College at Arizona State University. A.T.O. was supported by by the Strategic Initiative Funds, Office of the President, Arizona State University to the Institute of Human Origins' DNA and Human Origins at Arizona State University project. L.A.F.F. was supported by a Junior Research Fellowship (Wolfson College, University of Oxford) and L.A.F.F. and A.Carmagnini were supported by the Wellcome Trust (210119/Z/18/Z). S.G. was supported by Carlsbergfondet grant CF14–0995 and Marie Skłodowska-Curie Actions grant 655732-WhereWolf. M.T.P.G was supported by ERC Consolidator grant 681396-Extinction Genomics. B.S. and J.K. were supported by IMLS MG-30-17-0045-17 and NSF DEB-1754451. A.H-B. was supported by the Leverhulme Trust. A.Cooper, K.J.M., and H.H. were supported by the Australian Research Council. A.S. and G.G. were supported by Australian Government Research Training Program Scholarships. **Author contributions:** A.R.P., K.J.M., A.M., R.K.W., G.L., L.A.F.F., and A. Cooper conceived of the project and designed the research; A.R.P. and K.J.M. coordinated the sample collection efforts with input from R.K.W., G.L., L.A.F.F., K.M.H., and A. Cooper; A.R.P., K.J.M., A.H.-B., J.M., C.A., J.C.C., A.E., J.K., A.L., A.O., S.P., B.W.S., M.T., M.J.C., M.-M.H.S., M.T.P.G., A.C.S., B.S., B.V.V., R.W.K., and A. Cooper provided and/or collected samples; A.R.P., K.J.M., K.W., A.M., C.S., J.H., A.J., A.T.S., P.B., and H.H. conducted the genomic laboratory work; K.J.M., A.M., G.G., G.L., L.A.F.F. and A. Cooper conducted the analyses of the genomic data; S. A.-C., A.H.-B., J.M., C.A., and A.E. conducted the morphological analyses; A.R.P., K.J.M., A.M., S.A.-C., B.V.V., R.K.W., G.L., L.A.F.F. and A. Cooper wrote the paper with input from all other authors. **Competing interests**: The authors declare no competing interests. **Data and materials availability:** The reads generated for this study have been deposited at the European Nucleotide Archive (ENA) with project number PRJEB31639. Geomorphometric data and collagen sequence were deposited onto a Dryad archive.

Supplementary Materials:

Materials and Methods

Figs. S1 to S20

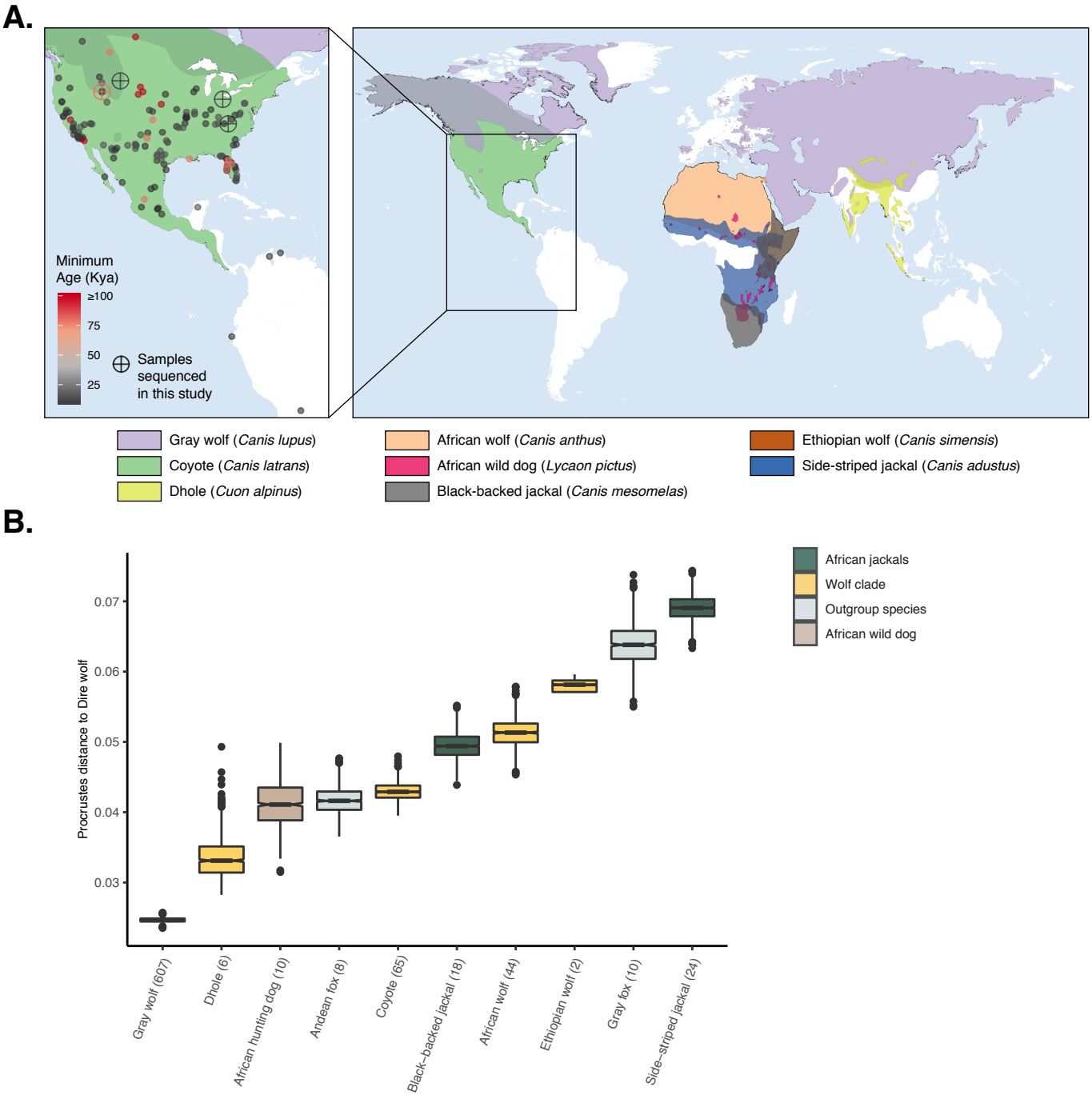
Tables S3; S15 to S22

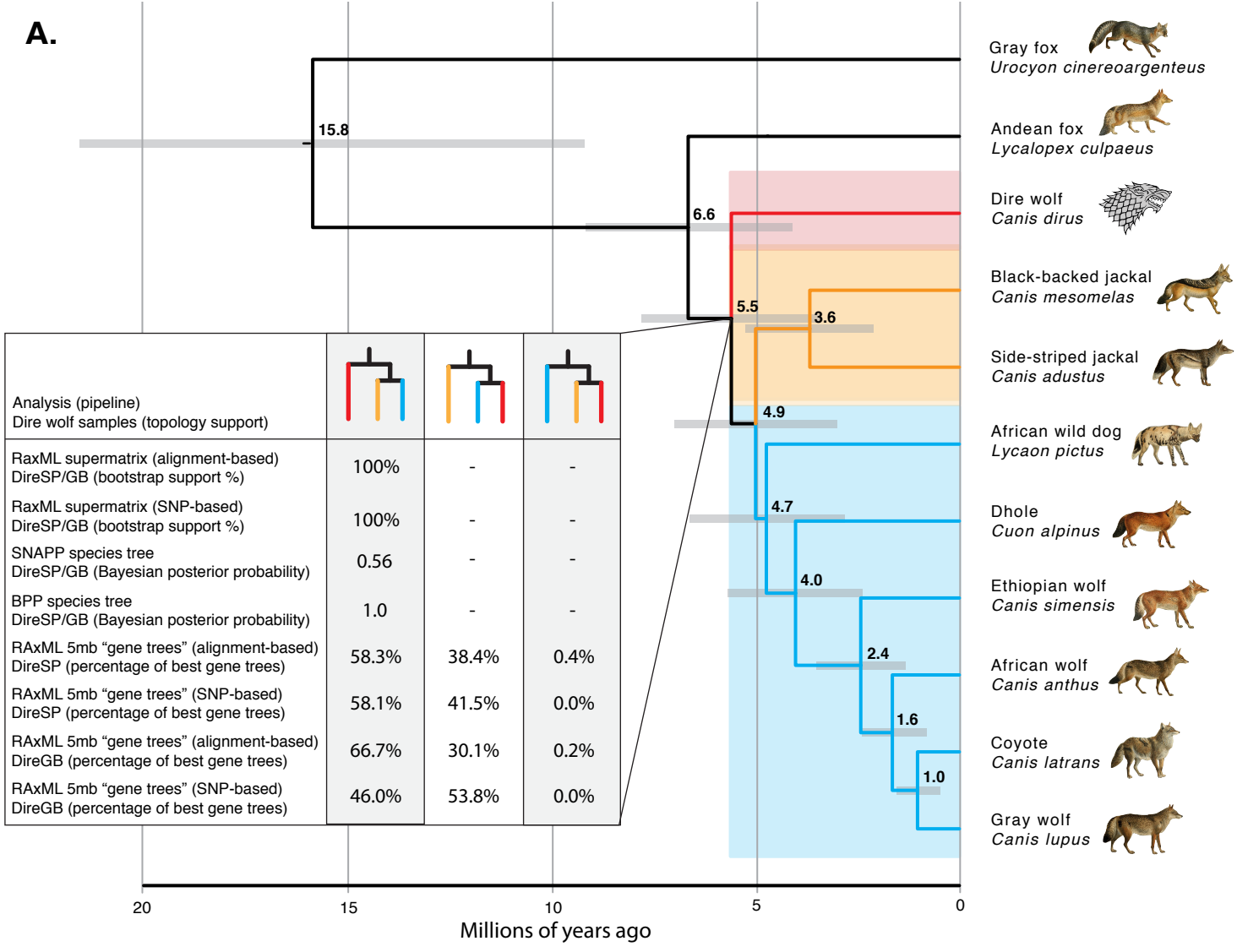
References 26 to 134

Fig. 1. Map of dire wolf remains and morphological differentiation with wolf-like canids A. Right: map representing the geographic range (obtained from IUCN, https://www.iucnredlist.org) of the canid species investigated in this study. Left: map representing the distribution of sites in the Americas where dire wolf remains (*Canis dirus*) were identified (*21*) (Table S3[1]). Colored circles represent the location and approximate age of the remains, with crossed circles representing the five samples from Idaho, Ohio, Tennessee, and Wyoming that yielded sufficient endogenous DNA to reconstruct both mitochondrial genomes and low-coverage nuclear genome sequences. **B.** Procrustes distance between the combined mandible and M1 shape of dire wolf and other extant canid species. Pairwise procrustes distances were calculated by superimposing landmarks from molar and mandibular shapes

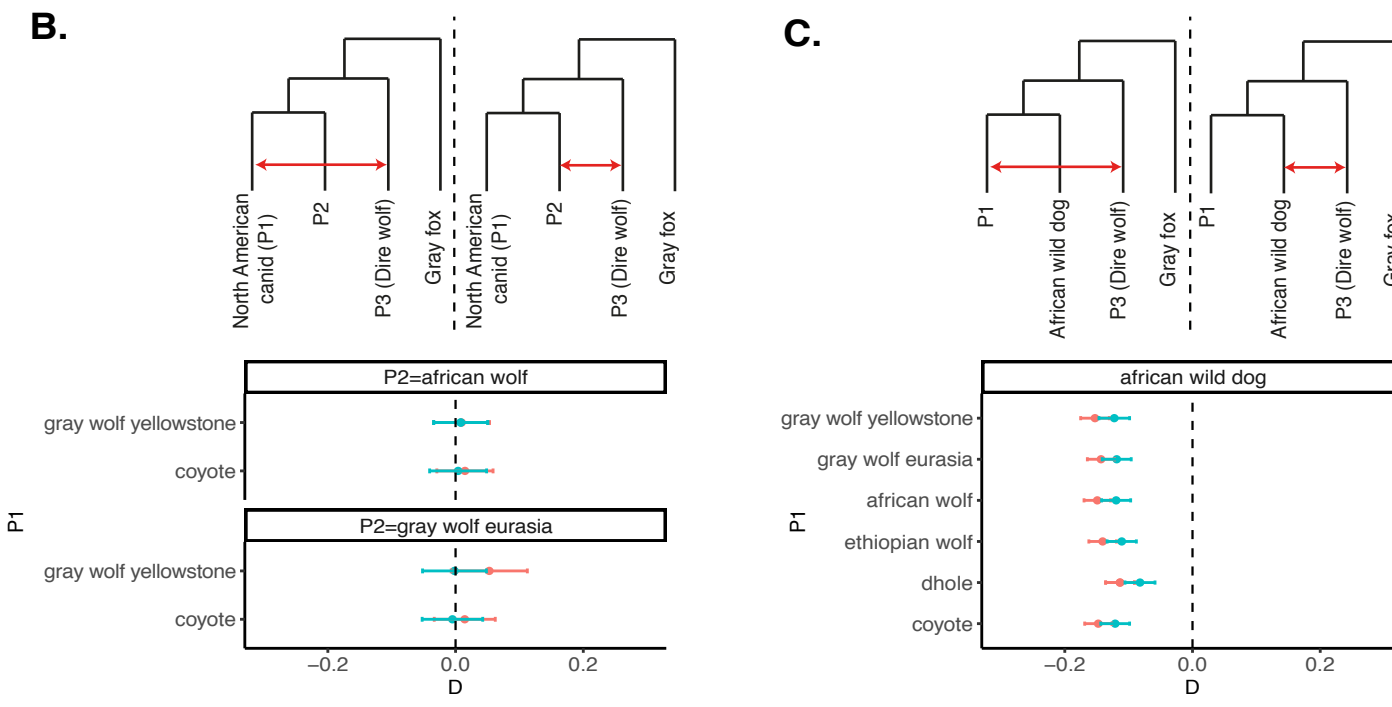
between pairs of specimens and by computing the square root of the squared differences between the coordinates of corresponding landmarks (*21*).

Fig. 2. Relationships among living and extinct wolf-like canids A. Time-scaled nuclear phylogeny generated in MCMCtree based on the best species tree topology obtained from BPP and SNAPP. Values associated with nodes are mean age estimates (millions of years before present) while bars represent 95% Highest Posterior Densities. The inset table shows levels of support for the three possible arrangements of the dire wolf (red), the African jackals (orange), and the remaining wolf-like canids (blue) that we obtained under different analytical frameworks (*11*) when including either one or both of our two highest coverage dire wolf samples (DireSP and DireGB). While only one dire wolf branch is depicted in the tree, when multiple dire wolf individuals were included they formed a monophyletic clade (Figure S12-13, and S15). **B.** Results of D statistics used to assess the possibility of gene flow between the dire wolf and extant North American canids (Yellowstone wolves and coyotes). Each dot represents the mean D calculated along the genome and the error bar represents 3 standard deviations. These plots show that the dire wolf genomes do not share significantly more derived alleles with extant North American canids compared to African wolves or Eurasian wolves (values of D not significantly different to zero), suggesting that no hybridization occurred between the dire wolf and the ancestor of extant North American canids. **C.** Results of D statistics showing the existence of an ancient gene flow event between the ancestor of the dhole, Ethiopian wolf, African wolf, gray wolf and coyotes and the lineage of the dire wolf (consistently non-zero values of D regardless of P1).


A.
Minimum
Age (Kya)
≥100
75
50
25
Samples
sequenced
in this study
Gray wolf (Canis lupus)
Coyote (Canis latrans)
Dhole (Cuon alpinus)
African wolf (Canis anthus)
African wild dog (Lycaon pictus)
Black-backed jackal (Canis mesomelas)
Ethiopian wolf (Canis simensis)
Side-striped jackal (Canis adustus)
B.
African jackals
Wolf clade
Outgroup species
African wild dog
Procrustes distance to Dire wolf
0.07
0.06
0.05
0.04
0.03
Gray wolf (607)
Dhole (6)
African hunting dog (10)
Andean fox (8)
Coyote (65)
Black−backed jackal (18)
African wolf (44)
Ethiopian wolf (2)
Gray fox (10)
Side−striped jackal (24)



Analysis (pipeline) Dire wolf samples (topology support)			
RaxML supermatrix (alignment-based) DireSP/GB (bootstrap support %)	100%	-	-
RaxML supermatrix (SNP-based) DireSP/GB (bootstrap support %)	100%	-	-
SNAPP species tree DireSP/GB (Bayesian posterior probability)	0.56	-	-
BPP species tree DireSP/GB (Bayesian posterior probability)	1.0	-	-
RAxML 5mb "gene trees" (alignment-based) DireSP (percentage of best gene trees)	58.3%	38.4%	0.4%
RAxML 5mb "gene trees" (SNP-based) DireSP (percentage of best gene trees)	58.1%	41.5%	0.0%
RAxML 5mb "gene trees" (alignment-based) DireGB (percentage of best gene trees)	66.7%	30.1%	0.2%
RAxML 5mb "gene trees" (SNP-based) DireGB (percentage of best gene trees)	46.0%	53.8%	0.0%



Science
AAAS

Supplementary Materials for

Dire wolves were the last of an ancient New World canid lineage

Angela R. Perri[1,*§], Kieren J. Mitchell[2,*§], Alice Mouton[3,*], Sandra Álvarez-Carretero[4,*], Ardern Hulme-Beaman[5,6], James Haile [7], Alexandra Jamieson[7], Julie Meachen[8], Audrey T. Lin[7,9], Blaine W. Schubert[10], Carly Ameen[11], Ekaterina E. Antipina[12], Pere Bover[13], Selina Brace[14], Alberto Carmagnini[4], Christian Carøe[15], Jose A. Samaniego Castruita[15], James C. Chatters[16], Keith Dobney[5], Mario dos Reis[4], Allowen Evin[17], Philippe Gaubert[18], Shyam Gopalakrishnan[15], Graham Gower[2], Holly Heiniger[2], Kristofer M. Helgen[19], Josh Kapp[20], Pavel A. Kosintsev[21], Anna Linderholm[7, 22], Andrew T. Ozga[23, 24], Samantha Presslee[25], Alexander T. Salis[2], Nedda F. Saremi[20], Colin Shew[3], Katherine Skerry[24], Dmitry E. Taranenko[26], Mary Thompson[27], Mikhail V. Sablin[28],Yaroslav V. Kuzmin[29, 30], Matthew J. Collins[31, 32], Mikkel-Holger S. Sinding[15, 33], M. Thomas P. Gilbert[15, 34], Anne C. Stone[23 ,24], Beth Shapiro[20, 35], Blaire Van Valkenburgh[3], Robert K. Wayne[3], Greger Larson[7], Laurent Frantz[4, 7#§], and Alan Cooper[2#§]

Correspondence to: angela.r.perri@durham.ac.uk (A.R.P.); kieren.mitchell@adelaide.edu.au (K.J.M.); laurent.frantz@qmul.ac.uk (L.A.F.F.); alanjcooper42@gmail.com (A.C.)

This PDF file includes:

Materials and Methods
Figs. S1 to S20
Tables S3; S15 to S22
References 25 to 134

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S2 and S4 to S14 (Excel)

Materials and Methods

Site descriptions

This section details the site descriptions for the five dire wolf samples from which we obtained mitochondrial and low-coverage nuclear genomes (Table S1 and Table S2). *Site name, state (sample repository; **repository code; lab code**; **publication code**; sample provenience)*

*Natural Trap Cave, Wyoming (University of Kansas; **KU48130; ACAD5529**; **DireNTC**; fragment from mandible)*

Natural Trap Cave (NTC), a late Pleistocene fossil site, is a 25-meter-deep karst sinkhole on the western side of the Big Horn Mountains in north-central Wyoming. It includes a rich vertebrate biota, of mostly open-habitat associated species, including both Pleistocene-only taxa, and taxa that survived the extinction event, such as pronghorn antelope (*Antilocapra americana*) and bighorn sheep (*Ovis canadensis*) (*26*). The earliest known radiocarbon date of wolves at Natural Trap Cave is approximately 25,800 years before present (*27*). The wolves at Natural Trap Cave were identified as both gray wolves (*Canis lupus*) and dire wolves by (*26*). But upon further morphological investigation, most of the wolves at this site have recently been identified as the extinct *Canis lupus* morph, the Beringian wolf (*2*). The specimen that yielded dire wolf DNA (KU48130) was the anterior portion of a right mandible (with several premolars and molars still *in situ*) and has a date of 21773 calBP (19970 ± 110 BP; OxA-37752).

*Gigantobison Bay, Idaho (Idaho Museum of Natural History; **IMNH 48001/52; ACAD18742**; **DireGB**; petrous from complete cranium)*

Gigantobison Bay is an open Pleistocene fossil bearing site located within the American Falls Reservoir, Idaho. Seasonal changes in water level in the reservoir periodically expose and inundate the site, occasionally revealing fossils in the sediment (most commonly extinct long-horned bison - *Bison latifrons* - for which the site is named). While remains from the site have proven intractable to radiocarbon dating, a minimum age can be assigned to all specimens based on regional changes in geology. Fossils from Gigantobison Bay are found in old fluvial (river) sediments that were overlain by lacustrine (lake) sediments deposited following the damming of the Snake River by the volcanic Cedar Butte Basalt, which is dated to 72 ± 14 ky (*28*). This sample is from a complete cranium with most teeth still *in situ*. We removed the left petrosal from the specimen, which successfully yielded aDNA.

*Sheriden Pit, Ohio (Cincinnati Museum Center; **VP1737; ACAD1735; DireSP**; incisor root)*

The Sheriden Pit is part of a local cave system accessed via sinkhole at Indian Trail Caverns in Wyandot County, northwestern Ohio. As opposed to the glacial till derived from the late Wisconsinan ice sheet that covers and fills most regional karst features, the sink contained alluvium. Glacial retreat from the region is estimated at 14,100 to 12,500 years B.P. (*29*) thus the sink, formed from subsurface dissolution and roof collapse, was open to the surface sometime during and/or after this period. The site was excavated by Cincinnati Museum Center (CMC) between 1990 and 1996 and has produced a diverse faunal assemblage of late Pleistocene (Rancholabrean) taxa, including extinct and extralimital species. Over 75 species have been identified to date from the site. Extinct taxa include *Arctodus*, *Castoroides*, *Cervalces*, and *Platygonus*. The assemblage is indicative of a mosaic habitat comprising open woodland and grassy ecosystems with one or more shallow, marshy ponds and streams in the area (*30–32*).

The sinkhole is developed in a stromatolitic reef of Middle Silurian Lockport Group dolomite. Approximately 10m deep, it formed a natural trap into which animals fell, were washed, or were carried by predators. Sediments indicate the past presence of running water into the sink and out through a horizontal cave connection at its bottom. When this connection became choked with sediment, a layer of laminated silts and clays indicative of standing water formed in the upper part of the pit. Four stratigraphic units can be identified, from top to bottom: a laminated diamicton, a layer of gray- and red- banded clay rhythmite, a lower diamicton, and a base unit of dolostone cobble colluvium. The rhythmite is the primary bone-bearing unit, although smaller amounts of fossils were recovered from the lower diamicton (*33*).

Although the *Canis dirus* specimen has no precise stratigraphic data associated, it almost certainly derives from the primary bone-bearing middle layer of rhythmite. Other bone material from this layer has produced radiocarbon dates of 11,060 ± 60 years BP (CAMS 10349) and 11,710 ± 220 years BP (PITT-0892).

Guy Wilson Cave, Tennessee (McClung Museum, University of Tennessee; ***MMNHC 0013****;* ***RW001****;* ***DireGWC****; 4th lower premolar root. Picture here https://viewshape.com/shapes/dfomd4trmbn)*

Guy Wilson Cave is a terminal Pleistocene fossil-bearing cave site located in Sullivan County, northeastern Tennessee. At least nine extinct large mammals have been recovered from the site: dire wolf (*Canis dirus*), caribou (*Rangifer tarandus*), flat-headed peccary (*Platygonus compressus*), long-nosed peccary (*Mylohyus* sp.), tapir (*Tapirus* sp.), horse (*Equus* sp.), ground sloth (*Megalonyx* sp.), mammoth (*Mammuthus* sp.), and mastodon (*Mammut americanum*). A sample from the cave was taken by Charles Coney (1970) and the material was donated to the McClung Museum, University of Tennessee, Knoxville (Schubert and Wallace 2007). Many bones are carnivore-gnawed, and it has been suggested that dire wolves used the site as a den (*34*).

A root of a lower right 4th premolar from a dire wolf (MMNHC 0013) was submitted to DirectAMS for radiocarbon dating (D-AMS 26659) and underwent standard gelatin extraction procedures. A portion of the dentine was digested in 0.5 molar HCl for three days at 4 °C with daily changes of acid, producing a strong, collagen pseudomorph. After multiple rinses in deionized H2O, the collagen was soaked in 5 g/L KOH at 4°C to remove organic contaminants and again underwent multiple rinses. It was then completely dissolved to gelatin in 0.05 molar HcL at 90°C, after which it was filtered through a 0.5 micron filter and lyophilized. The dried sample was then combusted in a vacuum and the resulting CO2 isolated and graphitized.This process was conducted twice on this sample in batches four months apart. Graphite targets were measured by an NEC Peletron accelerator mass spectrometer.

Resulting ages on two separate preparations were 10933 ± 44 and 10955 ± 25 years BP, which average to 10944 ± 22 years BP (12820-12720 cal BP) representing the latest reliable date for dire wolves prior to extinction.Two subsamples of the second extraction were submitted to the Washington State University Stable Isotope Laboratory for elemental and stable 13C and 15N analysis (sample G-168765). The extract had average C:N ratio of 3.0955, which is well

within the range expected for well-preserved collagen.The ∂13C of -20.08 and ∂15N of 10.28 are indicative of a terrestrial carnivore in a C3 food chain.

American Falls Reservoir, Idaho (Idaho Museum of Natural History; ***IMNH 255/8007; AJ66; DireAFR;*** *petrous from partial cranium)*

Remains from the American Falls Reservoir locality are located nearby to the Gigantobison Bay locality from the same site, having the same geology. Despite multiple attempts, the dire wolf material from American Falls Reservoir has also proven intractable to radiocarbon dating. These remains were found in the same fluvial (river) sediments overlain by lacustrine (lake) sediments deposited following the damming of the Snake River by the volcanic Cedar Butte Basalt, which is dated to 72 ± 14 ky (*28*). This sample is from a partial juvenile cranium with some teeth still *in situ* and was collected by Howard Emry in May 1988. We removed the right petrosal from the specimen, which successfully yielded aDNA.

Rancho La Brea Tar Seeps, California (La Brea Tar Pits and Museum; **LACMP23-1619***;* ***DireRLB****; left tibia fragment)*

The Rancho La Brea Tar Seeps (RLB), located in the Los Angeles Basin, California, are the result of asphalt originating from oil sands pushing to the surface and forming seeps that can reach several square meters in area and 9-11 m in depth (*35*). The resulting pooled asphalt has led to the entrapment of local fauna over tens of thousands of years. Over two million skeletal elements have been recovered and housed at the George C. Page Museum of La Brea Discoveries in Los Angeles. Of these, over 4,000 dire wolf individuals have been excavated - the most common species within the RLB assemblage (*35*). The specimen that yielded type I collagen (COL1) for this analysis comes from the La Brea Tar Pits and Museum. It was sourced from RLB locality P23-1, Grid B-1, Level 3.

Geometric Morphometrics

Geometric morphometric data

Geometric morphometrics was carried out on the mandibular first molar morphology of 735 specimens and the mandible morphology of 810 specimens representing 11 species (see Table S3). Mandibles were photographed by C.A., A.H-B., A.E. and A.P. using a Nikon reflex camera with 60mm fixed lens. Positioning of the mandible standardized with a spirit level. First mandibular molars were photographed by A.H-B. and A.E.. Mandibles were recorded using 15 fixed landmarks (Figure S1A), while first molars were recorded with 3 fixed landmarks and two curves of sliding semilandmarks; one to capture the anterior outline with 29 equidistant points and the other capturing the posterior of the outline using 19 equidistant points (Figure S1B). TpsDig 2.3 software was used for digitising landmarks (36) and data was imported into the R environment (37) for further processing and analyses.

Geometric morphometric analyses

Shape data was standardized using Generalised Procrustes analysis (38) which scales, translates and rotates the data to minimize the least squared distances among the configuration of landmarks for each specimen (39). Full Procrustes distance is regarded as the simplest and truest representation of similarities/relationships between shapes. This is because for highly dimensional data with low levels of covariance a plot of 2 PCs rarely presents an accurate picture (38). At this stage, sliding semilandmarks of the molar shape dataset were slid using the

minimum bending energy method (40). The superimposition and sliding procedures were carried out using the Morpho package (41). Size, defined as the square root of the sum of squared distances from each landmark to the configurations centroid (centroid size; CS), was extracted from the configuration of landmarks for each specimen. CS was then plotted as boxplots and tested using a pairwise t-test. Mean shapes were calculated for each species (hereafter group) and full Procrustes distances were calculated among group means. Full Procrustes distances were also calculated among specimens to their respective group means and among all specimens in the dataset. Full Procrustes distances were calculated with the Shapes package (42).

Morphological variance was calculated as the dispersions around the group mean using the full Procrustes distance after Foote (43). Homogeneity of dispersions between groups was tested pairwise for significant differences following Anderson (44). All multiple comparison p-values were adjusted using the false discovery rate (FDR) method (45). Differences among group shapes were tested using a Procrustes ANOVA (39). Allometry, defined as shape variation that co-varies with size, was assessed using a Procrustes ANCOVA. Homogeneity of allometric slopes among groups was also assessed and if found to be sufficiently parallel allometry was removed from the dataset by regression. These tests and procedures were carried out using the Geomorph package (46).

The following analyses were carried out on both shape and allometry removed shape. A principal component analysis was carried out on the aligned coordinates of the specimens. Linear discriminant analyses (LDA) paired with leave-one-out correct cross validation (CCV) procedures were applied to subsets of principal components (PCs) (47). The sub setting of PCs was carried out using a stepwise approach after (48) combined with resampling to equal sample size following (47). These LDAs were applied both on a multiclass basis across the entire dataset, and also on a pairwise basis among groups. In each case of applying the LDA the number of PCs used was determined using the aforementioned methods, with numbers of PCs selected to optimise identification. The cross validation percentage for each LDA was used to assess the accuracy of identification among groups.

To further assess which groups were closest to the dire wolf, the distance from each group mean to all other means was assessed using bootstrapping and plotted as one standard deviation of bootstrapped values around the mean. This was achieved by resampling the specimens of each group by bootstrap procedures, then recalculating the group mean shape and finally calculating the full Procrustes distance among the mean shapes. This procedure was carried out for 999 iterations. Full Procrustes distances of all specimens to the mean of each group were visualized with a violin and boxplot (Figure 1B) to assess the differences in groups sampling and variance and its effect on the results.

Morphological-molecular comparisons and tests

Size of both datasets (in the form of CS) was tested for phylogenetic signal. Unlike the multivariate shape data, for which there are few tests available, we could test the continuous univariate CS data against multiple evolutionary models following procedures similar to Meloro and Raia (49) and Piras et al. (50). We compared the CS data to the following models: Brownian (51), Ornstein-Uhlenbeck (52), accelerating-decelerating (53), Pagel's lambda (54), and a white noise model (following (50)). This was carried out using the Geiger package (55). We chose the

model with the lowest Akaike Information Criterion (AIC) as the best fit. We also ran a test of significance for phylogenetic signal in size data using the K-statistic (56) with functions from the package phytools (57).

To assess the mandibular morphology for phylogenetic signal we used a multivariate generalized K-statistic (Kmult, (58)). This method compared the phylogenetic distances constructed from both mtDNA and nuclear DNA with the full Procrustes distance tree calculated among species' mean shapes. This test is a multivariate approach to the method developed by Blomberg et al. (56), which assesses the data for a phylogenetic signal according to what is expected under a Brownian motion model of evolution. The phylogeny was then mapped to the PCA of mean shapes to visualise the correspondence among shape variables and phylogenetic distances (59). As the correspondence was particularly poor among some species, and rapid divergence and/or convergence is known from canids (e.g. African wolves and jackals,(60)), and the phylomapping approach to identifying which species did not conform was visually difficult to decipher, it was therefore useful to investigate this further by examining each pairwise comparison.

To identify the species that deviated the most from how the morphological and molecular distances corresponded, the ratio of each respective pairwise distance was calculated. This metric presented a method for assessing molecular-morphological incongruence and was applied just to Canina species with outgroups (Andean fox and gray fox) removed. Raw incongruence scores could be interpreted as follows: lower scores between species represented high morphological similarity relative to a high genetic divergence; in contrast high incongruence scores represented morphologically dissimilar species that had a relatively low genetic divergence. For the purposes of plotting these values for visual assessment a heatmap was used. However, as the incongruence among some species was particularly large the distribution of the incongruence scores was skewed rendering the finer detailed incongruence poorly visualised. Log transforming these values adjusted the extremes of the incongruence scores for ease of visual assessment. Furthermore, as these values are only understandable relative to the dataset being examined we featured scaled them. The resulting visualization of the scores could therefore be interpreted as follows: distance ratios closer to 1 were more morphologically similar than expected given their genetic distance (possible convergence or stabilising selection); distance ratios closer to 0 were more morphologically different than expected (possible rapid divergence).

Centroid Size

Significant differences in centroid size were identified among many groups (FDR adjusted $p<0.05$), but not all (TableS4). Dire wolves were the largest species in both datasets, particularly in molar size, but the range, particularly in mandible size, overlapped with wolves substantially (Figure S2).

Morphological variance

Morphological diversity varied among taxa and numerous pairwise differences were found to be significant (at the FDR corrected $p<0.05$ level, see Table S5 for pairwise comparisons). This likely reflects the differing levels of ecological plasticity among some species (e.g. wolves are highly plastic and return the highest Procrustes variance scores). However, the variance may also be influenced by sampling, as wolves are extensively sampled across their entire range,

whereas some other species, particularly dire wolves, were represented by specimens from only a fraction of their range. Such variation in morphological diversity should be investigated further in future studies for other species, but here we focus on dire wolves. Dire wolves were less morphologically diverse than gray wolves in both mandibular morphology and first mandibular molar morphology, but only mandible morphology was significantly different ($p<0.01$, see Table S5). However, this could reflect the extensive spatial sampling of gray wolves (their entire range) compared with the restricted sampling of dire wolves (primarily La Brea).

Group differences and allometry

All Procrustes ANOVA pairwise comparisons were found to be significant (at the FDR corrected $p<0.05$ level see Table S6 for pairwise comparisons). Of particular note, dire wolves and gray wolves were different, but as their variance differed and gray wolves occupied a greater area of morphospace in both mandible and first molar datasets, it is likely that this characteristic resulted in different central tendencies. CCV percentages varied across both datasets (Table S7 & S8), but all were high (87.5–100%) and well above the CCV percentage achievable by random chance alone; this indicates that all species are morphologically different and distinguishable to a high degree of accuracy. Dire wolves and gray wolves could be identified between with a high level of confidence (90–97% CCV).

Allometry was found to be significant in both the mandibular and first molar datasets ($p<0.01$, Figure S3). Allometric slopes were found to be parallel (at the $p<0.01$ level) in both datasets and could therefore be removed. Removal of allometry increased the already high ability to identify between dire wolves and gray wolves in the first molar dataset (from ~91% to 100% Table S9), but decreased the identification accuracy in the mandibular dataset (from 100% to ~98% Table S10). The change in mandibular results is relatively small considering the already high level of accuracy. These allometric dental difference may indicate that large tooth size has been selected for in dire wolves, which may correspond with their specialization in consumption of megafauna and hard foods.

Dire wolf mean shape was closest to both gray wolf mean tooth shape and mandible shape; gray wolf mean shapes was found to be the second nearest neighbor in both datasets (Figure 1B & Figure S4). Dhole are also very close to dire wolf mandibular mean shape, but the sample size for dhole is also extremely small and the variability of shape is lower. When allometry is removed the allometry free dire wolf mandible mean shape was found to be closest to the black-backed jackal, while the first molar mean shape was closest to the African hunting dog. It was possible to calculate procrustes distances among both combined mean shapes (i.e. the distance between each species based on the combination of superimposed mandible and first molar landmark configurations), but because of the preservation of the dire wolf specimens the mean shapes for each morphological dataset was constructed from different sets of individuals and as a result it was not possible to calculate the combined shape distance of individuals to species means. The results from the combined shape and allometry free shape, identify wolves and black-backed jackals respectively, as the closest to dire wolf combined mean shape.

Phylogenetic signal and morphology

A brownian motion model best fit the phylogenetic signal in the CS data, however it was found to be non-significant in both datasets (mandible: $K=0.8109$, $p=0.092$; first molar:

K=0.8078, p=0.095). Almost all of the results of phylogenetic signal testing in the morphological datasets were non-significant (p>0.05) except when allometry was removed from the mandibular dataset and the combined mandibular-molar dataset, which produced significant results (Kmult 0.4082, FDR adjusted p=0.0168 and Kmult 0.54247, FDR adjusted p=0.0168 respectively). This could indicate that selection pressures on size are strong and the tempo of this selective pressure variable, which could then obscure the phylogenetic signal in morphology. Such a scenario would not be surprising given the element being examined (mandibles) and its role in feeding behavior, which is highly variable among these species. None of the phylomapping exercises were clear and the majority had extensive overlap of internal phylogeny nodes (Figure S5). Furthermore, the phylomapping approach was not able to indicate which pairs of species deviated from the phylogenetic signal the most from visual inspection alone.

To investigate this further we examined the incongruence between the molecular distances compared with the morphological distances. These incongruences are the result of multiple evolutionary mechanisms and as such a low score can be the result of either stabilising selection or convergence, whereas a high score may be the result of rapid divergence. Therefore we do not attempt to use these scores as any form of test or to determine the process, rather we use them as a descriptive tool for examining the dataset (Tables S11-13). The morpho-molecular incongruence score revealed that when raw shape was analyzed the molar and combined datasets for the dire wolves and gray wolves had the lowest incongruence score, with high morphological similarity, despite deep genetic divergence (Tables S11 & S13). In the mandibular dataset dire wolves vs. dhole had the lowest incongruence score and wolves vs. dire wolves were the second lowest (Table S11, Figure S4 & S6). On the other side of the incongruence spectrum, where species were more morphologically dissimilar relative to their depth of genetic divergence and the resulting incongruence scores were high, coyotes v gray wolves were found to be the most incongruent in the mandible and combined datasets (Tables S11 & S13), while in the molar dataset African wolves v gray wolves had the highest incongruence scores (Table S11). When allometry was removed gray wolves v black-backed jackals consistently the lowest incongruence across all datasets (Tables S12-13). Gray wolves v coyotes had the highest incongruence scores in the allometry free mandible dataset (Table S12). Coyotes v African wolves produced the highest incongruence scores in the molar and combined allometry free datasets (Table S12-13). Incongruence scores between gray wolves and dire wolves in the allometry free datasets continue to be low, particularly for mandible morphology (Table S12), suggesting that the increase in phylogenetic signal found in these datasets were more likely the result of improved correspondence of genetic and morphological distances among other species.

Proteomics

Collagen extraction and MS/MS

The *Canis dirus* (LACMP23-1619) purified collagen sample (extracted using the method stated in Fuller et al. (*61*), was prepared for proteomic analysis following a slightly modified version of the ZooMS protocol outlined in Welker et al. (*62*). The collagen was removed from the Eppendorf and stored in the freezer, and 100 µl 50 mM Ambic was added to the empty Eppendorf and heated at 65 °C for 1 hour. This was followed by digestion overnight at 37 °C; 50 µl of the heated sample was digested using 1 µl of 0.5 µg/µl porcine trypsin in trypsin resuspension buffer (Promega, UK) and the other 50 µl was dried down and resuspended in 50 µl 100 mM Tris solution to be digested with elastase (Worthington; USA) at the same concentration

in 10% Tris solution. Two different enzymes were used to increase the protein sequence coverage for LC-MS/MS (*63*, *64*). Digestion was stopped by the addition of trifluoroacetic acid (TFA) at a concentration of 0.5-1% of the total solution. Peptides were desalted using zip-tips (*65*) and eluted in 100 µl of 50% acetonitrile (ACN)/0.1% TFA (v/v).

The extracted peptides were analysed at the Discovery Proteomic Facility at Oxford University. The sample was analyzed on a Q-Exactive employing an Easyspray column (ES803, 500mmx75µm, Thermo) and a gradient of 2%-35% ACN in 0.1% FA/5%DMSO over 60 minutes. The MS1 resolution was set to 70,000 with an AGC target of 3E6. MS2 spectra for up to 15 precursors were acquired with a resolution of 17,500 and an AGC target of 1E5 for up to 128ms and 28% normalized collision energy (higher-energy collision dissociation). The precursors were excluded for 27 seconds from re-selection.

The LC-MS/MS raw files were converted to MGF files using Proteowizard (*66*) and searched against a mammal collagen database which included common contaminants (http://www.thegpm.org/crap/) in PEAKS v7.5 (*67*, *68*). Mass tolerances were set at 0.5Da for the fragment ions and 10ppm for precursor ions and up to 3 missed cleavages were permitted. Searches allowed various post-translational modifications (PTMs) including oxidation (MHW) and hydroxylation of proline (both +15.99), deamidation (NQ; +0.98) and pyro-glu from E (-18.01) and a fixed PTM of carbamidomethylation (+57.02) which occurs due to sample preparation. A maximum of 3 PTMs was allowed per peptide. Protein tolerances were set at 0.5% false discovery rate (FDR), >50% average local confidence (ALC; de novo only) and -10lgP score ≥ 20.

Sequences of both COL1A1 and COL1A2 were concatenated using previously published mammal collagen consensus sequences. Telopeptides very rarely survive in fossil samples and so these were removed from all sequences. Isoleucine and leucine cannot be differentiated using low energy tandem mass spectrometry and *de novo* sequencing as both amino acids are isobaric. Therefore, the identification of leucine/isoleucine was consistent with the consensus sequences. Once a potential collagen sequence was compiled for *Canis dirus*, the sequence was added to the collagen database and the sample was re-analysed using PEAKS to check for coverage and sequence substitutions. Any differences noted in the consensus were inspected manually. In order for a difference to be considered authentic, it had to occur in more than 1 product ion spectrum and be covered by both b and y ions.

Phylogenetic analysis

Here we built a phylogenetic tree using the COL1A1 and COL1A2 sequence generated by MS/MS analysis from a dire wolf as well as representatives Carnivora species. We first obtained the amino acid sequence of COL1A1 and COL1A2 from a redfox (*Vulpes vulpes*; XP_025851655, XP_025859557), a dog (*Canis familiaris*; NP_001003090, NP_001003187), a dingo (*Canis familiaris dingo*; XP_025295726, XP_025327115), a grizzly bear (*Ursus arctos*; XP_026368913, XP_026361636), a northern seal (*Callorhinus ursinus*; XP_025715155, XP_025728689) and a cat (*Felis catus*; XP_003996748, XP_003982813) from genbank. We also obtained amino acid sequence from the genome of all modern species used in this study (gray fox, andean fox, black-backed jackal, side-striped jackal, african wild dog, dhole, ethiopian wolf, African wolf, Eurasian gray wolf, and Yellowstone gray wolf). To do so we downloaded the

GFF file from Ensembl (v95) for canFam3.1. We then extracted and translated the coding sequence of COL1A1 and COL1A2 using gffread v0.10.5.

The amino acid sequence was then concatenated for each species and then aligned using MAFFT v7.123b (*69*). We then masked all leucine (L) and isoleucine (I) from the alignment as these are isobaric and low-energy MS/MS sequencing is not capable of discriminating between them. A Phylogenetic tree was generated using MrBayes version 3.2.1 (*70*) with the amino acid model estimated from the data (prset aamodellpr = mixed). We ran two runs of four chains each with 1,000,000 generations. Convergence was assessed by ensuring all ESS for all parameters where higher than 100. The resulting phylogeny indicates that gray wolves, African wolves, and coyotes, as well as dogs, all form a monophyletic clade (posterior >0.7) that excludes the dire wolf (Figure S7).

Genomics

Ancient DNA

Five samples (DireNTC, DireGB, DireSP, DireGWC), out of 46 specimens, possessed enough endogenous DNA for deeper sequencing (Table S1). Here we describe the protocol used by the six laboratories which attempted DNA extraction and sequence from dire wolves remains (Oxford, Copenhaguen, UCSC, UCLA, ASU and Adelaide). The samples which were successful (selected for deep sequencing) and unsuccessful are mentioned in the title of each section. Additional metadata for these samples can be found in Table S1 and site description for the five samples that were selected for deep sequencing can be found in the section above.

University of Oxford (PalaeoBarn)

Successful sample(s): DireAFR; unsuccessful sample(s): JH167-JH180/AJ62-AJ56.

DNA extraction

DNA was extracted from tooth or bone samples (see Table S1) in a dedicated ancient DNA laboratory using appropriate sterile techniques and equipment. Extraction was carried out following the Dabney extraction protocol (*71*) but with the addition of a 30 minutes pre-digest stage (*72*).

Library preparation

Illumina libraries were built following (*73*), with the addition of a six base-pair barcode added to the IS1_adapter.P5 adapter. The libraries were then amplified on an Applied Biosystems StepOnePlus Real-Time PCR system to check that library building was successful, and to determine the optimum number of cycles to use during the indexing amplification PCR reaction. A six base-pair barcode was used during the indexing amplification reaction resulting in each library being double-barcoded with an "internal adapter" directly adjacent to the ancient DNA strand and which would form the first bases sequenced, and an external barcode that would be sequenced during Illumina barcode sequencing.

Capture

The lysates were sent to Arbor Biosciences. They used myBaits® Whole Genome Enrichment (WGE) (37) procedure following myBaits® manual version 3.0 using a probe set generated from a supplied C. lupus genomic DNA.The baits can be produced from any pure

eukaryotic or prokaryotic genomic DNA source, for example modern gray wolf (for capture of DireGWC and DireAFR). Arbor Biosciences' unique process converts this gDNA into a pool of biotinylated RNA baits. These baits can then enrich corresponding molecules from a user-supplied NGS library, for example from an ancient or extinct species such as the dire wolf, via the process of myBaits in-solution hybridization capture. This enriched library allows for orders-of-magnitude more efficient NGS than would be possible using the full, non-enriched library.The supplied DNA extracts were used to build Illumina TruSeq-style libraries, using blunt-end adapter ligation and a uracil non-stalling polymerase for 8 to 12 cycles of indexing amplification. Extracts of sample AJ66 were sonicated prior to library preparation, and size-selected to retain all fragments < 500nt (Fraction I-S). Fragments >500nt consequent this sonication were further sonicated to <300nt and converted to libraries separately (Fraction I-L). Arbor then employed the gray wolf WGE bait set to enrich between 270 and 700ng available library of Fraction I-S in two ~40hour rounds of hybridization capture. The first round used 55C for the hybridization and wash temperatures, and the entirety of the enriched product was amplified 8 cycles before being taken to a second round that used 60C temperatures. After the second round, half of each enriched library was amplified between 8 and 12 cycles and then sequenced. The libraries from Fraction I-L were sequenced without enrichment.

Sequencing

Libraries were screened on an Illumina HiSeq 2500 (Single End 80bp) sequencer at the Danish National High-Throughput Sequencing Centre and on a Illumina MiniSeq at the AMIS laboratory in Toulouse. Based on this data we selected DireAFR for deeper sequencing because of its preservation. Deeper sequencing was conducted on Illumina HiSeq 4000 (Paired End 150bp) at Novogene (Novogene Corporation Inc CA 91914, USA).

University of California Los Angeles (UCLA)
Successful sample(s): DireGWC; unsuccessful sample(s): None.

DNA extraction

DNA from sample DireGWC (Table S1) was extracted from a lower right 4th premolar in a designated ultra-clean facility at UCLA using the appropriate sterile techniques and equipment in keeping with standard aDNA practice. Extraction was carried out following the protocol of (*74*) based on a silica-column based protocol.

Capture

The lysates were sent to Arbor Biosciences. They used myBaits® Whole Genome Enrichment (WGE) (37) procedure following myBaits® manual version 3.0 using a probe set generated from a supplied *C. lupus* genomic DNA. They used the DNA extracts to build Illumina TruSeq-style libraries, using blunt-end adapter ligation and a uracil non-stalling polymerase for 8 to 12 cycles of indexing amplification. Both extracts were taken to library prep without any treatment.

Library preparation

Arbor Biosciences employed the gray wolf WGE bait set to enrich between 270 and 700ng available library of the sample in two ~40hour rounds of hybridization capture. The first round used 55C for the hybridization and wash temperatures, and the entirety of the enriched product

was amplified 8 cycles before being taken to a second round that used 60C temperatures. After the second round, half of each enriched library was amplified between 8 and 12 cycles.

Sequencing

Libraries were sent for sequencing on a HiSeq4000 (Paired-end 100bp) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Australian Centre for Ancient DNA, University of Adelaide

Successful sample(s): DireNTC, DireGB and, DireSP; unsuccessful sample(s): R46006, R30312, R13470, R21446, and P232517.

DNA extraction for successful sample(s): DireNTC, DireGB and, DireSP

Sample IMNH48001/52 (DireGB; Table S1) was subjected to a silica-based DNA extraction method (*75*). This protocol included decalcification in 4 mL 0.5 M EDTA for 1 hour at 37 °C under constant rotation, after which the EDTA was removed and replaced with 4 ml of fresh 0.5 M EDTA and incubated overnight under constant rotation at 37 °C. A final incubation with an additional 60 µl of Proteinase K (20 mg/mL) was then performed for 2 hours at 55°C, following which the supernatant (digestion buffer) was mixed with a modified QG buffer (15.5 mL QG buffer [Qiagen], 1.3% Triton X-100 [Sigma-Aldrich], 25 mM NaCl [Sigma-Aldrich], 0.17 M Sodium Acetate [Sigma-Aldrich]) and bound to silicon dioxide particles, which were then washed with 80% ethanol. Bound DNA was eluted in 200 uL of TE buffer (10 mM Tris, 1 mM EDTA, pH 8).

Sample KU48130 (DireNTC; Table S1) was subjected to an alternative DNA extraction protocol optimised for recovery of short DNA fragments (*71*). Bone samples were decalcified in 1 mL 0.5 M EDTA for 1 hour at 37 °C under constant rotation, after which the EDTA was removed and replaced with 980 µl of fresh 0.5 M EDTA and 20 uL of Proteinase K (20mg/mL) then incubated overnight under constant rotation at 55 °C. The digestion buffer was mixed with 13 mL of a modified PB buffer (12.6 mL PB buffer [Qiagen], 6.5 µL Tween-20, and 390 µL of 3M Sodium Acetate) and bound to silicon dioxide particles, which were then washed with 80% ethanol. Bound DNA was eluted in 100 µL of TE buffer.

Sample VP1737 (DireSP; Table S1) was extracted following a third protocol, beginning with decalcification under rotation overnight in 10 mL of 0.5 M EDTA at room temperature. The decalcified material was then digested under rotation overnight in 3 mL of 100 mM Tris-HCl, 100 mM NaCl, 0.5 mg/mL proteinase K, 10 mg/mL dithiothreitol (DTT), and 1% sodium dodecyl sulphate (SDS) at 55 °C. Following digestion, 3 mL of Tris-saturated phenol was added and mixed under rotation for 10 min at room temperature, followed by centrifugation at 1500 g for 5 min. The aqueous phase was then transferred to a new tube. This process was repeated twice: once with 3 mL of Tris-saturated phenol, and once with 3 mL of chloroform. The final aqueous phase was de-salted with sequential additions of DNA-free water to an Amicon Ultra-4 Centrifugal Filter Unit (Millipore) and concentrated to a final volume of 200 µL.

DNA extraction for unsuccessful sample(s): R46006, R30312, R13470, R21446, and P232517.

Samples R46006, R30312, and R13470 were extracted following the same protocol as described above for DireGB. Sample R21446 and two subsamples of P2325171 were extracted following the same protocol as described above for DireNTC.

Library preparation for successful sample(s): DireNTC, DireGB and, DireSP

We used Gansauge *et al.* (76)'s single-stranded protocol to create Illumina sequencing libraries from the DNA extracts of IMNH48001/52, KU48130, and VP1737. After library preparation, we performed a real-time PCR assay to determine how many cycles of PCR were required to optimise library quantity and complexity (77). Duplicate real-time PCR assays were performed for each library in a final volume of 10 µL, each comprising: 1 µL of a 1:5 dilution of library, 1 x Platinum Taq DNA Polymerase High Fidelity buffer (ThermoFisher Scientific), 2 mM MgSO4 (ThermoFisher Scientific), 0.25 mM of each dNTP (ThermoFisher Scientific), 0.4 µM each of a P5 and P7 indexing primer (designed based on Meyer & Kircher, 2010), 0.004 x ROX (Life Tech), 0.2 x SYBR (Life Tech), 0.56 M DMSO (Sigma-Aldrich), and 0.2 U of Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific), in laboratory grade water. Real-time PCRs were performed on a LightCycler 96 (Roche) with the following cycling conditions: 94 °C for 6 min; 40 cycles of 94 °C for 30 s, 60 °C for 30 s, 68 °C for 40 s; followed by a high-resolution melt. The libraries were then amplified using conventional PCR. In order to maintain library complexity and minimise PCR bias, each library was amplified in eight separate 25 µL reactions, each comprising: 3 µL of undiluted library, 1 x Platinum Taq DNA Polymerase High Fidelity buffer (ThermoFisher Scientific), 2 mM MgSO4 (ThermoFisher Scientific), 0.25 mM of each dNTP (ThermoFisher Scientific), 0.4 µM each of a P5 and P7 indexing primer, and 0.2 U of Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific), in laboratory grade water. Cycling conditions for the PCR were as follows: 94 °C for 6 min; a number of cycles of 94 °C for 30 s, 60 °C for 30 s, 68 °C for 40 s as determined using rtPCR (8 for KU48130, 9 for VP1737, 9 for IMNH48001/52); and 68 °C for 10 min. PCR products from each library were pooled and purified using AMPure (Agencourt), and resuspended in 30 µL of buffer comprising 10 mM Tris, 0.1 mM EDTA, and 0.05% Tween-20.

Library preparation for unsuccessful sample(s): R46006, R30312, R13470, R21446, and P232517.

Extracted DNA was enzymatically repaired and blunt-ended, and had custom adapters ligated following the protocol of Meyer and Kircher (2010). Adapter sequences featured unique barcodes in order to allow identification and exclusion of any downstream contamination. Libraries were subjected to a short round of PCR in order to increase the total quantity of DNA using primers complementary to the adapter sequences. Cycle number was kept low (exact number determined by rtPCR) and template was split into eight separate PCRs per library in order to minimise PCR bias and maintain library complexity. Each individual PCR (25 µL) contained 1× PCR buffer, 2.5 mM MgCl2, 1 mM dNTPs, 0.5 mM each primer (Table S3), 0.1 U AmpliTaq Gold and 2 µL DNA. Cycling conditions were as follows: 94 °C for 12 min; 12-13 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 40 s (plus 2 s/cycle); and 72 °C for 10 min. PCR products were purified using AMPure magnetic beads (Agencourt). Commercially synthesised biotinylated 80-mer RNA baits (MYcroarray, MI, USA) were used to enrich the libraries for placental mammal mitochondrial DNA.

Sequencing for successful sample(s): DireNTC, DireGB and, DireSP

The libraries from DireGB and DireSP were diluted to 1.5 nM and each was run on one lane of an Illumina HiSeq X Ten using 2 x 150 bp PE (300 cycle) chemistry. The library from DireNTC was diluted to 2 nM and run on an Illumina NextSeq flow cell using the 2 x 75 PE (150 cycle) High Output chemistry. Due to the modified structure of the adapters used in (*76*) Gansuage *et al*.'s library protocol, these sequencing runs used a custom R1 sequencing primer (CL72; see (*78*)) instead of the default Illumina primer included in the kits.

Sequencing for unsuccessful sample(s): R46006, R30312, R13470, R21446, and P232517.

The enriched libraries were sequenced on Illumina high-throughput sequencing platforms. Only 21 sequencing reads in total (among all six libraries) could be mapped against the gray wolf mitochondrial reference, all of which could be excluded as common lab contaminants (e.g. *Homo sapiens*, *Bos taurus*) mapping to conserved gene regions. Consequently, these samples were excluded from further analysis.

University of Copenhagen

Successful sample(s): None; unsuccessful sample(s): Bt2, Bb2, Bt3C, and Bb3C.

DNA extraction

Teeth and bone were subsampled from a lower jawbone from collections in UCLA (Bt1, Bb1, Bt2, Bb2, Bt3C, Bb3C, see Table S1), originally found in the La Brea tar pits. Bone and teeth powder was digested in a EDTA, urea and proteinase K buffer as in (*79*) and purified as in (*71*) using a modified binding buffer as in (*80*). Out of a total of 6 individual extractions (1 ml), 3 teeth and 3 bone sub-samples, 2 bone and 2 teeth were purified as above and the other bone and tooth involving an additional Phenol Chloroform purification step as in before purification as in (*81*).

Library preparation

Double stranded Illumina libraries were made on a tooth and a bone (without Phenol Chloroform treatment) extract using the commercial NEBNext DNA Sample Prep Master Mix Set 2 (E6070, New England Biolabs Inc., Beverly, Massachusetts, USA) and using MinElute (Qiagen) purifications in between reactions with 5x PB buffer (*81*). Single stranded Illumina libraries were made on the remaining 4 purified extracts (including 2 Phenol Chloroform treated extracts) strictly according to (*78*). qPCR was performed using IS7/8 primers (*73*) and Lightcycler 480 reagents using 1 uL 10x diluted library on an Mx3005 instrument (Agilent). This was done to check for successful library preparation and estimate the number of cycles for indexing amplification. Index PCR was performed in 100 uL reactions using Platinum Taq polymerase (Invitrogen) with 6-base indexed full length P5 primers and common non-indexed full length P7 primer. Samples were given 10-25 cycles in PCR and subsequently purified using MinElute columns (Qiagen). Quantity and quality of libraries was measured using a Qubit 2.0 fluorometer and a 2100 Bioanalyzer (Agilent).

Sequencing

While double strand libraries failed in amplification, single-strand libraries were screened using an Illumina HiSeq 2500 (SR 80 mode) platform at the Danish National High-Throughput Sequencing Centre in Copenhagen. After mapping, all reads were found to be non-mammalian and hence non-endogenous for the sample.

University of California Santa Cruz (UCSC)

Successful sample(s): None; unsuccessful sample(s): JK376.

DNA Extractions

DNA from sample JK376 (Table S1) was extracted and libraries were built in a dedicated aDNA lab at the University of California, Santa Cruz. The coprolite was first crushed to expose undigested bone and then the sample was divided into two separate tubes for extraction, one with bone and one with all other material. The bone fraction was powdered using a Retsch MM 400 ball mill. DNA extraction was performed following the method outlined in Dabney et al. (2013) on 120mg of bone fraction and 250mg of non-bone material was extracted with MoBio's Powerlyzer kit.

Library Preparation

We prepared Illumina libraries from both extractions using the double stranded DNA library preparation protocol outlined in Meyer and Kircher using 20uL of extract (73). Libraries were amplified for 25 cycles using Amplitaq Gold hot start polymerase (2U Polymerase, 0.2uM each primer, 0.25mM each dNTP, 2.5mM MgCl2, 1X Amplitaq Buffer) and were SPRI purified using a factor of 1.75x.

Capture

Target enrichment was performed on the post amplified library from the non-bone fraction using Arbor Biosciences myBaits® Custom. The RNA probe set includes 75 mammalian mitochondria and was designed after (82).

Sequencing

Shotgun and enriched libraries were sequenced separately on an Illumina MiSeq (2x75bp) at the University of California, Santa Cruz.

Arizona State University (ASU)

Successful sample(s): None; unsuccessful sample(s): DW01-DW19.

DNA extraction

DNA was extracted from teeth and bones (samples DW01-DW19, Table S1) in a dedicated ancient DNA laboratory at Arizona State University using previously established protocols (*83*).

Library preparation

Libraries were constructed by following a modified version of a protocol created by Meyer and Kircher (2010). Double stranded libraries were constructed from DNA extracts and amplified in duplicate using AmpltiTaq Gold Polymerase (Applied Biosystems). Samples were

purified using a MinElute PCR Purification kit (Qiagen) and quantified using an Agilent 2100 Bioanalyzer (Agilent).

Capture

Mitochondrial DNA was captured using previously established protocols. Modern DNA was extracted from cheek swabs of two dogs using a standard phenol-chloroform extraction (*84*). The mitochondrial genome was amplified in two separate PCRs using primers and methods from previously published sources (85). In brief, the PCR products were pooled into equimolar amounts for primer sets A and B, which were constructed based on previous work on the mitochondrial genome of Canis familiaris (7, 85). The bait DNA was then sheared to 200-300bp using a Covaris sonicator and attached to beads that were thoroughly washed (86).

Ancient samples were captured using these modern baits in accordance with previously published protocols (87). In summary, a hybridization mixture was created which included the indexed libraries, an Agilent blocking agent, and an Agilent hybridization buffer among other reagents. The beads were then incubated in this mixture for two nights before being washed, removing everything that was not bound to the beads, and eluted. This enriched library elution was purified using a MinElute column and quantified using a KAPA Library Quantification Kit by Illumina Platforms. Based on these results, some samples were amplified or diluted to ensure they reached a concentration of at least 4 nM, as requested for sequencing.

Sequencing

Amplified libraries were pooled in equimolar amounts and sequenced on an Illumina MiSeq (paired-end 150bp).

Modern DNA

Black-backed jackal - UCLA

DNA from a black-backed jackal (C. mesomelas) (Table S14) was extracted from a blood sample collected in 1987 (Soysambu, Eastern Africa) using the Qiagen DNeasy Blood and Tissue Kit. The DNA extract was sent to ArborSciences for library preparation. The DNA was sonicated and size-selected to 300nt modal fragment lengths and converted to two identical Illumina TruSeq-style libraries using standard a-tail chemistry and 6 indexing cycles, and subsequently sequenced at Novogene (Novogene Corporation Inc CA 91914, USA) on an Illumina Hiseq4000 (150bp paired-end).

Side-striped jackal - Copenhagen

A ~37.5x nuclear genome (Table S14) was sequenced from a tissue sample (T-1252) from a Side-striped jackal (C. adustus) collected in 2002 from Guinea, (mtDNA-sequenced was previously sequenced in (88)). The DNA was extracted using a Kingfisher Duo extraction robot, using Cell and Tissue DNA Kit from ThermoFisher Scientific using manufacturer's protocol (ThermoFisher Scientific, Waltham, MA). The DNA was fragmented into 400-600bp molecules using a Bioruptor NGS device (Diagenode, Liège, Belgium). A next generation sequencing library was made using the commercial NEBNext DNA Sample Prep Master Mix Set 2 (E6070, New England Biolabs Inc., Beverly, Massachusetts, USA) in combination with BGIseq adaptors (89), and was sequenced on a lane of PE100 on a BGIseq500 platform by BGI-Europe.

Data processing

Modern data

Modern samples sequenced in this study as well as those downloaded from public repositories (Table S14) were aligned to canFam3.1 (dog reference genome) and VulVul2.2 (red fox assembly; accession: GCA_001887905.1) using BWA mem (90), with a realignment step as implemented in GATK (91). For each sample, we then computed depth of coverage using the following command in ANGSD (92):

```
angsd -doCounts 1 -i input.bam -doDepth 1 -out sample_cov -minQ 0 -minMapQ 0
```

The output was used to build a cumulative distribution of per sample depth of coverage. All regions, within each sample, that fell within the 5% highest and lowest coverage were excluded from further analysis. This procedure ensured that no abnormally covered (e.g. repetitive regions or copy number variation) region were included in the analysis. The likelihood of each possible genotype, in single sample for every base of the reference genome (excluding those in abnormal coverage regions) was then computed using the GATK genotype likelihood function as implemented in ANGSD:

```
angsd -GL 2 -out file_name -doCounts 1 -setMinDepth min -setMaxDepth max -i input_bam_file -doGlf 4
```

At each base the genotype was encoded as missing (N) unless the likelihood of the highest genotype was ten fold higher than the next best possible genotype. To accommodate the difference in coverage among modern and ancient samples in our D-statistics analysis (see below), we also called genotypes by randomly sampling a single read of 20 base pair minimum and with a mapping quality (MAQ) and base quality (BQ) of at least 30 at each covered position in the genome (93–95).

Ancient data

Raw reads were filtered allowing one mismatch to the indices used in library preparation. Adapter sequences were removed using AdapterRemoval (96). Reads were aligned using Burrows-Wheeler Aligner (BWA) version 0.7.17 (90) to canFam3.1 (dog reference genome) and VulVul2.2 (red fox assembly; accession: GCA_001887905.1) with the following parameters ("-l 1024,-n 0.01, -o 2") (97). FilterUniqueSAMCons (98) was then used to remove duplicates. BAM files from different sequencing lanes were merged using the MergeSamFiles tool from Picard v1.137 (http://broadinstitute.github.io/picard/). To accommodate the low coverage of the nuclear genome of the dire wolf samples, genotypes were called by randomly sampling a single read of 20 base pair minimum and with a mapping quality (MAQ) and base quality (BQ) of at least 30 at each covered position in the genome, excluding bases within 5bp of the start and end of a read (93–95). Molecular damage was assessed using MapDamage2.0 using default parameters (99) (Figure S8; S9). The damage plots from DireAFR and DireGWC are expected when building double-stranded libraries from ancient DNA, as C to T misincorporations are observed on both strands (with damage on the complementary strand manifesting as G to A misincorporations). In contrast, single-stranded libraries (78) were created from the remaining dire wolf samples, meaning that damage from the complementary strand is not observed, resulting in only C to T misincorporations.

Ascertainment

Specific analyses, such as the supermatrix phylogeny based on SNPs (using ascertainment correction implemented in RAxML) as well as for the D-statistics, necessitate a list of pre-defined SNPs. Here, we used the genome consensus (see genotype calling section above) of all high coverage modern genomes (see column used-for-ascertainment in Table S14) to obtain a list of SNPs for these analysis. All variable positions, with a minor allele found in at least in two high coverage modern genomes (as heterozygous or homozygous) were kept for further analyses. On canFam3.1, this resulted in ~46M SNPs, ~13M of which were transversions and ~38M SNPs on VulVul2.2, ~11M of which were transversions.

Phylogenetic analyses - mitochondrial genomes

We used htsbox (https://github.com/lh3/htsbox) to generate mitochondrial majority consensus sequences from the bam files using BQ>=30 and MAPQ>=30 while excluding bases within 5bp of the start and end of a read to limit the incorporation of deamination in the analysis. To reconstruct the mitochondrial phylogeny, we retrieved the modern mitochondrial genomes from several extant canids from the BAM files aligned to the dog reference genome, including the genome of the gray wolf (C. Lupus), the North American endemic coyote (C. latrans), the African wolf (C. anthus), Ethiopian wolf (C. simensis), and wild dog (Lycaon pictus); and the dhole (Cuon alpinus), the black-backed (C. mesomelas) and the side-striped African jackals (C. adustus) (Table S14). Additional mitochondrial genomes were obtained from NCBI for the Arabian wolf (C. lupus arabs), the dog (C. lupus familiaris), the red wolf (C. rufus) and the Great Lake wolf (C. lycaon) were added as well in the analyses. The gray fox (Urocyon cinereoargenteus) and Andean fox (Lycalopex culpaeus) were used as outgroups. These were aligned together using MAFFT. Accession numbers are listed in Table S15.

Partitioning of the mitochondrial genome into 13-protein coding and two rRNA genes (12S,16S) was performed using AMAS (100). For this analysis, we removed any genes which were not covered in at least one dire wolf (Table S16). PartitionFinder2 (101) implemented in the Cipres web server (102) was subsequently used to determine the optimal substitution model for the 9 remaining genes (Table S17). Bayesian phylogeny reconstructions were done in MrBayes 3.2.6 (70) as implemented in the Cipres web server (102). Markov chain Monte Carlo (MCMC) sampling was performed with 4 chains run for 5 x 106 generations with one tree sampled every 1000 generations. A strict majority rule consensus tree was built from the combined chains, excluding the 25% first iterations as burn-in.

The amount of missing data ranged between 49% to 85% for the dire wolf specimens (see Table S16). DireAFR and DireNTC have the highest amount of missing data with 79% and 85% respectively. To address potential bias introduced by the missing data (103, 104), we performed several tree analyses. Firstly with the data covered in 9 genes (10587bp) in three dire wolves (Table S6), excluding the two dire wolves with the most missing data (DireAFR and DireNTC). The phylogeny recovered was similar to those recently obtained by (14) and (105) (Figure S10 A), including a monophyletic clade that contains Eurasian and American wolves (C. lupus), the dog (C. familiaris) and the African wolf (C. anthus). The Great Lake wolf (C. lupus lycaon) clusters with the coyote (C. latrans) and the red wolf (C. rufus), most likely due to admixture events between gray wolves and coyotes (106), (107). While we obtained relatively lower for the node (posterior probability ~0.5) leading to the Ethiopian wolf (C.simensis) and the golden

jackal (C. aureus) yet their position is consistent with the tree obtained by (105). The dire wolves form a strong monophyletic cluster, basal to all extant Canis, except for the C. adustus (the side-striped jackal) and L. pictus (African wild dogs) (Figure S10 A).

The second phylogenetic reconstruction included all five dire wolves (9 genes, 10587bp; Table S17; Figure S10 B). We retrieved similar relationships except for the deep nodes in the phylogeny, which could not be resolved. The five dire wolf specimens, however, still formed a highly supported monophyletic cluster. The third phylogenetic reconstruction excluded all the missing data from the alignment (566bp; Figure S10 C). C.adustus (side-striped jackal) is now basal to all canids and the Andean fox. This dramatically reduced our power to resolve deeper nodes within the phylogeny, yet the five dire wolves and the wolf like candis (wolves, coyote, and dog) clustered in two separate, highly supported clades (posterior probability > 0.9).

Phylogenetics - nuclear DNA (ascertainment free pipeline)

Filtering

We used bedtools (108) to obtain all regions of the canFam3.1 and VulVul2.2 assemblies that were covered by at least 1 read, in each dire wolf, excluding repetitive elements and CpG islands (see Table S18). We then extracted the sequence of these loci from the consensus sequences obtained of each modern species (see genotyping). We then merged the modern data with each dire wolf separately, filtering out loci that 1) had >20% missing data in any species 2) were shorter than 30 bp. DireNTC was excluded from this analysis due to its low coverage (Table S1).

Supermatrix

For each data set (multiple, and single dire wolf data sets) we built a maximum likelihood tree, with 100 bootstrap replicates using the GTR+G model as implemented in RAxML (109) by concatenating all loci into a "supermatrix". Analyses of single dire wolf samples, as well as combined samples (DireSP and DireGB; Table S18) provided enough power to retrieve a well supported topology (Figure S10). In all cases, the dire wolf samples were basal to all Canis, Lycaon and Cuon species.

We assessed whether the basal placement of the dire wolf could be due to the excessive long external branch induced by singleton DNA damage (Figure S8). To do so, we first estimated the proportion of "true" singletons in each dire wolf sequence by comparing the length of branches in other species to the root of Caninae based on the RAxML trees. We then randomly edited the direwolf sequences back to the ancestral state with a probability equal to the excess branch length (see Table S18) and built a maximum likelihood tree with the resulting alignment. This correction dramatically reduced the external branch length of both samples, but did not affect the topology (Figure S11). This analysis correction was also applied to alleviate potential issues of deamination affecting coalescent-based analysis and molecular dating (see below).

Species tree

We first built species trees based on the multispecies coalescent model using both SNAPP (110) and BPP (111). These methods have the advantage of taking either very short loci (BPP) or SNPs (SNAPP) rather than well resolved gene tree topologies (e.g. such as produced by

ASTRAL (112)) as an input. The latter are difficult to obtain for this study due to the highly fragmented nature of the dataset (see Table S18).

BPP

BPP uses the multispecies coalescent, to jointly estimate the species tree topology, divergence time (τ, in coalescent unit) and nucleotide diversity (θ). Here we used BPP to obtain a tree topology (A01 analysis). We used a uniform species tree prior, a diffuse inverse Gamma distribution (3, 0.015) for τ (roughly corresponding to ~10Myr root age) and a diffuse inverse Gamma distribution (3, 0.001) for θ.

We ran BPP on the two highest coverage direwolf samples (DireSP and DireGB). To do so we used every contiguous region of the genome of >30bp in which both direwolf were covered (see filtering section above). We used a burnin of 2,000 samples, sample frequency of 2, and collected 20,000 samples. The topology inferred placed the dire wolves as basal to Caninae (all species except the gray fox; Figure S12 A.). To assess whether basal placement could be due to an excess of singletons (see section Supermatrix above) we ran the same analysis (with the same priors) on the corrected data (see above). When removing the excess of singleton the resulting tree was identical to the one inferred by RAxML (Figure S12 B.).

SNAPP

We also analysed our SNP data using a multispecies coalescent approach implemented in SNAPP (110). As for BPP, this analysis was restricted to the two highest coverage direwolf samples (DireSP and DireGB). For each locus, we extracted biallelic SNPs, converting these to 0 (ancestral allele; using the gray fox to polarise allele), 1 (heterozygote) and 2 (derived allele), excluding any site that was missing in a single species. This resulted in 2,433 SNPs without singleton correction and 1,711 SNPs with singleton correction (see section Supermatrix above). We sampled u and v mutation rates parameters from a normal distribution (mean and sigma of 1.0). A uniform distribution (0, 1) was used for the species tree prior, and a Gamma distribution (3, 1000) for θ. This analysis was repeated on the corrected data-set (with correction for singleton excess in dire wolves), with the same priors.

As for BPP, the topology inferred by SNAPP put the dire wolf as basal to Caninae (all species except the gray fox; Figure S13 A&B). But similarly, this basal placement was most likely based on excess of singleton as demonstrated by the tree obtained after correcting for excess singletons (Figure S13 C&D).

Discordance visualization using Discovista

We then inferred maximum likelihood trees, using RAxML, in 500kb, 1Mb and 5Mb sliding windows across the genome (100kb overlap). This analysis was conducted for each dire wolf sequence separately. For each bin we concatenated all sequence that were covered in the dire wolf, using the same filters as for the supermatrix analysis. Only bins with at least 2kb of coverage in a dire wolf were considered (Table S19). The frequency, and support, of different topologies were visualised using discovista (113). Overall this analysis strongly rejects a topology in which the dire wolf is sister species to wolves (dire/can in Figure S14). Topologies in which the dire wolves (dire/out) or the black jackal/side-striped jackal (dire/in) were the most

basal group in the phylogeny were almost equally supported - although the dire/in topology obtained slightly higher support in this analysis (Figure S14).

Nuclear DNA (pipeline based on SNPs ascertained in modern genomes)

SNP-calling and variant filtering

We created indexed VCF files for each BAM file using samtools v0.1.18 (mpileup; part of the SAMtools package, (114)) and BCFtools v0.1.17 (call, index; part of the SAMtools package). We used Parallel v20170822 (115) to process each BAM file in parallel. We then converted the autosomal biallelic variants from the VCF files to random "pseudo-haploid" eigenstrat formatted variants using vcf2eig (part of eig-utils; https://github.com/grahamgower/eig-utils) with the following options: -m (include monomorphic sites), -s (include singleton sites), and -t (exclude transitions). The eigenstrat formatted files were then filtered to contain only variants from the list ascertained in the high-coverage modern samples (see Ascertainment section above; see Table S14 for a list of genomes that were used for the ascertainment) using eigreduce (part of eig-utils; https://github.com/grahamgower/eig-utils) with the following options: -m (include monomorphic sites) and -s (include singleton sites). We then used eigreduce (-i, -m, -s) to create eight sets of samples: all samples, modern (non-dire wolf) samples only, modern samples plus DireSP, modern samples plus DireGB, modern samples plus DireSP and DireGB, modern samples plus DireNTC, modern samples plus DireGWC, and modern samples plus DireAFR.

Supermatrix

The filtered eigenstrat formatted files for each of the eight sets of samples (see above) were then transposed to a PHYLIP file using eig2phylip (part of eig-utils; https://github.com/grahamgower/eig-utils). For each set of samples we created supermatrix trees (i.e. all variants concatenated) using the rapid bootstrapping algorithm in RAxML v8.2.4 (116)(-f a, -m ASC_GTRCAT) with 100 bootstrap replicates (-# 100) and the Felsenstein ascertainment correction based on the number of invariant sites (--asc-corr=felsenstein), which was calculated as the total ungapped length of the canFam3.1 autosomes minus the length of the alignment.

Discordance visualization using Discovista

For each of the eight sets of samples (see above) we further broke the genome down into 422 non-overlapping 5Mb windows using eigreduce (-R). For each of the 422 windows we then created a tree (i.e. all variants concatenated) using the rapid bootstrapping algorithm in RAxML v8.2.4 (-f a, -m ASC_GTRCAT) with 100 bootstrap replicates (-# 100) and the Felsenstein ascertainment correction based on the number of invariant sites (--asc-corr=felsenstein), which was calculated as the length of the window (5 million bases) minus the length of the alignment. As for the ascertainment free pipeline (see above) we then summarised and visualised the frequency and support of different tree topologies using discovista (113). Ultimately, we obtained comparable results to those from the ascertainment free pipeline (Figure S15; Figure S16).

Molecular dating - MCMCtree

Fossil calibration

Two fossil calibrations were used to calibrate the canid phylogeny so the absolute divergence times could be estimated with MCMCtree (part of the PAML suite v4.9 (117)):

1. We calibrated the root of the tree (Urocyon versus the other samples) using a uniform distribution with a minimum of 10.3 Ma and a maximum of 20 Ma. The uniform distribution had soft bounds, implemented as described by (118), such that the true age is between 10.3 Ma and 20 Ma with the left and right tail probabilities being 0.025.. The minimum bound was based on the occurrence of Metalopex macconnelli at the end of the Clarendonian NALMA (dated to 10.3 Ma; (119)), which appears to be closer to Urocyon than to Canis (Tedford, Wang, and Taylor, 2009). The maximum bound was set close to the end of the Harrisonian NALMA (dated to 20.6 Ma; (119)), which allows for the possibility that some Leptocyon species could have been early stem members of an extant canid lineage.

2. We calibrated the divergence between Andean fox (Cerdocyonina) and Canina using a soft-bounded uniform distribution (implemented as above) with a minimum of 4.9 Ma and a maximum of 10.3 Ma. The minimum bound of 4.9 Ma was based on the occurrence of Canis ferox, which is likely a stem member of Canina (7). The maximum bound was set at the end of the Clarendonian NALMA (dated to 10.3 Ma; (119)), based on the absence of recognised crown members of Canini prior to this point in time.

Molecular data

We used the same data as for the BPP analysis above (corrected and uncorrected). The data was concatenated and used to construct the molecular alignment. For this analysis we used only one representative of each species, randomly selecting one African wild dog and using the Eurasian wolf (ptw, SRS661492 see Table S14) instead of the Yellowstone wolf genome (based on overall coverage). As for the species tree, this analysis was restricted to the two highest coverage direwolf samples (DireSP and DireGB). We ran the analysis on both corrected and uncorrected data, i) without direwolf sequences (10 species), ii) direwolf sequence with the highest coverage(DireGB; 11 species), and (iii) with both direwolf sequences (DireGB and DireSP; 12 species).

Simulated data

We carried out a simulation study to investigate the impact that incorporating deamination (from aDNA) can have on divergence time estimates. To do so, we simulated molecular alignments based on a set tree topology, and gradually increased the number of external substitutions on one or two (i.e., if either one or two ancient sequences were simulated to be included in the alignment) branches in the tree.

We first ran BASEML (part of the PAML suite v4.9 (117)) to estimate the (i) best-scoring maximum-likelihood (ML) tree for the molecular data under the HKY+Γ5 substitution model, the (ii) transition/transversion ratio, the (iii) base frequencies, and the (iv) shape parameter α for the discrete-gamma model for rate heterogeneity. These parameters were estimated for the corrected molecular alignments (alignments with one and two direwolf sequences, respectively) under both tree topologies (“A” and “B”), being then used to simulate the corresponding molecular alignments under each topology to which substitutions were subsequently added.

Alignments were simulated using seqgen (120) as implemented in the phyclust R package (121). We used the parameters estimated by BASEML for the corrected molecular alignments under each tree topology evaluated. This means that the simulated alignments had the same

sequence length and biological properties than the corrected alignments. We then added an increasing proportion of "errors", erradded = 0%, 0.5%, 1%, 3%, and 5%; on the simulated direwolf sequence(s). Altogether, this resulted in 10 simulated alignments (5 error rates x 2 topologies) with 12 species (two direwolf) and 10 simulated alignments with 11 species (one direwolf).

Bayesian model selection for molecular clock and tree topology

We used the autocorrelated-rates, also known as geometric Brownian diffusion (GBM (118, 122)), and the independent log-normal rates (ILN (123, 124)) models. The analysis was run on different topologies for both alignments: "A" (dire/out in Figure S6) and "B", (dire/in in Figure S14). Altogether, we evaluated the fit of 4 models clock/topology combinations for both corrected and uncorrected data sets, prior to the dating analysis: (i) model 1: GBM + topology A, (ii) model 2: GBM + topology B, (iii) model 3: ILN + topology A, and (iv) model 4: ILN + topology B.

For each model, the likelihood values collected during the MCMC for each model were used to estimate marginal likelihoods with the mcmc3r R package (125). Marginal likelihoods were used to compute Bayes factors (BFs) and posterior probabilities, which then were used to select the best-fitting model for each alignment. This analysis supported the combination of topology A + clock GBM (dire/out in Figure S6) as the model that best fit the data (Bayes Factor=5.53; Posterior probability= 0.941) and was used for the subsequent Bayesian inference analyses. The results obtained for each model tested are shown in Table S20.

Priors and substitution model selection

Estimation of divergence times for the molecular alignments was carried out using MCMCtree under the GBM model (see above) and the HKY+Γ substitution model with 5 discrete-gamma categories. We used a uniform prior for node ages using the birth-death (BD) process (118) with the following parameter values: λBD = 1 (birth-rate), μBD = 1 (death-rate), and ρBD = 0.1 (sampling fraction for extant species).

We used the gamma-Dirichlet distribution for the rate (r) prior as implemented in MCMCtree (126). The shape parameter was set to $\alpha = 2$, which corresponds to a diffuse prior. The scale parameter β was chosen based on the estimated molecular branch lengths of the phylogeny. To do so, we first ran RAxML v8.2.10 (116) on the concatenated molecular alignment. The resulting best-scoring ML tree was used to estimate the distance from the tips to the root (i.e., the number of substitutions from the tips to the root), broot-tips. Given that broot-tips = r x troot and that the mean rate of the gamma-Dirichlet distribution is defined as $r = \alpha / \beta$, then, α / β = broot-tips / troot. We can therefore use broot-tips to estimate the scale parameter as $\beta = (\alpha \times \text{timeroot}) / \text{broot-tips}$. Table S21 lists the priors used for each analysis. Lastly, the prior on $\sigma i2$ was defined using a diffuse gamma-Dirichlet distribution, $\sigma i2 \sim \Gamma(2,2)$.

Results – simulated and real data sets

The divergence times were estimated for both real and simulated data sets using the preferred model according to BFs: autocorrelated-rates + topology "A". The results for the estimated mean age of the (i) root, the (ii) Vulpini-Canini divergence, the (iii) direwolf divergence from the jackals and the rest of canids, the (iv) jackals split from the rest of canids,

and (vi) the divergence between the two dire wolves are shown in Table S22. Note that the posterior divergence times for alignments without direwolf sequences were only estimated for the corrected and uncorrected real data sets. This is because the simulations required alignments in which at least one direwolf sequence was present, so aDNA deamination could be simulated.

Altogether, the results with the simulated data sets indicate that the presence of errors slightly affected divergence times, although the observed effect was minimal (e.g., the time of divergence between dire wolves and other canids increased by 3% at 0.5% error and 6% at 1% error). Nevertheless, introducing errors dramatically affected divergence time between the two ancient sequences on which we added errors; i.e., the time of divergence between the two dire wolves increased by ~two fold (Figure S17). Lastly, we also ran the analysis without the dire wolf sequences, which yielded highly similar results (see column "Estimated divergence times for the alignment without dire wolf specimens" in Table S22). The results presented in Figure 2A are based on the alignment with only one dire wolf specimen (DireGB).

D-statistics

We used D-statistics, as implemented in Admixtools (127), to detect gene-flow from the dire wolf into other canid lineages. We only used the two highest coverage dire wolf genomes for this analysis (DireSP and DireGB). We computed D of the form D(gray_fox, (dire wolf, (P1, P2)), where P1 or P2 can be any extant canid genome. We did not compute these statistics using the black-backed or side-striped jackal given the uncertainty in the topology at the root of Canis (see above). We used randomly sampled reads per site instead of genotype calls for all genomes, including in high coverage modern genomes, in order to account for different depth of coverage (same procedure as for the dire wolf; see above). The SNPs used in this analysis were ascertained in modern high coverage sequences (see section ascertainment above). We used a weighted block jackknife procedure over 1Mb blocks to assess significance ($|Z|>3$). This analysis was repeated on datasets aligned to both the Red fox (VulVul2.2) and dog (Canfam3.1) references genomes to avoid potential issues arising from a reference bias. We only used scaffolds longer than 1Mb.

We found no evidence of an excess of shared derived allele between dire wolves and North American wolves (Yellowstone gray wolf) or coyotes (Figure S18; Figure 2B), since the most recent common ancestor of African wolves, gray wolves (using canFam3.1 as reference:

P3	P1	P2	D	Z	Reference
DireSP	african_wolf	gray_wolf_yellowstone	-0.0013	-0.096	canFam3.1
DireSP	african_wolf	coyote	-0.0256	-1.91	canFam3.1
DireSP	gray_wolf_eurasia	coyote	-0.0179	-1.112	canFam3.1
DireSP	coyote	gray_wolf_yellowstone	0.0348	2.194	canFam3.1
DireSP	gray_wolf_eurasia	gray_wolf_yellowstone	0.0194	1.118	canFam3.1
DireSP	african_wolf	gray_wolf_yellowstone	-0.0013	-0.096	canFam3.1

DireGB	african_wolf	coyote	-0.0085	-0.623	canFam3.1
DireGB	gray_wolf_eurasia	coyote	-0.019	-1.199	canFam3.1
DireGB	coyote	gray_wolf_yellowstone	0.0216	1.33	canFam3.1
DireGB	gray_wolf_eurasia	gray_wolf_yellowstone	-0.0011	-0.059	canFam3.1
DireGB	african_wolf	gray_wolf_yellowstone	0.0085	0.598	canFam3.1

And VulVul2.2:

P3	P1	P2	D	Z	Reference
DireSP	african_wolf	gray_wolf_yellowstone	0.0094	0.644	VulVul2.2
DireSP	african_wolf	coyote	0.0147	1.002	VulVul2.2
DireSP	gray_wolf_eurasia	coyote	0.0144	0.904	VulVul2.2
DireSP	coyote	gray_wolf_yellowstone	0.0293	1.834	VulVul2.2
DireSP	gray_wolf_eurasia	gray_wolf_yellowstone	0.0528	2.657	VulVul2.2
DireGB	african_wolf	gray_wolf_yellowstone	0.0081	0.574	VulVul2.2
DireGB	african_wolf	coyote	0.004	0.269	VulVul2.2
DireGB	gray_wolf_eurasia	coyote	-0.0049	-0.31	VulVul2.2
DireGB	coyote	gray_wolf_yellowstone	0.0133	0.782	VulVul2.2
DireGB	gray_wolf_eurasia	gray_wolf_yellowstone	-0.0017	-0.102	VulVul2.2
DireGB	african_wolf	gray_wolf_yellowstone	0.0081	0.574	VulVul2.2

These results indicate that our data does not support the existence of gene flow between dire wolf and extant populations of North American canids (coyotes and gray wolves) taking place since they diverge from their most closely related Eurasian gray wolves and African wolves counterparts.

Our results, however, are consistent with admixture between dire wolf lineage and the ancestor of the dhole, wolves and coyotes (Figure S19). This is demonstrated by the fact that the dire wolf genome share significantly more derived allele with the genome of the gray wolves, coyote, African wolf, ethiopian wolf, and dhole, than with the genome of the African wild dog (Figure S19). This result was replicated using both the dog reference genome (canFam3.1):

P3	P2	P1	D	Z	Reference
DireSP	african_wild_dog1	coyote	0.1865	28.877	canFam3.1
DireSP	african_wild_dog1	dhole	0.1442	21.156	canFam3.1
DireSP	african_wild_dog1	ethiopian_wolf	0.1836	28.62	canFam3.1
DireSP	african_wild_dog1	african_wolf	0.1915	30.079	canFam3.1
DireSP	african_wild_dog1	gray_wolf_eurasia	0.1885	29.267	canFam3.1
DireSP	african_wild_dog1	gray_wolf_yellowstone	0.1923	30.176	canFam3.1
DireSP	african_wild_dog1	dhole	0.1442	21.156	canFam3.1
DireGB	african_wild_dog1	coyote	0.1997	30.403	canFam3.1
DireGB	african_wild_dog1	dhole	0.145	20.783	canFam3.1
DireGB	african_wild_dog1	ethiopian_wolf	0.1914	29.31	canFam3.1
DireGB	african_wild_dog1	african_wolf	0.2022	31.039	canFam3.1
DireGB	african_wild_dog1	gray_wolf_eurasia	0.2023	30.732	canFam3.1
DireGB	african_wild_dog1	gray_wolf_yellowstone	0.2041	31.131	canFam3.1
DireGB	african_wild_dog1	dhole	0.145	20.783	canFam3.1

and the red fox reference genome (VulVul2.2):

P3	P2	P1	D	Z	Reference
DireSP	african_wild_dog1	coyote	0.1476	20.87	VulVul2.2
DireSP	african_wild_dog1	dhole	0.1135	15.094	VulVul2.2
DireSP	african_wild_dog1	ethiopian_wolf	0.1407	19.531	VulVul2.2
DireSP	african_wild_dog1	african_wolf	0.1491	21.52	VulVul2.2

DireSP	african_wild_dog1	gray_wolf_eurasia	0.1433	20.128	VulVul2.2
DireSP	african_wild_dog1	gray_wolf_yellowstone	0.1527	20.406	VulVul2.2
DireSP	african_wild_dog1	dhole	0.1135	15.094	VulVul2.2
DireGB	african_wild_dog1	coyote	0.1211	16.086	VulVul2.2
DireGB	african_wild_dog1	dhole	0.0822	10.49	VulVul2.2
DireGB	african_wild_dog1	ethiopian_wolf	0.1108	14.514	VulVul2.2
DireGB	african_wild_dog1	african_wolf	0.1196	15.765	VulVul2.2
DireGB	african_wild_dog1	gray_wolf_eurasia	0.1187	15.675	VulVul2.2
DireGB	african_wild_dog1	gray_wolf_yellowstone	0.1225	15.321	VulVul2.2
DireGB	african_wild_dog1	dhole	0.0822	10.49	VulVul2.2

This admixture between ancestral lineages could be partly explain why our phylogenetic analyses could not resolve the root of this phylogeny.

D-statistics also suggests the existence of gene-flow between the dire wolf and the ancestor of all wolves and coyotes, using DireGB:

P3	P2	P1	D	Z	Reference
DireGB	ethiopian_wolf	gray_wolf_eurasia	0.0416	3.666	canFam3.1
DireGB	ethiopian_wolf	gray_wolf_yellowstone	0.0431	3.798	canFam3.1
DireGB	ethiopian_wolf	coyote	0.0305	2.687	canFam3.1
DireGB	ethiopian_wolf	african_wolf	0.0408	3.396	canFam3.1

yet this signal became non-significant when using either DireSP:

P3	P2	P1	D	Z	Reference
DireSP	ethiopian_wolf	gray_wolf_eurasia	0.0195	1.758	canFam3.1
DireSP	ethiopian_wolf	gray_wolf_yellowstone	0.0247	2.234	canFam3.1
DireSP	ethiopian_wolf	coyote	0.0085	0.754	canFam3.1
DireSP	ethiopian_wolf	african_wolf	0.0292	2.46	canFam3.1

and when using the VulVul2.2 assembly:

P3	P2	P1	D	Z	Reference
DireGB	ethiopian_wolf	gray_wolf_eurasia	0.0288	2.633	VulVul2.2
DireGB	ethiopian_wolf	gray_wolf_yellowstone	0.0317	2.767	VulVul2.2
DireGB	ethiopian_wolf	coyote	0.0241	2.098	VulVul2.2
DireGB	ethiopian_wolf	african_wolf	0.0277	2.278	VulVul2.2
DireSP	ethiopian_wolf	gray_wolf_eurasia	0.0086	0.726	VulVul2.2
DireSP	ethiopian_wolf	gray_wolf_yellowstone	0.028	2.378	VulVul2.2
DireSP	ethiopian_wolf	coyote	0.0155	1.333	VulVul2.2
DireSP	ethiopian_wolf	african_wolf	0.0277	2.3	VulVul2.2

This suggests that this signal might be due to reference bias toward CanFam3.1.

Our results are also consistent with admixture signal described in (*14*); namely admixture between wolf and golden jackal (*e.g.* D(gray fox (golden jackal, (Eurasian wolf, coyote))=-0.104, Z<-3 using canFam3.1 and D=-0.25, Z<-3 using VulVul2.2) and between African wolf and ethiopian wolf (*e.g.* D(gray fox(ethiopian wolf, (Eurasian wolf gray wolf))=-0.35, Z<-3 using canFam3.1 and D=-33, Z<-3, using VulVul2.2).

We also found evidence of gene flow between side-striped jackal and the ancestor of the dhole, wolves and coyotes (Figure S20). At first, this signal seem to involve the ancestor of the dhole, wolves and coyotes as well as african wild dog (see result on canFam3.1 in Figure S20). This signal, however, disappeared when the analysis using the VulVul2.2 reference genome (Figure S20), suggesting the existence of a reference bias in the African wild dog genomes toward the dog (canFam3.1) reference genome.

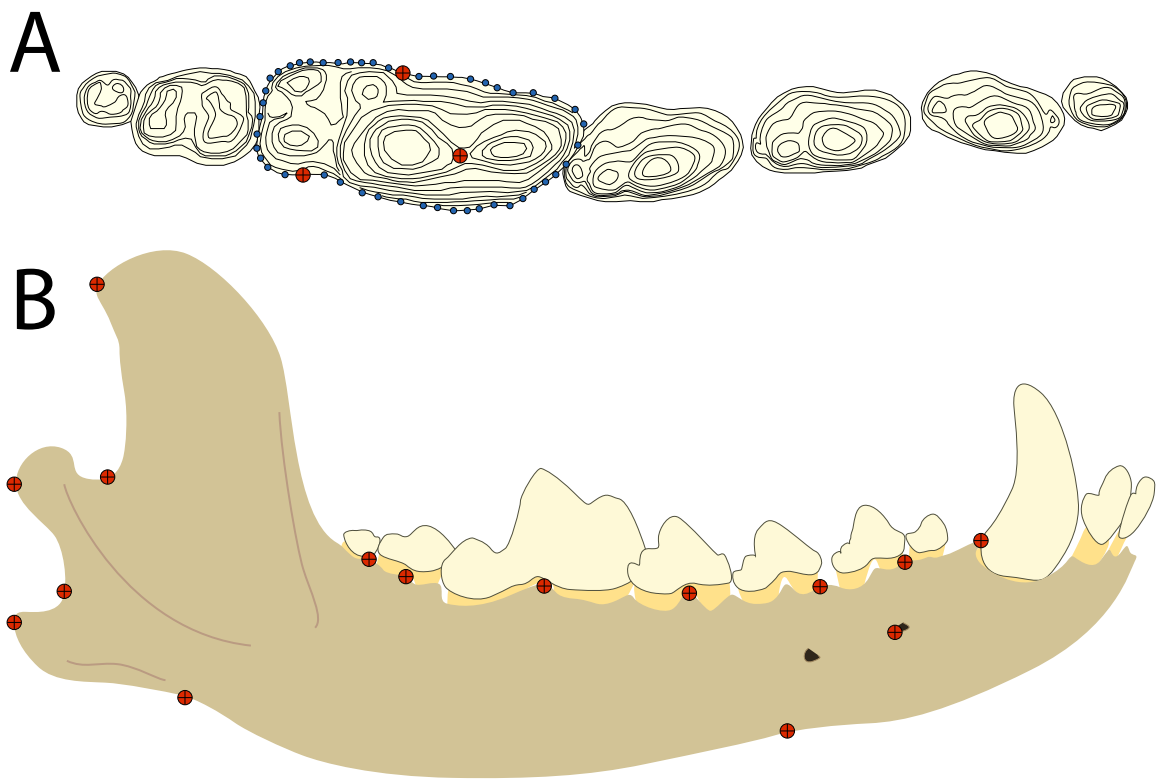


Fig. S1.

Landmark configuration for geometric morphometric protocol. Landmarks are marked in red while sliding semi landmarks are marked in blue. **A.** First mandibular molar with both landmarks and sliding semilandmarks configurations; **B.** Mandible with landmark configuration.

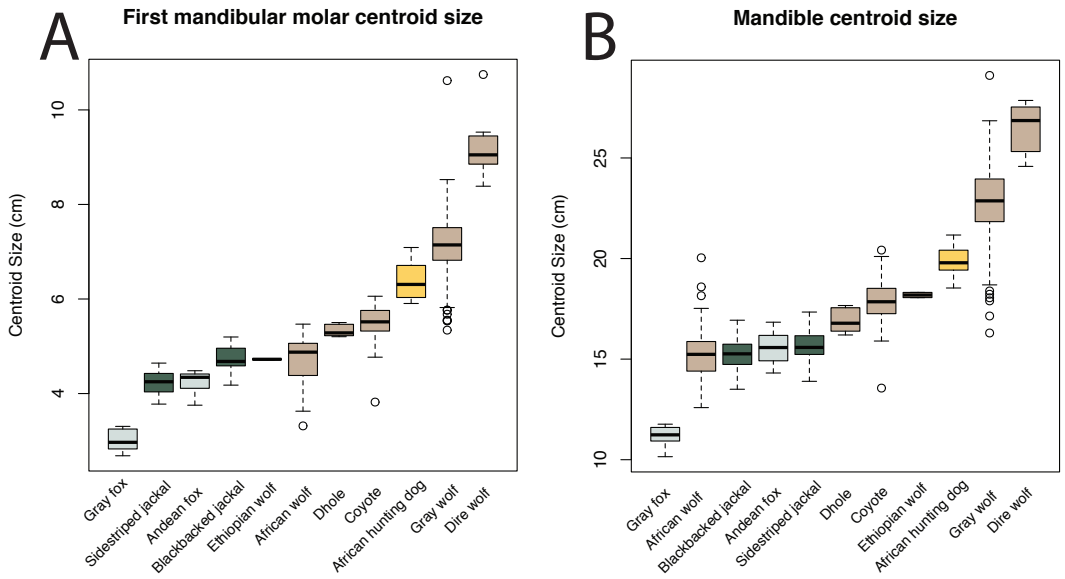
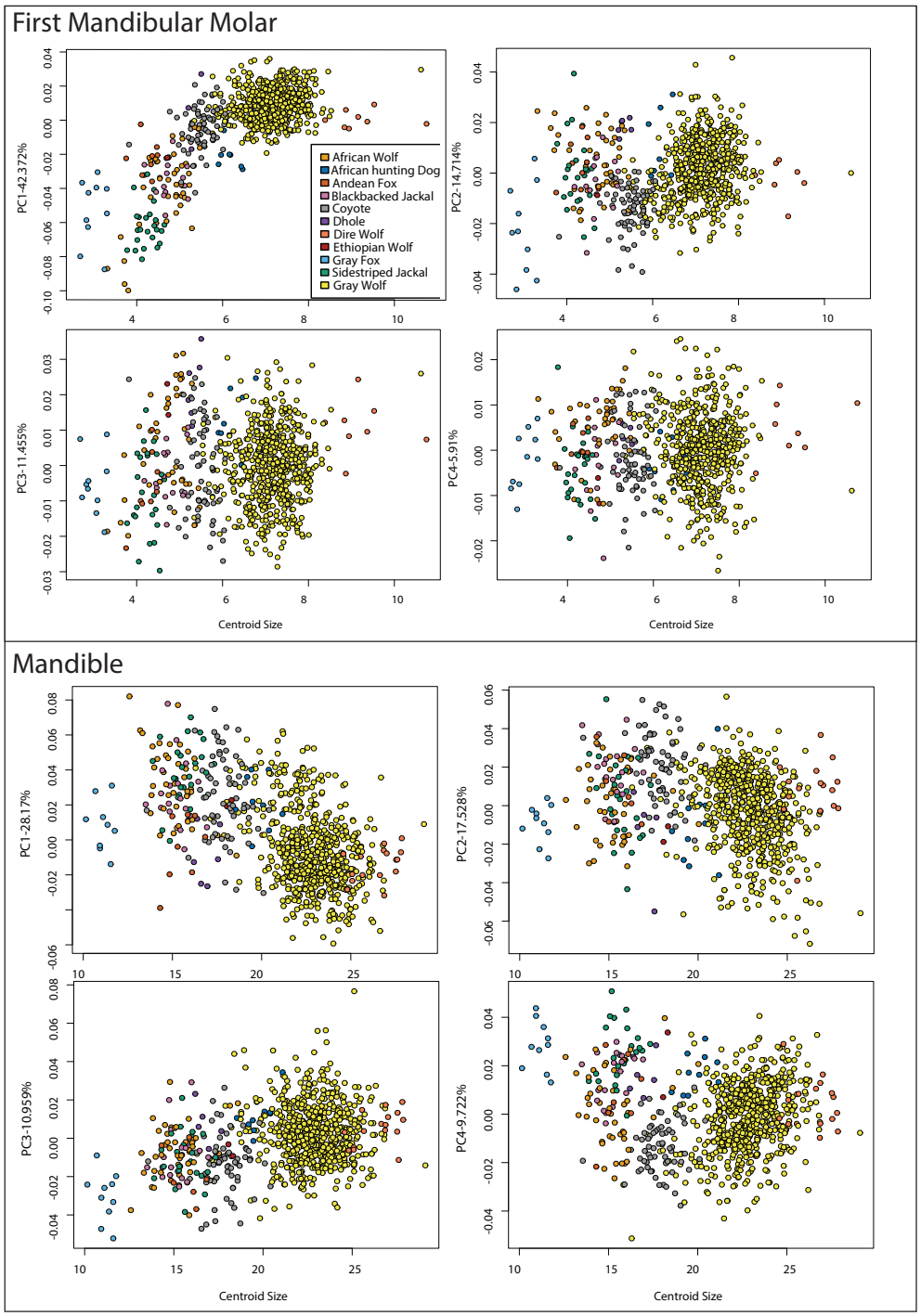


Fig. S2.

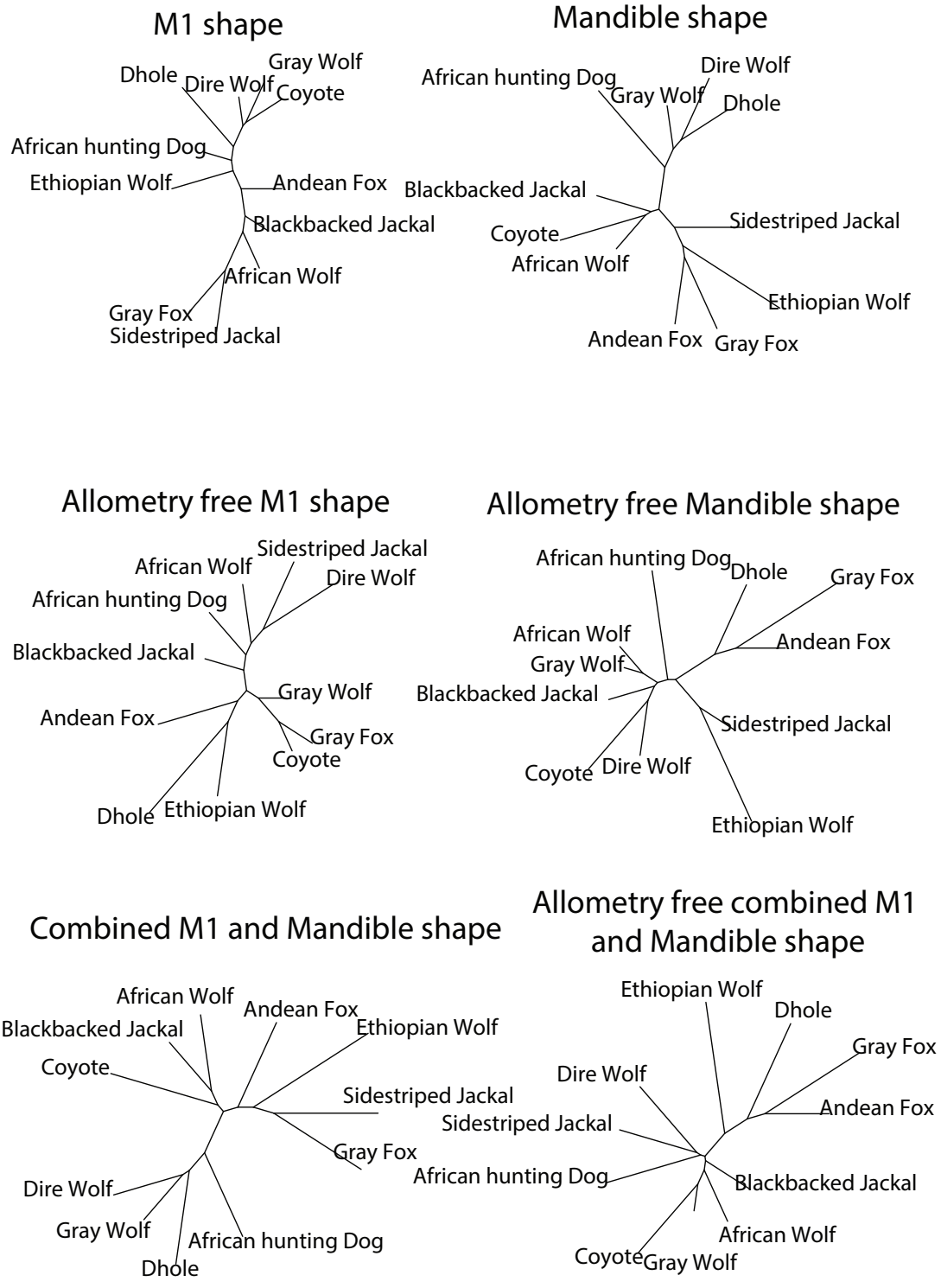
Boxplot of centroid size comparisons. Colours correspond to: Gray=outgroups; Green=jackals; Brown=wolf-like canids; Yellow=African hunting dog. A. Mandible centroid size; B. First molar centroid size.



1145

1146 **Fig. S3.**

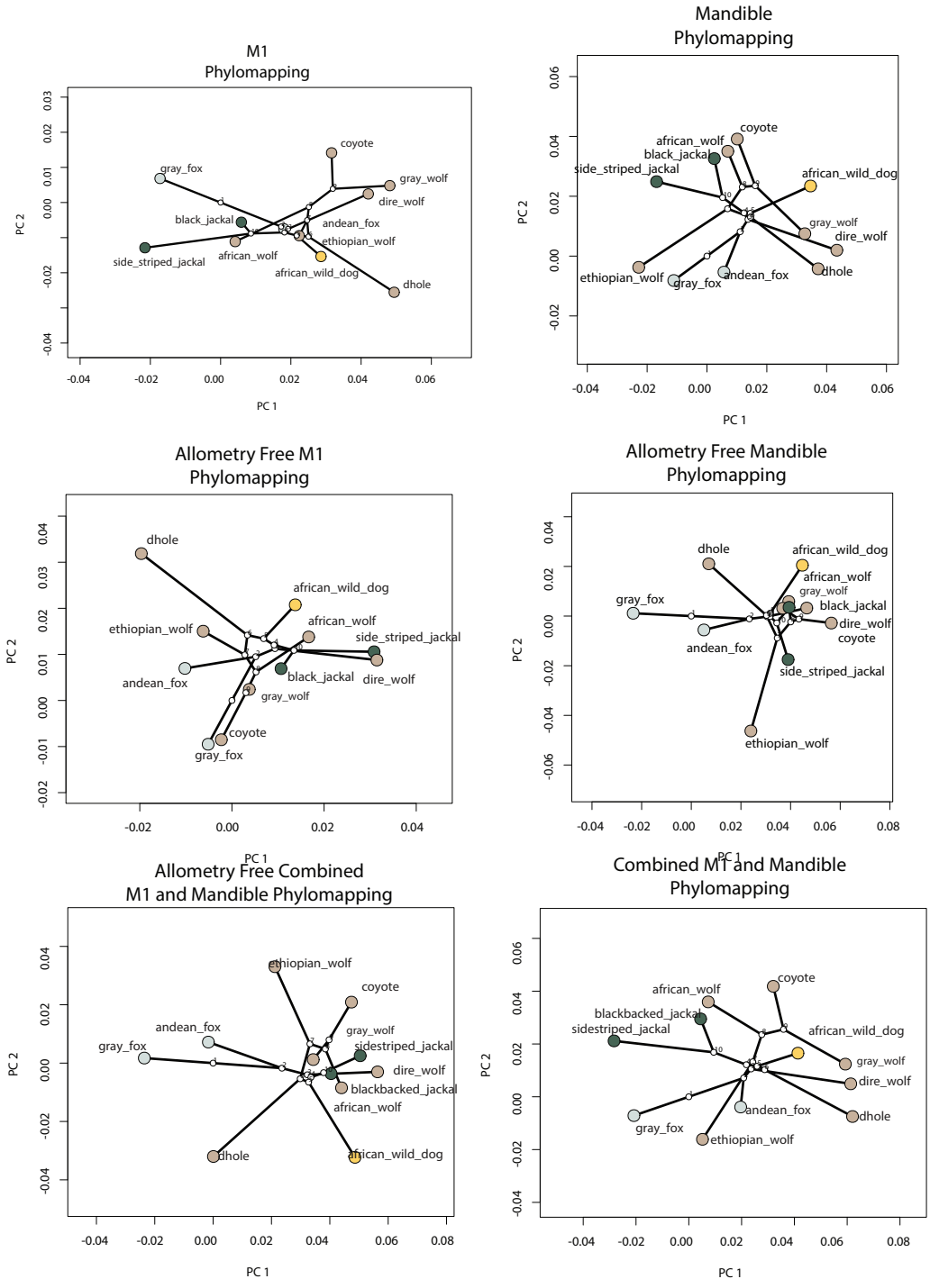
1147 **Panel of first two PCs versus CS in both mandibular and first molar datasets.** CS is on x
1148 axis for all plots. Plots show PCs 1–4 for each morphological data set versus CS. PC variance is
1149 listed on the y axis of each plot.



1150

1151 **Fig. S4.**

1152 **Panel of unrooted neighbour joining trees based on different metrics.**



1153

1154 **Fig. S5.**

1155 **Panel of PCAs with phylogenies mapped to mean shapes.** Colours correspond to:
1156 Gray=outgroups; Green=jackals; Brown=wolf-like canids; Yellow=African hunting dog.

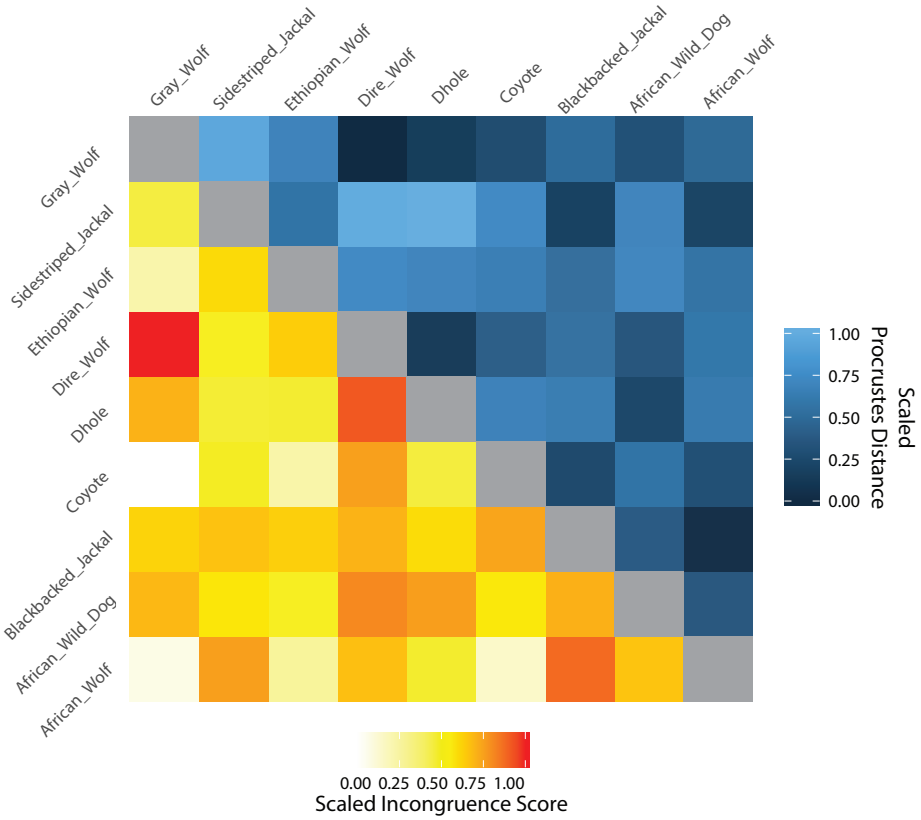


Fig. S6.

Heatmap visualisation of incongruence scores and full Procrustes distances. The lower triangle, scaled in red to white consists of incongruence scores constructed from combined morphological datasets versus genetic distances and transformed and scaled for ease of visualisation. Red scores represent highly similar morphologies with distant genetic divergence, whereas whiter scores represent distant morphologies with recent genetic divergence. The upper triangle, scaled from blue to black, represents feature scaled full Procrustes distances, with darker blues colours representing shorter distances and lighter blues representing greater distances.

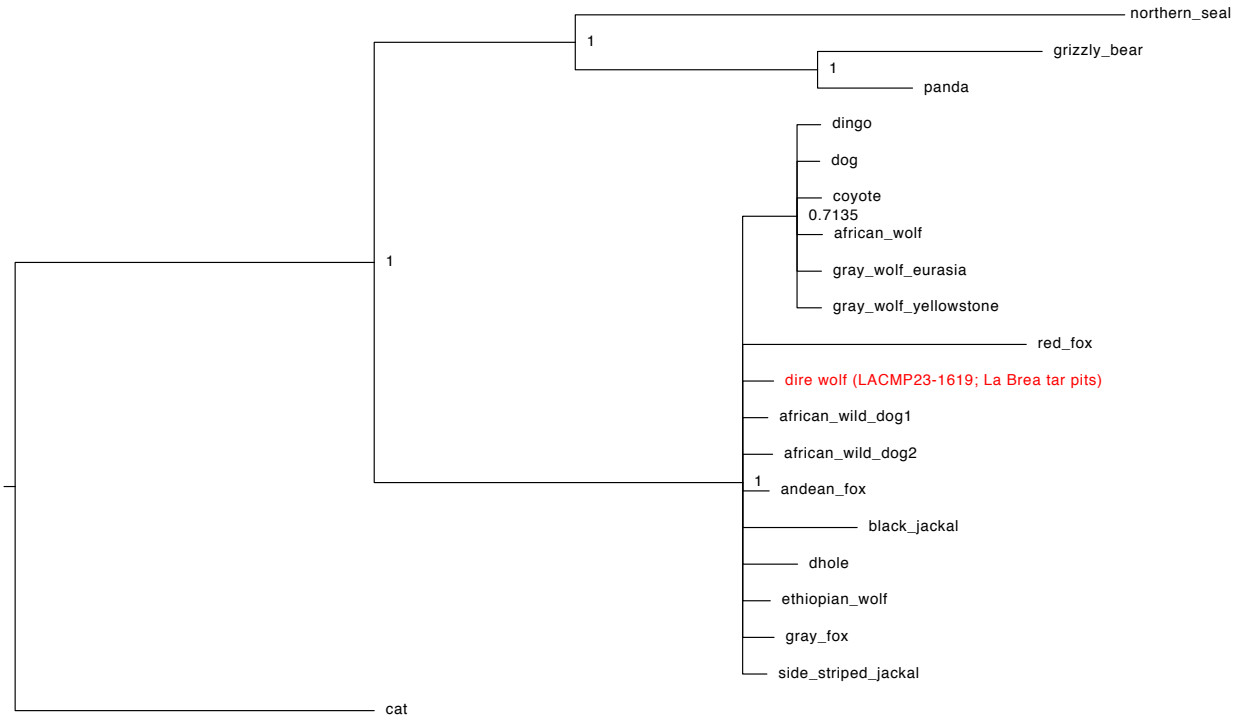


Fig. S7.

Bayesian phylogeny (MrBayes) of the concatenated COL1A1 and COL1A2 amino acid sequence.

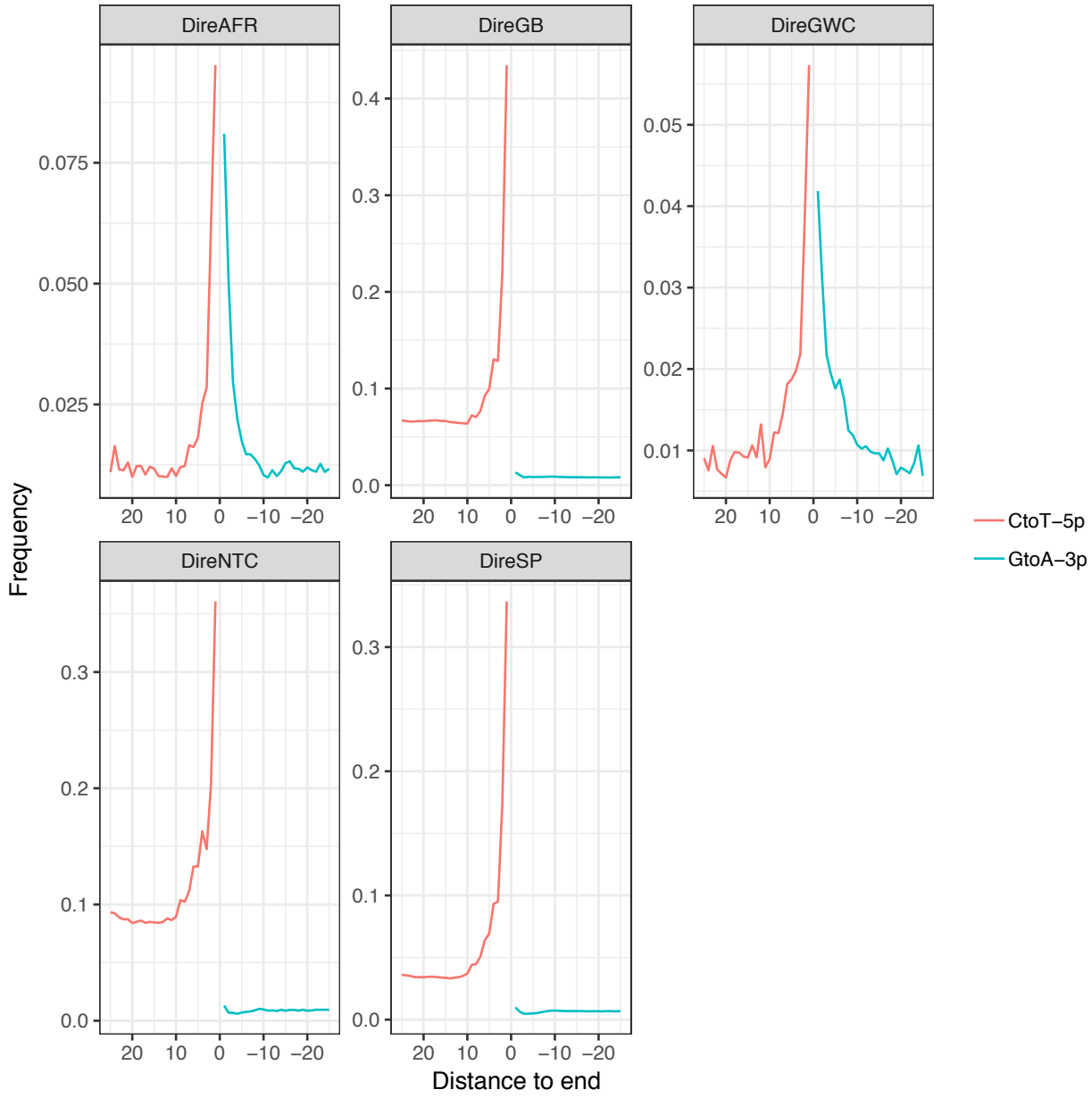


Fig. S8.

Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in nuclear and mitochondrial genome analyses.

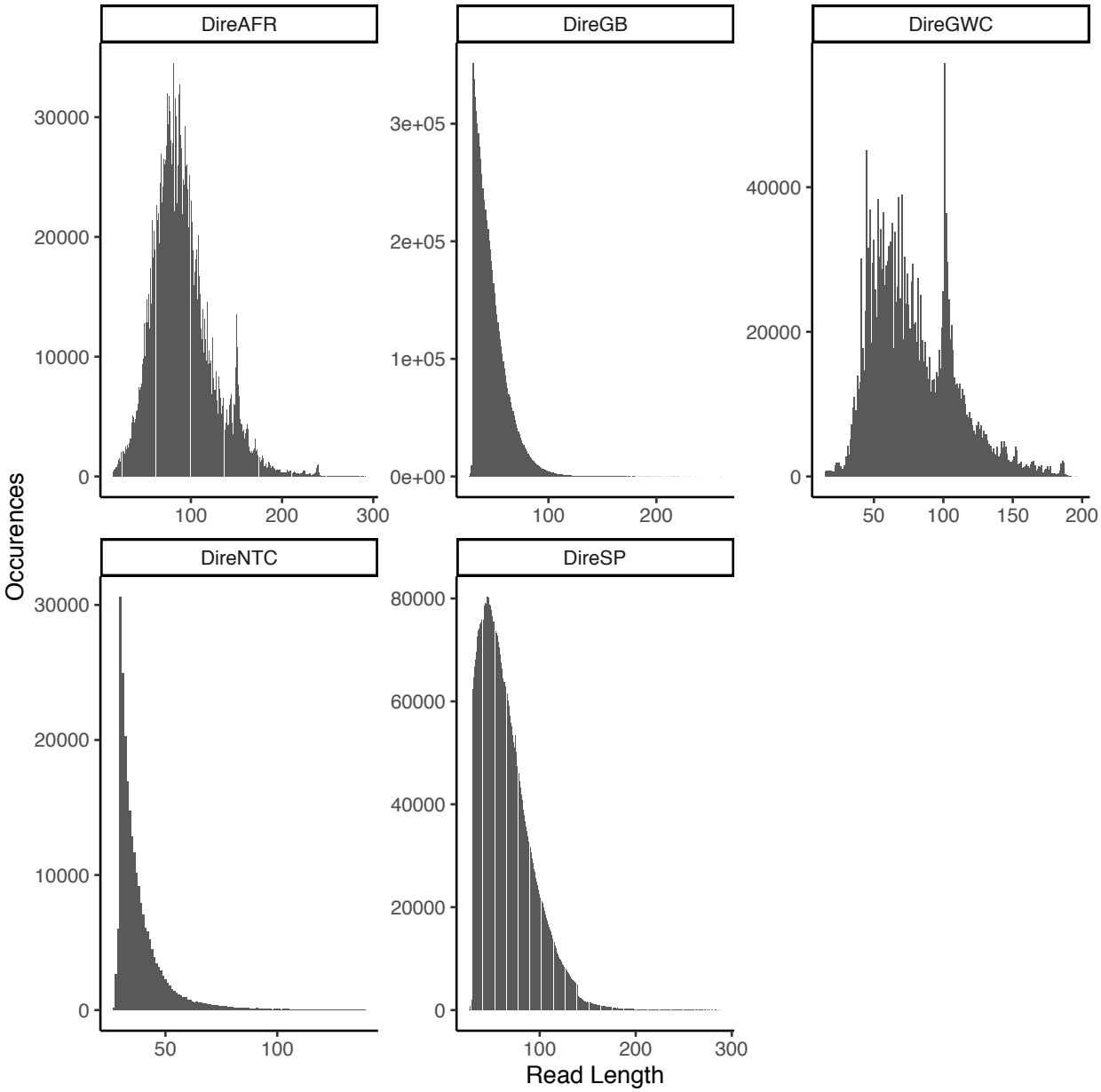


Fig. S9.

Per library read length distribution.

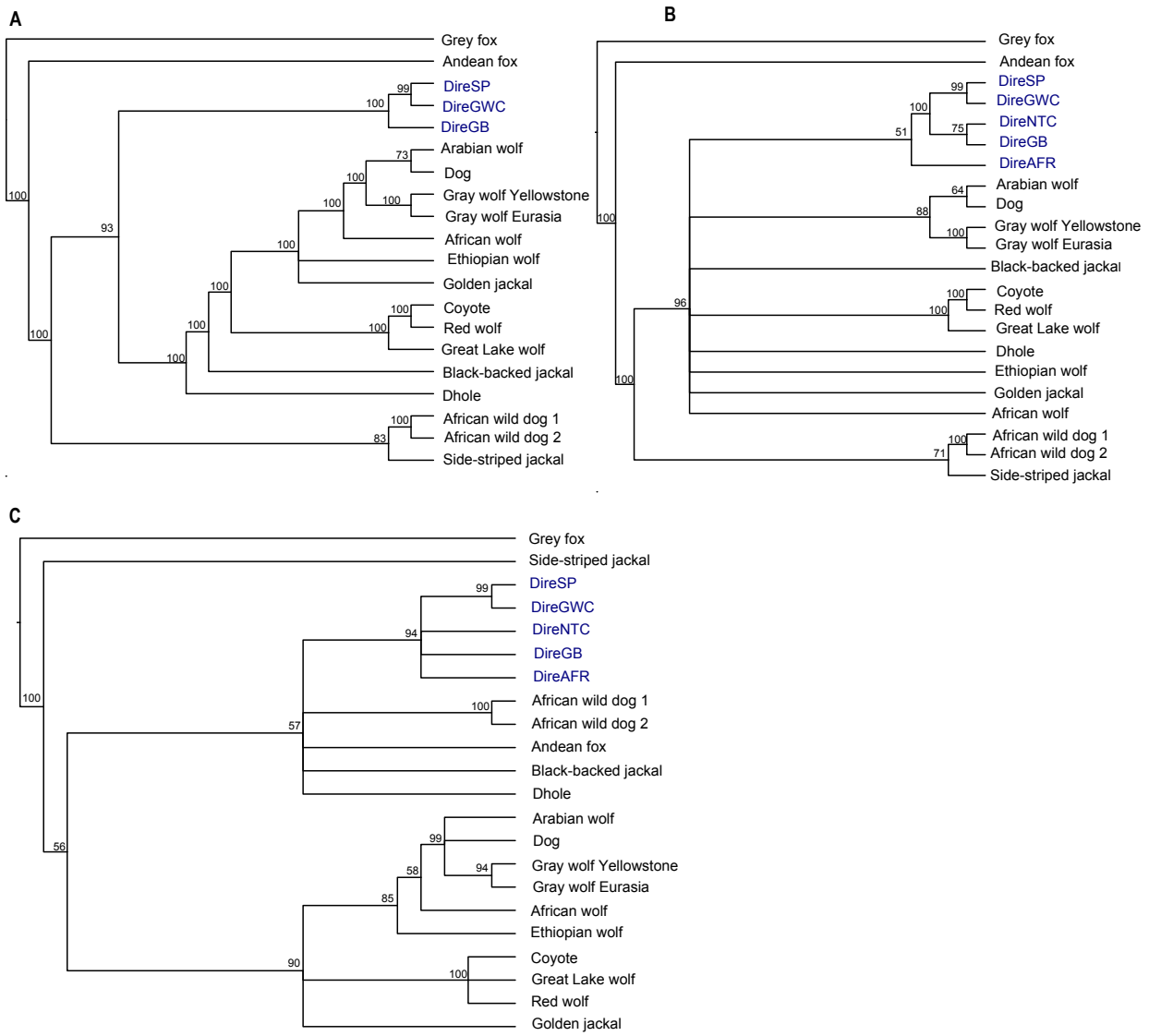
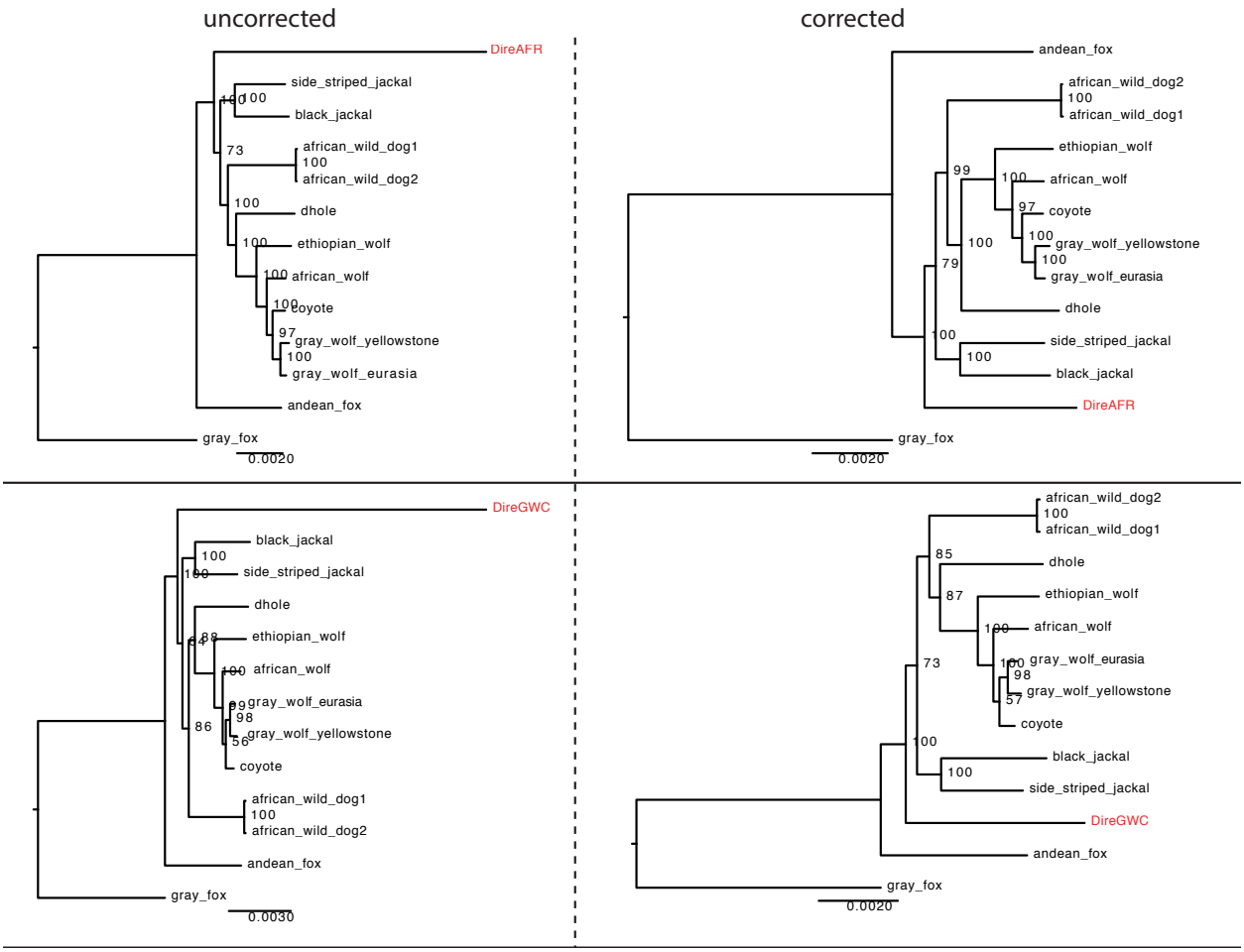


Fig. S10.

Bayesian phylogeny (MrBayes) of mitochondrial DNA. A. Three dire wolves were included - based on 9 mitochondrial genes (10587bp) (DireNTC, DireAFR were excluded) **B.** Same as A but with the five dire wolves. **C.** All five direwolves specimen without any missing data (566bp).



1187

1188 **Fig. S11.**

1189 **Maximum likelihood (RAxML) trees based on concatenation of sequence covered in each**
1190 **dire wolf, with and without branch correction scheme (removing singleton in dire wolf; see**
1191 **section supermatrix).** Dire wolves are marked in red.

1192

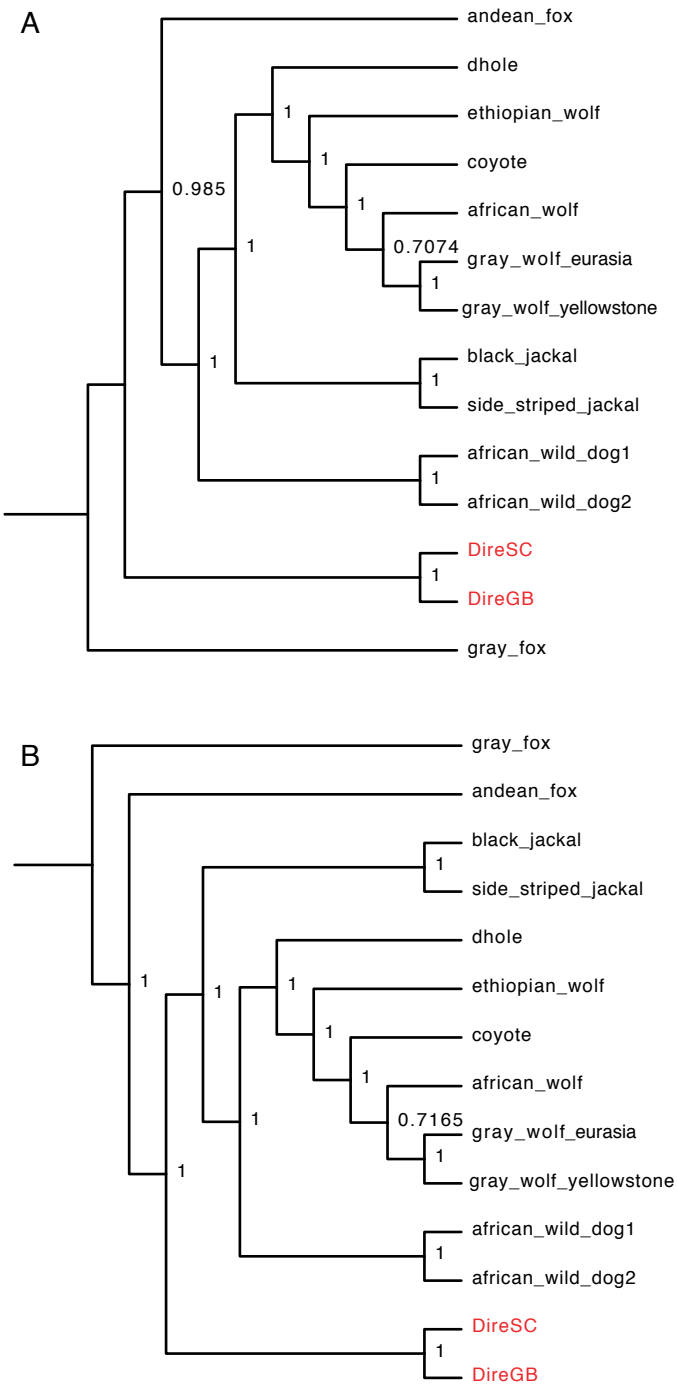


Fig. S12.

Phylogeny tree built using BPP. A. Without branch correction scheme **B.** With branch correction scheme. Dire wolves are marked in red.

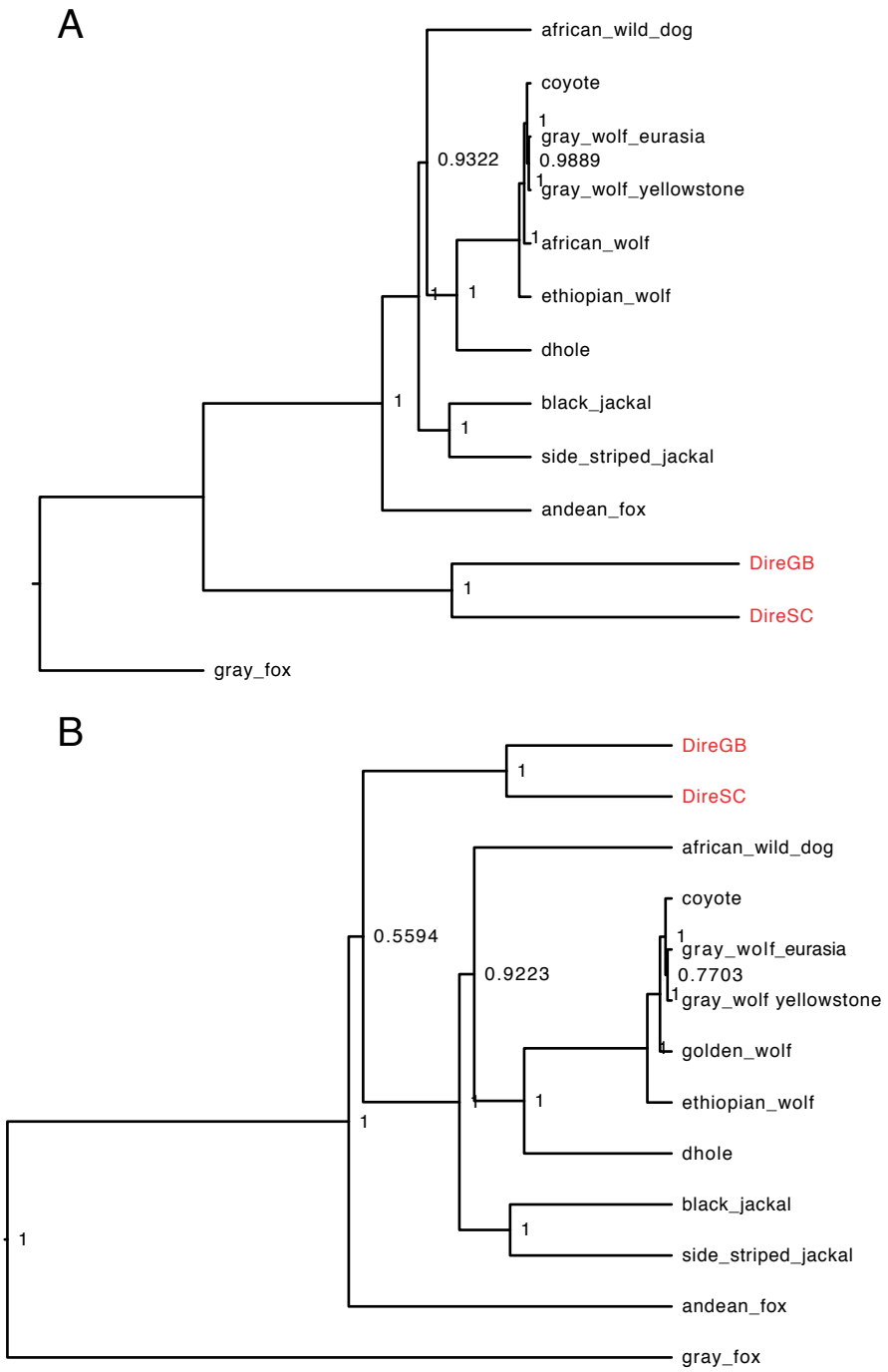


Fig. S13.

Phylogeny based on SNPs, built using SNAPP. A. Without branch correction scheme **B.** With branch correction scheme. Node labels correspond to posterior probabilities. Dire wolves are marked in red.

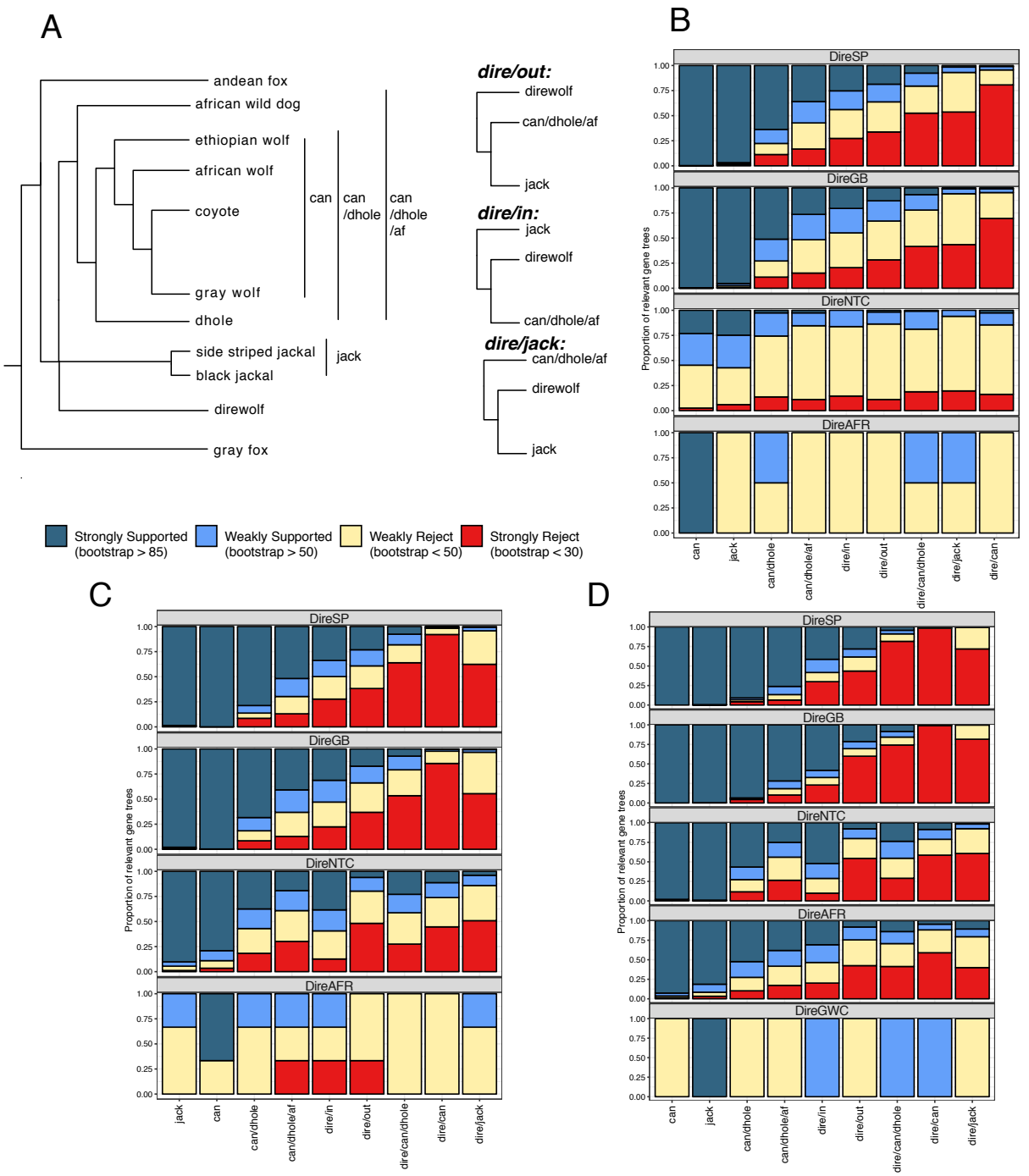


Fig. S14.

Discordance visualisation of maximum likelihood trees build from different bins size. A. Tree defining clades (jack, can, cand/dhole, can/dhole/af) for which we measured support. In addition, the three possible arrangements of the dire wolf, African jackal, and wolf-like canid lineages are displayed (dire/out, dire/in, dire/jack). **B.** Clades and/or alternative topologies displayed in panel A are listed on the x-axis. The y-axis depicts the proportion of individual trees (made from 500kb bins along the genome) that strongly support (>85% bootstrap support; dark blue), weakly support (⩽85% bootstrap support; light blue), weakly reject (best tree favours

alternative arrangement with ⩽85% bootstrap support; orange), or strongly reject that clade/topology (best tree favours alternative arrangement with >85% bootstrap support; red). **C.** As for panel B, but using 1Mb bins. **D.** As for panel B, but using 5Mb bins.

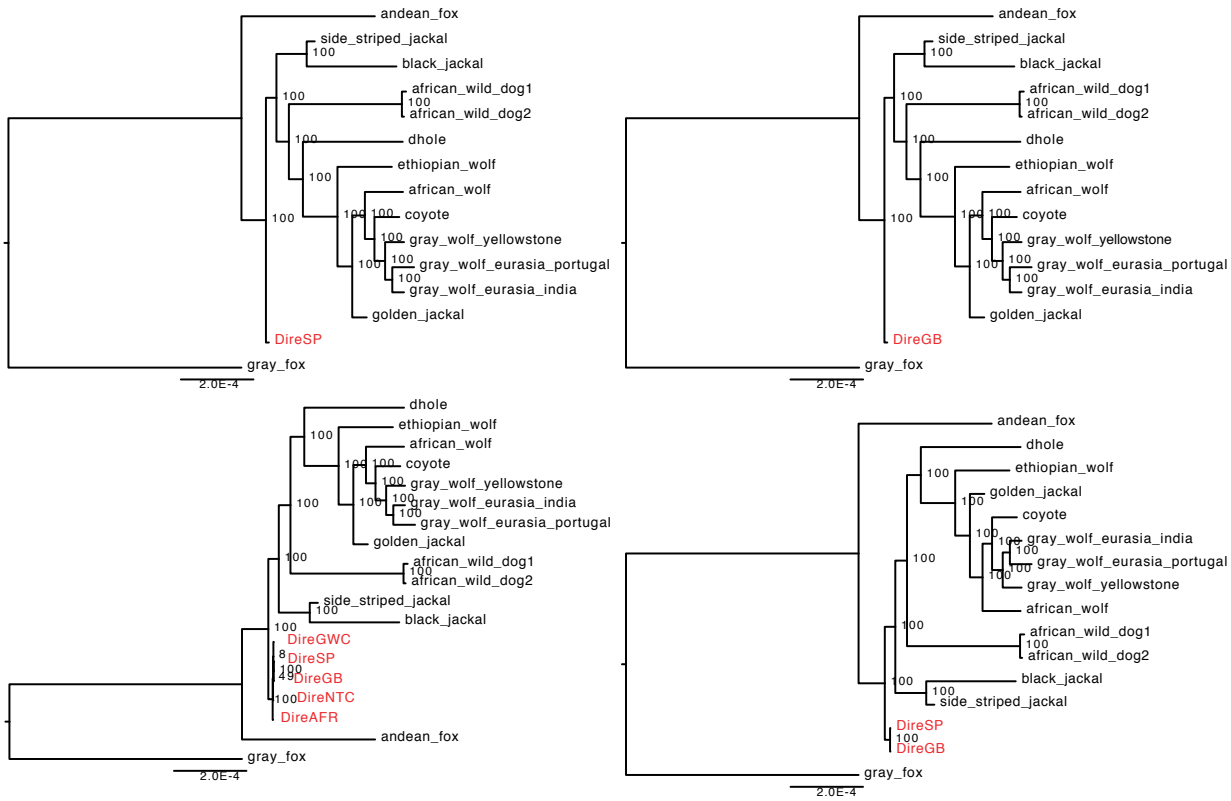


Fig. S15.
Maximum likelihood trees built using ascertainment correction as implemented in RAxML (based on pre-ascertained SNPs in modern canids; see section *supermatrix*). Samples in red represents dire wolves.

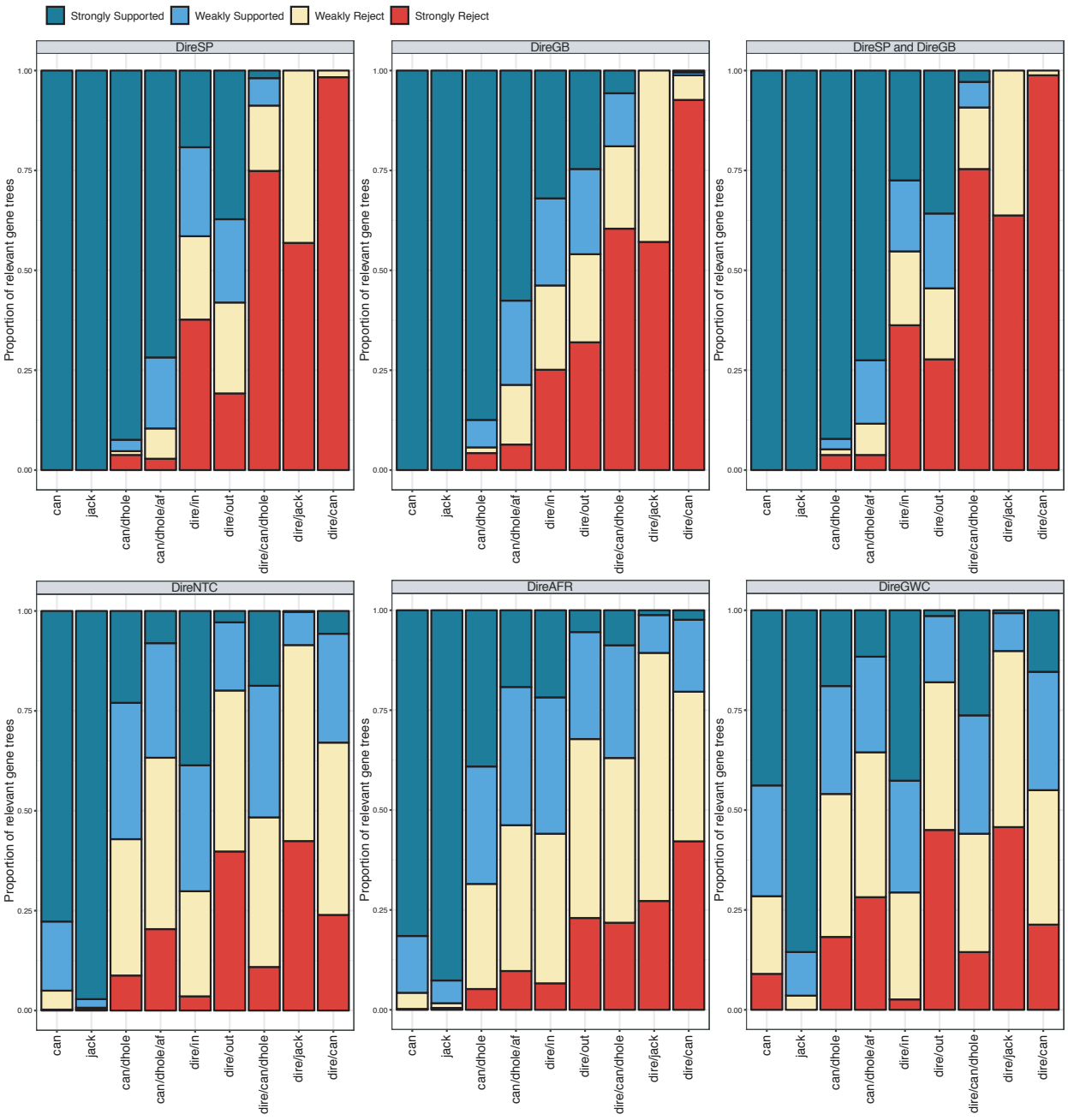


Fig. S16.

Discordance visualisation of maximum likelihood trees build in 5Mb window size, and using the ascertainment ascertainment correction as implemented in RAxML. Clades and/or alternative topologies displayed in Figure S14 are listed on the x-axis. The y-axis depicts the proportion of individual trees that strongly support (>85% bootstrap support; dark blue), weakly support (⩽85% bootstrap support; light blue), weakly reject (best tree favours alternative arrangement with ⩽85% bootstrap support; orange), or strongly reject that clade/topology (best tree favours alternative arrangement with >85% bootstrap support; red).

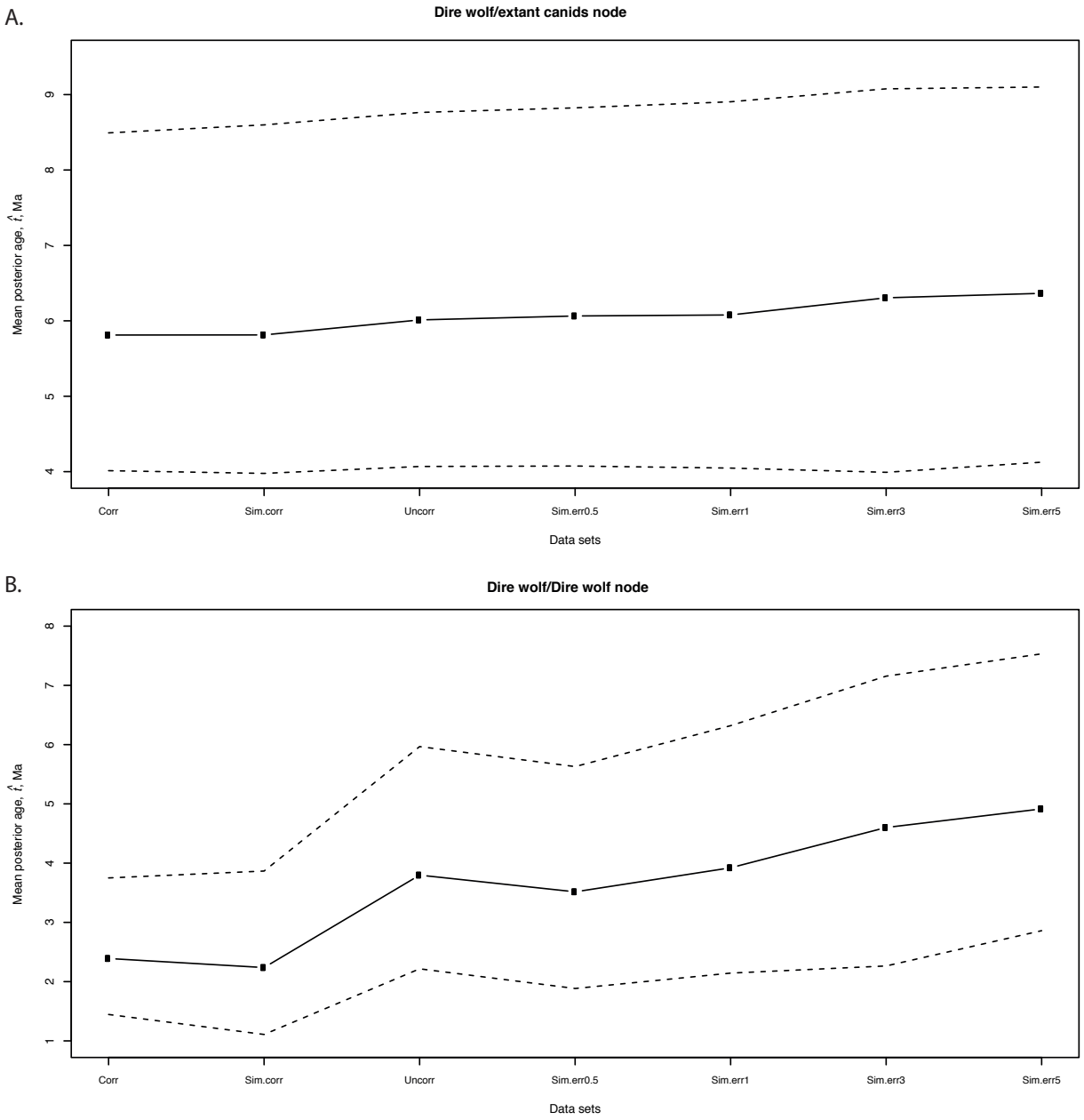


Fig. S17.

Different of age node across different error rate. Age of node (y-axis), under different data sets (x-axis; Corr=corrected data; Sim.corr=simulated data based on corrected tree; Uncorr=uncorrected data; Sim.err0.5-5=simulated with 0.5%-5% error on the dire wolf branch), representing the ancestor of **A.** dire wolf and extant canids **B.** two dire wolves. Dashed line represent 95% HPDI. This figure demonstrates that the error either introduced via simulation (Sim.errX) or due to deamination (Uncorr) on the branch of the dire wolf 1) does not affect the age of the node representing the ancestor of the dire wolf and extant canids (A.) 2) inflates the age of the node representing the ancestor of two dire wolves (B.). We note that the real data (Uncorr) is most similar to simulated data with 0.5% singleton in the dire wolf sequence (Sim.err0.5).

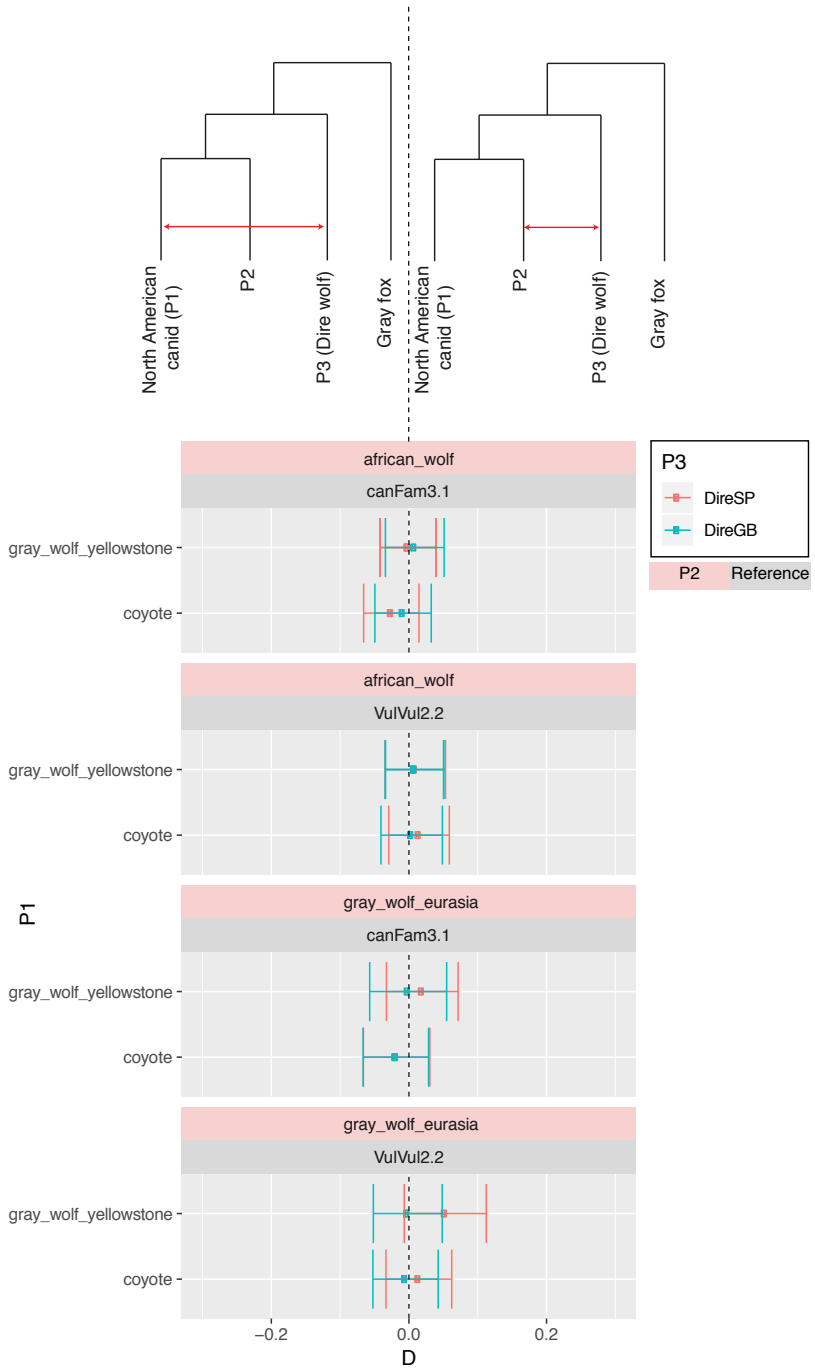


Fig. S18.

Results of D statistics used to assess the possibility of gene flow between the dire wolf and extant North American canids (Yellowstone wolves and coyotes). This figure is similar to the one presented in Figure 2B, but with results using two different genome assemblies. Each dot represents the mean D calculated along the genome and the error bar represents 3 standard deviations. These plots show that the dire wolf genomes do not share significantly more derived alleles with extant North American canids compared to African wolves or Eurasian wolves (values of D not significantly different to zero), suggesting that no hybridization occurred between the dire wolf and the ancestor of extant North American canids.

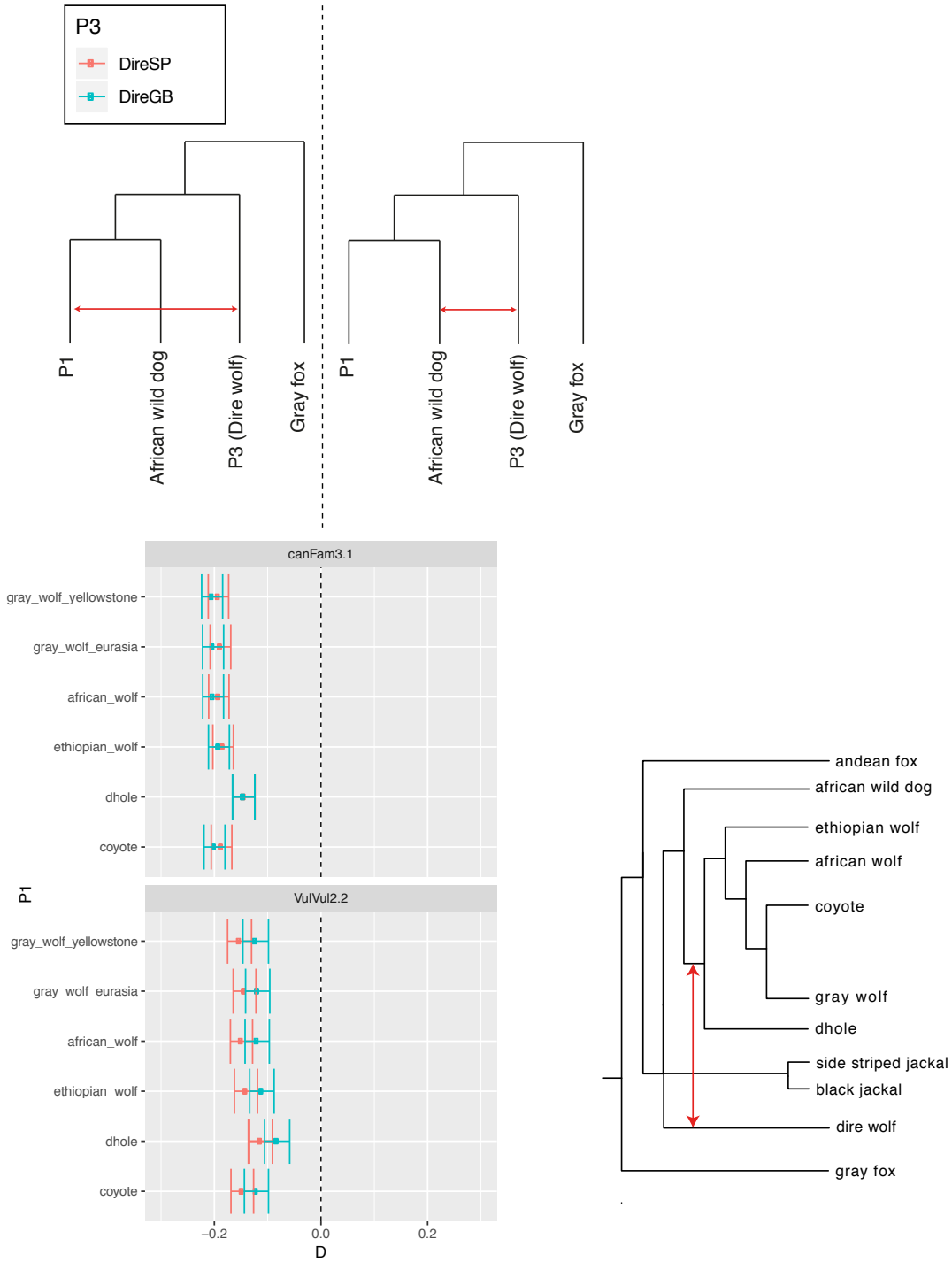


Fig. S19.

Results of the D-statistics of the form D(gray fox; dire wolf (African wild dog, P1)). P1 on the y-axis represent any wolf like canids (coyotes, wolves etc.), or the dhole. Each dot represents the mean D calculated along the genome and the error bar represents 3 standard deviations. The analysis was run on both Dog (canFam3.1) and Red fox assembly (VulVul2.2). This analysis suggests that there was an admixture event between the dire wolf lineage and the ancestor of the dhole, wolves and coyotes.

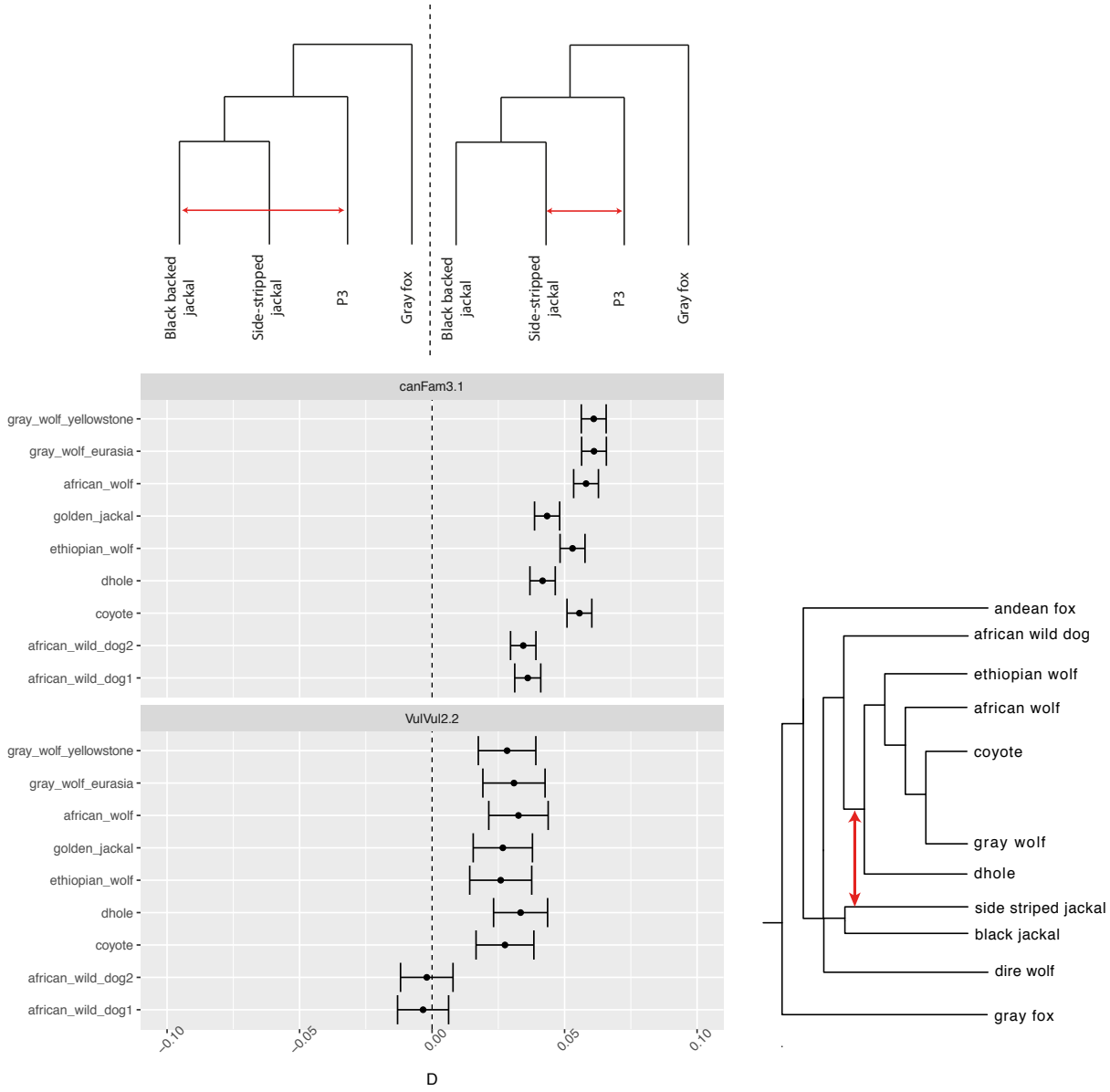


Fig. S20.
Results of the D-statistics of the form D(gray fox; P3(side-striped jackal, black-backed jackal)). P3 label can be found on the x-axis. Each dot represents the mean D calculated along the genome and the error bar represents 3 standard deviations. The analysis was run on both Dog (canFam3.1) and Red fox assembly (VulVul2.2). This analysis suggests that there was an admixture event between the ancestor of the wolf like canids (dhole, wolves, ethiopian wolves etc) and the side-striped jackal.

Table S1

Information about ancient samples sequenced in this study including provenance, age (radiocarbon or stratigraphic information), and sequencing statistics (endogeneous content etc.). * represents duplicate samples (available as an excel file online).

Table S2

Locations and age of dire wolf remains in the Americas, obtained from (1, 7, 128–134) (available as an excel file online).

Table S3.

Sample sizes for geometric morphometric datasets

Species	**Mandible**	**Mandible M1**
African wolf	44	43
African hunting dog	10	10
Andean fox	8	8
Black-backed jackal	18	16
Coyote	65	67
Dhole	6	6
Dire wolf	16	8
Ethiopian wolf	2	2
Golden jackal	27	26
Gray fox	10	10
Side-striped jackal	24	24
Gray wolf	607	541

Table S4.

Pairwise CS testing. All p values have been adjusted for multiple comparisons using the FDR method. The upper triangle are first molar results. Lower triangle are mandibular results (available as an excel file online).

Table S5.

Pairwise morphological variance comparisons by species. All p values have been adjusted for multiple comparisons using the FDR method. The upper triangle are first molar results. Lower triangle are mandibular results. The Procrustes variance scores are listed at the side for the first molar and below for the mandible (available as an excel file online).

Table S6.

Pairwise Procrustes ANOVA results by species. All p values have been adjusted for multiple comparisons using the FDR method. The upper triangle are the results from the first molar dataset. The lower triangle are results from the mandibular dataset. The Ethiopian wolf sample size (2) was too small for testing (available as an excel file online).

Table S7.

First molar pairwise CCV results from LDA optimised with reduced dimensionality. Upper triangle are numbers of PCs used to achieve maximum discrimination of balanced groups. Lower triangle contains the maximum CCV result achieved. Ethiopian wolf sample was too small for LDA procedures (available as an excel file online).

Table S8.

Mandibular pairwise CCV results from LDA optimised with reduced dimensionality. Upper triangle are numbers of PCs used to achieve maximum discrimination of balanced groups. Lower triangle contains the maximum CCV result achieved. Ethiopian wolf sample was too small for LDA procedures (available as an excel file online).

Table S9.

Allometry free first molar pairwise CCV results from LDA optimised with reduced dimensionality. Upper triangle are numbers of PCs used to achieve maximum discrimination of balanced groups. Lower triangle contains the maximum CCV result achieved. Ethiopian wolf sample was too small for LDA procedures (available as an excel file online).

Table S10.

Allometry free mandibular pairwise CCV results from LDA optimised with reduced dimensionality. Upper triangle are numbers of PCs used to achieve maximum discrimination of balanced groups. Lower triangle contains the maximum CCV result achieved. Ethiopian wolf sample was too small for LDA procedures (available as an excel file online).

Table S11.

Incongruence scores from full shape. Upper triangle are scores calculated from first molar morphological data. Lower triangle contains incongruence scores calculated from mandibular data (available as an excel file online).

Table S12.

Incongruence scores from allometry free shape. Upper triangle are scores calculated from first molar morphological data. Lower triangle contains incongruence scores calculated from mandibular data (available as an excel file online).

Table S13.

Incongruence scores from combined mandibular and first molar datasets. Upper triangle are scores calculated from allometry free morphological data. Lower triangle incongruence scores are calculated from full shape (available as an excel file online).

Table S14.

Table containing information (coverage, accession etc.) of modern whole genomes used in this study (available as an excel file online).

Table S15.

Species used for the mitochondrial phylogenies and their accession numbers. Species with accession numbers in bold were used as well in the nuclear analyses.

	Accession number
Gray fox (*Urocyon cinereoargenteus*)	**SRS1937014**
Andean fox (*Lycalopex culpaeus*)	**SRS523207**
Yellowstone gray wolf (*Canis lupus*)	**SRS661496**
Eurasia gray wolf (*Canis lupus*)	**SRS661492**
Dog (*Canis lupus familiaris*)	canFam3.1
Arabian wolf (*Canis lupus arabs*)	DQ480507
Coyote (*Canis latrans*)	**SRS661477**
Red wolf (*Canis rufus*)	SRS661493
Great Lake wolf (*Canis lycaon*)	SRS661486
African wolf (*Canis anthus*)	**ERS3334821**
Ethiopian wolf (*Canis simensis*)	**SRS3929738**
Black-backed jackal (*Canis mesomelas*)	**ERS3216353**
African wild dog 1 (*Lycaon pictus*)	**SRR2971425**
African wild dog 2 (*Lycaon pictus*)	**SRR2971441**
Side-striped jackal (*Canis adustus*)	**ERS3216352**
Golden jackal (*Canis aureus*)	**SRS1025419**
Dhole (*Cuon alpinus*)	**SRS3929739**

Table S16.

Amount of missing data (in %) for each direwolf specimen. Genes with 100% of missing data in at least one specimen are in bold.

	DireGB	DireSP	DireNTC	DireGWC	DireAFR
ATP6 (681bp)	77%	62%	97%	63%	**100%**
ATP8 (204bp)	72%	42%	**100%**	42%	75%
COX1 (1545bp)	55%	65%	91%	65%	67%
COX2 (684bp)	37%	24%	68%	4%	68%
COX3 (784bp)	95%	91%	97%	83%	98%
Cytb(1140bp)	76%	83%	91%	60%	97%
ND1(951bp)	70%	78%	91%	71%	96%
ND2(1042bp)	62%	60%	88%	63%	**100%**
ND3 (346bp)	86%	77%	**100%**	74%	71%
ND4 (1378bp)	68%	60%	87%	49%	64%
ND4L (297bp)	74%	59%	87%	37%	**100%**
ND5 (1821bp)	66%	70%	87%	60%	78%
ND6 (528bp)	65%	50%	73%	55%	88%
S12 (954bp)	33%	34%	69%	7%	66%
S16 (1580bp)	38%	35%	74%	11%	64%
% missing data	61.8%	60.2%	85.5%	49.3%	79.7%

Table S17.

Best model for each partition in the mitochondrial genome as inferred using partitionFinder2 (*101*).

Best Model	Sites	Partition names
TRN+I+G	4692	ATP8_pos1, ND2_pos1, CYTB_pos1, ATP8_pos2, S16_pos1, S12_pos1, ND5_pos1, ND4_pos1, ATP6_pos1
HKY+I+G	3733	COX1_pos2, COX2_pos2, RCND6_pos2, COX3_pos2, ATP6_pos2, ND1_pos2, ND4_pos2, CYTB_pos2, ND5_pos2, ND2_pos2, ND4L_pos2, ND3_pos2
TIM+I+G	3110	ND1_pos3, CYTB_pos3, ND2_pos3, COX3_pos3, ATP6_pos3, COX2_pos3, ND4_pos3, ND5_pos3, ATP8_pos3, ND4L_pos3, ND3_pos3
TRNEF+I+G	1715	RCND6_pos1, ND3_pos1, COX1_pos1, ND4L_pos1, COX2_pos1, COX3_pos1, ND1_pos1
TRN+G	515	COX1_pos3
HKY+G	176	RCND6_pos3

Table S18.

Total size of alignment, number of loci and correction factor used in the supermatrix analysis using RAxML(*116*) (Figure S3).

Sample	Total Size	Number of loci	Correction factor
DireAFR (AJ66)	139817	1395	0.77
DireGWC (RW001)	67487	815	0.803
DireSP (ACAD1735)	28646071	477919	0.728
DireGB (ACAD18742)	28593900	595622	0.78
Combined (DireSP+DireGB)	412476	7295	NA

Table S19.

Number of bins, and average number of coverage for different bin size used in the DISCOVISTA analysis presented in Figure S6.

Sample	Average number of base covered per bin	Total coverage	Total number of bins with coverage	Total number of bins with coverage >2000bp	Bin size
DireAFR	154	901979	5839	2	500kb
DireAFR	363	946090	2606	4	1Mb
DireAFR	2063	1238090	600	305	5Mb
DireGWC	47	272110	5839	0	500kb
DireGWC	123	321804	2606	0	1Mb
DireGWC	824	494514	600	4	5Mb
DireSP	40367	235704849	5839	5517	500kb
DireSP	80380	209469949	2606	2463	1Mb
DireSP	390380	234228275	600	569	5Mb
DireGB	42599	248734410	5839	5516	500kb
DireGB	84820	221040684	2606	2465	1Mb
DireGB	411956	247173801	600	569	5Mb
DireNTC	1159	6770271	5839	134	500kb
DireNTC	2335	6084926	2606	1909	1Mb
DireNTC	11440	6864245	600	558	5Mb

Table S20.

Bayesian model selection of clock and tree topology for the molecular data sets.

Data[a]	**Model**[b]	**log m*L* ± S.E**[c]	**Pr (CI-2.5%, CI-97.5%)**[d]	**Bayes factor**
ALN1 (1 dw)	**Topo A + GBM + BD prior**	**-176267.70 ± 0.04**	**0.942 (0.935, 0.948)**	5.58
	Topo A + ILN + BD prior	-176270.50 ± 0.04	0.058 (0.052, 0.065)	
	Topo B + GBM + BD prior	-176282.10 ± 0.04	$5.42 \cdot 10^{-7}$ ($4.85 \cdot 10^{-7}$, $6.03 \cdot 10^{-7}$)	
	Topo B + ILN + BD prior	-176284.30 ± 0.05	$5.55 \cdot 10^{-8}$ ($4.92 \cdot 10^{-8}$, $6.22 \cdot 10^{-8}$)	
ALN2 (2 dws)	**Topo A + GBM + BD (priorD)**	**-178325.40 ± 0.05**	**0.969 (0.965, 0.972)**	6.87
	Topo A + ILN + BD prior	-178328.90 ± 0.05	0.031 (0.028, 0.035)	
	Topo B + GBM + BD prior	-178341.30 ± 0.05	$1.22 \cdot 10^{-7}$ ($1.08 \cdot 10^{-7}$, $1.37 \cdot 10^{-7}$)	
	Topo B + ILN + BD prior	-178344.70 ± 0.05	$3.99 \cdot 10^{-9}$ ($3.51 \cdot 10^{-9}$, $4.50 \cdot 10^{-9}$)	

[a] ALN*X*: molecular alignments with one direwolf (ALN1) or two direwolf specimens (ALN2).

[b] Topo A: tree topology that supports "dire/out" (Figure S6), Topo B: tree topology that supports "dire/in" (Figure S6), GBM: autocorrelated-rates model, ILN: independent log-normal rates model, BD: birth-death prior.

[c] log mL ± S.E : log marginal likelihood ± standard error; calculated with the mcmc3r R package.

[d] Pr (CI-2.5%, CI-97.5%): posterior probabilities for the model and the corresponding equal trail bootstrap confidence intervals; calculated with the mcmc3r R package.

Table S21.

Priors used for the Bayesian divergence times estimation and Bayesian model selection analyses.

Analysis	Prior on rates[a]	Prior on root age[b]	Prior on Andean fox – Canini split
Divergence times	Γ(2,3040)	U(10.3, 20.0)	U(4.9, 10.3)
Bayes factors	Γ(2,200)	U(0.999, 1.001)	U(4.9, 10.3)

[a] The prior on the rates used is a diffuse gamma distribution regardless of the analysis and the data. Therefore, the shape parameter is always $\alpha = 2$. The scale parameter β changes according to the analysis performed. For divergence times estimation analyses with both corrected and uncorrected molecular alignments, $b_{root\text{-}tips} = 0.0093 \sim 0.01$ and $t_{root} = 15.2$ Ma, thus $\beta = (\alpha \times t_{root}) / b_{root\text{-}tips} = (2 \times 15.2) / 0.01 = 3040$. For Bayes factors analyses, we fix $t_{root} = 1$ Ma because we are not interested in estimating absolute divergence times, only in selecting the best fitting-model. Therefore, $\beta = (2 \times 1) / 0.01 = 200$.

[b] The prior on the root age is required when using MCMCtree. For divergence time estimation, we use a fossil calibration on the root ranging from the earlier part of the Miocene until the end of the Clarendonian, U(10.3,20.0). For the Bayes factors analyses, we cannot use a fossil calibration on the root as we are not interested in obtaining absolute divergence times. Therefore, as $t_{root} = 1$ Ma, we construct a narrow uniform distribution with soft bounds U(0.999,1.001).

Table S22.

Estimated mean divergence times of four external nodes for the molecular alignments. Using the model selected according to the Bayes factors analysis: autocorrelated-rates model and topology "A". Results for both real and simulated molecular data have been included.

Tree node + Data type	Estimated divergence times for the alignment without direwolf specimens	Estimated divergence times for the alignment with 1 direwolf specimen	Estimated divergence times for the alignment with 2 direwolf specimens
Root.corr	**16.277 (10.485,22.796)**	**15.812 (10.327,22.674)**	**15.977 (10.397,22.723)**
Root.simcorr	-	15.781 (10.381,22.595)	15.898 (10.326,22.764)
Root.uncorr	15.982 (10.381,22.763)	15.811 (10.277,22.733)	16.067 (10.323,22.832)
Root.err0.5	-	15.902 (10.275,22.760)	15.592 (10.219,22.725)
Root.err1	-	15.582 (10.249,22.687)	15.685 (10.244,22.740)
Root.err3	-	15.127 (10.193,22.472)	15.483 (10.212,22.721)
Root.err5	-	15.881 (10.512,22.674)	15.243 (10.190,22.594)
Vulpini-Canini.corr	**6.925 (4.853,10.024)**	**6.629 (4.801,9.860)**	**6.917 (4.852,10.029)**
Vulpini-Canini.simcorr	-	6.583 (4.814,9.714)	6.845 (4.814,10.006)
Vulpini-Canini.uncorr	6.646 (4.795,9.793)	6.892 (4.806,10.086)	7.013 (4.848,10.064)
Vulpini-Canini.err0.5	-	7.095 (4.818,10.139)	7.079 (4.861,10.131)
Vulpini-Canini.err1	-	6.779 (4.824,10.091)	7.073 (4.853,10.174)
Vulpini-Canini.err3	-	7.485 (4.887,10.255)	7.273 (4.861,10.202)
Vulpini-Canini.err5	-	7.151 (4.855,10.151)	7.332 (4.879,10.179)

Dw-(jk+rest).corr	**-**	**5.573 (3.958,8.328)**	**5.811 (4.013,8.492)**
Dw-(jk+rest).simcorr	-	5.487 (3.945,8.128)	5.813 (3.977,8.598)
Dw-(jk+rest).uncorr	-	5.971 (4.046,8.878)	6.012 (4.069,8.763)
Dw-(jk+rest).err0.5	-	5.969 (3.908,8.686)	6.064 (4.075,8.824)
Dw-(jk+rest).err1	-	5.710 (3.922,8.674)	6.078 (4.047,8.905)
Dw-(jk+rest).err3	-	6.376 (4.047,9.038)	6.305 (3.991,9.076)
Dw-(jk+rest).err5	-	6.139 (4.058,8.875)	6.364 (4.125,9.101)
jk-rest.corr	**5.283 (3.617,7.746)**	**4.980 (3.505,7.471)**	**5.203 (3.571,7.624)**
jk-rest.simcorr	-	4.912 (3.513,7.263)	5.205 (3.513,7.746)
jk-rest.uncorr	5.054 (3.503,7.542)	5.472 (3.654,8.194)	5.451 (3.643,8.025)
jk-rest.err0.5	-	5.417 (3.495,7.941)	5.485 (3.624,8.045)
jk-rest.err1	-	5.202 (3.503,7.973)	5.524 (3.611,8.190)
jk-rest.err3	-	5.854 (3.667,8.410)	5.735 (3.455,8.362)
jk-rest.err5	-	5.622 (3.620,8.235)	5.810 (3.674,8.501)
dw1-dw2.corr	-	-	**2.390 (1.446,3.750)**
dw1-dw2.simcorr	-	-	2.238 (1.109,3.866)
dw1-dw2.uncorr	-	-	3.798 (2.216,5.967)
dw1-dw2.err0.5	-	-	3.516 (1.885,5.630)
dw1-dw2.err1	-	-	3.921 (2.144,6.320)
dw1-dw2.err3	-	-	4.601 (2.265,7.155)

dw1-dw2.err5	-	-	4.915 (2.860,7.532)

*- Abbreviations: **corr**: corrected alignment (real data), **uncorr**: uncorrected alignment (real data), **simcorr**: simulated alignment using the corrected alignment with real data, **err0.5**: simulated alignment with 0.5% added error, **err1**: simulated alignment with 1% added error, **err3**: simulated alignment with 3% added error, **err5**: simulated alignment with 5% added error, Vulpini-Canini: Divergence between Vulpini and Canini, **Dw-(jk+rest)**: Divergence between the direwolf and the clade with the split between jackals and the rest of canids, **jk-rest**: Divergence between the jackals and the rest of canids. **dw1-dw2**: Divergence between the two dire wolf specimens.

Chapter 11, "Specialised sledge dogs aided Inuit dispersal across the North American Arctic"

Summary

Different dog populations may have arrived into the Americas alongside successive waves of migrating humans. In this paper, using dog remains that predate European contact, we provide evidence for a secondary dispersal of dogs into the Americas. Genetic and morphometric analyses reveal that a specific genetic lineage of dogs possessing a unique morphology accompanied Inuit migrants into and across the North American Arctic, and whose modern descendants are indigenous Arctic sledge dogs. Dogs were, and are, both technologically and culturally important in the Arctic and the arrival of unique dogs with the Inuit provides evidence for their role in human migrations throughout the region over the last 2,000 years.

Authors

Carly Ameen, Tatiana Feuerborn, Sarah K. Brown, Anna Linderholm, Ardern Hulme-Beamanb, Ophélie Lebrasseur, Mikkel-Holger S. Sinding, Zachary T. Lounsberry, ***Audrey T. Lin****, Martin Appelt, Lutz Bachmann, Kate Britton, Matthew Betts, John Darwent , Rune Dietz, Merete Fredholm, Shyam Gopalakrishnan, Olga I. Goriunova, Bjarne Grønnow, James Haile, Jón Hallsteinn Hallsson, Ramona Harrison, Mads Peter Heide Jørgensen, Rick Knecht, Robert J. Losey, Edouard Masson-MacLean, Thomas H. McGovern, Ellen McManus-Fry, Morten Meldgaard, Åslaug Midta, Madonna L. Moss, Iurii G. Nikitin, Tatiana Nomokonova, Albína Hulda Pálsdóttir, Angela Perri, Aleksandr N. Popov, Lisa Rankin, Joshua D. Reuther, Mikhail Sabline, Anne Lisbeth Schmidt, Scott Shirar, Konrad Smirarowski, Christian Sonne, Mary C. Stiner, Mitya Vasyukov, Gro Birgit Ween, Sanne Eline Wennerberg, Catherine F. West, Øystein Wiig, James Woollett, Love Dalén, Anders J. Hansen, M Thomas P Gilbert, Benjamin Sacks, Laurent Frantz, Greger Larson, Keith Dobney, Christyann M. Darwent, and Allowen Evin.*

Status

Published (2019) Proceedings of the Royal Society B.
doi:10.1098/rspb.2019.1929

My contribution

For this paper, I partitioned the mitochondrial genomes for the BEAST2 analysis. I also selected the parameters and made the xml file for the Bayesian skyline analysis in BEAST2. The effective population size was inferred from the Bayesian Skyline plots (Figure S3 & S9) generated from these BEAST2 analyses. The Bayesian tree dates the origin of the so-called X-clade to ~5,300 years BP. The result of this work can be seen in Figure S7-S8.

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Ameen C *et al*. 2019 Specialized sledge dogs accompanied Inuit dispersal across the North American Arctic. *Proc. R. Soc. B* **286**: 20191929. http://dx.doi.org/10.1098/rspb.2019.1929

Received: 19 August 2019
Accepted: 5 November 2019

Subject Category:
Palaeobiology

Subject Areas:
evolution, genomics, palaeontology

Keywords:
archaeology, geometric morphometrics, ancient DNA, migration, *Canis lupus familiaris*, circumpolar

Author for correspondence:
Carly Ameen
e-mail: c.ameen@exeter.ac.uk

†C.A, T.R.F., S.K.B. and A.L. contributed equally to this study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4740998.

THE ROYAL SOCIETY PUBLISHING

Specialized sledge dogs accompanied Inuit dispersal across the North American Arctic

Carly Ameen[1,2,†], Tatiana R. Feuerborn[3,4,6,7,8,†], Sarah K. Brown[9,11,12,†], Anna Linderholm[13,14,†], Ardern Hulme-Beaman[2,14,17], Ophélie Lebrasseur[2,14,18], Mikkel-Holger S. Sinding[5,6,20,23], Zachary T. Lounsberry[11], Audrey T. Lin[14,16], Martin Appelt[24], Lutz Bachmann[20], Matthew Betts[25,26], Kate Britton[27,28], John Darwent[9], Rune Dietz[29,31], Merete Fredholm[19], Shyam Gopalakrishnan[4,5], Olga I. Goriunova[32], Bjarne Grønnow[24], James Haile[14], Jón Hallsteinn Hallsson[33], Ramona Harrison[34], Mads Peter Heide-Jørgensen[35], Rick Knecht[27], Robert J. Losey[36], Edouard Masson-MacLean[27], Thomas H. McGovern[37,38], Ellen McManus-Fry[27], Morten Meldgaard[4,6], Åslaug Midtdal[39], Madonna L. Moss[40], Iurii G. Nikitin[41], Tatiana Nomokonova[42], Albína Hulda Pálsdóttir[21,33], Angela Perri[43], Aleksandr N. Popov[41], Lisa Rankin[44], Joshua D. Reuther[45], Mikhail Sablin[46], Anne Lisbeth Schmidt[24], Scott Shirar[45], Konrad Smiarowski[38,47], Christian Sonne[29,30,48], Mary C. Stiner[49], Mitya Vasyukov[50], Catherine F. West[51], Gro Birgit Ween[22], Sanne Eline Wennerberg[52], Øystein Wiig[20], James Woollett[53], Love Dalén[7,8], Anders J. Hansen[4,6], M. Thomas P. Gilbert[5,54], Benjamin N. Sacks[10,11], Laurent Frantz[55], Greger Larson[14,15], Keith Dobney[2,27,56], Christyann M. Darwent[9] and Allowen Evin[57]

[1]Department of Archaeology, University of Exeter, Exeter, Devon, UK
[2]Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, Merseyside, UK
[3]Department of Archaeology and Classical Studies, Stockholm University, Stockholm, Sweden
[4]Centre for GeoGenetics, and [5]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[6]The Qimmeq Project, University of Greenland, Nuussuaq, Greenland
[7]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
[8]Centre for Palaeogenetics, Stockholm, Sweden
[9]Department of Anthropology, [10]Department of Population Health and Reproduction, and [11]Mammalian Ecology and Conservation Unit of the Veterinary Genetics Laboratory, University of California Davis, Davis, CA, USA
[12]Washington Department of Fish and Wildlife, Olympia, WA, USA
[13]Department of Anthropology, Texas A&M University, College Station, TX, USA
[14]The Palaeogenomics and Bio-archaeology Research Network, Research Laboratory for Archaeology and History of Art, [15]School of Archaeology, and [16]Department of Zoology, University of Oxford, Oxford, UK
[17]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK
[18]GCRF One Health Regional Network for the Horn of Africa (HORN) Project, Institute of Infection and Global Health, Liverpool, UK
[19]Department of Veterinary and Animal Sciences, and [20]Natural History Museum, and [21]Centre for Ecological and Evolutionary Synthesis (CEES) Department of Biosciences, University of Oslo, Oslo, Norway
[22]University of Oslo Museum of Cultural History, Oslo, Norway
[23]Greenland Institute of Natural Resources, Nuuk, Greenland
[24]National Museum of Denmark, Copenhagen, Denmark
[25]Canadian Museum of History, Gatineau, Quebec, Canada
[26]Department of Anthropology, University of New Brunswick, Fredericton, New Brunswick, Canada
[27]Department of Archaeology, University of Aberdeen, Aberdeen, UK

© 2019 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

[28]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Sachsen, Germany
[29]Arctic Research Centre, and [30]Department of Bioscience, Aarhus Universitet, Aarhus, Denmark
[31]Department of Bioscience Roskilde, Aarhus Universitet, Roskilde, Denmark
[32]Laboratory of Archaeology and Paleoecology of the Institute of Archaeology and Ethnography (Siberian Branch of Russian Academy of Science), Irkutsk, Russian Federation
[33]Faculty of Agricultural and Environmental Sciences, The Agricultural University of Iceland, Reykjavik, Iceland
[34]Department of Archaeology, History, Cultural Studies, and Religion, University of Bergen, Bergen, Hordaland, Norway
[35]Birds and Mammals, Greenland Institute of Natural Resources, Copenhagen K, Denmark
[36]Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada
[37]Department of Anthropology, Hunter College CUNY, New York, NY, USA
[38]The Graduate Center, City University of New York, New York, NY, USA
[39]Holmenkollen Ski Museum, Oslo, Norway
[40]Department of Anthropology, University of Oregon, Eugene, OR, USA
[41]Museum of Archaeology and Ethnography at the Institute of History, Archaeology and Ethnography of the Peoples of the Far East (Far Eastern Branch of Russian Academy of Science), Vladivostok, Russian Federation
[42]Department of Archaeology and Anthropology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[43]Department of Archaeology, Durham University, Durham, UK
[44]Department of Archaeology, Memorial University of Newfoundland, St John's, Canada
[45]Department of Anthropology, University of Alaska Museum of the North, Fairbanks, AK, USA
[46]Zoological Institute of Russian Academy of Sciences, St Petersburg, Russian Federation
[47]Section for Cultural Heritage Management, Department of Cultural History, University Museum of Bergen, Bergen, Norway
[48]School of Forestry, Henan Agricultural University, Zhengzhou, China
[49]School of Anthropology, University of Arizona, Tucson, AZ, USA
[50]Department of Biological Diversity and Sustainable Use of Biological Resources, Russian Academy of Sciences, Moskow, Russian Federation
[51]Department of Anthropology and Archaeology Program, Boston University, Boston, MA, USA
[52]Government of Greenland, Veterinary and Food Authority, Nuuk, Greenland
[53]Département des Sciences Historiques, Université Laval, Quebec, Canada
[54]Norwegian University of Science and Technology, University Museum, Trondheim, Norway
[55]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
[56]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia, Canada
[57]Institut des Sciences de l'Evolution–Montpellier, CNRS, Université de Montpellier, IRD, EPHE, Montpellier, Occitanie, France

CA, 0000-0002-4580-2125; AH-B, 0000-0001-8130-9648; OL, 0000-0003-0687-8538; ATL, 0000-0003-2505-1480; SG, 0000-0002-2004-6810; AE, 0000-0003-4515-1649

Domestic dogs have been central to life in the North American Arctic for millennia. The ancestors of the Inuit were the first to introduce the widespread usage of dog sledge transportation technology to the Americas, but whether the Inuit adopted local Palaeo-Inuit dogs or introduced a new dog population to the region remains unknown. To test these hypotheses, we generated mitochondrial DNA and geometric morphometric data of skull and dental elements from a total of 922 North American Arctic dogs and wolves spanning over 4500 years. Our analyses revealed that dogs from Inuit sites dating from 2000 BP possess morphological and genetic signatures that distinguish them from earlier Palaeo-Inuit dogs, and identified a novel mitochondrial clade in eastern Siberia and Alaska. The genetic legacy of these Inuit dogs survives today in modern Arctic sledge dogs despite phenotypic differences between archaeological and modern Arctic dogs. Together, our data reveal that Inuit dogs derive from a secondary pre-contact migration of dogs distinct from Palaeo-Inuit dogs, and probably aided the Inuit expansion across the North American Arctic beginning around 1000 BP.

1. Introduction

Dogs (*Canis lupus familiaris*) played a critical role in early human adaptation to circumpolar high-latitude environments. Early dog specimens from Late Pleistocene to Early Holocene sites throughout northeastern Asia [1–3], bear witness to this early association between humans and dogs in the Arctic. Recent genetic analyses indicate that the earliest dogs found in the Americas belonged to a now extinct lineage of Arctic dog that was introduced from Eurasia at least 10 000 years ago [4]. Aside from this initial peopling of the Americas, the North American Arctic has experienced additional human migration episodes of genetically distinct populations [5–10], which were accompanied by potentially distinct dog populations. The importance of dogs during these migrations, however, remains largely unknown [4,11]. Investigating whether or not these new groups brought genetically differentiated dog populations with them into the North American Arctic, and the relationship between these dogs and those already present in the region, is crucial for understanding the history of dogs in the Americas.

Archaeological evidence suggests that dogs were relatively rare in the North American Arctic prior to the Inuit period [12,13]. The Inuit emergence in Alaska beginning approximately 2000 BP brought large-scale changes in lifeways, subsistence practices and material culture to the North American Arctic. Their subsequent expansion translocated this culture out of Alaska eastward to Greenland, and along the coast of subarctic Eastern Canada starting in 1000 BP [14,15]. The rapid expansion of the Inuit is attributed in part to their exploitation of advanced transportation technologies, including the development and widespread usage of the umiak and kayak for sea travel, and the dog sledge for use on land and ice [16–18]. Despite the ubiquitous association of dog sledging with North American Arctic peoples, and the presence of dog sledge technology in Siberia by 9000 BP [1,19], there remains little firm evidence for dog sledging in the Americas before 1000 BP [12]. Today, sledge dogs remain culturally and economically crucial to indigenous lifeways in the Arctic, but dog numbers are declining rapidly due to changing climate, the recurrence of parvovirus and distemper, the preferential use of snowmobiles, and the culling of indigenous dogs during both historical and modern times [20,21]. Furthermore, the relationship between modern Arctic dogs and archaeological dogs from the Inuit and preceding Palaeo-Inuit periods remains unclear.

Previous studies have suggested that dogs associated with modern Arctic populations are the direct descendants of the populations that were brought by the Inuit [22–24], reaffirming the intrinsic technological role of dogs for modern Inuit who continue to occupy the North American Arctic. However, more recent introductions of dogs attributed to the European settlement of the Eastern Arctic beginning in the eighteenth century, and Alaskan settlers at various times in the nineteenth century, largely in connection to mining and settlement [22,23], probably also contributed to the genetic make-up of modern

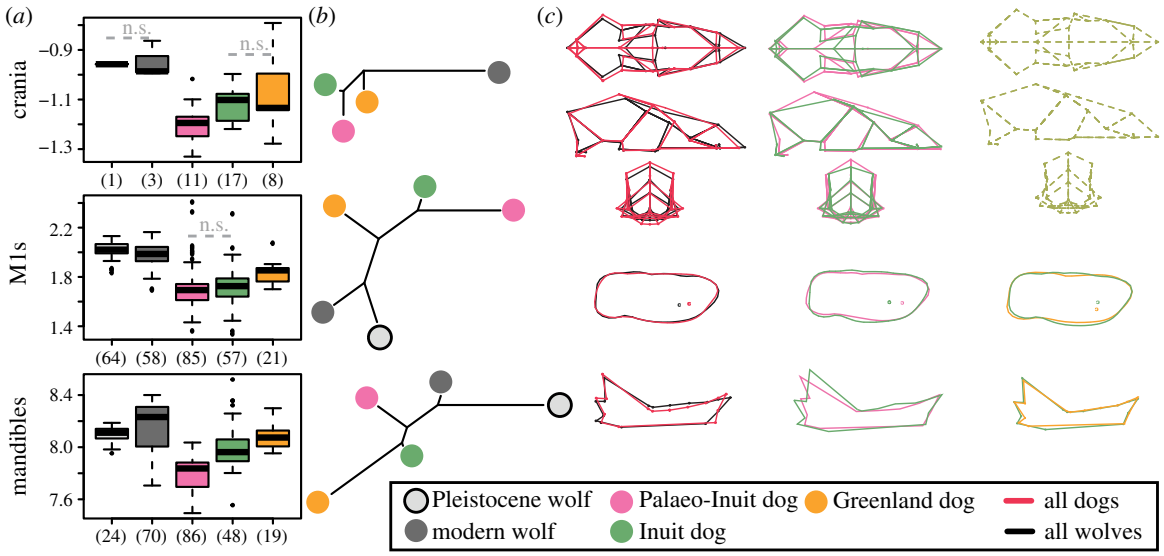


Figure 1. Morphometric variation of Arctic dogs and wolves. (*a*) Size variation of Pleistocene and Modern wolves, and Palaeo-Inuit, Inuit, historic and recent Greenland (Historic and modern Greenland breeds, see electronic supplementary material, text) dogs. Boxplot of the log-transformed centroid size with sample size shown in brackets. 'n.s.' highlight non-significant pairwise comparison (Wilcoxon's test) between neighboring groups (table 2). (*b*) Overall shape differentiation between groups shown as neighbour-joining networks derived from Mahalanobis distances for each element separately. (*c*) Visualization of the cranial (top), first lower molar (middle) and mandible (bottom) shape differences between: wolves (black) and all domestic dogs (red); Palaeo-Inuit dogs (pink) and Inuit dogs (green); and Greenland dogs (orange), and Inuit dogs (green). Shape differences are visualized along the discriminant axis between the groups. Wireframes with dashed lines indicate non-significant differences.

dog populations across the region. In order to disentangle the spatio-temporal patterns in past North American Arctic dog populations and their relationship to recent Arctic dogs, it is necessary to establish the morphological and genetic distinctiveness of these groups.

To do so, we examined the phenotypic changes associated with the arrival of Inuit dogs using geometric morphometric (GMM) analyses [25] to explore cranial, lower first molar and mandibular morphometric variation. We then analysed complete and partial mitochondrial genomes of pre-contact Arctic dogs alongside publicly available data for pre-contact dogs from across the Americas [4,26], as well as historical and modern Arctic sledge dogs [22,23]. Integrating these datasets allowed us to investigate the history of pre-contact dogs in the North American Arctic, contextualize the origins of Inuit dogs relative to earlier Palaeo-Inuit dogs, and clarify the relationship between these archaeological dogs and modern Arctic dog populations.

2. Results and discussion

(a) The origins and legacy of pre-contact Arctic dogs

To better understand the diversity of ancient Arctic dog morphologies, we investigated the phenotypic variation between Palaeo-Inuit, Inuit and more recent Greenland dog (historical Arctic and modern Greenland breed, see electronic supplementary material, text), as well as Arctic wolf populations. To do so, we used GMM on the crania, lower first molars and hemi-mandibles (figure 1 and table 1; electronic supplementary material, figure S14). Overall, wolves were easily distinguishable from dogs (table 2). In particular, wolves were less morphologically diverse in lower M1 size and shape, and possessed consistently larger measurements, and narrower and lower braincases relative to our sampled dogs (figure 1).

Analyses of the dogs found that Inuit dogs differ from both Palaeo-Inuit and recent Greenland dogs (figure 1 and table 2). Inuit dogs tended to be larger than Palaeo-Inuit dogs, but smaller than recent Greenland dogs with no allometric repatterning (except for the mandible between Palaeo-Inuit and Inuit; figure 1*a* and table 2). Comparatively, Inuit dogs possess a proportionally narrower cranium, a less elevated braincase, a wider lower M1 and a more developed mandibular ascending ramus than Palaeo-Inuit dogs (figure 1*c*). Where the differences are significant, the mean cross-validation between Palaeo-Inuit and Inuit dogs ranges from 71.0% to 83.8% depending on the element (table 2).

We detected no differences in the crania between the Inuit and recent Greenland dogs (table 2). However, Inuit and Greenland dogs did differ in the size, shape and form of their lower M1 and mandible (mean cross-validation ranging from 63.5% to 85.0%), with no change in size-shape relationship (allometries) nor variance. Compared to historical and modern Arctic dogs, Inuit dogs showed a proportionally wider molar and a more convex bend to the mandibular body (figure 1*c*). Inuit dogs exhibit a similar cranial shape to the recent Arctic dogs while differing from the Palaeo-Inuit dogs, though the Inuit dogs differ equivalently from both Palaeo-Inuit and recent Arctic dogs in their molar and mandible (table 2).

Given that the GMM analyses suggested that the dogs associated with the different cultural groups were morphologically divergent, we assessed whether Palaeo-Inuit, Inuit and historical and modern Arctic dogs were also genetically distinct. We obtained mitochondrial genomes from 186 samples with a minimum average coverage of threefold to investigate the mitochondrial diversity of Arctic dogs through time and trace patterns of migration. An additional 40 samples were assigned to specific haplogroups by generating mitochondrial control region sequences using Sanger sequencing (table 1). Globally, modern and ancient domestic dogs group into four major (A–D), and two minor (E–F) mitochondrial clades [4,26,29]. Our phylogenetic analyses revealed that nearly all of the sampled dogs belonged to the mitochondrial A clade (electronic supplementary material, figure S4). We also identified four major subclades within the A clade: A1a, A2a, A1b and

Table 1. Number of samples generated or used in study. (*a*) Morphometric sample size: number of samples per group and per element analysed. (*b*) Genetic sample size: numbers of samples successfully sequenced per group and per type of sequencing/number attempted.

		Palaeo-Inuit	Inuit	historical	modern	modern wolf	Pleistocene wolf	total
(*a*)								
GMM	crania	11	17	6	2	3	1	40
	mandible	86	48	5	14	70	24	247
	lower M1	85	57	6	14	58	64	284
	no. of spec.[a]	124	92	12	16	70	77	391
		Palaeo-Inuit	**Inuit**	**historical**	**modern**	**Siberin Holocene**	**other[b]**	**total**
(*b*)								
DNA	Sanger[c]	23/87	128/261	20/51	9/14	13/30	0/36	193/479
	3 × coverage[d]	12/41	84/197	62/126	14/14	14/27	1/14	199/431
	10 × coverage[d]	2/41	28/197	37/126	14/14	3/27	0/14	94/431
	no. of spec.[a]	92	289	147	24	40	36	628[a]

[a]no. of spec.: number of specimens. Some specimens underwent Sanger and next-generation sequencing or multiple elements from an individual were included in the GMM analyses.

[b]Cultural affiliations including medieval and historical Iceland, Norse and unknown.

[c]D-loop haplotypes obtained via Sanger sequencing/samples sequenced with Sanger sequencing from specified culture.

[d]Mitogenomes obtained with indicated mean coverage/samples sequenced with next-generation sequencing from specified culture.

A2b, three of which had been previously identified in the Arctic (A1b, A2a [23] and A2b [4]) (figure 2). By contrast, most modern European dogs have haplotypes in the A1a clade.

All previously sampled pre-contact dogs from the Americas, including the earliest specimens known from the continent, exclusively carried A2b haplotypes [4,11] (figure 2). The uniformity of this clade in early contexts and the arrival of dogs carrying different haplotypes in later periods suggests that specific dog populations were associated with different cultural group. Thus, our data show that early human migration(s) (prior to 10 000 years ago) into the Americas were associated with dogs carrying only A2b haplotypes, while later migrations (after 5500 years ago) into the North American Arctic introduced dogs carrying haplotypes belonging to A1a, A1b and A2a subclades (figure 2).

Of the 92 specimens that were extracted and sequenced from Palaeo-Inuit contexts, only 12 possessed sufficient DNA to allow them to be assigned to haplogroups. Despite this sample size, these data represent the only currently available genetic information from Palaeo-Inuit dogs. While most Palaeo-Inuit dogs possessed A2b haplotypes (83.3%), two Palaeo-Inuit specimens from Alaska possessed A2a haplotypes (CK-H37-M1, SEL-33-0057b) suggesting that the A2a haplotype was present in North American Arctic prior to the Inuit period (figure 2). We found haplotypes immediately basal to this A2 subclade in Siberia dating to several thousand years prior to the Inuit colonization of the North American Arctic, suggesting that the appearance of these lineages on Palaeo-Inuit sites was the result of the Siberian ancestry of the dogs and people (electronic supplementary material, figure S4). Thus, though the Inuit were likely not responsible for the first appearance of this dog lineage in the Americas, they were responsible for the considerable geographical expansion of this lineage into the Eastern Arctic, where, during the Inuit period, they became the most common haplotypes across the entire North American Arctic (figure 2).

Temporal shifts in mitochondrial haplotype frequencies suggest a near-replacement of Palaeo-Inuit dogs in the North American Arctic coinciding with the Inuit expansion from Siberia (figure 2 and electronic supplementary material, figure S11). Demographic analyses (as reconstructed from Bayesian skyline plots) show apparent founder effects coinciding with the timing of the Inuit expansion into the eastern around 1000 BP (electronic supplementary material, figure S9). The pattern of haplotype frequencies between locations and time periods provide strong evidence that Inuit migrants brought dogs with them from Siberia (figure 2; electronic supplementary material, figure S11). In particular, the frequencies of haplotypes from different subclades differed strongly between the Palaeo-Inuit and Inuit Arctic samples ($F_{ST} = 0.33$, $p < 0.001$), but were similar between Siberian and North American Inuit samples ($F_{ST} = 0.04$, $p > 0.10$; see electronic supplementary material, text and figure S11). These results are consistent with the results of previous human genetic analyses which have linked Inuit groups to both Siberian and earlier Alaskan populations [30,31].

Inuit and historical populations also possessed different haplotype frequencies ($F_{ST} = 0.33$, $p < 0.001$; figure 2). In particular, the A1a haplotypes increased over the past 300 years. A more recent expansion of subclades A1a and A1b compared to those of A2a was further supported by a network analysis and Bayesian skyline plots (see electronic supplementary material, text and figure S12). The European exploration of both Greenland and the Canadian Arctic during the nineteenth and twentieth centuries, and the nineteenth century Alaskan Gold Rush increased the interaction between indigenous Arctic groups and Europeans, facilitating longer distance travel of both dogs and people [24]. This cultural mixing brought Eurasian dogs to the region in large numbers [22], probably contributing to the increase in A1a haplotype frequency. On the other hand, two instances of A1a subclade haplotypes from northern Alaska in the Inuit archaeological sample suggest the possibility of drift as a partial explanation for the increase in frequency of this subclade over the past 300 years. Lethal epidemics in indigenous dogs have also led to large-scale population turnover and replacement by European breeds in many regions [24,32]. While the

Table 2. Pairwise comparisons between dogs and wolves, Palaeo-Inuit and Inuit dogs, and Inuit and recent Greenland dogs. Differences are assessed using Wilcoxon's test for size, MANOVAs for shape and form and MANCOVAs for allometries. Differences in variances are tested following [27] for shape and Fligner–Killeen tests for size. Leave-one-out cross-validation percentages were obtained from 100 linear discriminant analyses based on balanced samples and dimensionality reduction [28] and are presented as the mean and 90.0% confidence interval. Shaded boxes are non-significant results after correction of the *p*-values for multiple comparisons.

		Dogs versus wolves		Palaeo-Inuit dog versus Inuit dog		Inuit dog versus Greenland dog	
		comparison	cross-validation (%)	comparison	cross-validation (%)	comparison	cross-validation (%)
crania	size	$W = 137$, $p = 8 \times 10^{-4}$	91.2 (62.5–100%)	$W = 141$, $p = 0.02$	71% (63.6–77.3%)	$W = 75$, $p = 0.7$	38% (12.2–50%)
	shape	$F_{4,35} = 9.16$, $p = 3.58 \times 10^{-5}$	86.7% (50–100%)	$F_{2,25} = 6.3$, $p = 6 \times 10^{-3}$	75.5% (68.2–86.4%)	$F_{2,22} = 2.3$, $p = 0.1$	56.4% (43.7–68.7%)
	form	$F_{4,35} = 2.89$, $p = 0.036$	85.5% (50–100%)	$F_{3,24} = 3.807$, $p = 0.023$	76.2% (63.6–82%)	$F_{3,21} = 2.03$, $p = 0.14$	54.4% (37.5–75%)
	allometries	$F_{35,2} = 1.19$, $p = 0.56$		$F_{23,2} = 0.9374$, $p = 0.64$		$F_{20,2} = 1.163$, $p = 0.56$	
	size variance	$x^2 = 1.5$, $p = 0.22$		$x^2 = 4 \times 10^{-4}$, $p = 0.98$		$x^2 = 0.49$, $p = 0.48$	
	shape variance	$d_{var} = 0.0012$, $p = 0.10$		$d_{var} = 3 \times 10^{-4}$, $p = 0.57$		$d_{var} = 5 \times 10^{-5}$, $p = 0.91$	
M1	size	$W = 18\,576$, $p < 2.2 \times 10^{-16}$	89.2% (88–90.6%)	$W = 2776$, $p = 0.1418$	55% (37.6–61.4%)	$W = 906$, $p = 5 \times 10^{-4}$	68.9% (62–76%)
	shape	$F_{17,267} = 29.2$, $p < 2 \times 10^{-16}$	91.9% (90.1–93.5%)	$F_{7,134} = 17.4$, $p = 2 \times 10^{-16}$	81.1% (77.2–85.1%)	$F_{5,72} = 12.317$, $p = 2 \times 10^{-8}$	81.4% (75–90%)
	form	$F_{12,272} = 46.6$, $p < 2 \times 10^{-16}$	94.8% (93.8–95.9%)	$F_{8,133} = 15.591$, $p = 4 \times 10^{-16}$	81.2% (78–84.3%)	$F_{8,69} = 10.4$, $p = 8 \times 10^{-8}$	82.9% (76–90%)
	allometries	$F_{98,184} = 1.41$, $p = 0.022$		$F_{98,41} = 0.84$, $p = 0.76$		$F_{73,2} = 1.06$, $p = 0.6$	
	size variance	$x^2 = 16.8$, $p = 4 \times 10^{-5}$ (Dogs > Wolves)		$x^2 = 0.25$, $p = 0.62$		$x^2 = 3.3$, $p = 0.07$	
	shape variance	$d_{var} = 0.0001$, $p = 0.01$ (Dogs > Wolves)		$d_{var} = 7 \times 10^{-5}$, $p = 0.281$		$d_{var} = 4 \times 10^{-5}$, $p = 0.6$	
mandibula	size	$W = 12\,156$, $p < 2 \times 10^{-16}$	79% (77.6–80.3%)	$W = 3435$, $p = 2 \times 10^{-10}$	78.5% (76–81.3%)	$W = 655$, $p = 5 \times 10^{-3}$	63.5% (57.9–71%)
	shape	$F_{14,232} = 22.79$, $p < 2 \times 10^{-16}$	87.2% (85.1–89.4%)	$F_{6,127} = 12.7$, $p = 3 \times 10^{-11}$	78.6% (74–82.3%)	$F_{6,60} = 10.7$, $p = 4 \times 10^{-8}$	85% (76.3–92.1%)
	form	$F_{15,231} = 39.08$, $p < 2 \times 10^{-16}$	93.7% (92.6–95.2%)	$F_{7,126} = 18.104$, $p < 2 \times 10^{-16}$	83.8% (80.2–87.5%)	$F_{5,61} = 12.68$, $p = 2 \times 10^{-8}$	84.2% (78.9–89.5%)
	allometries	$F_{26,218} = 4.52$, $p = 1.8 \times 10^{-10}$		$F_{26,105} = 4.85$, $p = 3 \times 10^{-9}$		$F_{26,38} = 1.42$, $p = 0.15$	
	size variance	$x^2 = 0.04$, $p = 0.83$		$x^2 = 2.94$, $p = 0.08$		$x^2 = 2.68$, $p = 0.101$	
	shape variance	$d_{var} = 0.0004$, $p = 0.026$ (Dogs > Wolves)		$d_{var} = 3 \times 10^{-4}$, $p = 0.202$		$d_{var} = 5 \times 10^{-4}$, $p = 0.178$	

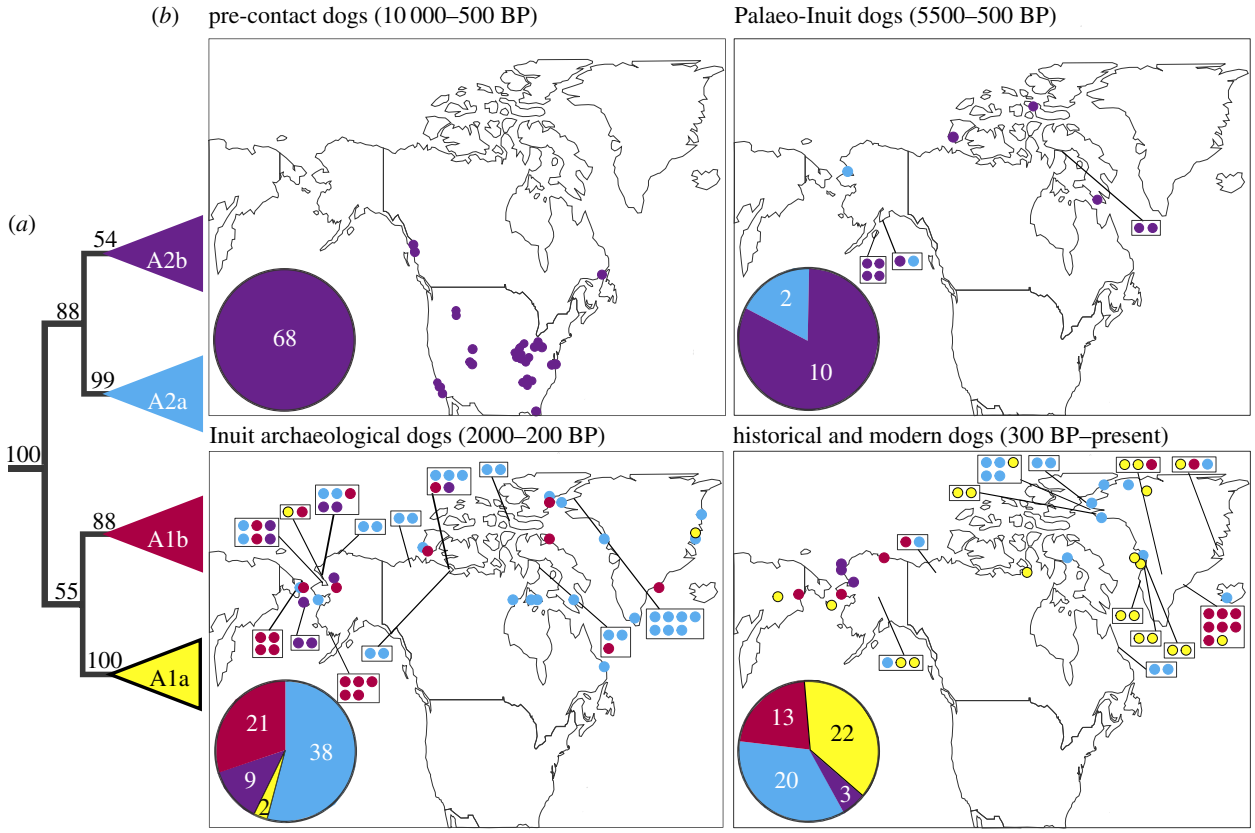


Figure 2. Phylogenetic topology and geographic distribution of haplotypes through time. (*a*) The A-clade mitochondrial haplotypes of dogs inferred by maximum-likelihood analyses depicting the four subclades discussed in the text with their respective bootstrap support (for the whole tree see electronic supplementary material). (*b*) Geographical origin of North American dog samples and cultural affiliation. Pie charts indicate the abundance of subclades. Sites with more than one sample are shown in boxes with representation of sample number and haplotype. Modern samples outside of the North American Arctic were excluded from the map and pie chart. Culture dates represent the earliest and latest appearance of each group in the North American Arctic within this dataset [6]. (Online version in colour.)

Inuit dog lineage have been preserved in both modern Canadian Inuit and Greenland Sledge Dogs, the mitochondrial haplotype diversity of the population has changed. The low frequency of the A1a haplotypes in Inuit dog populations (2.9%) stands in stark contrast to the high frequency of Ala haplotypes in recent Arctic dogs (37%). We observed no A1a haplotypes in Palaeo-Inuit dog populations. Our results show that although modern North American Arctic dogs are descended in large part from Inuit period populations there has been regional European A1a introgression during the historical/modern era contributing to the increase in frequency of A1a.

(b) Canids as tools for the adaptation of humans to Arctic environments

Dogs are an important cultural symbol in the North American Arctic where sledge dogs and dog traction are highly visible components of Arctic identity [33–35]. Dogs were used not only for traction and sledging, but also for hunting, clothing and occasionally as food either preferentially, or during periods of famine [36]. Wolves and other wild canids were additionally exploited for their pelts which protect from frost buildup, or killed as a precaution against potential predation or conflict [37,38].

While most Arctic assemblages primarily consist of domestic canid specimens, there is evidence that wolves were also exploited during the Inuit period. On the basis of size, a large mandible from Nunalleq—a Thule Inuit site in southwestern Alaska [39]—was provisionally identified as a grey wolf (AL2797), alongside a large number of dog specimens [38,40]. Phylogenetic analysis of the mitochondrial genome from this individual places it among modern and historical wolves from Alaska and Canada (electronic supplementary material, figure S3). This identification was further confirmed by multiple analyses based on low coverage nuclear genome obtained from this individual (electronic supplementary material, figure S10). Two additional Alaskan samples (TRF.02.27, TRF.02.28) taken from clothing made of canid pelts held in ethnographic collections are also likely made from wolf pelts on the basis of their mitochondrial genomes (electronic supplementary material, figure S4) indicating that the use of wolf pelts continued into at least the historical period despite easy access to dogs.

Both nineteenth century Arctic explorers and twentieth century anthropologists reported that Arctic groups often encouraged hybridization of their dogs with wolf populations in order to maintain and strengthen their lineages [41–44]. In the wild, asymmetric bias has been seen with evidence for female wolf–male dog hybridization being dominant and only rare instances of male wolf–female dog hybridization being observed [45]. While the nature of female wolf–male dog hybridization would result in the offspring carrying wolf mitochondrial genomes it also would more than likely exclude these offspring from being in a domestic archaeological context. On the other hand, intentional hybridization of dogs and wolves would likely have been biased towards male wolves mating with female dogs through the deliberate picketing of female dogs in oestrus [46]. The sexual asymmetry in wolf hybridization remains unclear, although explanations for the limited male wolf–female dog hybridization linked to

biological and behavioural constraints, such as male wolf aggression towards dogs, social compatibility and fertility cycles [45,47], mitochondrial introgression from wolves into dogs is improbable, and we did not observed it in our mitochondrial dataset. Anecdotes of hybridization between wolves and dogs are nevertheless widespread in Greenland today, and these reports suggest that hybrids often make poor sledge dogs, and that wolf traits are reportedly selected against [44,46]. This, combined with the observed infrequency of gene flow between wolves and dogs over thousands of years [48,49] makes it unlikely (though not impossible) that any similarities observed between Inuit dogs and Arctic wolves is the result of systematic hybridization, and an analysis of nuclear DNA will clarify this. In addition, the larger cranial and mandible size of the Inuit dogs, when compared with those from the Palaeo-Inuit period, would have been an advantage for their role in transportation and traction during this period though the overall robusticity of the dogs is difficult to detect from skull elements alone. More extensive analyses of postcranial elements could further quantify the unique morphology required for prolonged specialization in sledge pulling.

(c) Novel X-clade dogs in Eastern Siberia and Alaska

Phylogenetic analysis also revealed seven canid mitogenomes (03.P04.H1.1024, AL2990, AL2991, AL3004, AL3053, AL3007, TRF.02.29) clustering with modern Eurasian wolves, outside any known domestic dog clades, forming a novel subclade referred to in this study as X clade (electronic supplementary material, figures S2, S4–S6 and S8). Three of the samples in X clade are derived from archaeological contexts: two from Neolithic Siberian canids excavated at the Boisman II site in Khasansky District of Primorsky Krai, Russia (AL2990, AL2991; see electronic supplementary material, table S5 for radiocarbon dates) and one dog from the Birnirk site, Pajpel'gak in Chukotka (03.P04.H1.1024). Additionally, this novel subclade contains four historical dogs from Kamchatka, Chukotka, Bering Island and Alaska. While the origins of the fur from the ethnographic Alaskan clothing (TRF.02.29) cannot be conclusively confirmed as dog based on the sample record alone, the ethnographic samples of aboriginal dogs collected from Kamchatka, Chukotka and Bering Island have a definitive classification, having been collected from dogs during the twentieth century (AL3004, AL3053, AL3007; M. Sablin 2019, personal communication).

While the X-clade clusters with modern grey wolves, the inclusion of several recent dogs with known origin eliminates the possibility of this clade representing grey wolves. Although, as previously discussed, interbreeding between dogs and wolves is generally biased towards the mating of male wolves with female dogs which would not result in the passing of mitochondrial genomes from wolves to dogs. The X clade could represent an ancient introgression event of mitochondrial wolf haplotypes into dogs. Additionally, the ancient origin of the clade is demonstrated by sample AL2991 (BOIS 5), which was directly dated to between 6660 and 6495 cal. BP, falling within the clade. This ancient origin is corroborated by Bayesian tree, generated with BEAST2.4 and calibrated with dated samples, estimating the origin of the clade between 5300 and 10 000 years BP (electronic supplementary material, figure S8). Due to constraints on coverage used in the Bayesian analysis sample AL2991, the oldest dated sample in the clade, was not included in the analysis. The inclusion of AL2991 may have pushed the age of the clade backwards as AL2991 lies well within rather than basal to the clade and is a few millenia older than the samples included in the Bayesian analysis. Even AL2991 included in the Bayesian analysis the sample age falls within the estimated window of origin for the X clade. The comparatively recent origin of this clade, compared to some of the more prominent mitochondrial clades, suggests that the X clade was not present in dogs at the time of domestication. The X clade lineage also appears in historical dogs, demonstrating the continuation of the lineage until at least 75 years ago. The single individual carrying an X clade mitochondrial genome in Alaska may indicate a relatively recent introduction of the lineage to the North American Arctic after the arrival of the Inuit, explaining why the lineage did not spread across the North American Arctic during the Inuit migration and expansion like A1a and A1b haplotypes. The absence of the X-clade haplotypes in published studies speaks to the apparent low frequency and restricted distribution of these haplotypes to far eastern Siberia and Alaska. Furthermore, the absence of these haplotypes in modern dogs may reflect the lack of systematic sampling in the region to date, the low frequency or the disappearance of these haplotypes.

3. Conclusion

The phenotypic and genetic data presented here suggest that a novel dog population that was morphologically divergent from, and genetically more diverse than earlier Palaeo-Inuit dogs, accompanied Inuit migrants into and across the North American Arctic. The Inuit migration represents a significant episode in the history of dogs in the North American Arctic, and the dispersal of Inuit culture is mirrored in the dispersal of its genetically distinct dogs. More specifically, our data indicate that though dogs that possessed A2a signatures were present in the North American Arctic prior to the arrival of the Inuit, they were responsible for dominance and eastward expansion of the A2a subclade. The settlement of the North American Arctic by Inuit cultures brought a more mitochondrially diverse and morphologically distinct dog population, and the subsequent European colonization of the North American Arctic further influenced the mitochondrial diversity in more recent centuries. Despite the fact that dog sledging is widely associated with the North American Arctic today, sledging was probably less common prior to the Inuit period. The preservation of these distinctive Inuit dogs is likely a reflection of the highly specialized role that dogs played in both long-range transportation and daily subsistence practices in Inuit society. The legacy of these Inuit dogs survives today in Arctic sledge dogs, making them some of the last remaining descendant populations of a pre-European dog lineage in the Americas.

4. Material and methods

(a) Ancient and modern specimens

Archaeological and ethnographic materials were sampled from Palaeo-Inuit, Inuit, historical and modern contexts across the North American Arctic, subarctic Eastern Canada, Iceland and eastern Siberia. A total genetic dataset of 628 specimens was composed of 186 novel sequences (table 1) and 221 GenBank entries (electronic supplementary material, table S1 and S2). A total of 512 specimens were analysed using GMM, and the dataset

includes 40 crania, 247 mandibles and 284 lower first molars (59 individual specimens were analysed by more than one element; electronic supplementary material, table S3).

(b) Ancient DNA extraction and sequencing

All samples were processed in facilities dedicated to ancient DNA analyses and all PCRs were performed in separate facilities [50]. Ancient DNA laboratory work was conducted at four institutions: the Swedish Museum of Natural History, the Centre for GeoGenetics at the Natural History Museum of Denmark, University of Oxford's PalaeoBARN Lab, and the Veterinary Genetics Laboratory (VGL) ancient DNA facility at the University of California, Davis (UC Davis).

Total genomic DNA was extracted from 628 samples from across Russia and Arctic North America using modified DNA extraction methods from previous studies ([51,52], see electronic supplementary material). Complete mitochondrial genomes were obtained through either shotgun sequencing (422 samples) or mitochondrial capture approaches (308 samples). Details of the approaches and library preparation are in the electronic supplementary material, text. Following sequencing, the reads from each sample were mapped to the CanFam3.1 reference genome with BWA aln, and aligned reads with a mapping quality score lower than 30 were filtered out of the resulting bam files [53,54]. Subsequently, consensus sequences were called and the sequences were aligned with a panel of reference sequences.

Complete mitochondrial genomes were obtained for 147 samples that had a minimum of threefold read depth over 80% of the mitochondrial genome, and maximum-likelihood phylogenies with constructed using RAxML [55,56]. Additionally, 94 samples had genomes with 10-fold mean coverage, this was reduced to 76 samples with a minimum mean of 10-fold read depth of at least 80% of sites covered. These 76 mitochondrial genomes were then used for further demographic analyses. More robust phylogenetic trees were constructed from the 10-fold dataset with RAxML and BEAST2.4 [55–57]. Effective population size was inferred from Bayesian Skyline plots generated with BEAST2.4 [55–57]. Additionally, fragmentary mitochondrial sequences were obtained from 40 additional samples using Sanger sequencing, all of which possessed sufficient information to assign the resulting D-loop sequences to specific haplotypes. These data were combined with previously published data from pre-contact dogs from the Americas, modern dogs, and modern wolves acquired from GenBank for phylogenetic and demographic analyses (electronic supplementary material, table S2) [4,26].

(c) Geometric morphometrics

Geometric morphometric analyses were performed on a total of 571 elements (MNI = 401). Mandible and lower M1 shape were analysed in two-dimensional from photographs (see electronic supplementary material, text for protocol details) digitized with 15 landmarks, and two landmarks and 49 sliding semilandmarks, respectively, using tpsDig2 (electronic supplementary material, figure S14) [58]. Crania were analysed in three-dimensional using models built by photogrammetry [59] in Agisoft PhotoScan (Agisoft LLC, St Petersburg, Russia), with 30 landmarks digitized using morphoDig [60] (electronic supplementary material, figure S14). Prior to analyses, coordinates were superimposed with a generalized procrustes analysis (GPA) [61,62] using the Procrustes distance criterion for optimizing semilandmarks position, and symmetrizing left and right landmarks for the crania.

Size variation was tested with Wilcoxon test and visualized with boxplots showing the log-transformed centroid size. Shape and form (size + shape) variation were explored with principal component analysis (PCA, electronic supplementary material, figure S15), before the differences were tested with MANOVA and visualized using canonical variate analyses (electronic supplementary material, figure S16) based on a reduced number of PCA scores [28]. Differences between groups were also depicted using neighbour-joining networks based on Mahalanobis distances and visualization of shape changes along the discriminant axis. Cross-validation percentages were calculated following [28] and are reported as the mean and 90% confidence interval of 100 discriminant analyses based on ressampled balanced group size. Allometries (size and shape relationships) were tested using MANCOVA. Analyses of Procrustes variance was performed following [27] for shape and Fligner–Killeen tests for size. All statistical analyses were performed in R v. 3.4.3 [63] primarily with the package Morpho [64].

Data accessibility. Mitochondrial sequence alignments have been deposited at the European Nucleotide Archive (ENA) with project no. PRJEB31489.

All datasets are available in the electronic supplementary material files and mitochondrial sequence alignments have been deposited at the European Nucleotide Archive (ENA) with project no. PRJEB31489.

Authors' contributions. C.A., T.R.F., S.K.B, K.D., L.F.F., G.L., B.S, C.M.D., and A.E. designed research; C.A., T.R.F., A.L., S.K.B., A.H.-B., Z.L., O.L., M.-H.S.S., J.H. and A.E. performed research; C.A., T.R.F., S.K.B, A.H.-B., M.-H.S.S., Z.L., A.T.L., L.F.F. and A.E. analysed data; and C.A., T.R.F., A.L. S.K.B., L.F.F., G.L. and A.E., wrote the paper with contributions from all authors.

Competing interests. We declare we have no competing interests.

Funding. The following people and institutions supported this study: Doug Anderson; Mogens Andersen; Peter M. Bowers; Joanne Bird; Stephen Brown; Fawn Carter; Neil Duncan; Max Friesen; Stacey Girling-Christie; Kristian Gregersen; John Hoffecker; Claire Houmard; Susan Kaplan; Brian Kooyman; Reinhardt Møbjerg Kristensen; Genevieve LeMoine; Owen Mason; David Morrison; Georg Nyegaard; Anne Mette Olsvig; Jeff Rasic; Knud Rosenlund; Karen Ryan; Chelsea Smith; Kevin P. Smith; Doug Stenton; Daniel Thorleifsen; Andrew Tremayne; Karen Workman; Bill Workman; the Maxwell Museum of Anthropology; Qanirtuuq Inc., Quinhagak, AK; the Nunalleq Archaeology and Culture Center, and Wolfson College, University of Oxford. This research was supported by the following grants: AHRC (grant no. AH/K006029/1), AHRC-LabEx (grant no. AH/N504543/1), European Research Council grant (grant no. ERC-2013-StG-337574-UNDEAD), Natural Environmental Research Council grants (grant nos. NE/K005243/1, NE/K003259/1 and 2210 GG005 RGA1521), National Science Foundation's Office of Polar Programs (grant nos. NSF-ARC-1108175 and NSF-PLR-1304810), the EU-funded ITN project ArchSci2020 (grant no. 676154), Marie Skłodowska-Curie action WhereWolf (grant no. 655732), the Qimmeq project, the Velux Foundations, the Aage og Johanne Louis-Hansens Fond and the Wellcome Trust (grant no. 210119/Z/18/Z).

Acknowledgements. We acknowledge the participation of ZIN RAS (state assignment no. AAAA-A17-11702281019503), and the support from Science for Life Laboratory, the National Genomics Infrastructure, and UPPMAX (project no. b2014314) for providing assistance in massive parallel sequencing and computational infrastructure.

References

1. Pitulko VV, Kasparov AK. 2017 Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. *J. Archaeol. Sci. Rep.* **13**, 491–515. (doi:10.1016/j.jasrep.2017.04.003)
2. Lee EJ *et al.* 2015 Ancient DNA analysis of the oldest Canid species from the Siberian Arctic and genetic contribution to the domestic dog. *PLoS ONE* **10**, 1–13. (doi:10.1371/journal.pone.0125759)
3. Goebel T, Slobodin SB, Waters MR. 2010 New dates from Ushki-1, Kamchatka, confirm 13,000 cal BP age for earliest Paleolithic occupation. *J. Archaeol. Sci.* **37**, 2640–2649. (doi:10.1016/j.jas.2010.05.024)

4. Ní Leathlobhair M *et al.* 2018 The evolutionary history of dogs in the Americas. *Science* **361**, 81–85. (doi:10.1126/science.aao4776)
5. Duggan AT *et al.* 2017 genetic discontinuity between the maritime archaic and beothuk populations in Newfoundland, Canada. *Curr. Biol.* **27**, 3149–3156.e11. (doi:10.1016/j.cub.2017.08.053)
6. Raghavan M *et al.* 2014 The genetic prehistory of the New World Arctic. *Science* **345**, 1255832. (doi:10.1126/science.1255832)
7. Rasmussen M *et al.* 2010 Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762. (doi:10.1038/nature08835)
8. Scheib CL *et al.* 2018 Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **360**, 1024–1027. (doi:10.1126/science.aar6851)
9. Sikora MS *et al.* 2019 The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188. (doi:10.1038/s41586-019-1279-z)
10. Flegontov P *et al.* 2019 Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236–240. (doi:10.1038/s41586-019-1251-y)
11. Perri A, Widga C, Lawler D, Martin T, Loebel T. 2018 New evidence of the earliest domestic dogs in the Americas. *bioRxiv*. See https://www.biorxiv.org/content/early/2018/06/11/343574.abstract.
12. Morey DF, Aaris-sørensen KIM. 2002 Paleoeskimo Dogs of the Eastern Arctic. *Arctic* **55**, 44–56. (doi:10.14430/arctic689)
13. Howse L. 2018 Hunting technologies and archaeofaunas: societal differences between hunter-gatherers of the Eastern Arctic. *J. Archaeol. Method Theory*. **26**, 88–111. (doi:10.1007/s10816-018-9366-2)
14. McGhee R. 2009 When and why did the Inuit move to the eastern Arctic. In *The Northern World, AD 900–1400* (eds Herbert Maschner, Owen Mason), pp. 155–164. Salt Lake City, UT: Univerity of Utah Press.
15. Friesen TM, Arnold CD. 2008 The timing of the Thule migration: new dates from the Western Canadian Arctic. *Am. Antiq.* **73**, 527–538. (doi:10.1017/S0002731600046850)
16. Mathiassen T. 1927 Archaeology of the central Eskimos: the Thule culture and its position within the Eskimo culture. Report of the Fifth Thule Expedition 1921–24. Copenhagen, Denmark: Gyldendaslke Boghandel.
17. McGhee R. 1996 *Ancient people of the Arctic*. Vancouver, Canada: UBC Press.
18. Gulløv HC. 1997 *From middle ages to colonial times: archaeological and ethnohistorical studies of the Thule Culture in South West Greenland 1300–1800 AD*. Copenhagen, Denmark: Dansk Polar Center.
19. Pitul'ko VV, Kasparov AK. 1996 Ancient arctic hunters: material culture and survival strategy. *Arctic Anthropol.* **33**, 1–36.
20. Lévesque F. 2018 Threatening the fantasy of an Arctic welfare state. In *Dogs in the north* (ed. RJ Losey), pp. 172–190, 1st edn. London, UK: Routledge.
21. Sonne C *et al.* 2018 Greenland sled dogs at risk of extinction. *Science* **360**, 1080.
22. Brown SK, Darwent CM, Wictum EJ, Sacks BN. 2015 Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. *Heredity* **115**, 488–495. (doi:10.1038/hdy.2015.49)
23. Brown SK, Darwent CM, Sacks BN. 2013 Ancient DNA evidence for genetic continuity in arctic dogs. *J. Archaeol. Sci.* **40**, 1279–1288. (doi:10.1016/j.jas.2012.09.010)
24. MacRury IK. 1991 *The Inuit dog: its provenance, environment, and history*. Harwinton, CT: Inuit Sled Dog International.
25. Webster M, Sheets DH. 2010 A practical introduction to landmark-based geometric morphometrics. *Paleontol. Soc. Pap.* **16**, 163–188. (doi:10.1017/S1089332600001868)
26. Thalmann O *et al.* 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874. (doi:10.1126/science.1243650)
27. Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004 14 - Morphometrics and systematics. In *Geometric morphometrics for biologists* (eds ML Zelditch, DL Swiderski, HD Sheets, WL Fink), pp. 363–381. San Diego, CA: Academic Press.
28. Evin A, Cucchi T, Cardini A, Strand Vidarsdottir U, Larson G, Dobney K. 2013 The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743. (doi:10.1016/j.jas.2012.08.005)
29. Pang J-F *et al.* 2009 mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* **26**, 2849–2864. (doi:10.1093/molbev/msp195)
30. Tackney J, Jensen AM, Kisielinski C, O'Rourke DH. 2019 Molecular analysis of an ancient Thule population at Nuvuk, Point Barrow, Alaska. *Am. J. Phys. Anthropol.* **168**, 303–317. (doi:10.1002/ajpa.23746)
31. Raghavan M *et al.* 2015 Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884. (doi:10.1126/science.aab3884)
32. Hjortlund S. 1907 Om de smitsomme Hundesygdomme i Nordgrønland og deres Bekæmpelse. *Meddelelser fra Direktoratet for den kongelige grønlandske.* **1907**, 57–81.
33. Andersen DB. 1992 *The use of dog teams and the use of subsistence-caught fish for feeding sled dogs in the Yukon river drainage, Alaska*. Juneau, AK: Alaska Department of Fish and Game, Division of Subsistence.
34. Park RW. 1987 Dog remains from Devon Island, NWT: Archaeological and osteological evidence for domestic dog use in the Thule culture. *Arctic* **40**, 184–190.
35. Lykke LP. 2017 Arctic Nomads. Sisimiut; Nebula.
36. Masson-Maclean E, McManus-Fry E, Britton K. 2019 The archaeology of dogs at the precontact Yup'ik site of Nunalleq, Western Alaska. In *Beyond domestication: archaeological investigations into the human-canine connection* (eds A Burtt, B Bethke). Gainesville, FL: University of Florida Press.
37. Issenman BK. 2011 *Sinews of survival: The living legacy of inuit clothing*. Vancouver, Canada: UBC Press.
38. McManus-Fry ET. 2015 Pre-contact ecology, subsistence and diet on the Yukon-Kuskokwim delta: an integrated ecosystem approach. PhD thesis, University of Aberdeen.
39. Ledger PM *et al.* 2018 Three generations under one roof? Bayesian modeling of radiocarbon data From Nunalleq, Yukon-Kuskokwim Delta, Alaska. *Am. Antiq.* **83**, 505–524. (doi:10.1017/aaq.2018.14)
40. McManus-Fry E, Knecht R, Dobney K, Richards MP, Britton K. 2016 Dog-human dietary relationships in Yup'ik western Alaska: the stable isotope and zooarchaeological evidence from pre-contact Nunalleq. *J. Archaeol. Sci. Rep*. **17**, 964–972. (doi:10.1016/j.jasrep.2016.04.007)
41. Frobisher SM. 1577 The second voyage of Master Martin Frobisher. In *VOYAGES: VOL V (EVERYMAN'S LIBRARY NO 338)* (ed. R Hakluyt). London, UK: Hakluyt Society.
42. Whitbourne R. 1622 *A discourse and discovery of New-found-land with many reasons to prove how worthy and beneficial a plantation may there be made*. London, UK: Felix, Kyngston.
43. Allen G. 1920 *Dogs of the American Aborigines*. Bulletin of the Museum of Comparative Zoology at Harvard College, vol. 63. Cambridge, MA: Harvard College Press.
44. Degerbøl M, Freuchen P. 1935 *Mammals*. Report of the Fifth Thule Expedition 1921–24. The Danish expedition to Arctic North America in charge of Knud Rasmussen, PhD. Copenhagen, Denmark: Gyldendaslke Boghandel.
45. Hindrikson M, Männil P, Ozolins J, Krzywinski A, Saarma U. 2012 Bucking the trend in wolf-dog hybridization: first evidence from europe of hybridization between female dogs and male wolves. *PLoS ONE* **7**, e46465. (doi:10.1371/journal.pone.0046465)
46. Coppinger R, Coppinger L. 2001 *Dogs: a startling new understanding of canine origin, behavior and evolution*. Chicago, IL: Simon and Schuster.
47. Vila C, Wayne RK. 1999 Hybridization between wolves and dogs. *Conserv. Biol.* **13**, 195–198. (doi:10.1046/j.1523-1739.1999.97425.x)
48. Wang G-D, Xie H-B, Peng M-S, Irwin D, Zhang Y-P. 2014 Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* **2**, 65–84. (doi:10.1146/annurev-animal-022513-114129)
49. Frantz LAF, Larson G. 2018 A genetic perspective on the domestication continuum. In *Hybrid communities* (eds C Stépanoff, J-D Vigne), pp. 41–56. London, UK: Routledge.
50. Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. 2005 Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544. (doi:10.1016/j.tree.2005.07.005)
51. Ersmark E *et al.* 2015 Population demography and genetic diversity in the Pleistocene cave lion. See https://www.openquaternary.com/jms/article/view/oq.aa.

52. Dabney J *et al.* 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15 758–15 763. (doi:10.1073/pnas.1314445110)
53. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. See http://arxiv.org/abs/1303.3997.
54. Lindblad-Toh K *et al.* 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819. (doi:10.1038/nature04338)
55. Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. (doi:10.1093/bioinformatics/btl446)
56. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
57. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537. (doi:10.1371/journal.pcbi.1003537)
58. Rohlf FJ. 2015 The tps series of software. *Hystrix It. J. Mamm.* **26**, 9–12. (doi:10.4404/hystrix-26.1-11264)
59. Evin A, Souter T, Hulme-Beaman A, Ameen C, Allen R, Viacava P, Larson G, Cucchi T, Dobney K. 2016 The use of close-range photogrammetry in zooarchaeology: creating accurate 3D models of wolf crania to study dog domestication. *J. Archaeol. Sci. Rep.* **9**, 87–93. (doi:10.1016/j.jasrep.2016.06.028)
60. Lebrun R. 2018 MorphoDig, an open-source 3D freeware dedicated to biology. Paris, France: IPC5. (doi:10.2307/2992207)
61. Rohlf FJ, Slice D. 1990 Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst. Biol.* **39**, 40–59.
62. Adams DC, Rohlf FJ, Slice DE. 2013 A field comes of age: geometric morphometrics in the 21st century. *Italian J. Mammal.* **24**, 7–14.
63. R Core Team. 2018 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
64. Schlager S. 2017 Morpho and Rvcg - Shape Analysis in R. In *Statistical shape and deformation analysis* (eds GZ, SL, GS), pp. 217–256. New York, NY: Academic Press.

Supplementary Information for

Specialised sledge dogs aided Inuit dispersal across the North American Arctic

Carly Ameen, Tatiana R. Feuerborn, Sarah K. Brown, Anna Linderholm, Ardern Hulme-Beaman, Ophélie Lebrasseur, Mikkel-Holger S. Sinding, Zachary T. Lounsberry, Audrey T. Lin, Martin Appelt, Lutz Bachmann, Matthew Betts, Kate Britton, John Darwent, Rune Dietz, Merete Fredholm, Shyam Gopalakrishnan, Olga I. Goriunova, Bjarne Grønnow, James Haile, Jón Hallsteinn Hallsson, Ramona Harrison, Mads Peter Heide-Jørgensen, Rick Knecht, Robert J. Losey, Edouard Masson-MacLean, Thomas H. McGovern, Ellen McManus-Fry, Morten Meldgaard, Åslaug Midtdal, Madonna Moss, Iurii G. Nikitin, Tatiana Nomokonova, Albína Hulda Pálsdóttir, Angela Perri, Aleksandr N. Popov, Lisa Rankin, Joshua D. Reuther, Mikhail Sablin, Anne Lisbeth Schmidt, Scott Shirar, Konrad Smirarowski, Christian Sonne, Mary C. Stiner, Mitya Vasyukov, Catherine F. West, Gro Birgit Ween, Sanne Eline Wennerberg, Øystein Wiig, James Woollett, Love Dalén, Anders J. Hansen, Tom Gilbert, Benjamin Sacks, Laurent Frantz, Greger Larson, Keith Dobney, Christyann M. Darwent, Allowen Evin

Corresponding Author: Carly Ameen
Email: C.Ameen@exeter.ac.uk

This PDF file includes:

Supplementary text
Figs. S1 to S17
Captions for databases S1 to S6
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S6

Table of Contents:

\section*{A note on cultural terminology:} 2

\section*{A note on modern and historical samples:} 3

\section*{Ancient DNA Methods} 3

\subsection*{Sampling and DNA Extraction} 4

\subsection*{DNA Library Building} 5

\subsection*{Sequencing and Data Preparation} 7

\subsection*{Maximum Likelihood Tree Construction} 7

\subsection*{Bayesian Skyline Analysis} 8

\subsection*{Haplotype Determination} 8

\section*{Ancient DNA Expanded Results} 9
\subsection*{A-clade Subclades and Haplotypes} 9
\subsection*{Demographic Histories of Inuit and Historical Dogs} 11
\subsection*{Analyses of AL2797 Nuclear Genome} 12

\section*{Geometric Morphometrics Methods} 13
\subsection*{2D hemi-mandible and lower M1} 13
\subsection*{Building 3D models of crania using photogrammetry} 13
\subsection*{GMM Statistical Analysis} 14

\section*{Stable Isotope Analysis} 15
\subsection*{Stable Isotope Analysis Methods} 15
\subsection*{Stable Isotope Analysis Results} 15

\section*{Archaeological Site Descriptions} 16
\subsection*{North American Paleo-Inuit sites} 16
\subsection*{North American Inuit Sites} 27
\subsection*{Siberian Archaeological Sample Descriptions} 50
\subsection*{Icelandic and Norse Greenlandic Archaeological Sample Descriptions} 54
\subsection*{Museum and Ethnographic Sample Descriptions} 58

SI Figure Captions 65

Captions for Databases S1 - S7 67

SI References 67

\section*{A note on cultural terminology:}

The terms ‘Inuit’ and ‘Paleo-Inuit’ are used throughout this manuscript in the following way; dogs from archaeological sites from the more recent archaeological traditions of the North American Arctic, from ~ 1000 BP onwards, are referred to collectively as ‘Inuit’ and include specimens from sites attributed to Eskimo, Neo-Eskimo, Northern Maritime, Thule, Old Bering Sea (OBS), Punuk, and Birnirk cultures. The Northern Maritime tradition includes OBS, Punuk, Birnirk, and Thule, with the development of toggling harpoons and drag floats originating with OBS ~2000 BP in the Bering Strait region. The Thule migration out of Alaska into the Eastern Arctic of Canada and Greenland is generally agreed to date to ca. AD 1300 (Friesen and Arnold 2008), thus “Inuit” origins are much older in the Bering Strait than in the Eastern Arctic. The term ‘Paleo-Inuit’ here refers to dogs associated with Arctic traditions prior to ~1000 BP, and includes specimens from sites attributed to the Allutiiq, Arctic Small Tool, Denbigh Flint Complex, Choris, Dorset, Ipiutak, Kachemak, Norton, and Saqqaq cultures. The use of the term ‘Paleo-Inuit’ should not be considered as acknowledgment of any direct ancestral relationship between these groups (human or canid) and the Inuit culture. In some cases, there is no direct contact between the Inuit peoples and those who occupied certain parts of this macro-region prior to ~ 1000 BP. However, the Inuit

migration represents a pivotal moment in the bio-cultural history of this entire region, and in relation to the dogs included in this analysis, the samples classified as 'Paleo-Inuit' should be interpreted only as chronologically predating the Inuit period.

The cultural terminology used in this manuscript follows the advice of the Inuit Circumpolar Council (ICC Resolution 2010 – 01), which represents the majority of indigenous groups of the Arctic macro-region from Chukotka to Greenland. A summary of these terms and their relationship to Arctic archaeological research is outlined by Friesen (T. M. Friesen 2015), and the separation of dogs sampled here into Inuit and Paleo-Inuit groups is supported by both the archaeological record and ancient DNA analysis of human remains (Raghavan et al. 2014; Duggan et al. 2017; Flegontov et al. 2017; Jensen 2016; Tackney et al. 2019). We recognise that no simplified naming system is completely suitable to cover the multiple, diverse and unique indigenous groups that make up both the modern and archaeological occupants of the North American Arctic and Subarctic regions. We have adopted the above classification system which is supported by modern Indigenous representatives of the region as well as regional archaeological experts, in an attempt to find a suitable (although not perfect) terminology for discussing these large scale, pan-regional trends.

\section*{A note on modern and historical samples:}

Throughout this manuscript, 'Modern' refers to samples collected from living dogs within the last two decades. Samples referred to as historical include those which are older than 20 years, collected during ethnographic expeditions from living dogs, or skeletal remains of dogs from surface contexts with no known archaeological dates. Samples categorised as historical may represent dogs of archaeological origin as a result of their surface context rather than sub-surface and the slow rate of sediment deposition in the region; however, without radiocarbon dating the age of the samples cannot be precisely determined. Samples of historical origin are separated from Inuit samples in this study due to the contact between Europeans and Arctic communities introducing European dogs to the region. Modern, historical, and Inuit age dogs from the sledge district of Greenland represent the traditional Greenland Sledge Dog. To be comparable with samples across the Arctic region, Greenland Sledge Dogs are divided by the age of the sample. The Canadian Inuit Dog and Greenland Sledge Dog are considered to represent the same breed, with the key differences being the location and size of the population (S. K. Brown et al. 2015). Similarly, this may be the case for historical dogs in Siberia related to the Siberian husky and Alaskan Malamutes with historical Alaskan dogs; however, they show a greater extent of admixture with Eurasian dog populations than Greenland Sledge Dogs (S. K. Brown et al. 2015). Alternatively, the Alaskan Husky represents a breed originating after contact between native Alaskans and Europeans with the crossing of Alaskan village dogs, imported Siberian Huskies, and European dogs producing Alaskan Huskies (S. K. Brown et al. 2015).

\section*{Ancient DNA Methods}

All samples used in the study were sampled, extracted, and setup for PCR in laboratories dedicated specifically to working with genetic samples of ancient origin, using appropriate sterile techniques and equipment, see SI Table 1 (A. Cooper and Poinar 2000; Gilbert et al. 2005). Ancient DNA lab work was conducted in four facilities, the Centre of GeoGenetics at the Natural

History Museum for Denmark, the Swedish Natural History Museum, the University of Oxford's PalaeoBarn, or the Veterinary Genetics Laboratory (VGL) ancient DNA facility at the University of California, Davis (UC Davis).

\subsection*{Sampling and DNA Extraction}

\subsubsection{Centre for GeoGenetics and Swedish Natural History Museum}*

For teeth and bone, all specimens were initially 'cleaned' through the removal of the surface through superficial drilling with a Dremel drill. Following the removal of the external layer the desired 30-80mg of bone power was drilled out of the bonee. For the samples which were derived from teeth the outer surface was removed and the cementum layer of the tooth was drilled to a powder for subsequent extraction, amounting to between 15 and 30 mg per sample. 51 samples were skin samples originating from clothing items or museum hides, the hairs on the skin were removed with a sterile scalpel from the subsample taken from the hide of approximately 30 mg. Bone, tooth, and skin samples extracted at the Swedish Natural History Museum underwent a pre-digestion step consisting of a 45 minute incubation in 315 µL of 0.5M EDTA and 7.5 µL of 10 ng/µL proteinase K. After incubation the samples were spun in a centrifuge at 13,000 rpm for 5 minutes, following which the supernatant was removed. From this stage the remaining undigested material was extracted overnight with 630 µL of 0.5M EDTA, 70 µL of 1M UREA, and 15 µL of 10 ng/µL proteinase K as per (Ersmark et al. 2015). The skin samples were also incubated overnight in a lysis buffer as per (Gilbert et al. 2007). Purification of extracts was performed according to (Ersmark et al. 2015).

\subsubsection{University of California, Davis}*

UC Davis bone and tooth specimens were extracted via the silica extraction method (Rohland and Hofreiter 2007; Rohland, Siedel, and Hofreiter 2010) with slight modifications, as per (Sarah K. Brown, Darwent, and Sacks 2013). The modifications include, 1) prior to lysis, samples were subject to a decalcification solution in 2 ml of 0.5 M EDTA solution and kept rotating at room temperature for two days, and 2) subsequent to extraction, DNA was ethanol precipitated to remove inhibitors as per (Kemp, Monroe, and Smith 2006), and eluted in a final volume of 100 µl.

\subsubsection{University of Oxford's PalaeoBARN}*

All the bone or teeth samples were "cleaned" by removing the outer surface layer, after this a small piece was cut and the bone/tooth was pulverised using a mixer mill generating around 50-70 mg powder, all of this was undertaken in a separate room in the dedicated aDNA lab. Extraction was carried out using the silica method following the Dabney extraction protocol (Dabney et al. 2013) but with the addition of a 30 minutes pre-digest stage following (Damgaard et al. 2015).

\subsection*{DNA Library Building}

\subsubsection{Centre for GeoGenetics and Swedish Natural History Museum}*

Libraries were built at the Swedish Natural History Museum using the blunt ending single□tube library preparation, 'BEST' method specifically designed for working with degraded DNA (Carøe et al. 2018). The loss of DNA during the library build process is reduced through the limitation in the number of purifications the sample undergoes. The extract initially undergoes blunt end repair followed by a denaturing step at 65°C, without purification the sample proceeds to the adapter ligation step where Illumina adapters are ligated to the ends of the DNA fragments. Finally, after the adapter ligation reaction has incubated, the nicks in the strands are filled in and following incubation each sample is then purified using Qiagen MinElute columns.

\subsubsection{University of California, Davis}*

The libraries constructed at UC Davis were prepared using the NEBNext Ultra DNA kit (New England Biosystems). The manufacturer's protocol was modified by reducing the reaction volume in half, including the volume of DNA. Library preparation was conducted in a dedicated laboratory equipped with a laminar flow hood with UV lights for sterilization. DNA was subject to end repair and dA-tailing and incubated for 30 minutes at 20°C and then 30 minutes for 65°C. Next, ligation to the NEBNEXT Ultra Universal adapter was performed by adding 32.5 µL of the end repair/dA-tailing reaction with 1.25 µL of 1.5 µM adapters (diluted tenfold for ancient DNA), 0.5 µL Ligation enhancer, and 7.5 µL of Blunt/TA Ligase Master Mix. This reaction was incubated at 20°C for 15 minutes. USER enzyme (1.5 µL) was added to the reaction to cleave the loop adapters, incubated at 37°C for 15 minutes, and then purified with 1X Ampure beads according to protocol (no size selection). Libraries were uniquely barcoded through PCR amplification with indexing primers by mixing 11.5 µL of adapter ligated DNA with 12.5 µL Kapa Biosystems Uracil + master mix (Kapa: Woburn, MA), and 0.5 µL of each 25 µM primer (Universal NEB primer and Indexed primer (Supplemental Appendix A). The PCR cycling conditions were as follows: 98°C for 45s; 18 cycles of 98°C for 15s, 65°C for 30s, 72°C for 30s; 72°C for one minute. Libraries were then gel electrophoresed and quantified using a Qubit fluorometer (Life technologies; Grand Island, NY). PCR amplified libraries were then purified with 1X Ampure beads and eluted in 33 µL of 10mM Tris-HCl. Cleaned libraries were then quantified using a Qubit fluorometer (Life technologies; Grand Island, NY).

Target enrichment for the domestic dog mitochondrial genome was performed on the amplified and cleaned library using MYcroarray's custom MYBaits enrichment kit. The custom RNA baits were developed from the *Canis familiaris* reference mitochondrial genome (Kim et al. 1998) in 4x tiling of 80mer bases. Forty eight uniquely barcoded DNA samples were pooled in equimolar amounts (10 ng each, 480 ng total), and concentrated in a speed vacuum until dry and then reconstituted in 3.4 µL of DI RNase free water. Concentrated DNA libraries were hybridized according to manufacturer's instructions, in a final concentration of 5.2 x SSPE, 5.2 x Denhardt's, 5mM EDTA and 0.1\% SDS. This solution was incubated with the DNA libraries at

48°C for 37.5 hours. Hybridized DNA was then immobilized on magnetic streptavidin beads, washed at 48°C and then eluted with 100mM NaOH to destroy the RNA baits, and subsequently neutralized with neutralization buffer (1M Tris-HCl, pH 7). The captured DNA was purified with a Qiagen Qiaquick spin column and eluted in 30 µL of elution buffer. This product was PCR amplified for 10 cycles using 5 µL of capture product, 18 µL of nuclease free water, 25 µL of Kapa HiFi Uracil + master mix, 1 µL 10uM universal NEB primer and 1µL sequencing non-indexed primer (5' CAAGCAGAAGACGGCATACGAGAT 3'). The capture products were purified with 1X Ampure beads and eluted in 33 µL of 10mM Tris-HCl, and then DNA was quantified on the Qubit analyzer.

As per (Enk et al. 2014), the first round of capture underwent a second hybridization, to increase the amount of mitochondrial DNA captured. Cleaned capture product from the first round of capture was concentrated on a speed vacuum until dry and then reconstituted with 3.4 µL of DI RNase free water. The capture product was then hybridized according to manufacturer's protocol at 55°C for 48 hours, and subsequently cleaned via Streptavidin beads. Captured libraries were subsequently cleaned, and amplified an additional 15 cycles. Amplified capture libraries were cleaned via 0.8X Ampure bead clean up and delivered to the UC Davis Genome Center for sequencing on the Illumina Miseq. Sequencing varied between runs: Plate 1 and 2 = 250 Paired End, Plate 3 = 150 bp SE, Plate 4 = 100 bp SE.

\subsubsection{PaleoBARN, University of Oxford}*
Illumina libraries were built at the University of Oxford following (Meyer and Kircher 2010), with the addition of a six base-pair barcode added to the IS1_adapter.P5 adapter. The libraries were then amplified on an Applied Biosystems StepOnePlus Real-Time PCR system to check that library building was successful, and to determine the optimum number of cycles to run during the indexing amplification PCR reaction. A six base-pair barcode was added during the indexing amplification reaction. This resulted in each library being double-barcoded with an "internal adapter" directly adjacent to the ancient DNA strand and which would form the first bases sequenced, and an "external barcode" that would be sequenced during Illumina barcode sequencing. Libraries were sequenced on an Illumina HiSeq 2500 (Single End 80bp) sequencer at the Danish National High-Throughput Sequencing Centre, and on an Illumina NextSeq 500 (Single End 80bp) at the Natural History Museum in London.

Mitochondrial captures were performed using the myBaits manual version 3.01 (August 2015), with a 55°C incubation for 24 hours. The heat elution step was skipped, and the captures were amplified using KAPA Hi-Fi polymerase following the myBaits manual instructions for 23 cycles. The PCR reaction was cleaned using Minelute columns.

\subsection*{Sequencing and Data Preparation}

The samples were shotgun sequenced on the Illumina HighSeq2500 in pools of 40 samples per lane with paired end sequencing. The number of reads sequenced for each sample ranged from 3,000 to 27,551,904.

Adapters were trimmed and paired reads were merged with SeqPrep (https://github.com/jstjohn/SeqPrep). The UC Davis samples were demultiplexed and trimmed adapters using MiSeq Reporter version 2.3.32 (Illumina). The merged reads were then mapped using BWA aln, Burrows-Wheeler Aligner, to the CanFam3.1 dog reference genome (Li 2013; Lindblad-Toh et al. 2005). The reads were then filtered using SAMtools with the quality filter set to 30 (Li et al. 2009) and PCR duplicates were removed with the MarkDuplicates application in PicardTools (http://broadinstitute.github.io/picard/). Reads that mapped to the mitochondrial genome were then extracted from the bam file and the read depth of the mitochondrial genome was calculated with QualiMap bamqc. A consensus sequence was created for each sample from these extracted reads using htsbox (https://github.com/lh3/htsbox). The consensus sequence for each of the samples were merged together into a single file and the sequences were aligned using the Mafft algorithm using the default parameters (Katoh and Standley 2013).

Default parameters on MapDamage2.0 were used to assess the molecular damage patterns of the samples (Jónsson et al. 2013). The expected damage patterns at the ends of fragments were seen on ancient samples (Fig. S1). Alternatively, the historical samples from within the most recent 150 years displayed much lower levels of damage which is also as would be expected for these samples (Fig. S1).

\subsection*{Maximum Likelihood Tree Construction}

Maximum likelihood phylogenetic trees were constructed with RAxML from the aligned consensus sequences using a General Time Reversible model with a Gamma distribution (Stamatakis 2014, 2006). Coyote *(Canis latrans)* was used as an outgroup for each of the phylogenetic trees, in addition to a reference panel of published mitochondrial genomes (Thalmann et al. 2013; Pang et al. 2009; Imes et al. 2012; Frantz et al. 2016; Botigue et al. 2016; Ní Leathlobhair et al. 2018). Each phylogeny was constructed with 1,000 bootstraps to ensure the robustness of each of the nodes of the tree. Separate maximum likelihood trees were constructed from aligned samples with a mean of 1x read depth, a minimum mean of 3x read depth of 80\% of positions in the mitochondrial genome, and a minimum mean of 10x read depth of 95\% of positions for a robust representation of phylogenetic relationships (Fig. S2). The output from RAxML was visualised using FigTree v1.4.3 (Rambaut: http://tree.bio.ed.ac.uk/software/figtree/).

\subsection*{Bayesian Skyline Analysis}

Bayesian Skyline Plots (BSPs) were computed using BEAST2.4 software to infer effective population sizes. Analyses were run for subsets of the data divided by phylogenetic clade affiliation and by cultural affiliation. A minimum threshold of 10X mean read depth was set for all bayesian analyses. Skyline analyses were run to compare differences in population size across different populations of dogs. The first Bayesian skyline analysis contained a reference panel of published dogs from (Ní Leathlobhair et al. 2018; Thalmann et al. 2013) in addition to the 76 new complete mitochondrial genomes with over 10x read depth (Fig. S3). BSPs were also run for subsets of the data based on their age to assess differences in the effective population size based on their cultural affiliation or sample age. The same Three analyses were undertaken to address the difference: 1) Inuit dogs (n=22) from across the North American Arctic, 2) Recent dogs (n=39) from 300 years before present to the current day in the North American Arctic, and 3) Recent dogs (n=28) from 300 years ago until 50 years ago. The third analysis was undertaken to separate the influence the impact of unbalanced representation of samples in the dataset in the recent populations. All samples from the North American Arctic which pass the 10x read depth threshold are exclusively from Greenland and all but two carry related haplotypes which could reflect an accurate representation of haplotype frequencies over the last five decades or alternatively the small sample size and/or sampling bias. Due to the small sample size in the Paleo-Inuit dataset (n=2), this population was not not included in the culture/age based analyses.

To run these analyses xml files were created using BEAUti, a graphical interface function of BEAST. The datasets imported into the xml file using BEAUti were alignments of the samples partitioned based on codon position, tRNA, rRNA, and control region after removing repeating regions. Each partition was assigned a different substitution model on the basis of the recommendations given by partitionfinder: coding regions (HKY), tRNA (HKY), rRNA (TN93+G), and control region (HKY). Within BEAUti individual priors were set for each of the samples based on their ages with a uniform distribution for the age of the tip. A strict molecular clock was used as recommended in previous studies (Thalmann et al. 2013). 50 million MCMC (Markov Chain Monte Carlo) chains were run on the dataset with trees being samples every 5000 iterations with a 10\% burnin. Tracer v.1.6 was utilised to evaluate convergence and visualisation of the effective population size of the dataset. Trees created with BEAST were visualised using Maximum Clade Credibility tree type using TreeAnnotator v1.8.3 with a 10\% burnin. Topology agreement was found between the tree generated by RAxML and the whole dataset tree generated by BEAST.

\subsection*{Haplotype Determination}

D-loop haplotypes were determined by extracting the 609 bp HVI portion from all sequenced mitogenomes. The haplotypes were determined by using the Canis mtDNA HVI database http://chd.vnbiology.com/ (Thai, Chung, and Tran 2017). Sequences that contained too many

ambiguous bases were haplotyped via BLAST, and or by alignment to known haplotypes (Savolainen et al. 2002).

A portion (N =193/480) of the UCD samples were Sanger sequenced as per (Sarah K. Brown, Darwent, and Sacks 2013). Of those samples that were Sanger sequenced, only N = 35 did not have mitogenome sequences, and thus the D-loop haplotype information was used to classify the individual to a clade.

On the basis of phylogenetic subclades found in all iterations of phylogenetic trees generated in this study, four haplogroups within the A-clade were identified. For the purposes of this study these clades representing haplogroups were identified as A1a, A1b, A2a, and A2b. To confirm the topology and obtain the most robust confidence in the phylogenetic relationships the 10x read depth dataset utilised for bayesian and maximum likelihood analyses were adjusted to exclude samples basal to the clades to assess for support at these nodes. The naming of the main dog clades (A-F) follows the convention of previous studies (Savolainen et al. 2002; Pang et al. 2009); however, the nomenclature used in this study to refer to the subclades (A1a = haplotypes A18, A20, A22, A1b = haplotype A29, A2a = haplotype A31, A2b = haplotype A3) was created for this study.

\section*{Ancient DNA Expanded Results}

\subsection*{A-clade Subclades and Haplotypes}

A maximum likelihood tree was constructed from all samples which had a minimum of 3x read depth of the mitochondrial genome (Fig. S4, Fig. S5). All Inuit and Palaeo-Inuit samples fall in the previously identified Dog A-clade, together with the historical dogs from Greenland, Canada, and Alaska. Three samples from Russia included in this study, a Neolithic dog from the Boisman II site (AL2989), an ethnographic sample collected from near the Anadyr River (AL3032), and a 2,400 year old individual from Pospelov (AL2992) all fall basal to the A-clade. Four subclades (A1a, A1b, A2a, A2b) were identified in our dataset within the A-clade. Haplotypes identified from partial and complete mitochondrial genomes were assigned to each of the four subclades identified.

The A1a subclade contains modern non-Arctic dogs of a global distribution, modern Greenland Sled Dogs (n=10) and historical dogs found across the Arctic from Chukotka through to East Greenland (n=12). There is no appearance of pre-contact dogs or Paleo-Inuit dogs, however, two Thule dogs from Alaska and Greenland carry A1a haplotypes. The dog (ZMK 113a/1950-P27:TRF-01-23) from the Paleo-Inuit site at Deering, Alaska, was directly radiocarbon dated to 287±24BP, or 435 to 353 cal.BP (see Deering Paleo-Inuit site description and Table S5). The date shows that this dog was deposited during the Thule era, possibly related to the contemporary Western Thule to the east of the Ipiutak site. The other Thule dog carrying the A1a haplogroup

(P68/2016KMG:TRF-01-09) was excavated from a site in northeast Greenland called Dødemandsbugt. This dog was recovered from a house from the early phase of the site, dating to the 15th to 17th centuries CE. The appearance of the A1a haplotype during the Thule Inuit era in Northeast Greenland could help support hypotheses about the Thule settlement of northeastern Greenland, with the isolated location of this sample may be related to the arrival of the Thule Inuit to northern Greenland from the Thule District via Peary Land (Johnson 1933; Sørensen and Gulløv 2012).

Two further haplogroups, A1b and A2a are isolated to Siberian and American Arctic, with the exception of an imported Greenland Sledge Dog in Denmark, a historical (1600-1800 CE) dog from Iceland, and a dingo which are known to be of Asiatic origin (Savolainen et al. 2004). Haplogroup A1b contains modern individuals, including a previously published dingo and Siberian husky, modern Greenland Sledge Dogs (n=2), as well as historical dogs from Siberia to East Greenland (n=11). Additionally, clade A1b includes Thule Inuit dogs (n=16) from sites across the American Arctic, dogs (n=5) from Birnirk Culture sites in Chukotka, and two published 1,700 year old dogs from Aachim Lighthouse, Siberia (Lee et al. 2015). Haplogroup A2a is found in seven of the published dogs from Zhokhov site fall in the earliest branches of the clade (Lee et al. 2015; Ní Leathlobhair et al. 2018). Historical dogs from Alaska, Canada, and Greenland (n=16), as well as one historical dog from Iceland, are found within clade-A2a, along with Thule Inuit dogs from Canada and Greenland (n=36), two Birnirk culture dogs from the Walakpa site in Alaska, and one Paleo-Inuit individual from the Chugachik in Alaska which is attributed to the Kachemak Culture from approximately 2,000 BP.

Finally, A2b haplotypes are found in all precontact dogs (n=68), consistent with the work by Ní Leathlobhair and colleagues, a monophyletic lineage is formed by all previously sampled precontact American dogs from the past ~9,000 years (Ní Leathlobhair et al. 2018) (Fig. S4). However, the clade does not exclusively contain precontact dogs, with a temporal and geographic range of Arctic dogs also falls into this subclade. A2b contains historical dogs from Alaska (n=1) and Greenland (n=1), Thule dogs from Alaska and Canada (n=8), Punuk and Old Bering Sea Culture dogs from St. Lawrence Island on the coast of Alaska (n=3), Paleo-Inuit dogs from Canada and Kodiak Island, Alaska (n=10), and dogs from the Okhotsk Culture of the Kurils (n=2). The only dogs from Greenland which fall into the the A2b clade are from north western Greenland in the Thule District from a site called Iita thought to date to the late 19th century CE.

Interestingly, no haplogroup was restricted to a particular region of the North American Arctic, which is consistent with a model of rapid range expansion of sledge dogs. However, there seems to be a geographic pattern in the dataset; for example 60\% of East Greenland dogs from the Inuit period to present show haplotypes from the A1b subclade, making it the most abundant haplogroup in East Greenland. However, it is currently difficult to assess if such pattern reflects sampling bias.

High read depth phylogenetic trees were constructed using samples with at least 10x read depth of the mitochondrial genome. Maximum likelihood (ML) trees were constructed with RAxML and a bayesian Maximum Clade Credibility tree with BEAST (Fig. S6, Fig. S7, Fig. S8). As a result of this increased read depth requirement only 22 Inuit and five Paleo-Inuit samples were included in the alignment. The appearance of the four A-clade subclades is maintained in the 10x ML trees with greater support for the nodes. The novel subclade of historical dogs and Neolithic canids within a Eurasian wolf clade is also preserved in this phylogeny with more robust bootstrap values; however, the minimum threshold cutoff has reduced this to three samples (TRF.02.29, AL3007, 03.P04.H1.1024). Of the three dogs basal to A-clade in the 3x phylogeny, two met the increased 10x threshold (AL2992, AL3032), and these samples retain their positions basal to the clade.

\subsection*{Demographic Histories of Inuit and Historical Dogs}

Effective population size of populations divided by age was inferred from BSPs generated with BEAST2.4 (Fig. S9).

When one continuous population is assumed for dogs starting in the Inuit era of North America and continuing to present day a decrease in the effective population size can be seen between 15,000 and 5,000 years ago with a drop in the effective population size. This is followed by a small increase in the mean effective population size 2,000 years ago, returning to a similar population size as seen at 15,000 years ago. Coinciding with the Thule migration around 1,000 years ago, the effective population size decreased and is estimated to have continued to decrease until present day.

Bayesian estimation of the effective population size of only Inuit dogs, excluding all historical and modern Arctic dogs, shows a steady population until approximately 5,000 years ago when the effective population size decreased gradually until plateauing around 3,000 years ago. Another, more abrupt, decrease in the effective population size occurred approximately 1,000 years ago, dropping until 500 years ago. This decline corresponds to the timing of the Thule Inuit expansion across the North American Arctic and reflects the bottleneck in the expanding population.

The demographic analysis of North American Arctic dogs from the last three centuries contrasts with that of the Inuit dogs. The historical dogs show no decrease in the effective population size before 3,000 years ago; however, the recent dog populations show a decrease 2,500 and 1,200 years ago. This decline likely represents the Thule Expansion, however the dates are not completely in sync due to the absence of Inuit dogs in the analysis to calibrate the age of the decline. Starting around 250 years ago the population underwent an additional, more rapid decline in the effective population size. When comparing the results of the two analyses (Analysis-1 of dogs from 300 years ago until present and Analysis-2 of dogs from 300 to 50 years ago) the key difference is seen in the estimated population size at time 0 or present. When

the 11 modern Greenland Sledge Dogs are included in the analysis the mean effective population size of the current population decreases by 27%. Analysis-1 estimated the mean effective population size at time 0 to be 675 while Analysis-2 estimated a mean effective population size of 2,463.

When the dataset is subdivided based on cultural affiliation/age a change in the effective population size can be seen in connection with the bottleneck experienced by Inuit dogs during the Thule Expansion. Additionally, a decline in the effective population size has also occurred over the last 250 years, coinciding with increased urbanisation in the Arctic following the arrival of Europeans as well as a changing climate. While the sampling strategy of modern dogs may have underrepresented the diversity in modern Arctic dogs from the most recent 50 years, there is no doubt that there has been a marked decline in the effective population of North American Arctic dogs over the last three centuries. This decrease in population size is corroborated with census data from dogs in Greenland which have experienced a drop in numbers from 36,521 dogs in 1998 down to 14,126 dogs in 2016 (Source: Greenland Statistics).

Bayesian Skyline simulations can be informative about the general trends which have occurred over time in a population; however, as demonstrated there can be limitations to the strict interpretation of the results. For example, the census information of the current population size of Greenland Sledge Dogs shows the stark difference between the simulated effective population size and the genuine population size. This is a recognised phenomena, in which the effective population size is an underestimate of the observed population size (xx). As a result of the underestimation inherent in effective population size simulations the general trends are the noteworthy aspects of the estimates rather than the estimated population size itself.

\subsection*{Analyses of AL2797 Nuclear Genome}

We obtained enough reads (>700,000 reads mapping to CanFam3.1) from a sample obtained from the site of Nunalleq (AL2797) to conduct some basic nuclear analysis. We used the same SNPs set and same samples as in (Ní Leathlobhair et al. 2018)see Table S6 for a list of genomes with accessions). Using smartpca (Patterson, Price, and Reich 2006) we performed Principal Components Analysis (PCA) projecting AL2797. This analysis indicates that AL2797 is more closely related to wolves than dogs (including Arctic dogs; Fig. S10). We then used plink v1.9 (Purcell et al. 2007) to compute an Identity By State (IBS) matrix using all SNPs. This matrix was used to build a neighbour joining tree (NJ) using the R package “ape” (E. Paradis, Claude, and Strimmer 2004). The tree captures the deep split between East Asian and Western Eurasian dogs (Frantz et al. 2016) and demonstrate the close affinity of AL2797 with North American wolves (Fig. S10).

We then computed D-statistics of the form D (AndeanFox, AL2797, Pop3, Pop4) using ADMIXTOOLS (Patterson et al. 2012) where Pop3 was fixed as either European dogs, Asian dogs, Arctic dogs, or CTVT and Pop4 represented possible other sample. We plotted, as box

plots, the results of these combinations (Fig. S10). Positive values imply that the AL2797 share more derived alleles with the population on the y axis, while negative values imply that pre-contact dogs are closer to the other populations. The results indicate that AL2797 is most closely related to North American wolves.

\section*{Geometric Morphometrics Methods}

Morphometric analysis of the dog crania in 3 dimensions (3D) and hemi-mandible and lower M1 in 2 dimensions (2D) was performed on specimens from Paleo-Inuit and Inuit archaeological sites as well as ethnographic materials were sampled from historical, and modern contexts (Table S3) using landmark and sliding semi-landmark based geometric morphometric approach. Modern and Pleistocene wolves from the North American Arctic were also included in the analyses to provide a baseline wild reference morphology (Fig. S13).

\subsection*{2D hemi-mandible and lower M1}

Data was collected from 2D digital photographs taken of the labial surface of the mandible and the occlusal view of the lower M1. Photographs were taken with a reflex camera paired with a fixed 60mm micro lens following a standardised protocol where the position of the mandible was maintained as parallel to the focal plane of the camera using a bubble level and supported by plasticine if necessary. The camera was positioned directly above the specimen and also levelled using a bubble level to maintain consistency in camera positioning. Each picture contained a scale on the same plane as the element of interest. Mandible photographs were taken by CA, AH-B and AE. Photographs of the lower M1 were taken by AH-B. Coordinates of landmark and sliding semi-landmarks were digitized on photographs of the hemimandible and lower M1 using the software tpsDIG2 (F. J. Rohlf 2004). The mandible was captured using a series of 15 landmarks while the lower M1 protocol used 2 landmarks and 49 sliding semi-landmarks along the outline of the tooth (Fig. S14). All mandibles were digitised by CA, and all lower M1s by AH-B.

\subsection*{Building 3D models of crania using photogrammetry}

Photographs of the canid crania were taken following the semi-automated protocol outlined in the supplementary information of (Evin et al. 2016). Crania were placed at the centre of a round turntable which also contained a scale. Photos were taken using a reflex camera paired with a fixed 60mm micro lens from three angles relative to the specimen, at approximately 40°, 15° and 0°. At each height, 36 photographs were taken at every ~10°, covering an entire 360° rotation. This process was repeated for both the dorsal and the ventral side of the specimen, so that for each specimen a total of approximately 216 photographs were taken (i.e. [~36 photos x 3 angles] x 2 sides). These images were then used to build 3D models of the canid crania using Agisoft Photoscan Professional (version 1.4.1, 2018). 3D coordinates of 30 landmarks were acquired on

the models using MorphoDig (version 0 beta 5, 2018) with 30 landmarks (Fig. S14). All models were built and digitised by AE.

\subsection*{GMM Statistical Analysis}

Sets of coordinates were superimposed independently using General Procrustes Analyses (F. James Rohlf and Slice 1990; Goodall 1995). During the superimposition of the lower M1 data, sliding semi-landmarks were allowed to slide following the Procrustes distance minimization criteria (Sampson et al. 1996). The analysis of GMM data was performed in RStudio using R version 3.4.3 (R Core Team 2018) with base functions, and the libraries Morpho (Schlager 2017), Geomorph (Adams, Collyer, and Kaliontzopoulou 2018), ape (Emmanuel Paradis, Claude, and Strimmer 2004), and function "mevolCVP" (Evin et al. 2013).

Differences in size between the three dog groups and the modern and pleistocene wolves were tested with a Wilcoxon test for size and visualised with boxplots (Fig. 1.A). Variation in shape was visualised with Principal Component Analyses (PCA) (Fig. S15) and the differences tested with multivariate analyses of variance (MANOVA) and Canonical Variate Analyses (CVA) (Fig. S16) paired with leave-one-out cross-validation, performed with the R function "mevolCVP" (Evin et al. 2013). Cross Validation Percentages (CVP) obtained from 1000 resampled balanced designs were expressed as a mean CVP and a 90\% confidence interval (Evin et al. 2013). Because discriminant analyses are sensitive to unbalanced sample size between groups and to the number of variables used as group predictors compare to the number of specimens per group we used a resampling approach based on 100 discriminant analyses performed on balanced sample size (randomly selecting specimens in the larger group to match the size of the smaller group), this was paired with a dimensionality reduction approach selecting the N first principal components of the PCA that maximise the differences between groups (Baylac and Friess 2005). Following the CVA, neighbour-joining networks were constructed using Mahalanobis D^2 distances (Fig. 1.B). Differences in shape were visualized along the discriminant axes calculated by multivariate regressions (Monteiro 1999)(Fig. 1.C).

Allometries (the size and shape relationship) were analysed by testing for homogeneity of within-groups allometric patterns using MANCOVAs with shape as the dependent variable, the log centroid size as a covariate and group as a factor. Only the interaction term of the Mancovas was provided (table 1). Variance analysis, also called morphological disparity, was performed using the Procrustes variance (Zelditch et al. 2004) using the morphol.disparity function of the geomorph R package (Adams, Collyer, and Kaliontzopoulou 2018) for shape and Fligner-Killeen test for size.

The recent Greenland dog category was composed of Historical Arctic Dogs and Modern Greenland dogs. The Historical Arctic dogs came from sites across the North American Arctic while the Modern Greenland dogs correspond to modern Greenland Sledge Dog Breed

specimens from Bern, Natural History Museum (see SI). Based on our relatively small sample size we found very little morphometric differences between these two groups, so they were pooled for all the analyses included in this article. Dogs of these two categories did not differ in centroid size for the crania (W=4, p= 0.64), lower first molar (W=35, p=0.60) nor the mandible (W=43, p=0.50). We did not detect differences in the cranial shape (F(5,2)=1.86, p=0.39) nor mandible shape (F(9,9)=1.39, p=0.31). We found differences in lower M1 shape (F(4,15)=3.94, p=0.02) but since the cross validation percentages between the two groups was relatively low (67\% (58.3-83.3\%) it did not prevent the pooling of the specimens.

\section*{Stable Isotope Analysis}

\subsection*{Stable Isotope Analysis Methods}

Previous stable isotope analysis of carbon and nitrogen was conducted on the dogs from the Thule site of Nunalleq (GDN-249) near Quinhagak, Alaska (E. McManus-Fry et al. 2016; Forbes, Britton, and Knecht 2015; Ledger et al. 2016). To investigate the relationship between domestic and wild canids, nine new specimens were sampled for stable isotope analysis, including the specimen AL2729, confirmed by analysis of nuclear DNA to be a wolf (main text, SI text). Samples were taken from mandibles and were prepared at the Department of Archaeology, University of Aberdeen, where collagen extraction was performed following a standard protocol (Britton et al. 2013; Longin 1971; T. A. Brown et al. 1988; M. J. Collins and Galley 1998). First specimens were demineralised in 0.5 M hydrochloric acid (HCl) at 4°C, which was changed at regular intervals until demineralisation was complete, usually once or twice per specimen. Demineralisation was complete for all specimens after 14 days, and the demineralised samples were then rinsed with deionised water and gelatinised for 48h at 70°C in a solution containing three drops of 0.5M HCl in deionised water to make a weakly acidic solution (pH 3). These residues were then filtered using an E-zee-filter TM (Elkay, UK), followed by ultrafiltration to isolate the >30 kDa fraction, ensuring that only the longest collagen chains were used for analysis. Finally, the purified filtrate was frozen and lyophilised, and resulting samples were weighed into tin capsules for carbon and nitrogen isotope analysis. Isotopic measurements were performed in duplicates using EA/CF-IRMS by Iso-Analytical (Cheshire, UK) with analytical precision of $\pm$0.2 \% (1-sigma). Integrity of the archaeological proteins preserved in the specimens was established using standard collagen quality control indicators, which all samples met (Ambrose 1990; van Klinken 1999).

\subsection*{Stable Isotope Analysis Results}

The results were plotted alongside the 28 previously sampled dog specimens from (E. McManus-Fry et al. 2016) (Fig. S17; Table S4). The patterns of isotopic variation indicate that the protein in the diet of the wolf specimen AL2729 was more terrestrially-based than the dogs at Nunalleq.

Interpretation of this result suggests that the wolf specimen was not living in a domestic context alongside the dogs at Nunalleq which were provisioned with a diet higher in marine proteins (likely salmon). Instead, this specimen likely represents a wild wolf, possibly killed to prevent predation, or for its fur, further evidence for the use of both dogs and wolves at Arctic sites. For specimens that were suitable for both GMM and isotope analysis, a comparison of the relationship between shape and size and diet was performed, but results indicated no relationship between these morphological variables and the diet of the Nunalleq dogs.

\section*{Archaeological Site Descriptions}

Below is a description of the archeological sites where canid remains were sampled for this study. Further information about the specific samples can be found in Tables S1 and S3.

\subsection*{North American Paleo-Inuit sites}

\subsubsection{Agvik, Banks Island, Northwest Territories}*

Extract ID: OKRN-1:TUR

Agvik (OkRn-1), also referred to as the Fish Lake site, is located on the southern coast of Banks Island overlooking Amundsen Gulf (Hodgetts and Eastaugh 2017) ; it was initially mapped and tested in 2009 (Arnold 2010). A juvenile dog radius was sampled from faunal remains recovered from a test unit at the head of a gully, which is stored at the Prince of Wales Northern Heritage Centre in Yellowknife. Associated caribou bones were submitted for radiocarbon dating: 420$\pm$40 BP (Beta-264000) and 380$\pm$40 BP (Beta-263998). The results fit with the estimated age of the site based on dwelling types and artifacts at AD 1350–1550 (Hodgetts and Eastaugh 2017).

\subsubsection{Angyaruaq (Jungjuk) Port Site, Middle Kuskokwim}*

Extract ID: SLT-094-09-00507

The Angyaruaq (49-SLT-094) is located in southwestern Alaska near the village of Crooked Creek along the north bank of the Kuskokwim River, approximately 230 km northeast of Bethel. Angyaruaq is the Yup'ik placename roughly translated as "toy canoe" that figures prominently in oral accounts of Crooked Creek's origins. The site, excavated by Northern Land Use Research, is composed of three loci with multiple features. The canid remains come from a semi-subterranean house pit, Feature 1, in Locus 1. Feature 1 dates between AD 140-320 (1810-1630 cal BP) and has an artifact assemblage that shows affinities to the Norton tradition (Hays et al. 2012).

\subsubsection{Cape Addington Rockshelter, Alaska}*

Extract ID: 49-CRG-188

Cape Addington Rockshelter (49-CRG-188) is located on Noyes Island in the outer Prince of Wales Archipelago of southeast Alaska. Dates range from ~2000 to 500 cal BP. Throughout the site's 1500-year

occupational history, the site was used seasonally. The most common vertebrates were deer, harbor seal, halibut, salmon, Pacific cod, and eagle, with fewer remains of Steller sea lion, northern fur seal, and a variety of seabirds. California mussels were also heavily used. Dog remains were limited to three specimens (Moss 2004); a premolar tooth from Trench 1, Unit 2, Level III-B (ca. AD 450-920) was tested in this study for mtDNA (Moss 2004).

\subsubsection{Cape Krusenstern, Kotzebue Sound, Alaska}*
Extract ID: CK-H10-H1, CK-H10-M1, CK-H22-P1, CK-H37-M1, CK-H40-C1, CK-H40-C2 ; P96/2013KMG-C.Krusenstern-A1

Cape Krusenstern was designated as an Archaeological District (49-NOA-002) in 1973 because it contains a historical record of human habitation of these beach ridges for more than 5,000 years. Discovery of this rich archaeological record, known locally as Sealing Point, came late in the 1958 field season for J.L. Giddings (J. Louis Giddings and Anderson 1986). Excavations were undertaken under the leadership of Giddings from 1959 to 1961, and then after his death in 1964, Anderson returned to continue excavation in 1965 on the Ipiutak beach ridges (J. Louis Giddings and Anderson 1986). All artifacts and faunal remains, with the exception of a portion of the Old Whaling locality are housed at the Haffenreffer Museum of Anthropology at Brown University.

A total of nine Ipiutak dwellings were excavated across four beach ridges at Cape Krusenstern between 1959 and 1965 and were the focus of Doug Anderson's PhD dissertation (J. Louis Giddings and Anderson 1986). C. Darwent reanalyzed the faunal remains from three of the Ipiutak houses—a sample of one house from each of three house clusters: House 10, House 37, and House 40. The houses were generally rectangular with driftwood roof posts and sod and wood walls. The mandible of a juvenile dog and humerus of an adult was sampled for DNA from House 10, which dates to AD 690–950; an adult mandible was selected from House 37, which dates to AD 390–780; and two crania from House 40, which dates to AD 390–780.

Five semi-subterranean, lobed dwellings and five shallow, tent-like dwellings were excavated at the Old Whaling locality between 1959 and 1960 (J. Louis Giddings and Anderson 1986). Radiocarbon dates from recent excavations at Old Whaling indicate the site was likely inhabited between 1150 and 850 BC (J. Darwent and Darwent 2005). Analysis of the faunal assemblage from Old Whaling stored at the Haffenreffer Museum (Christyann M. Darwent 2006)(Darwent 2006) revealed two *Canis* remains. A phalanx was selected by C. Darwent for DNA testing in 2012. The following year S. Brown traveled to the Natural History Museum in Copenhagen to find dog samples from Greenland; to our surprise she discovered approximately 120 bones from Old Whaling. In 1962–63, Giddings spent a sabbatical year at the National Museum of Denmark, and it appears that faunal material from excavations at Cape Krusenstern were brought with Giddings to Copenhagen for identification by zoologist Ulrik Møhl and then remained behind when he returned to Brown University. An astragalus was selected from 12 small *Canis* fragments in a bag labeled House 22 (P96/2013KMG-C.Krusenstern-A1). The Haffenreffer Museum was not aware of their existence until this project began.

\subsubsection{Choris Peninsula, Kotzebue Sound, Alaska}*

Extract ID: CP-H2-H1(CP2-1), CP-H2-H2(CP2-2), CP-H2-H3(CP2-3), CP-H2-H4(CP2-4), CP-H3-F1 (CP3-1), CP-H3-F2 (CP3-2)

Choris Peninsula is a finger-like projection extending into Eschscholtz Bay on the eastern shore of Kotzebue Sound. The Choris site (49-SLK-007) was identified in 1956, and excavated in 1957 under the direction of J.L. Giddings (J. L. Giddings 1957; J. Louis Giddings and Anderson 1986). It is considered the type site for the Choris culture, with its most defining characteristics being the remains of three large ovoid (communal?) dwellings and the presence of pottery (J. L. Giddings 1957; J. Louis Giddings and Anderson 1986). Five AMS radiocarbon dates were run on caribou bone from Houses 1–3, and Area 3 of the Choris site and cluster between 750 and 400 BC (Christyann M. Darwent and Darwent 2016). Faunal remains from this site were reanalyzed by C. Darwent (Christyann M. Darwent and Darwent 2016), and from this analysis four *Canis* left humerii were selected from House 2 and two right femora from House 3 were selected for DNA testing.

\subsubsection{Chugachik Island, Upper Kachemak Bay, Cook Inlet, Alaska}*
Extract ID: SEL-033-0057, SEL-033-1108, SEL-033-1219, SEL-033-1505, SEL-033-2250, SEL-033-2269, SEL-033-2519, SEL-033-2520, SEL-033-2907

The Chugachik Island Site (49-SEL-033) is located on the southern end of Chugachik Island in the Kachemak Bay State Wilderness Park, in Upper Kachemak Bay, Cook Inlet in southcentral Alaska. Karen Workman led an excavation in 1974 in association with the Alaska State Division of Parks and the Youth Conservation Corps (K. W. Workman 1975, 1978). Further excavations were conducted during University of Alaska Anchorage (UAA) field schools led by William Workman and John Lobdell in 1977 and 1981(K. W. Workman 1978; Lobdell 1980; W. B. Workman, E., and Workman 1980). Additional excavations were led by UAA-affiliate Peter Zollars in 1982 and 1984 (Zollars 1982; W. B. Workman and Zollars. 2002).

The Chugachik Island site was found to consist of an extensive and complex stratified shell midden, with the main component dating to the Kachemak Tradition. Samples of charcoal and birchbark basketry collected in 1974 yielded dates of 1705$\pm$65 radiocarbon years and 2310$\pm$65 radiocarbon years, respectively . Five samples collected in 1977 and 1981 yielding dates ranging from 1475$\pm$70 radiocarbon years to 2740$\pm$75 radiocarbon years (W. B. Workman, E., and Workman 1980; W. B. Workman 1998). The basal component was found to date to the Arctic Small Tool Tradition

During the 1974 excavation, a canid burial was recovered from a depth of 160 cmbs and additional individual candid skeletal elements were recovered from elsewhere in the site (K. W. Workman 1975). During the 1977 excavation, 126 canid skeletal elements, including another burial, were recovered ((Lobdell 1980):129) determined that “none of the recovered canid limb bones were long or large enough to be other than domestic dog of a stature much smaller than wolf.” Additional canid remains were recovered during later excavations; all but one of the samples cited in this paper were recovered in 1981.

A dog tooth from SEL-033-057, Level 26 was submitted by C. Darwent for radiocarbon dating by Arizona AMS (AA10285: 2723 $\pm$ 51 BP, d13C: -12.3‰), along with a dog rib from level 12

(AA102986: 2274 $\pm$ 49, d13C -13.3‰). Given the clear marine signature of the carbon isotope values for these specimens, a marine calibration was used to compensate for the reservoir effect using Calib 7.1 (Stuiver, Reimer, and Reimer 2019). Thus, these dog remains date to between ca. 700 BC and AD 200, which is consistent with the cultural affiliation of Kachemak.

\subsubsection{Coffman Cove}*
Extract ID: 49-PET-067

The Coffman Cove site has undergone significant damage over the years, and much of the site has been destroyed. Excavations in 2006 aimed to mitigate the adverse effects of transfer of the site from State to City ownership (Moss et al. 2016). The site is located on the northeast side of Prince of Wales Island in southeast Alaska. It was occupied between 5500 and 1000 cal BP. The most abundant fish were salmon, Pacific cod, sculpin, rockfish and herring. Harbor seal and deer were the most common mammals, and scoters, the most common bird. Dog remains were limited to 59 specimens. One tooth sample was tested in this study for mtDNA.

\subsubsection{Coffman Cove Ferry Terminal}*
Extract ID: 49-PET-556

The Coffman Cove Ferry Terminal site is located on the northeast side of Prince of Wales Island in southeast Alaska (Reger et al. 2017). It is not as extensive as 49-PET-067, but it too has undergone damage due to construction (of a ferry terminal). It is located just 600 m away from 49-PET-067. 49-PET-556 was occupied between 3000-2000 cal BP. Although this site was occupied at the same time as one of the components at 49-PET-067, the faunal assemblages differ, with a greater emphasis on Pacific cod at the Ferry Terminal site. A total of 19 dog specimens were found at the site; a single tooth was tested for mtDNA.

\subsubsection{Cove Site, Knud Peninsula, Ellesmere Island, Nunavut}*
Extract ID: SGFM-5

Sites in the Bache Peninsula region were investigated by members of the Arctic Institute's Ellesmere Island Research Project under the direction of Peter Schledermann between 1977 and 1982 (Schledermann 1990). The Cove site (SgFm-5) is located on the northeast end of Knud Peninsula, Ellesmere Island; the main structure at this site is a Late Dorset (AD 800–1100) communal longhouse defined by a "wall" of boulders and stones (Schledermann 1990). Nearly the entire 14.5 x 5.5 m structure was excavated over four field seasons. Faunal remains from the dwellings' interior were dominated by birds, but they also included a mandibular carnassial tooth of an adult *Canis*, which was sampled for mtDNA at the Canadian Museum of History.

\subsubsection{Deering, Seward Peninsula, Kotzebue Sound, Alaska}*
Extract ID: TRF-01-23(ZMK 113a/1950-P27), TRF-01-24(ZMK 113a/1950-P28), TRF-01-27(ZMK 113a/1950-P31), P97/2013KMG-Deering-H1(ZMK 113a/1950-P31), P98/2013KMG-Deering-C2(ZMK 113a/1950-P27), P99/2013KMG-Deering-C3(ZMK 113a/1950-P28)

Deering is a northwestern Native Alaskan community situated on a spit within Kotzebue Sound on the northern coast of the Seward Peninsula. Several archaeological sites are located along the Deering spit that make up the Deering Archaeological District (Alaska Heritage Resources Survey number KTZ-000169) (Sloan 2014; Peter M. Bowers and Williams 2009). One such site, KTZ-00023, in the district contains an Ipiutak *qalegi* or *qargi,* ceremonial house, discovered and excavated by Helge Larsen in 1949 (VanStone and Lucier 1974).

This unusually large ceremonial structure, or 'men's house', was built with a foundation of non-locally sourced spruce (Larsen 2001). The site, dated to between AD 700 and 900, is situated just to the east of later Western Thule houses, which are dated to between AD 1030 and 1640 (Sloan 2014; Larsen 2001). The antechamber of the qalegi contained preserved organic materials such as wood shavings, animal bones, and dog faeces (Larsen 2001). Additionally, remains of sledges were preserved at the site which indicate that the sledge style used by the Ipiutak people at Deering was that of a built-up sledge which had previously not been seen in North America and was more reminiscent of similarly aged sledges in Sweden, Finland, and western Siberia (Larsen 2001). Larsen argues that the combination of dog remains and faeces as well as sledge materials presents evidence for the possibility of the use of the dogs as traction to pull the sledges (Larsen 2001). The left petrous bone was sampled from three individuals for the genetic portion of this study derived from the third and fourth layers of the site, these layers have been radiocarbon dated to approximately 1250 years before present (Larsen 2001). Two of the specimens (TRF-01-23:ZMK 113a/1950-P27, TRF-01-24:ZMK 113a/1950-P28) were directly radiocarbon dated for this study at the University of Oxford Radiocarbon Accelerator Unit. Both specimens returned dates which postdate the Ipiutak occupation and fall within the Thule era, see Table S5. Direct radiocarbon dating of sample TRF-01-24:ZMK 113a/1950-P28 had greater uncertainty as a result of the reservoir corrections required. TRF-01-24:ZMK 113a/1950-P28 showed a marine reservoir offset, therefore for calibration a mixed curve was used (INTCAL13 and MARINE13, as recommended by the University of Oxford Radiocarbon Accelerator Unit, after (Reimer et al. 2013). The marine proportion was estimated using interpolation, two endpoints of -21 per mille and -12.5 per mille, with a result of 74\% marine component with 10\% uncertainty.

\subsubsection{Icebreaker Beach Site, Truelove Lowlands, Devon Island, Nunavut}*
Extract ID: QKHN-13:1654

The Icebreaker Beach site (QkHn-13) is an early Predorset occupation located in the Truelove Lowland region of northern Devon Island. Radiocarbon dating of unburned caribou bone dated to approximately 2500–1900 BC (Helmer 1991). A large canid "quite possibly dog," was reported from this site. A distal humerus of an adult *Canis* was pulled from the Feature 15 (indeterminate feature type) faunal remains at the Canadian Museum of History and sampled for mtDNA.

\subsubsection{Kaleruserk, Igloolik, Nunavut}*
Extract ID: TRF-01-22(ZMK-123c/1955), P80/2013KMG-Igloolik-F1, P87/2013KMG-Igloolik-F2

Igloolik is an island just north of Canada's Melville Peninsula, within the Foxe Basin, and is associated with a Paleo-Inuit settlement (Maribeth Suzanne Murray 2005). The site was originally explored in the 1950s by Jørgen Meldgaard, and dog remains were recovered from the Káleruserk locality, also known as Perry Hill. These remains are assumed to be pre-Dorset because they are associated with deposited dated to the early fourth millennium B.P. (Morey and Aaris-Sørensen 2002; Jørgen Meldgaard 1960; Maribeth Suzanne Murray 2005). Up to as many as 29 intact and disturbed tent rings were identified on the site (Maribeth Suzanne Murray 2005). A crania from Igloolik was analysed for mtDNA analysis for this study, and two femora were also sampled for mtDNA.

\subsubsection{Kugzruk Island, Lopp Lagoon, Seward Peninsula, Alaska}*
Extract ID: KZ-7005-1

Kugzruk Island (49-TEL-149) is located northeast of Cape Prince of Wales on Seward Peninsula. Two sites were discovered by J. L. Giddings in the barrier beach area to the front of Lopp Lagoon, Agualak Island and Kugzruk Island, in 1959. On Kugzruk Island, they located two Norton-period house pits and a possible third that had eroded into the adjacent pond; he named this cluster Site 1: Kugzruk Pond, Houses 1 and 2 (J. Louis Giddings and Anderson 1986). Wood from the pond excavations produced a date of 2306 ± 38 BP (J. Louis Giddings and Anderson 1986), which calibrates to 400–200 BC and fits with the presumed Norton age based on house shape and artifacts. Fauna is described as coming from the "bottom of a small pond adjacent to the site and partially excavated from the mound itself" according to notes with the collection at the Haffenreffer Museum of Anthropology, Brown University. The bones, which included a phalanx, three metatarsals, five metacarpals, and a distal humerus were all stamped with the number 7005. A dog maxilla from this collection was sampled for DNA.

\subsubsection{Lagoon, Banks Island, Northwest Territories}*
Extract ID: OJRL-3:534

The Banks Island Archaeological Research Project, directed by Charles Arnold, focused on the excavation of the Lagoon site (OjRl-3) in 1976 and 1977 (Arnold 1980)(Arnold 1980). Area A at this site included a buried cultural component with dwelling remains and associated midden areas with excellent organic preservation. In addition to faunal remains and osseous artifacts, a wooden fragment of a bowl, a skin kamik and bag, a leather thong, knotted muskox hair, and knotted baleen. Subsistence at this site was focused on geese, ptarmigan, ringed seal, and to a lesser extent muskox and arctic fox.
This site, however, is best known for the remains of three *Canis* vertebrae, 12th thoracic to 1st lumbar; "the dorsal surface of the spinous process of each vertebra is flattened and lipped" ((Arnold 1979):264), and suggests the animal may have carried a pack. A small fragment of bone was collected at the Prince of Wales Northern Heritage Center from the elements for mtDNA sampling (OJRL-3:534; 982.50.534) Radiocarbon dates on three muskoxen scapulae from Area A yielded dates indicating this early Dorset site was likely occupied from 825–200 BC.

\subsubsection{Mill Island, Baffin Island, Nunavut}*
Sample ID: HMCZ_47059

Mill Island is located in Hudson Strait off the southwestern coast of Baffin Island. The Dorset site excavated there in the 1950s consisted of two semi-subterranean houses and their associated midden (Park 1993). While the architecture at the site is reminiscent of Thule type, all of the recovered artifacts were identified as Dorset, with the archaeologists concluding there was no contact between Dorset and Thule peoples at this site, nor any evidence of long term re-occupation of the structures, although it is likely that the Thule remodelled the house at some point. A single lower M1 from this site was included in this analysis.

\subsubsection{Nanook, Kimmirut (Lake Harbour), Baffin Island, Nunavut}*
Extract ID: KDDQ-9

The Nanook site (KkDq-9) is located near Kimmirut (Lake Harbour) in the Tanfield Valley of southern Baffin Island, Nunavut, and was excavated under the direction of Moreau Maxwell in 1962 (Maxwell 1973). The buried remains of an amorphous stone and sod structure with a slight central depression (approximately 2 x 3.5 m) were unearthed in a grassy gully described as Nanook, Component 2 (Maxwell 1973). Artifacts are characteristic of eastern Arctic early to middle Dorset sites, and calibrated radiocarbon dates suggest the site was occupied from approximately 420 BC to AD 120 (Maxwell 1973).

This site unearthed the first Dorset-age dog remains; measurements of the partial cranium have been suggested to fall within that of a large Inuit dog (Cleland and Haag 1973). The prominent sagittal crest is indicative of a male. Cleland appears to have refit and glued the fragmentary cranial pieces to undertake his analysis. Dark discoloration of the bone was from soil staining, as there were no signs of burning. A sample of the P4 root was taken at the Canadian Museum of History for mtDNA.

\subsubsection{Nipisat I, Sisimiut District}*
Extract ID: TRF-01-14(ZMK 136/1989), P86/2013KMG-Nipisat-H1, Pxx/2013KMG-Nipisat-H2

The Saqqaq site, Nipisat I, is located in the Sisimiut District of western Greenland (Morey and Aaris-Sørensen 2002; Gotfredsen and Møbjerg 2010). Nipisat I was continuously occupied between 4000 and 2500 BP, the Saqqaq site with the longest and well dated occupation (Morey and Aaris-Sørensen 2002). Excavations at Nipisat I yielded 17 bones and teeth belonging to Saqqaq dogs (Morey and Aaris-Sørensen 2002; Gotfredsen and Møbjerg 2010). The dog remains are all derived from the third phase of the site, during the most recent occupation, from 3000 to 2500 BP (Morey and Aaris-Sørensen 2002). The samples taken for this study were drilled from a canine tooth, and two humeri.

\subsubsection{North Point, Gulf of Alaska}*
Extract ID: SUM-025-74

The North Point site (49-SUM-025) is located in a small cove on the south shore of Port Houghton, approximately 35 km northeast of Kupreanof Island and 25 kn east of Admiralty Island, southeast Alaska (P. M. Bowers and Moss 2001; P. M. Bowers et al. 2001). The site is also known as the “North Point Wet site” due to it being situated in the intertidal zone, and partial inundation of water creating an excellent medium for organic preservation. The site was investigated in 1994, 1995 and 2005 by Northern Land

Use Research Inc. and US Forest Service archaeologists. The occupations at the North Point site dates between 2200-2800 BP (300 to 950 BC).

\subsubsection{Nunguvik, Navy Board Inlet, Baffin Island, Nunavut}*
Extract ID: PGHB-1:1535

The Nunguvik site (PgHb-1) is located on the west shore of Navy Board Inlet on northeastern Baffin Island. This is an extensive site, with more than 80 houses, that spans early and late Dorset and early and late Thule. Father Guy Mary-Rousselière excavated the remains of several house structures from both Dorset and Thule in the early 1970s (Mary-Rouselière 1976, 1979). The cranium of an old, female dog was recovered from House 71, a rectangular Dorset house that dates to AD 130–580. Mary- Rousselière (Mary-Rouselière 1976) reports that 700 bones were found during the 1971 season, primarily caribou and bird, but also seal and other mammals, including dog; he comments that the dog is "perhaps intrusive," suggesting that it may, in fact, be from later Thule occupation of the site. The cranium was missing its teeth post-mortem; however, she appeared to have experienced loss of at least one tooth pre-mortem with alveolar resorption, and she had clearly suffered a blow to the frontal. Because the sample failed to yield DNA, it was not submitted for direct radiocarbon dating.

\subsubsection{Point Hope, North Slope, Alaska}*
Sample ID: NMNH_241525

Point Hope is the type site for the Ipiutak culture, located on the western edge of Alaska, above the Arctic circle on the Chukchi sea. The site has been dated to between 500 - 900 AD, and consists of both seasonally occupied dwellings and a large cemetery (Hill 2018). While much of the excavations have focused on the human remains from the cemetery, a significant number of faunal remains were also collected, including the remains of a single dog used in this analysis, which includes a crania, a mandible and a lower M1.

\subsubsection{Port Joli , Nova Scotia}*
Extract ID: ALDF-30A-1568, ALDF-30A-911, ALDF-30A-912

These dog samples, which serve as a North American Arctic outgroup, are derived from the Area A shell midden at AlDf-30, a small Middle Maritime Woodland Period site in Port Joli, Nova Scotia (M. W. Betts, Burchell, and Schöne 2017). Dating to ca. AD 575-660, the site is located approximately 250 meters from the high tide mark, on a knoll in the middle of a small fen. The appendicular bones of the dog were found articulated, but no evidence of an intentional interment was present (the mandible was found in an adjacent level). The kitchen midden abutted multiple concatenated living surfaces (likely wigwam floors), and a sweat house structure. A phalanx and two premolars were sampled for mtDNA.

\subsubsection{Port Refuge, Cape Grinnell, Devon Island, Nunavut}*
Extract ID: RBJU-1:1107, RBJU-1:COLD

Archaeological research was undertaken by Robert McGhee, between 1972 and 1977, at the Port Refuge site (RbJu-1) situated along Port Refuge Bay on the south coast of Grinnell Peninsula, Devon Island (McGhee 1979). One component of the site, the Cold component, is located on a beach ridge between 22 and 25 masl approximately 600 m inland from the northern shore of Port Refuge Bay; faunal remains were excavated and collected from 12 features, which included a Canis humerus midshaft (RBJU-1:COLD) sampled for mtDNA. At the Upper Beaches component, located at the southwestern entrance to Port Refuge Bay, faunal material was recovered from three features, and two putative dog remains were identified. Both were examined at the Canadian Museum of History and were identified as a juvenile maxilla with P3 (P4 erupting), and a juvenile mandible with canine and P3 (P4 erupting). As these likely came from a single individual only one (RBJU-1:1107) was sampled for mtDNA. All materials from this site are housed at the Canadian Museum of History.

\subsubsection{Qajaa, Disko Bay}*
Extract ID: TRF-01-64(ZMK350/1982-P79), ZMK350/1982-P78, P81/2013KMG-Qaja-H1, P83/2013KMG-Qaja-T1

Qajaa is an archaeological site from the Saqqaq culture located on the eastern side of Disko Bay in western Greenland (Morey and Aaris-Sørensen 2002). Due to the location of the site along the shore much of the site has been eroded by the sea (Møhl 1986). The site is estimated to have accumulated over approximately 900 years, between 3600 and 2700 BP (J. Meldgaard 1983; Morey and Aaris-Sørensen 2002). Excavations at Qajaa uncovered 21 dog bones in the earliest phases of the site's habitation (Morey and Aaris-Sørensen 2002; Møhl 1986).

\subsubsection{Qeqertasussuk, Disko Bay}*
Extract ID: TRF-01-17(ZMK 70/1983), P82/2013KMG-QT-H1, P84/2013KMG-QT-T1, P85/2013KMG-QT-H2

Qeqertasussuk lies in the south extreme of Disko Bay in western Greenland and like Qajaa is an archaeological site attributed to the Saqqaq culture (Grønnow 1994, 2017). The Saqqaq occupation of Qeqertasussuk occurred between 3900 and 3100 BP (Morey and Aaris-Sørensen 2002; Grønnow 1994). 41 dog bones were excavated from the site from across the 800 years of site usage (M. Meldgaard 2004). Among the dog remains there were several acticulated dog limbs, some of which display cut marks on the bones as a result of processing by humans (M. Meldgaard 2004). However, despite the generally excellent organic preservation no sledge related artefacts were discovered at the site, indicating their rare or complete absence at Qeqertasussuk, in line with all other Saqqaq sites (Morey and Aaris-Sørensen 2002; M. Meldgaard 2004). The samples used in this study were taken from three dog humeri and one tibia.

\subsubsection{Rastoden, Bache Peninsula, Ellesmere Island, Nunavut}*
Extract ID: SFFL-10

Sites in the Bache Peninsula region were investigated by members of the Arctic Institute's Ellesmere Island Research Project under the direction of Peter Schledermann between 1977 and 1982. The Rastoden site is located on the southeastern tip of Thorvald Peninsula, south of Bache Peninsula, on a beach ridge 26 masl (Schledermann 1990). Bone from this late Predorset site (1200–1300 BC) was surface collected

from Camp Area B, which included several circular tent-ring features and midden areas. Because excavation was not undertaken, the faunal assemblage is small. A maxillary P4 from Area B-4 was sampled for mtDNA at the Canadian Museum of History.

\subsubsection{Sermermiut, Ilulissat}*
Sample ID: HMCZ_10539

The Sermermiut Valley is located in the World Heritage Site of Ilulissat Icefjord 2 kilometres south of the centre of the Disko Bay town of Ilulissat, Greenland. Excavations and survey since the 1950s have uncovered houses and middens from Saqqaq, Dorset and Thule periods in this region (Therkel Mathiassen 1958). Due to preservation conditions for organic materials on the site, bone and wood from the Saqqaq and Dorset layers did not survive (Therkel Mathiassen 1958). A lower M1 from a Thule era specimen was analysed in this study. This specimen is described by G. Allen (Allen 1920) as the type specimen used for his analysis of an 'Eskimo Dog', which was "exhumed by Dr. AI. P. Porsild from an old village site at Sermermiut", and is now held at Harvard's Museum of Comparative Zoology.

\subsubsection{Tikilik, Igloolik Island, Nunavut}*
Extract ID: NIHF-4:249C, NIHF-4:3422C, NIHF-4:4BSP, NIHF-4:9BSM

Tikilik (NiHf-4) is a Late Dorset site located on a peninsula jutting out along the southwestern coast of Igloolik Island. The site was excavated by Susan Rowley in 1993–94, and faunal remains from the Paleo-Inuit components were identified by Maribeth Murray. The date of this site was estimated between 1200 and 800 BP (AD 750–1150) based on beach-ride elevation, Dorset parallel harpoon heads, and typical Late Dorset-style artwork (Maribeth S. Murray 1996). However, there are Thule components to this site according to notes associated with the original excavation bags and unpublished documents for this site at the Prince of Wales Northern Heritage Centre in Yellowknife—discussed below. Bone recovered from the lower black soil layer is Late Dorset, and bone from the upper peat layer is associated with later Thule occupation. A cranial fragment (Unit -24N/9E; NIHF-4:249C), canine tooth (Feature 22, Unit -34N/-22E; NIHF-4:3422C), pelvis fragment (Feature 4, Unit 1N/-5E; NIHF-4:4BSP), and mandibular fragment (Feature 9, Unit 56N/-7E; NIHF-4:9BSM) were collected from the Prince of Wales warehouse storage facility for mtDNA sampling.

A portion of the dog mandible recovered from the lower black soil layer of Feature 9 (NIHF-4:9BSM) was submitted by C. Darwent for radiocarbon dating (Beta-349748: 960 $\pm$ 30 BP, d13C -17.2‰, d15N -6.9‰). The calibrated age for this specimen is AD 1020–1160, which fits with the Late Dorset designation based on stratigraphy and artifact styles.

\subsubsection{Uyak, Kodiak Island, Alaska}*
Sample ID: HMCZ_38342, NMNH_257528, NMNH_257533, NMNH_257534, NMNH_257535, NMNH_257537, NMNH_257540, NMNH_257542, NMNH_259180, NMNH_259181, NMNH_259182_1, NMNH_259182_2, NMNH_259182_3, NMNH_259182_5, NMNH_259182_6, NMNH_260310, NMNH_260311, NMNH_26033, NMNH_260318, NMNH_260320, NMNH_260321, NMNH_Uyak_C21753_3, NMNH_Uyak_C25, NMNH_Uyak_C28, NMNH_Uyak_C9, NMNH-256720-1, NMNH-258782-2, NMNH-258782-3, NMNH-258782-5, NMNH-258782-6, NMNH-Uyak-

C018, NMNH-Uyak-M006, NMNH-Uyak-M008, NMNH-Uyak-M012, NMNH-Uyak-M014, NMNH-Uyak-M015, NMNH-Uyak-M016, NMNH-Uyak-M018, NMNH-Uyak-M019, NMNH-Uyak-M021, NMNH-Uyak-M023, NMNH-Uyak-M026, NMNH-Uyak-M030, NMNH-Uyak-M051, NMNH-Uyak-M054, NMNH-Uyak-M055, NMNH-Uyak-M056, NMNH-Uyak-M057, NMNH-Uyak-M062, NMNH-Uyak-M071, NMNH-Uyak-M073, NMNH-Uyak-M075, NMNH-Uyak-M085, NMNH-Uyak-M086, NMNH-Uyak-M087, NMNH-Uyak-M088, NMNH-Uyak-M090, NMNH-Uyak-M091, NMNH-Uyak-M092, NMNH-Uyak-M095, NMNH-Uyak-M102, NMNH-Uyak-M108, NMNH-Uyak-M109, NMNH-Uyak-M111, NMNH-Uyak-M117, NMNH-Uyak-M118, NMNH-Uyak-M121, NMNH-Uyak-M123, NMNH-Uyak-M124, NMNH-Uyak-M127, NMNH-Uyak-M130, NMNH-Uyak-M131, NMNH-Uyak-M135, NMNH-Uyak-M139, NMNH-Uyak-M140, NMNH-Uyak-M143, NMNH-Uyak-M145, NMNH-Uyak-M147, NMNH-Uyak-M156, NMNH-Uyak-M159, NMNH-Uyak-M161, NMNH-Uyak-M162, NMNH-Uyak-M166, NMNH-Uyak-M168, NMNH-Uyak-M169, NMNH-Uyak-M171, NMNH-Uyak-M177, NMNH-Uyak-M178, NMNH-Uyak-M200, NMNH-Uyak-M201, NMNH-Uyak-M202, NMNH-Uyak-M203, NMNH-Uyak-M204, NMNH-Uyak-M206, NMNH-Uyak-M207, NMNH-Uyak-M208, NMNH-Uyak-M209, NMNH-Uyak-M211, NMNH-Uyak-M212, NMNH-Uyak-M213, NMNH-Uyak-M214, NMNH-Uyak-M215, NMNH-Uyak-M216, NMNH-Uyak-M217, NMNH-Uyak-M219, NMNH-Uyak-M22, NMNH-Uyak-M220, NMNH-Uyak-M221, NMNH-Uyak-M222, NMNH-Uyak-M223, NMNH-Uyak-M224, NMNH-Uyak-M225, NMNH-Uyak-M226, NMNH-Uyak-M227, NMNH-Uyak-M228, NMNH-Uyak-M229, NMNH-Uyak-M230, NMNH257526, NMNH257536, NMNH257538, NMNH257539, NMNH257541, NMNH260312
Extract ID: AL3198, 560028-01, 560028-02, 560028-03, 560028-04, 560028-05, 560028-06, 560028-11, 560028-22, 560028-28, 560028-07, 560028-08, 560028-09, 560028-10, 560028-12, 560028-13, 560028-14, 560028-15, 560028-16, 560028-17, 560028-18, 560028-19, 560028-20, 560028-21, 560028-23, 560028-24, 560028-25, 560028-26, 560028-29, 560028-30, 560028-27

The Uyak site (49-KOD-145) is a substantial prehistoric midden at least three meters deep, covering hundreds of acres and located on the western side of Kodiak Island, Alaska. The site was excavated by physical anthropologist Aleš Hrdlička from 1933-1936 for the United States National Museum (now the National Museum of Natural History), and consisted of house structures and hearths, stone and organic artefacts, human remains, and a large faunal assemblage (Steffian 1992; Heizer 1956). Hundreds of only the "best preserved" dog remains were collected, with excavators noting they were ubiquitous at the site. Excavation of the specimens was poorly recorded. Also, it is unknown in what context the dog remains were found, and whether they represent intentional burials, midden/refuse deposits, or some combination of deposition types. Thirty one specimens were sampled for DNA analysis, while 97 mandibles, 91 lower M1s and 15 crania from 133 unique specimens were used for the GMM analysis.

\subsubsection{Westwind, Lake Hazen, Ellesmere Island, Nunavut}*
Extract ID: TJFD-3:1

The Westwind (TjFd-3) site is an early Paleo-Inuit (Independence I) site related to the initial migration of humans into the high arctic regions of Canada and Greenland (ca. 3900 BP or 1950 BC). Survey of this area was undertaken by Patricia Sutherland within the boundaries of the Northern Ellesmere Island National Park Reserve between the late 1970s and the early to mid-1990s (Sutherland 1996). The site is located adjacent to Lake Hazen in the Eureka Upland, which is an intermontane plateaus is surrounded by extensive ice caps and mountains that range over 2700 m high. Faunal remains from this site and a nearby

site similar in age were identified by Darlene Balkwill of the Canadian Museum of Nature, who reported that the vast majority of the remains are of muskox, hare, and fox (Christyann Marie Darwent 2002). Two Canis canines were reported from Feature 5, the most complete of which was sampled for mtDNA at the Canadian Museum of History.

\subsection*{North American Inuit Sites}

\subsubsection{Aftermiut, Qaqortoq}*

Extract ID: P95/2013KMG-Aftermiut-M1, Pxx/2013KMG-Aftermiut-M2, Pxx/2013KMG-Aftermiut-M6, P78/2013KMG-Aftermiut-M5, P79/2013KMG-Aftermiut-M7

The Thule Inuit site of Aftermiut (ZMK 13/1934) is located at Qaqortoq (formerly Julianehåb) in the southern tip of Greenland. It was investigated by T. Mathiassen and E. Holtved in 1934 (T. Mathiassen and Holtved 1936), and is estimated to date to ca. AD 1350-1650 (Gullov 1997). A total of five dog mandibles were sampled for mtDNA at the Natural History Museum in Copenhagen.

\subsubsection{Ahteut, Kobuk River, Alaska}*

Extract ID: UA1-1941-1856

Ahteut (49-XBM-003) is a village site located in northwest Alaska along the middle portion of the Kobuk River between the present-day villages of Ambler and Kiana. The site has designated north and south components consisting of over 100 house features. James Louis Giddings excavated 12 houses at the site in the 1940s and dated a single house (3n) to a range between AD 1202 and 1250 based on five tree-ring dates (J. Louis Giddings 1952). House 10s is the only other dated feature from Ahteut. Two different caribou metapodial scrapers were selected from the museum collection, sampled, and radiocarbon dated returning nearly identical results of 800-680 cal BP (Shirar 2011). The single dog mandible sampled for DNA analysis and included in this study also originates from Giddings' House 10s.

\subsubsection{Beulah, Cape Krusenstern, Western Coronation Gulf, Northwest Territories}*

Extract ID: NCPF-12:2-18

The Beulah site (NcPf-12) is located on the west side of the Cape Krusenstern peninsula, on the mainland coast of western Coronation Gulf. The site is comprised of 12, single-roomed, ovate winter houses arranged in two clusters with the entrance tunnels oriented to the southwest (D. A. Morrison 1983). Limited excavation was undertaken in the "wall midden" of House 4, and a single dog mandible from was sampled for mtDNA at the Canadian Museum of History. Like the Clachan and Nuvuk sites, Beulah is early Thule (AD 1200–1400) in age (D. A. Morrison 1983).

\subsubsection{Bonasila, Lower Yukon River, Alaska}*

Sample ID: NMNH-254782-1, NMNH254781

The site of Bonasila (49-XHC-047) is located on the banks of the Yukon River below the village of Anvik, near the confluence of the Bonasila and Yukon rivers in central Alaska. The site was surveyed by

an expedition lead by Aleš Hrdlička, and is described as consisting of pit and tunnel dwellings with a cemetery (Hrdlicka 1928). Though the primary target of the survey was the collection of human remains, the fossilised remains of many animals were noted and collected during the survey, including dogs, though their context was not recorded. A mandible and a lower M1 from Bonasila were included in the GMM analyses.

\subsubsection{Cache Point, Richards Island, Mackenzie Delta, Northwest Territories}*
Extract ID: NHTS-2:C6-9-4, NHTS-2:C8-P1, NHTS-2:C8-P2

The Cache Point site (NhTs-2) is situated on a large bluff on the east coast of Richards Island in the mouth of the Mackenzie Delta, Northwest Territories; at least 23 semi-subterranean Thule houses were documented (T. M. A. X. Friesen 2009). To understand the history of Inuvialuit in this area, excavations were undertaken in 1996, 1998, and 1999 under the direction of Max Friesen. Diagnostic artifacts and radiocarbon dates place this site within the early Thule period at AD 1200–1300. House 6 was built of adzed logs and planks; it had a main room (the floor was compact sand rather than wooden planks), a side kitchen alcove, and an entrance tunnel (T. M. A. X. Friesen 2009). House 8 was a somewhat more complex wooden structure that had a roughly square main room, the back half which appears to have been a bench—the planks may have been removed when the house was abandoned; the floor in front of the bench was still covered with large driftwood logs (T. M. A. X. Friesen 2009; T. M. Friesen and Betts 2006). This house also had a kitchen tunnel connecting it to a kitchen, a side annex, and an external cache. Two right adult dog pelves—both carnivore gnawed—were selected from House 8, Level 3 at the Prince of Wales Northern Heritage Center in Yellowknife, and an adult femur from House 6, Level 3 was provided by Max Friesen for mtDNA sampling.

\subsubsection{Cape Espenberg, Seward Peninsula, Kotzebue Sound, Alaska}*
Sample ID: BELA-36034
Extract ID: BELA-34679, BELA-35135, BELA-36031, BELA-36034, BELA-36035, BELA-37372, BELA-37373, BELA-37374, BELA-37375, BELA-37376, BELA-37940, BELA-59973, BELA-59974, BELA-37370, BELA-37371, BELA-52965, BELA-52966, BELA-37367, BELA-37368, BELA-37369,BELA-59975, BELA-59976, BELA-59977, BELA-59978, BELA-59979, BELA-59980, BELA-59981, BELA-59982

Cape Espenberg is a long, narrow spit of land on the northwestern extent of the Seward Peninsula, the shores of which are bounded by the Chukchi Sea and Kotzebue Sound. It is part of Bering Land Bridge National Park and consists of a series of dune-covered beach ridges that jut out into the sound and have been building over the past 4000 years. Extensive mapping of the site was undertaken in 2007 and 2010 (J. Darwent et al. 2013); between 2009 and 2011, excavation of six Thule house depressions was directed by John Hoffecker and Owen Mason (Alix et al. 2015; J. Darwent et al. 2013; Norman et al. 2017). *Canis* specimens were sampled for mtDNA from across three sites at Cape Espenberg ((Sarah K. Brown, Darwent, and Sacks 2013):1)) 49-KTZ-304, Feature 12 and 21; 2) 49-KTZ-087, Feature 68A, 68B, and 87; 49-KTZ-088, Feature 33.

At KTZ-304, the oldest structure is Feature 12; a driftwood and whale-bone structure that revealed at least two to three occupations superimposed with periods of abandonment and rebuilding. It compares most closely to Birnirk structures described to the north at Barrow (Alix et al. 2015). The structure appears to have been built initially by Birnirk inhabitants, but then rebuilt or reused by early Thule occupants. Bone samples were collected from the upper levels (Level 1A-C and 2B) where associated radiocarbon dates span AD 1050–1300 (J. Darwent et al. 2013). Right mandibles were sampled from Level 1 (BELA-59976, 59977, 59978, 59980, 59982) and Level 2 (BELA-59975, 59979, 59981). Sample BELA-59981 is a large wolf mandible, which was confirmed by aDNA testing. Feature 21 at this site is a typical Thule structure built of driftwood logs and planks with a main room, a side-spur kitchen area full of burned and cemented fat (Buonasera et al. 2015), and long entrance tunnel. Based on radiocarbon dates the structure was occupied from AD 1250–1400. A wolf metacarpal (BELA-36031), three dog crania (BELA-36034, 36035, 37372), and four dog mandibles that did not match the crania (BELA-37373, 37374, 37375, 37376) from the lower levels of the tunnel, kitchen side spur, and house interior were all sampled for mtDNA.

The remains of three driftwood constructed Thule houses were excavated at KTZ-087. Features 68A and 68B were located adjacent to each other only one meter apart; 68A was excavated in 2010 and 68B in 2011; both structures radiocarbon date to AD 1450–1650 (Brown et al. 2013; Darwent et al. 2013). Feature 68A is a typical Thule dwelling with a square main room with planks forming the rear sleeping platform, a hearth area to the front, and then a 4-m long entrance tunnel. A separate external kitchen area with no connection to the house was also excavated (J. Darwent et al. 2013). A cranium and non-matching mandible from Level 3 of the floor of the tunnel/cold trap entrance were sampled (BELA-37370, 37371). Feature 68B is nearly identical in terms of the construction of the tunnel, but the main room was distinguished by a beam running perpendicular to the tunnel and a compacted sand floor visible in the wall profile. The main area of the house was full of wood-working debris, as well as bone and stone debitage, and thus has been interpreted as a communal men's house. Two crania were sampled from the tunnel floor (BELA-52965, 52966), one of which had evidence of chop marks to remove the parietal/occipital region and cut marks indicative of skinning on the frontal.

The third feature excavated at KTZ-087 was Feature 87, which radiocarbon dates to AD 1300–1450 (Sarah K. Brown, Darwent, and Sacks 2013; J. Darwent et al. 2013; Norman et al. 2017). This semi-subterranean wood-framed structure included a square main room with a wide rear platform and two narrow benches along the side walls (Norman et al. 2017). An amorphous area of burned material and cemented sand was located to the front of the house and was likely a kitchen alcove. A long tunnel served as the entrance. Faunal remains are dominated by small seal at 82\% followed distantly by caribou and dog at 3\% each (Norman et al. 2017). A left, fifth metacarpal from Level 2 of the midden area in front of the tunnel (BELA-59974) and from Level 3 of the main room/entrance (BELA- 59973) was sampled, along with a calcaneus from Level 2 outside and just to the east of the house (BELA-35135). Two modified maxillary canines that appear to have been used as pendants were also sampled: one from Level 2 (BELA-34679) at the rear of the house, and one from inside a ceramic vessel in the kitchen (BELA-37940).

Only Feature 33 was excavated at KTZ-088, which is a well-built Inupiat house dating to the 18th century based on radiocarbon dating of structural wood and caribou bone (Méreuze 2015; Sarah K. Brown,

Darwent, and Sacks 2013; J. Darwent et al. 2013). The structure was occupied pre-contact as no Russian-American trade items were recovered. Entrance to the main room was through a tunnel made of driftwood capped with a whale bone lintel. An external kitchen with thick layers of burned and cemented sand (Buonasera et al. 2015) was also excavated. Wood flooring was in place in the front of the room where a ceramic lamp also remained, but the sleeping platform planks appear to have been removed on abandonment. Three right ulnae were sampled from the tunnel (BELA-37367), kitchen (BELA-37368), and in front of the sleeping platform (BELA-37369) for mtDNA.

\subsubsection{Cape Garry, Somerset Island, Nunavut}*
Extract ID: PCJQ-5:H5-M1, PCJQ-5:H5-M3, PCJQ-5:H6-M1, PCJQ-5:H6-M2, PCJQ-5:H6-M3, PCJQ-5:H6-M4, PCJQ-5:H6-M6

Three Thule houses were excavated at Cape Garry (PcJq-5) on the southernmost extent of Creswell Bay, Somerset Island, under the direction of Allen McCartney in 1976. Faunal remains from this House 6 at this site, and House 1 at the Learmonth site, nearly 40 km to the north, were analyzed by Anne Rick (Rick 1980). House 6 was a roughly circular-shaped stone-and-sod Thule dwelling with a single platform; a small alcove had been added to the room at a later date. Two right mandibles from House 5 and five left mandibles from House 6 were sampled at the Canadian Museum of History for mtDNA.

A caribou rib from House 5 and a long-bone shaft fragment from House 6 were submitted by C. Darwent for radiocarbon dating at Arizona AMS. The results were nearly identical with calibrated ages of AD 1305–1445 (AA102351: 540 ± 45 BP, d13C -19.9‰) and AD 1300–1450 (AA102352: 528 ± 55 BP, d13C -18.4‰) respectively, which fits well with Classic Thule for this region.

\subsubsection{Cape Krusenstern, Kotzebue Sound, Alaska}*
Extract ID: CK-H4-M1, CK-H4-M2, CK-H4-M3, CK-H4-M4, CK-H4-M5, CK-H4-M6, CK-H5/7-M1, CK-H5/7-M2, CK-H6-M1, CK-H6-M2, CK-H6-M3, CK-H6-M4, CK-H6-M5, CK-H6-M6, CK-H6-M7, CK-H6-M8, CK-H6-M9, CK-H6-M10, CK-H8-M1, CK-H8-M2, CK-H8-M3, CK-H8-M4, CK-H8-M5, CK-H8-M6, CK-H8-M7, CK-H8-M8, CK-H8-M9, CK-H8-M10, CK-H8-M11, CK-H8-M12, CK-H8-M13, CK-H32-M1, CK-H32-M2, CK-H33-M1, CK-H33-M2, CK-H35-A1, CK-H50-M1

Cape Krusenstern was designated as an Archaeological District (49-NOA-002) in 1973 because it contains a historical record of human habitation of these beach ridges for more than 5,000 years. Discovery of this rich archaeological record, known locally as Sealing Point, came late in the 1958 field season for J.L. Giddings (J. Louis Giddings and Anderson 1986). Excavation of Thule-period dwellings was undertaken by Giddings from 1959 to 1961. Artifacts and faunal remains from these excavations are housed at the Haffenreffer Museum of Anthropology at Brown University.

The remains of nearly 30 Kotzebue period or late Thule houses were mapped along one of the youngest beach ridges at Cape Krusenstern. Excavation was undertaken in House 14, 31, 34, 35, and 50. House 50, which was part of a cluster of four houses, was a single-roomed dwelling with a long entrance tunnel and no evidence of a raised sleeping platform (J. Louis Giddings and Anderson 1986). A mandible was selected for DNA sampling. House 35 was clustered with the ruins of one other house and they may have

originally been joined; these are the smallest houses from this time period. House 35 was a nearly square single-roomed house with an entrance tunnel extending from the south end of the front wall ((J. Louis Giddings and Anderson 1986): 50). An atlas was selected for DNA testing. Radiocarbon dating and artifact styles indicate these houses were most likely occupied from AD 1400 to 1600.

Five house depressions were excavated that date to Early Western Thule settlement of Cape Krusenstern; houses 4 and 5 cluster on beach 10 and houses 6, 7, and 8 on beach 11 (J. Louis Giddings and Anderson 1986). Constructed of driftwood and sod, House 4 had a square main room, a short tunnel leading to a kitchen, and a longer entrance tunnel (J. Louis Giddings and Anderson 1986); five adult, and one subadult, dog mandibles were sampled for DNA. House 6 was somewhat larger, and better preserved than House 4, but the overall construction was similar; six adult, two subadult, and two juvenile dog mandibles were sampled for DNA. House 5 and House 7 were similarly constructed houses with a long tunnel entering a main room with a rear sleeping platform, a short tunnel and kitchen angling off the side of the main room, and a side room with planked floor was entered through a side tunnel off the main entrance tunnel (J. Louis Giddings and Anderson 1986). The faunal remains from houses 5 and 7 were commingled; two adult mandibles were selected for DNA sampling. House 8 consisted of one main room with a rear platform of driftwood planks, a large kitchen extending directly off of one of the front corners, and a long entrance tunnel (J. Louis Giddings and Anderson 1986); 12 adult, and one juvenile, dog mandibles were sampled for DNA. Based on radiocarbon dates, artifact styles, and comparison to similar sites in the region, the estimated age of these Early Western Thule houses is AD 1000–1250 (J. Louis Giddings and Anderson 1986).

House 32 and 33, which had been deeply buried by formation of the subsequent beach ridge deposit and only visible by two crescent-shaped depressions, were excavated in 1960 and 1961 respectively (J. Louis Giddings and Anderson 1986). These two dwellings compared with Birnirk structures excavated at other northern Alaska coastal sites, and radiocarbon dates indicate they were occupied AD 650–1150 (J. Louis Giddings and Anderson 1986). House 32 had a main room with a hearth area on the floor in front of the entrance tunnel; the tunnel which extended perpendicular to the dwelling had a planked floor and a small side alcove. House 33 had a similarly sized main room, but the entrance tunnel extended diagonally out from the main room and a plank-floored alcove extending from the front corner of the main room. Both houses had planked side platforms that extended across half of main room—in other words, oriented 180□ to later Western Thule-style dwellings. C. Darwent undertook analysis of the faunal remains from these houses, and two adult mandibles from each of the houses were sampled for mtDNA.

\subsubsection{Cape Nome, Seward Peninsula, Norton Sound, Alaska}*
Extract ID: UA2001-75-1, UA2001-75-2

Two dog skulls were collected by Otto Geist at Cape Nome in 1946. Primary documentation notes the specific locality as “Nome, recent Eskimo village, sandspit across from Nome,” but no other contextual details are provided. Both of these specimens were sampled for this project. The skulls are likely affiliated with the Inupiat ca. AD 1800-1900.

\subsubsection{Clachan, Cape Hearne, Western Coronation Gulf}*
Extract ID: NAPI-2:14, NAPI-2:15-38-21, NAPI-2:16, NAPI-2:19, NAPI-2:28-15-6, NAPI-2:29-1b-5, NAPI-2:33-17, NAPI-2:33-8, NAPI-2:C70-10

The Clachan site (NaPi-2) is located just south of Cape Hearne on the mainland coast facing Coronation Gulf, which is an area known locally as *Kugalyuk*. Archaeological investigation of this site (along with Beulah and Nuvuk) was undertaken by David Morrison between 1979 and 1981 (D. A. Morrison 1983). A cluster of three, stone-and-sod Thule winter houses were excavated, from which a sample of 18 individual dog mandibles were analyzed osteometrically (D. A. Morrison 1983). Morrison focused on carnassial length, tooth row length, and total mandible length. Bivariate scatterplots indicate the Clachan site dogs are smaller, on average, than modern Inuit sled dogs, and both are smaller than the northern Mackenzie wolf. Since all of these mandibles were stored together at the Canadian Museum of History, we were able to extract samples from nine of the 18 individuals. A series of 10 spruce wood and or willow twig samples were radiocarbon dated. Together the dates range from ca. AD 700 to 1200, which is 300–400 years too old for an early Thule site in Canada (D. A. Morrison 1983) and likely the result of seal oil contamination. Commingled with the dog mandible samples was a caribou premolar from the excavated houses, which C. Darwent submitted for radiocarbon dating from Arizona AMS: 700 $\pm$ 46 BP, d13C -17.5‰ (AA102348). The calibrated age fits with Early Thule sites for the western Canadian Arctic at AD 1240–1400.

\subsubsection{Deering, Seward Peninsula, Kotzebue Sound, Alaska}*
Extract ID: DRG-99-0043, DRG-99-0769, DRG-99-1196

Deering is a northwestern Native Alaskan community situated on a spit within Kotzebue Sound on the northern coast of the Seward Peninsula. Several archaeological sites are located along the Deering spit that make up the Deering Archaeological District (Alaska Heritage Resources Survey number KTZ-000169) (Peter M. Bowers and Williams 2009). Specimens DRG-99-0043, DRG-99-0769, DRG-99-1196 come from the Western Thule House 1 (KTZ-00300). The occupation of the Western Thule House 1 dates between AD 720 and 970.

\subsubsection{Dødemandsbugten, Clavering Island}*
Extract ID: TRF-01-09(ZMK 28/1932)

Dead Man's Bay, also known as Dødemandsbugten, is located on Clavering Island in North Eastern Greenland (Sørensen and Gulløv 2012). Settlement of the site took place between the fifteenth and nineteenth centuries CE (Morey 2010). Forty-three house ruins were identified by Danish archaeologist Helge Larsen, twenty-three of which were excavated (Morey 2010). Additionally, thirty graves were detected during the excavation as well as indications of seven interments within one of the houses (Larsen and Sørensen 1934). The site likely has two periods of occupation: stage 1, between the 15th and 17th centuries, and stage 2, between the 18th and 19th centuries CE (Gotfredsen 2010). The demise of the early occupation has been suggested to be the result of starvation seen through the recovery of humans remains within a house, the 'disproportionate' abundance of dog bones inside and around some houses, and gnaw marks attributed to humans on some of the well preserved dog remains which (Larsen and

Sørensen 1934; Morey 2010). Material from this site includes ancient DNA from the petrous bone of a skull from one of the early phase houses.

\subsubsection{Double Mer Point, Labrador}*
Extract ID: TRF-07-03(DMP04), TRF-07-04(DMP07), TRF-07-08(DMP15), TRF-07-12(DMP24)

The Double Mer Point site is an 18th century Inuit village located in central Labrador, Canada. The site was initially investigated by William Fitzhugh in 1968, and then tested by Richard Jordan of Bryn Mawr College in 1973. Since 2013 Lisa Rankin and graduate students from Memorial University of Newfoundland have undertaken a near complete excavation of the three semi-subterranean houses and surrounding middens in partnership with the nearby Inuit community of Rigolet, Nunatsiavut. Multiple dog bones have been recovered from the site. DNA was extracted from a mandible recovered from the midden between House 1 and House 2 ((Rankin 2014) and pers. comm. L. Rankin, 2019).

\subsubsection{Eskimobyen, Knud Peninsula, Ellesmere Island, Nunavut}*
Extract ID: SGFM-4:H14-C1, SGFM-4:H14-C2, SGFM-4:H14-C3, SGFM-4:H20-M

Sites in the Bache Peninsula region were investigated by members of the Arctic Institute's Ellesmere Island Research Project under the direction of Peter Schledermann between 1977 and 1982 (Schledermann and McCullough 2003). The Eskimobyen site (SgFm-4) is located on the Knud Peninsula and separated from the Bache Peninsula by Flagler Bay; 26 Thule winter stone-and-sod house structures were identified at this site (Schledermann and McCullough 2003). Three dog cranial fragments from House 14 and one mandibular fragment from House 20 were sampled for mtDNA at the Canadian Museum of History. Both of these houses are Ruin Island style and thus are associated with early Thule (AD 1200–1300) in the Smith Sound region of Ellesmere Island and northern Greenland. A caribou rib fragment from House 14 was submitted by C. Darwent for radiocarbon dating with Arizona AMS (AA102355: 1229 $\pm$ 47 BP, d13C -14.2). Although the rib was clearly from a caribou, the isotopic value was highly marine; the calibrated age for the sample using a marine correction is AD 1060–1273. One end of the correction fits with the estimated age of this site at AD 1200–1300.

\subsubsection{Fort Churchill, Manitoba}*
Extract ID: TRF-01-31(ZMK 28/2017), TRF-01-59(ZMK 27/2017)

Pre-Dorset, Dorset, and subsequent Thule settlements have all had phases of occupation in the Fort Churchill area on the west coast of the Hudson Bay in Manitoba, Canada (Ronald J. Nash 1972; Riewe, R., Suluk, L., & Brandon, L 1989). The pre-Dorset occupation is considered to be much later than other pre-Dorset sites, such as those in Igloolik, with radiocarbon dating placing the occupation around 945 BCE. (R. J. Nash 1976). The later Dorset occupation has been radiocarbon dated to approximately 130 BCE (R. J. Nash 1976). DNA was extracted from petrous bones from two dog skull recovered from a partially buried specimen and a specimen found on the surface of the Fort Churchill site, the presumed dates of the individuals are of more recent Thule occupations from within the last 200 years.

\subsubsection{Gambell, St. Lawrence Island, Alaska}*

Sample ID: NMNH_254868, NMNH_258749, NMNH_258751, NMNH_258758_1, NMNH_258761_1, NMNH_258761_2, NMNH_258761_3, NMNH258749, NMNH258761_1, NMNH258771-2, NMNH258771-8, NMNH258758_3

The Gambell site is located at the Northwest tip of St Lawrence Island in the Bering Sea. The site consists of a set of stratified mounds, with the oldest occupations dating to the Old Bering Sea cultural phase (J. L. Giddings 1960). Material from this site included in this study are 7 mandibles and 6 lower M1s from a total of 11 specimens.

\subsubsection{Geographical Society Islands, Northeast Greenland}*
Extract ID: P88/2013KMG-GeoSociety-M6, P89/2013KMG-GeoSociety-M5, P90/2013KMG-GeoSociety-M7, P91/2013KMG-GeoSociety-M4, P92/2013KMG-GeoSociety-M2, P93/2013KMG-GeoSociety-M1, P94/2013KMG-GeoSociety-M3

Several Thule period winter house structures (ca. AD 1500–1800) located in the King Oscar Fjord region of Northeast Greenland were excavated by P. V. Glob (Glob 1935) in the early 1930s (ZMK 17/1933). A total of seven right mandibles from adult dogs, originally analyzed by zoologist Magnus Degerbøl were sampled for DNA at the Natural History Museum in Copenhagen.

\subsubsection{Gutchiak, Eskimo Lakes, Mackenzie Delta, Northwest Territories}*
Extract ID: NHTN-1:2023M, NHTN-1:2036H, NHTN-1:2048H, NHTN-1:2099M, NHTN-1:2133H

The Gutchiak site (NhTn-1) is located on the edge of a low plateau, on the eastern side of a small tidal lagoon, at the tip of one of a series of long peninsulas that divide up the Eskimo Lakes like accordion folds. Gutchiak means “like a river” and is located in a place where strong tidal currents ensure late freeze-up and early break-up, and thus an ideal location for netting fish (D. Morrison 2000). The faunal remains—dominated by fish bone—were recovered from cultural deposits interpreted as a procurement/processing site; hearth areas and the remains of posts suggest the construction and use of a smoking tent for drying fish (D. Morrison 2000). The deposits span the precontact Thule/Inuvialuit period from approximately AD 1400 to 1800 (Matthew W. Betts 2008; D. Morrison 2000). Two adult mandibles and four juvenile humeri were sampled at the Prince of Wales Northern Heritage Centre in Yellowknife for mtDNA.

\subsubsection{Haa Island, Hayes Fjord, Ellesmere Island, Nunavut}*
Extract ID: SGFQ-1:H9-1, SGFQ-1:H9-2; SGFQ-1:H24-F1

Sites in the Bache Peninsula region were investigated by members of the Arctic Institute’s Ellesmere Island Research Project under the direction of Peter Schledermann between 1977 and 1982 (Schledermann and McCullough 2003). The Haa Island site (SgFq-1) is described as a precipitous rocky island located at the junction of Beitstad and Jokel Fjords within the larger Hayes Fjord on eastern Ellesmere Island, Nunavut (Schledermann and McCullough 2003). On the west side of the island a broad grassy slope holds the remains of 20 stone-and-sod Thule winter houses; it would have been a prime location for breathing-hole sealing. Metacarpals III-IV from an adult dog and a cranium/mandible from a juvenile dog were sampled from House 9, and a fibula was sampled from House 24 for mtDNA at the

Canadian Museum of History. A radiocarbon date on heather recovered from House 24 yielded a corrected age of ca. AD 1450 (Schledermann and McCullough 2003). Both of these dwellings date to the post-Ruin Island phase of Thule occupation in this region, or what would be called Classic or middle Thule elsewhere.

\subsubsection{Hebron, Labrador}*
Sample ID: HMCZ_7406

Hebron is located on the Northeastern coast of Labrador, Canada. Archaeological excavations at Hebron indicate the occupation of Moravian Inuit peoples during the 19th century (Loring and Arendt 2009), though there is indication that the site was occupied prior to the historic period, as Moravian missionaries in the early twentieth century uncovered large sod houses, stone burial chambers, and Inuit and Paleo-Inuit artefacts which were transported to museums and collections (Loring and Arendt 2009). In his book "Dogs of the American Aborgonies" Glover Allen mentions "the skull of an Eskimo Dog from Hebron, Labrador, collected in 1897" ((Allen 1920) pg. 443), which is now held in the collections at the Harvard Museum of Comparative Zoology. Both the cranium and mandible were used in this study for GMM analysis.

\subsubsection{Herschel Island, Yukon}*
Sample ID: HMCZ_57575

Herschel Island lies off the Yukon Coast, west of the mouth of the Mackenzie River in Northwestern Canada. Evidence of human occupation comes from a stretch of rapidly
eroding shoreline on the northeast edge of the island, including the Thule period Washout Site (NjVi-2) (Yorga 1980). Excavations in 1978 recovered a large quantity of well preserved midden materials, and included round and chipped stone tools, pottery, bone artefacts, and a wide range of organics including bird and fish bones, scales, feathers, hair, hide, gut, sinew, coprolites, and baleen (Morgan et al. 1983). A single lower M1 from this site was included in the GMM analysis, and is now held at the Harvard Museum of Comparative Zoology.

\subsubsection{Hopedale, Labrador}*
Sample ID: HMCZ_44669

The Hopedale region of Labrador, Canada lies on the southeastern coast of the province. Archaeological survey and excavations began there as early as 1935 with who examined five sites in the region, and excavated 45 of the recorded 88 Inuit sod houses, providing early evidence for the Thule Inuit occupation in Labrador (Arendt 2013). Since then, various excavations throughout the twentieth century have refined our understanding of the Thule occupation of Labrador, and Hopedale specifically, which radiocarbon dates indicating that the Thule arrived in the region during the 16th century, and genetic evidence indicating that they displaced early pre Thule peoples (Raghavan et al. 2014). A large number of faunal remains were found in and around house structures, as were the remains of a dog sled (Arendt 2013). For this analysis two specimens were analysed for GMM analysis, one providing a mandible, lower M1 and crania, and the other a single crania.

\subsubsection{Iglulualuit, Cape Bathurst Peninsula, Mackenzie Delta, Northwest Territories}*
Extract ID: NLRU-1:623C2, NLRU-1:625M1, NLRU-1:626M2

Iglulualuit (Nlru-1) is located on the Cape Bathurst Peninsula near the mouth of the Horton River Delta on Franklin Bay in the Western Canadian Arctic; it was excavated by David Morrison in 1987 (D. Morrison 1990), and the collections are stored at the Canadian Museum of History. Iglulualuit means "many houses" and is one of the largest sites in the Canadian Arctic with at least 30 Thule/Inuvialuit winter houses spread across 800 m of coast line. The two excavated houses, House 11 and House 20, were constructed mainly out of driftwood timbers with the occasional bowhead whale element or rocks (D. Morrison 1990). *Canis* remains were recovered from excavation of House 11 (AD 1500–1800), but comprised only 1\% of the faunal assemblage (Morrison 1990). A fragmentary adult cranium and two sets of juvenile mandibles were sampled for mtDNA.

\subsubsection{Iita (Etah), Foulke Fjord, Inglefield Land, Greenland}*
Extract ID: KNK2643x1838, KNK2643x1839, KNK2644x1119, KNK2644x1120, KNK2644x1121

Iita is a multicomponent Inughuit (Inuit) site located on an alluvial fan jutting from the northern shore of Foulke Fjord, Inglefield Land, northwestern Greenland. This site, also known as Etah in the historical literature, was the nearest Inughuit community to a number of over-wintering expeditions, the earlier being those of Elisha Kent Kane (1856) and Isaac Israel Hayes (1867) in the mid-1800s, and the latter associated with Robert Peary's North Pole expedition (1898–1902). Donald MacMillan's Crocker Land expedition 1913–17 was set up at Iita itself. The surface of the alluvial fan has over 180 visible features present—including the remains of nine winter houses, two of which were excavated in 2006 (LeMoine and Darwent 2010). Most of the surface features relate to late prehistoric and early historical period Inughuit occupations, but subsurface components are also present due to unique geological features that have created a series of buried strata and include occupations extending to AD 1000 and the Late Dorset period (Paleo-Inuit)—highly unusual for the High Arctic. However, dog remains were only recovered from Thule-Inughuit components (Sarah K. Brown, Darwent, and Sacks 2013; Johansen 2012). Two juvenile dog mandibles were sampled for DNA from winter house Feature 1: KNK2643x1838 from early Thule levels dated to AD 1275–1650, and KNK2643x1839 from historical Inughuit levels dating from 1900–1910, coinciding with Peary's expedition. Mandibles from a juvenile, adult, and old adult dog were sampled for DNA from historical Inughuit levels in winter house Feature 2 dated to AD 1850–1920 (KNK2644x1119, x1120, and x1121). Thus these samples span the early Thule to Historic habitation of this site by the Inughuit and their ancestors.

\subsubsection{Inugsuk, Upernavik, Greenland}*
Extract ID: TRF-01-25(ZMK 53b/1966), TRF-01-26(ZMK 53a/1966), TRF-01-45(ZMK 20/2017), P100/2013KMG-Inugsuk-C9, P101/2013KMG-Inugsuk-C1, P102/2013KMG-Inugsuk-C5, P103/2013KMG-Inugsuk-C8, P104/2013KMG-Inugsuk-C6, P106/2013KMG-Inugsuk-C4

Inugsuk is an island approximately 20 kilometers north of Upernavik in western Greenland which contains a Thule age site of the same name. Much of the site, primarily the middens, had been washed away to sea through erosion before the excavation (Therkel Mathiassen 1930a). The topmost layer of the

site contains materials from occupation phases in the eighteenth and nineteenth centuries CE(Therkel Mathiassen 1930a). Conditions on the site allowed for preservation of organic materials such as wood, baleen, narwhal tusk, and whale bone, among the preserved materials sledge crossbars, trace buckles and whip handles were recovered (Therkel Mathiassen 1930b, [a] 1930). Samples were included from the site were derived from the oldest phase, a surface find, and a partially buried skull from the upper layers of the site.

\subsubsection{Iyatayet, Cape Denbigh, Norton Sound, Alaska}*
Extract ID: IYAT-00192

The Iyatayet site (49-NOB-002), located on Cape Denbigh in Norton Sound, Alaska, was investigated by Andrew Tremayne in 2013 (Tremayne et al. 2018). The dog mandible was recovered from beach deposits eroding from the Nukleet (Thule) layers and dates to ca. AD 1100-1400.

\subsubsection{Kitnepaluk, St. Lawrence Island, Alaska}*
Extract ID: KP-1, KP-2, KP-3

Kitnepaluk is located on the west coast of St. Lawrence Island approximately 25 km south of Gambell. The site consists of a small rocky midden in addition to two cultural mounds that include midden deposits as well as house and cache features. In 1972 and 1973, as part of a multiyear research program run through the University of Bern, Hans-Georg Bandi excavated two trenches designated as K1 and K2. K1 was thirteen meters long, 1-1.5 meters deep, and was excavated through the mound on the northwest part of the site. K2 was twenty-five meters long, 2.9 meters deep, and was excavated through the mound on the southeast part of the site. Chronologically the site is estimated to date to the Punuk time period based on stylistic attributes of artifacts and two radiocarbon dates run on driftwood samples collected from two different house features revealed during excavation of K2 (Bandi and Blumer 2002; Mason 2016). Three complete dog mandibles collected from K2 were sampled for this study.

\subsubsection{Kittigazuit, Tuktoyaktuk Peninsula, Northwest Territories}*
Extract ID: NITR-2:T1

Investigation of the Thule site known as Kittigazuit site (NiTr-2) on Tuktoyaktuk Peninsula in the Mackenzie Delta region of the Northwest Territories was undertaken by Robert McGhee who excavated two Thule winter houses and associated midden areas here in 1969–70 (McGhee 1975; Matthew W. Betts 2008). A tibia was sampled for mtDNA at the Canadian Museum of History. Previous radiocarbon dates suggest the site was occupied between AD 1300 and 1660; C. Darwent submitted a caribou metapodial fragment for dating at Arizona AMS (AA102349: 375 ± 45, d13C -19.1‰), which produced a calibrated age of AD 1445–1635.

\subsubsection{Kotzebue, Kotzebue Sound, Alaska}*
Extract ID: KOTZ-10-0118A

The Kotzebue Archaeological District (49-KTZ-036) is located within the city of Kotzebue, or Qikiqtaġruk, on Baldwin Peninsula that juts into Kotzebue Sound. The district consists of at least 53 sites, 47 historic and 16 prehistoric, including the large villages of Old Kotzebue (49-KTZ-031) and Intermediate Kotzebue (49-KTZ-030) sites. The KOTZ-10-0118A dog remains were recovered from a Late Prehistoric semi-subterranean house pit floor (Housepit 10). Housepit 10 was documented by Northern Land Use Research Inc. in 2010 (Carlson et al. 2013). This structure is larger than many of the surrounding house pit structures and may represent a communal house, or \textit{*qalegi}*. The floor of Housepit 10 has yet to be radiocarbon dated. However, artifacts from the house floor reflect Late Prehistoric assemblages dating between AD 1400 and 1770.

\subsubsection{Kuk, Southampton Island, Nunavut}*
Extract ID: TRF-01-32(ZMK 25/2017), TRF-01-33(ZMK 24/2017)

The Kuk site (LjHp-2, 3, 4) is a Thule period site in Duke of York Bay on northwestern Southampton Island, Nunavut, Canada. Excavated initially by Therkel Mathiassen in the early 20th century, the site consists of three main groupings of semi-subterranean houses, containing traditional and well preserved Thule artefacts (Therkel Mathiassen 1927b). The site is considered to be contemporary with other local Classic Thule sites dating to 1200 CE (Clark 1980) and canid specimens were sampled for mtDNA sequencing.

\subsubsection{Kukulik, St. Lawrence Island, Alaska}*
Extract ID: UA1-1933-9631, UA1-1933-9632, UA1-1937, UA1-1939-1497, UA1-1939-1497-1, UA1-1939-1497-2, UA1-1939-1497-3, UA1-1939-1497-4, UA1-1939-1497-5, UA2001-078-0004

Kukulik is located on the northern coast of St. Lawrence Island and was the focus of extensive archaeological excavation in the early half of the 1930s. This work was conducted under the direction of Otto William Geist and Froelich Rainey (Geist and Rainey 1936) between 1931 and 1933 as the Alaska College Bering Sea Expedition and the between 1934 and 1935 as the United States Department of Interior-Alaska College Archaeological Expedition. A test cut 41 meters long was excavated through the site between 1931 and 1933 which revealed numerous house and cache features as well as extensive midden deposits. This test cut also revealed approximately five meters of deposition proving the Kukulik Site to be a large cultural mound built up during the course of approximately 2000 years of human occupation. The test cut was completed in 1933 and work continued at the site in 1934 and 1935 to excavate large portions of the mound north and east of the test cut (Houlette 2009). Dozens of dog specimens exist from this site and nine were sampled for DNA.

\subsubsection{Kuukpak (Gupuk), Richards Island, Mackenzie Delta, Northwest Territories}*
Extract ID: NITS-1:4794C, NITS-1:A2H1M, NITS-1:A2H1R, NITS-1:A2H1T

The Kuukpak site (NiTs-1), also known as Gupuk, is located on Richards Island facing the east channel of the Mackenzie River. The site is divided into six areas made up of spatially distinct clusters of houses; approximately 21 semi-subterranean house ruins extend across an 800 m stretch of beach (Matthew W. Betts 2008). House 1 in Area 1 is a cruciform-shaped semi-subterranean driftwood and sod structure that was excavated under the direction of Chuck Arnold in 1986 (Arnold 1994; T. M. Friesen and Arnold

1995); he returned in 2002 and placed several test pits to examine the impact of coastal erosion on the site. An adult canine was pulled from the faunal remains in Level 4 in Area 2, House 1's midden at the Prince of Wales Northern Heritage Centre in Yellowknife for mtDNA sampling. Area 2, House 1 and a large midden area was excavated in 1986, 1988, and 1989. Portions of the midden, although exceptionally well-preserved in permafrost, were eroding into the East Channel. Analysis of 28,000 specimens represented only 30% of the massive midden assemblage (Balkwill and Rick 1994). Faunal remains from Area 2, House 1 were searched for DNA samples and the following were selected: 1) adult metacarpal from Level 3, 2) adult incisor and canine from Level 5, and 3) juvenile rib from Level 5. Kuukpak was occupied from AD 1300–1500 based on a series of radiocarbon dates and temporally sensitive artifacts (Betts 2008).

\subsubsection{Lake Site, Southampton Island, Nunavut}*
Extract ID: KKHH-2:357

The origin and development of Sadlermiut culture on Southampton Island, Nunavut, was the reason for excavations at the Lake Site (KkHh-2) by Brenda Clark in 1978 (Clark 1980). Nine semi-subterranean Thule winter houses, with architectural similarities to those described for the northwestern Hudson Bay mainland (Allen P. McCartney 1977), were identified along a beach ridge facing Native Bay; two of these structures were excavated (House 2 and House 5). Walrus and seal dominate the faunal remains, but a few dog bones were recovered, which included an adult mandible that was tested for mtDNA. Two original dates for the site based on caribou bone suggest an occupation between AD 1300 and 1600 (154). A caribou long-bone shaft fragment was submitted by C. Darwent for radiocarbon dating by Arizona AMS (AA102343: 374 ± 45 BP, d13C -17.9‰), which suggests the site most likely falls between AD 1450 and 1650 or Classic (middle) Thule.

\subsubsection{Learmonth, Creswell Bay, Somerset Island, Nunavut}*
Extract ID: PEJR-1:H1-5201, PEJR-1:H5-M2, PEJR-1:H5-M3, PEJR-1:H5-M4

Three Thule houses were excavated at the Learmonth site (PeJr-1) located on the north shore of Creswell Bay, Somerset Island, under the direction of Allen McCartney in 1976. Faunal remains from House 1 at this site, and House 6 at the Cape Garry site nearly 40 km to the south, were analyzed by Anne Rick (Rick 1980). House 1 was a well-preserved semi-subterranean dwelling with two internal platforms, an entrance tunnel, and an external kitchen area. A humerus from House 1, and three left mandibles from House 5 were sampled at the Canadian Museum of History for mtDNA.

A caribou long bone shaft from House 1 and deciduous premolar from House 5 were submitted by C. Darwent for radiocarbon dating with Arizona AMS. House 1 produced a calibrated date of AD 1295–1420 (AA102353: 588 ± 45 BP, d13C -19.3‰), and House 5 a date of AD 900–1160 (AA102354: 1007 ± 46, d13C -20.9‰). The latter date is too early for initial Thule occupation of the Canadian Arctic, but the former fits well with other comparable early Thule sites in this region.

\subsubsection{McKinley Bay, Tuktoyaktuk Peninsula, Mackenzie Delta, Northwest Territories}*
Extract ID: OATI-1:F2-M1, OATI-1:F2-M2, OATI-1:H1R, OATI-1:H1S

The McKinley Bay site (OaTi1) is located on the northern margin of McKinley Bay on the Tuktoyaktuk coast, east of the Mackenzie Delta, Northwest Territories. At least 12 semi-subterranean houses and associated caches make up the site. In 1991, partial excavation of House 1 was undertaken by Chuck Arnold of the Prince of Wales Northern Heritage Centre; constructed of driftwood, this Thule house had a tunnel entering into a rectangular floor, but with no benches (Matthew W. Betts 2008). A scapula from Level 2 and a radius from Level 5 were sampled for mtDNA. Matthew Betts returned to the site in 2004 to undertake further testing in Area 1, Feature 2, and we sampled two right mandibles from Level 4 of this midden area. Radiocarbon dates and the associated artifact assemblage indicate site was occupied AD 1400–1650 (Betts 2008).

\subsubsection{Minto Lake Site, Minto Flats, middle Tanana Valley, Alaska}*
Extract ID: UA64-95

The Minto Lake Site is situated on a low grass-covered rise in Minto Flats and was investigated by John Matthews during the fall of 1964. Five test pits were excavated at the site and revealed concentrations of bone, hearth features, rolled pieces of birch bark, stone tools, flakes, worked bone and antler, and fire cracked rock. In test pit 2 a dog skull was collected approximately 4" below the ground surface in association with a rolled piece of birch bark. This dog specimen was sampled for this study.

\subsubsection{Misigtoq, Kangartik, Greenland}*
Extract ID: TRF-01-08(ZMK 18/1932)

Misigtoq is a Thule site located in Northeast Greenland in the along the Ammassalik Fjord (Therkel Mathiassen 1933; Morey 2010). The site was identified and excavated in 1931 by Therkel Mathiassen during his investigation into the 'Angmagssalik Eskimos' of East Greenland to elucidate the direction of migration to East Greenland (Therkel Mathiassen 1933). Settlement at Misigtoq has been suggested to have started as early as the 13th century CE and includes a total of nine house ruins (Gullov 1997; Therkel Mathiassen 1933). Mandibles on site with holes drilled in the ascending ramus possibly for use as toy sledges (Morey 2010). The sample obtained for this study from Misigtoq was taken from a petrous bone.

\subsubsection{Native Point, Southampton Island, Nunavut}*
Sample ID: NMNH_300707, NMNH_300773, NMNH_300774, NMNH_300775, NMNH_300776, NMNH_300777, NMNH300772, NMNH302619

Native Point (KkHh-1), is a Sadlermiut site on Southampton Island. The origins and relationship between Sadlermiut culture and other Inuit populations in the region remains ambiguous, with parts of their material cultural reflecting Dorset traits, and others more closely affiliated with the Thule (Ryan and Young 2013). Based on the dating of the human remains found at the site, it is likely that the samples included here date to the proto-historic period, prior to sustained European contact in the 19th century CE, and ~150-350 years prior to their final demise in 1902-1903 due to a devastating disease epidemic (Ross 1977). A total of eight individual specimens were analysed for GMM in this study, including eight lower M1s and six corresponding mandibles. These specimens are now housed at the Smithsonian National Museum of Natural History.

\subsubsection{Nelson River, Banks Island, Northwest Territories}*
Extract ID: OHRH-1:1426M1, OHRH-1:1426M2

The Nelson River site (OhRh-1) is situated on the southern coast of Banks Island, just west of the mouth of the Nelson River (Max Friesen and Arnold 2008). The semi-subterranean house was made of driftwood timbers and planks; it includes two distinct rooms, a raised earth platform interpreted as a sleeping area, and a separate external kitchen alcove make up this complex dwelling structure (Max Friesen and Arnold 2008). Several chronologically significant harpoon heads associated with the initial Thule migration into Canada were recovered, including Sicco, Natchuk, and Thule Type 2. Three new radiocarbon dates on muskox and caribou bone from this feature provide an age range of AD 1030–1300 (Max Friesen and Arnold 2008). Two lumbar vertebrae, one from a larger and one from a smaller variety of dog, were sampled for mtDNA at the Prince of Wales Northern Heritage Centre in Yellowknife.

\subsubsection{Nugarsuk, Upernavik District, Greenland}*
Sample ID: ZMK_P70
Extract ID: TRF-01-10(ZMK 139/1967-P69), TRF-01-11(ZMK 139/1967-P70)

In 1966 Jørgen Meldgaard recommended the excavation of a Thule site in the Upernavik District Greenland (Møhl 1979). The site of Nugarsuk is located on an island 20 kilometers from Upernavik, archaeological excavations began in 1967 by Jan Hjarnø (Møhl 1979). The settlement was occupied between approximately 1650 and 1850 CE (Møhl 1979). Six houses, both single family and communal house, were excavated on the site and it appears that only one house occupied at a time (Møhl 1979). Analysis of the faunal remains revealed that 96\% of the remains belonged to, ring seals (*Phoca hispida*) (Møhl 1979). The abundance of ring seal remains indicates that the community was depended on ring seals for subsistence (Møhl 1979). The preservation on the site was excellent, with skin and flesh present on bones due to the permafrost conditions (Møhl 1979). 155 bones or bone fragments were attributed to dogs, two petrous bones were sampled for this project along with one crania for GMM analysis.

\subsubsection{Nulato, Yukon River, Alaska}*
Sample ID: NMNH_A08222

The Nulato site is located on the lower Yukon River in the Yukon Valley, Alaska, and has been of interest to anthropologists and explorers due to its potential for sites dating to the early arrival of humans in the Americas. A number of sites in this region were found during surveys in the late 19th and early twentieth centuries, which were later ascribed to the Thule culture based on architectural design and artefact type (de Laguna 1936). Glover Allen reports that one of the largest dogs he examined was collected by Dr. W. H. Dall at Nulato, Alaska ((Allen 1920) pg 444), which is now housed at the Smithsonian National Museum of Natural History, and the mandible and associated lower M1 were included in this analysis.

\subsubsection{Nunalleq, Yukon-Kuskokwim Delta, Alaska}*
Sample ID: NUN_12001_21, NUN_12001_35, NUN_12003_30, NUN_12003_47, NUN_12003_57, NUN_12003_58, NUN_12003_90_C2, NUN_12004_90, NUN_12005_29, NUN_12006_16, NUN_12006_2_IS, NUN_12007_91_IS, NUN_12010_25_C2,

NUN_12012_3_IS, NUN_12012_30, NUN_12014_27, NUN_12014_28, NUN_12014_39, NUN_12014_41, NUN_12014_6_IS, NUN_12014_8_C1, NUN_12019_82_C1, NUN_12019_82_C2, NUN_13007_72, NUN_13015_56, NUN_13015_57_IS, NUN_13016_58, NUN_13016_59, NUN_13030_224, NUN_13030_227, NUN_13041_224, NUN_13085_228_C2_B, NUN_13092_20_C1, NUN_13092_20_C2A, NUN_13092_20_C2B, NUN_13102_26_IS, NUN_13131_223, NUN_13137_223, NUN_13137_224, NUN_13140_59, NUN_13146_229, NUN_14021_53, NUN_14048_89, NUN_14065_66_C1, NUN_14078_68_C1, NUN_14080_55, NUN_14083_91, NUN_14087_47, NUN_14097_20, NUN_15160_32_C1, NUN_3HF_11, NUN_3HF_11_C2, NUN_3HF_8_IS, NUN_3HF_9, NUN_3HF_9_IS, NUN_4_5, NUN_5cHF_90, NUN_HF_33
Extract ID: AL2784, AL2788, AL2791, AL2792, AL2794, AL2795, AL2797

The Nunalleq site (GDN-248) is located in the Yukon-Kuskokwim Delta region of Southwestern Alaska, between the Kanektok and Arolik rivers, south of the modern Yup'ik village of Quinhagak. Nunalleq was a large prehistoric, semi-subterranean settlement site dating from predominantly from c. AD 1450-1700 (Britton et al. 2013; Forbes, Britton, and Knecht 2015; Ledger et al. 2016), with the main occupation from AD 1570 - 1675 (Ledger et al. 2018). It consists of multiple, multi-layer domestic structures with thousands of organic artefacts and ecofacts preserved by the discontinuous permafrost covering of the site (Ledger et al. 2016), with dating further supported by the lack of any Euro-American artefacts, though a small amount of metal and metal modified artefacts have been found, consistent with trade for Eurasian metal prior to European contact (H. K. Cooper et al. 2016). Dog remains were the second most abundant mammal species found at Nunalleq after pinnipeds (E. T. McManus-Fry 2015), and their remains indicate a close relationship between humans and dogs at the site, based on a wide age range of dogs from the site from puppies to fully grown adults and their deposition within domestic spaces as well as middens (E. McManus-Fry et al. 2016; Forbes, Britton, and Knecht 2015). For this analysis, 52 lower M1s, 21 mandibles and five crania were used from a total of 62 individual specimens.

\subsubsection{Nuvuk, Cape Krusenstern, Western Coronation Gulf, Northwest Territories}*
Extract ID: NCPF-1:8-20

The Nuvuk site (NcPf-1) is located at the base of the Cape Krusenstern peninsula, on the mainland coast of western Coronation Gulf. It was an area known for tom-cod fishing, sealing, and caribou hunting and is mentioned by Diamond Jenness in his book on the Copper Eskimo (D. A. Morrison 1983). Houses 1 and 2 at the Nuvuk site were the remains of Thule winter houses with remnants of a poorly defined sleeping platform and an entrance tunnel (D. A. Morrison 1983) . A dog mandible from this house was sampled for mtDNA. The site is estimated to have the same early Thule date as the nearby Clachan site at AD 1200–1400 (D. A. Morrison 1983) .

\subsubsection{Nuvuk, Point Barrow, North Slope, Alaska}*
Extract ID: UA2001-76-1

The Nuvuk site (49-BAR-011) is a multicomponent site at Point Barrow in northern Alaska (Jensen 2009). The Thule occupation on Point Barrow, Alaska is known mostly from an occupation associated with a large Thule cemetery at Nuvuk (Jensen 2009). A single dog skull collected by Otto Geist at Point

Barrow was sampled for DNA. No date is listed but this specimen was likely collected in the 1950s or 1960s. No other contextual information exists for the cranium, which is housed at the University of Alaska Museum of the North.

\subsubsection{Nyboe Land, Greenland}*
Sample ID: ZMK-NYB-60-1985

This Northwestern most portion of Greenland is relatively unexplored, as it remains difficult to access. However, there are four known archaeological sites in this region, two dating to the Independence I/II cultures, and two to the Thule period (Gronnow and Jensen 2009). The Thule era site of Strømstedet is the only one which reports the recovery of any faunal remains, which are now held at the Danish Zoological Museum (Gronnow and Jensen 2009). At this site, a sled shoe made of whale bone was also found. A single dog mandible was included in the GMM analysis for this study.

\subsubsection{Onion Portage, Kobuk River, Alaska}*
Extract ID: OP-B1-M1, OP-B1-M2, OP-H13-M1, OP-H13-M2, OP-H13-M3, OP-H13-M4, OP-H13-M5, OP-H13-M6

The Onion Portage site (49-AMR-170) is located on the northern edge of the middle Kobuk River in northern Alaska (Anderson 1988). Excavated in several stages, J.L. Giddings initiated small-scale testing in 1941, 1961, and 1963 followed by deep vertical trench excavations in 1964 to discern the depth of the cultural layers (82). Douglas Anderson took over direction of the excavation after Giddings's death and expanded on his trenches to explore the horizontal extent of the site, along with testing of the surrounding area between 1966 and 1973 (82). Of the faunal remains recovered from the Onion Portage excavations, dogs were only identified from the Band 1 or Upper levels. This band spans the Arctic Woodland Culture from ca. AD 1000 to the 18th century (J. Louis Giddings 1952). Two dog mandibles are documented to have been recovered from Band 1, and six were recovered from within House 13, an Ahteut phase structure radiocarbon dated to ca. AD 1025–1225 (82). Faunal remains from this site are housed at the Haffenreffer Museum of Anthropology at Brown University.

\subsubsection{Peale Point, Frobisher Bay, Baffin Island, Nunavut}*
Extract ID: KKDO-1:2270, KKDO-1:2282, KKDO-1:2363

Excavations were undertaken at the Peale Point site (KkDo-1), a large Thule settlement located on a small island 13 km west of Iqaluit on southern Baffin Island, Nunavut, under the direction of Doug Stenton between 1981 and 1983 (Stenton 1987). A total of five semi-subterranean sod houses were excavated that produced a faunal assemblage of over 28,000 specimens. Each house was constructed following what Stenton refers to as the basic Thule design: a paved cold-trap entrance, a main room with sleeping platform, and cooking/storage areas. Samples were taken early Thule components (AD 1200–1400) from three dog ulnae at the Canadian Museum of History for mtDNA testing. A caribou molar was submitted by C. Darwent for radiocarbon dating with Arizona AMS, but unfortunately it did not produce a date.

\subsubsection{Piġniq (Birnirk), North Slope}*
Sample ID: NMNH_241526

Extract ID: UA2012-51-9215, UA2012-51-3947

Piġniq, the Birnirk type site, is located on the northern coast of Alaska at the intersection of the Chukchi and Beaufort Seas near Utqiaġvik (J. L. Giddings 1960). The site consisted of twenty anthropogenic mounds representing approximately 1500 years of human occupation. In the early 1950s Wilbert Carter excavated in three of these mounds (H, Q, and L) mostly focusing on a large trench in Mound H. The trench excavated through Mound H included forty-seven 6' x 6' squares (~1.8 m^2) and most of these were excavated to sterile gravels, although several areas did not thaw enough to find the bottom of the cultural layers (Carter 1966). Two dog skulls were sampled for this study and include one collected from the 5-5.5' (~1.52-1.67 m) level in Mound H and another collected from the 3-3.5' (~0.91-1.06 m) level in Mound L. Material used for GMM analysis from this site included a single crania, a mandible, and 2 lower M1s from collections held at the Smithsonian National Museum of Natural History.

\subsubsection{Pond Site, Richards Island, Mackenzie Delta, Northwest Territories}*
Extract ID: NITS-2:H1B, NITS-2:H1H, NITS-2:H2MC1, NITS-2:H2MC2

Located on Richards Island on the west bank of the East Channel, the Pond site (NiTs-2) is located about 1 km upriver from Kuukpak (NiTs-1) and adjacent to a small stream that flows into the Mackenzie River (Betts 2008). At least seven house ruins remain at the site, and during the summer of 1989 two houses and their associated midden areas were excavated under the direction of Chuck Arnold of the Prince of Wales Northern Heritage Centre (Matthew W. Betts 2008). House 1 at the Pond site was a semi-subterranean Thule-style dwelling with a rectangular main living area, a rear sleeping alcove, and a cold-trap entrance tunnel. House 2 appears to have had a similar configuration to House 1, but its architecture was not as well preserved at the rear. Both houses included extensive midden areas. From House 1 a dog baculum (with a healed midshaft fracture) and humerus were collected for DNA testing, and from House 2 an adult and a juvenile metacarpal. House 1 appears to be slightly older and dates to AD 1200–1400, whereas House 2 dates to AD 1300–1500 (Betts 2008); however, the ranges overlap considerably and the houses were likely reoccupied.

\subsubsection{Porden Point, Grinnell Peninsula, Devon Island, Nunavut}*
Extract ID: RBJR-1:1042.27, RBJR-1:1042.47, RBJR-1:642.56, RBJR-4:222.21, RBJR-4:222.34, RBJR-4:336.19, RBJR-5:99.1A, RBJR-5:99.1B, RBJR-5:99.4

Porden Point is located on the southeastern portion of the Grinnell Peninsula on northern Devon Island, Nunavut; it is made up of three sites (RbJr-1, RbJr-4, and RbJr-5) that include 17 Thule winter houses and over 100 other feature types that date to AD 1200 to 1400. Three houses were excavated by Robert McGhee in 1976 and 1977, and 13 additional houses were excavated by Robert Park in 1984 and 1985 (Park 1987). Based on cranial and mandibular elements of dogs from these sites, Park estimates between 16 and 21 dogs are represented, many of which have evidence of traumatic injuries such as broken canines and healed depressed fractures (Park 1987). We sampled for mtDNA across these cranial elements, which were boxed together at the Canadian Museum of History: three crania from RbJr-1, one cranium and two mandibles from RbJr-4, and two crania from RbJr-5.

\subsubsection{Point Spencer, Port Clarence, Norton Sound, Alaska}*
Extract ID: UA2001-72-2

A single dog skull and mandible were collected by an unknown individual at Point Spencer, Alaska and housed at the University of Alaska Museum of the North. No date is given for when these specimens were collected, but there is a remark in the primary documentation saying the "skull is recent prehistoric." The skull from this collection was sampled for DNA but the mandible was not.

\subsubsection{Punuk, Punuk Islands, Alaska}*
Sample ID: NMNH_258795_1, NMNH_258795_10, NMNH_258795_2, NMNH_258795_3, NMNH_258795_4, NMNH_258795_5, NMNH_258795_6, NMNH_258795_8, NMNH_258795_9, NMNH-258771-1, NMNH-258771-10, NMNH-258771-12, NMNH-258771-13, NMNH-258771-4, NMNH-258771-5, NMNH-258771-6, NMNH-258771-7, NMNH-258771-9, NMNH258771-11, NMNH258771-3, NMNH258795_11, NMNH258795_7

The Punuk Site (49-XSL-006) is located on the northeasternmost of the three Punuk Islands, just off the southeastern end of St. Lawrence Island in the Bering Sea. It was originally excavated by Henry Collins in 1928,(Hrdlicka 1930; W. W. Fitzhugh 2016) The precontact settlements on the island consist of houses and an extensive midden deposit, 16 feet deep (Rainey 1941). The midden contained what was considered typical refuse for the period and region, including large quantities of sea mammal and bird bones, stone tools and remnants of organic materials. 18 dog mandibles, 17 M1s and three crania from 24 individual dogs found at Punuk Island were used in the GMM analysis.

\subsubsection{Qaqaitsut West, Paris Fjord, Inglefield Land, Greenland}*
Extract ID: KNK492x36a (KNK492x33), KNK492x36b, KNK492x37, KNK492x38, KNK492x39, KNK494x1b, KNK494x2b, KNK494x3

Excavation of four early Thule winter houses and an associated cache feature was undertaken at Qaqaitsut by the Inglefield Land Archaeology Project in 2009—two on the east side of a pond and two on the west. Four to five Inughuit families from Qaanaaq resided here in the mid-1980s, and thus the site was well-known to the Greenland National Museum before we undertook mapping of the extensive historic and prehistoric features (LeMoine and Darwent 2010). Eroding into the west side of a pond were the remains of several Thule winter houses, one of these houses yielded an unusual concentration of dog remains in the kitchen area, most of which exhibited numerous cut marks and cracking for marrow extraction; however, few artifacts were recovered. The mandibles sampled for DNA represent seven adult dogs, and they were likely the remains of a sled-dog team consumed over a harsh winter. A fragment of a dog mandible from this house feature (KNK492) was refit with the rest of the element recovered more than 15 m away in a stone cache feature (KNK494); both features were used at the same time, and likely during a single winter. Radiocarbon dates on caribou bone recovered from the same house deposits date to AD 1275–1450 or the early Thule period (Sarah K. Brown, Darwent, and Sacks 2013).

\subsubsection{Qilalukan (Cockburn), Mittimatalik (Pond Inlet), Baffin Island, Nunavut}*
Sample ID: ZMK_TRF_01_34
Extract ID: TRF-01-34(ZMK23/2017), TRF-01-39(ZMK22/2017)

At the time of excavation the Qilalukan site was located two kilometres from the Hudson's Bay Company Station at Pond Inlet in northern Baffin Island (Therkel Mathiassen 1927a). Excavations were undertaken by Therkel Mathiassen in 1923 as a part of the Fifth Thule Expedition, and again by Peter Freuchen in 1924 (Therkel Mathiassen 1927a). The site is thought to have been an important narwhal (*Monodon monoceros*) hunting ground and as a result the site's name means narwhal (Therkel Mathiassen 1927a). Based on the similar typology in the archaeological assemblage, the decay of the ruins, and the uplift of land Mathiassen estimates this site like some of the other older Thule sites in the area could be up to 1000 years old (Therkel Mathiassen 1927a). Ethnographic accounts relate that until the turn of the 20th century Qilalukan remained an important ceremonial and winter residence to local populations (Therkel Mathiassen 1927a). Materials recovered from the site include many sledge remains such as sledge shoes, pegs, toggles, trace buckles, and crossbars (Therkel Mathiassen 1927a). Six dog bones were identified from the site, two of these were sampled for the mitochondrial portion of this study which were derived from contexts with human graves, with one of the crania also used for GMM analysis.

\subsubsection{Rama, Ramah Bay, Labrador}*
Sample ID: HMCZ_7407

The Thule arrived in northern Labrador, Canada around 700 - 600 BP, though the region surrounding Ramah Bay had been occupied by Palaeo-Eskimo groups since the Maritime Archaic period (7000 - 3500 BP). The region was known particularly for the highly prized and unique "Ramah chert", which is found extensively along the north Atlantic coast and sourced from Ramah Bay (Renouf 1993; Kaplan 1980), and there are a few Thule period sod houses from this region (Figure 2 in (Natcher and Procter 2012). A mandible and m1 from a single specimen, identified as an "Eskimo Dog" by the collector J.D. Sornborger, are including in the GMM analysis, and is currently held in collections at the Harvard Museum of Comparative Zoology.

\subsubsection{Resolute Bay, Cornwallis Island, Nunavut}*
Sample ID: NMNH_291088, NMNH_291091
The site at Resolute Bay, Cornwallis Island in the Canadian Arctic was first excavated by Henry Collins of the Smithsonian National Museum of Natural History in the late 1940s. The site consists of four semi-subterranean houses, and artefacts and the associated midden determined the site to be from the Thule period (H. B. Collins 1951). Many bones were also recovered, and included in this study are the remains of two dogs, with mandibles and lower M1s from both used in the GMM analysis.

\subsubsection{Saunaktuk, Eskimo Lakes, Mackenzie Delta, Northwest Territories}*
Extract ID: NGTN-1:6802M, NGTN-1:6802S, NGTN-1:81002T
Two sod and driftwood houses were identified at the Saunaktuk site (NgTn-2) on a peninsula jutting into the Eskimo Lakes, east of the Mackenzie Delta, Northwest Territories (Matthew W. Betts 2008). Excavations in Feature 1, 1 m high mound that had been disturbed by bulldozer activity, were undertaken in the late 1980s by Chuck Arnold (Betts 2008). The floor of the house was paved with flat stones, and the former structure appeared to be of an A-frame rather than typical four-post style of typical Thule houses in the west. From the undisturbed cultural layers, three specimens were sampled from Feature 1: a juvenile mandible and an adult scapula from one unit, and an adult incisor from another unit separated by

two meters. Typical Mackenzie Inuit-style artifacts included net fishing gear that post-dates AD 1400, along with radiocarbon dates, suggest an AD 1400–1500 occupation.

\subsubsection{Silumiut Island, Northwestern Hudson Bay, Nunavut}*
Extract ID: KKJG-1:H3-T1, KKJG-1:H4-S1, KKJG-1:H8-M1, KKJG-1:H13-M1, KKJG-1:H14-M1, KKJG-1:H14-M2, KKJG-1:H14-M3

Silumiut Island is one of a number of small islands that dot the coast of northwestern Hudson Bay, and at low tide is connected to the mainland. The Silumiut site (KkJg-1) was excavated under the direction of Allen McCartney between 1967 and 1969 (Allen P. McCartney 1977). A total of 28 Thule-style, stone-and-sod, semi-subterranean dwellings with cold-trap entrance tunnels were recorded, and seven of these houses were excavated (House 3, 4, 5, 6, 8, 13, 14). A massive quantity of faunal remains was recovered from this site, which is now housed in boxes at the Canadian Museum of History. We searched through 50 boxes of faunal remains (Staab 1979); four of these contained *Canis* bone for mtDNA sampling. House 3 midden, one adult tibia; House 4, one adult scapula; House 8, one adult mandible; House 13 midden, one adult mandible; House 14 midden, one juvenile and two adult mandibles. Only a single driftwood sample had been run originally (GAK-2759, 690 $\pm$ 90 BP), and therefore C. Darwent selected associated caribou bone, where possible, for radiocarbon dating (Table S5). Dates were obtained on four of the houses sampled and these span the period from AD 1400 to 1650. The site is about 150–200 years younger than had been estimated originally (Allen P. McCartney 1977) and is not associated with early Thule, but rather with middle or Classic Thule.

\subsubsection{Singauruk, Lopp Lagoon, Seward Peninsula, Alaska}*
Extract ID:SI-H2-1, SI-H2-2

Approximately 40 km north of the coastal village of Wales, Alaska, and situated between Lopp Lagoon and the Bering Sea, is an old channel known locally as Singauruk. Giddings first identified this location in 1959, and undertook excavations of part of a single house floor (J. Giddings 1973). Two Canis skeletal elements (adult metatarsal, old adult mandible) were submitted for mtDNA testing. This occupation dates to early Thule based on the presence of a Thule Type 2 harpoon head and the results of two radiocarbon samples submitted by C. Darwent on caribou bone from the faunal assemblage: 1) 942 ± 63 BP, d13C -17.3‰ (AA102341); 2) 880 ± 30 BP; d13C -18.8‰ (Beta-301397). Together these dates suggest an age of AD 1000–1200 for the Singauruk samples.

\subsubsection{Skraeling Island, Buchanan Bay, Ellesmere Island, Nunavut}*
Extract ID: SFFK-4:H10-H1, SFFK-4:H11-M1, SFFK-4:H11-M2, SFFK-4:H17-F1, SFFK-4:H20-H1, SFFK-4:H21-M1, SFFK-4:H22-T3, SFFK-4:H2-T2, SFFK-4:H3-F1, SFFK-4:H4-H1, SFFK-4:H4M-H1, SFFK-4:H6-H1, SFFK-4:H8-M1

Sites in the Bache Peninsula region were investigated by members of the Arctic Institute's Ellesmere Island Research Project under the direction of Peter Schledermann between 1977 and 1982 (McCullough 1989). The Skraeling Island site (SfFk-4) is located on the southern tip of Skraeling Island in Buchanan Bay, and adjacent to Kane Basin on the east coast of Ellesmere Island, Nunavut. In total, the site consists

of 23 early Thule (AD 1200–1400), Ruin Island-style winter houses (McCullough 1989). The massive faunal and artifact assemblage from these structures were analyzed by Karen McCullough. Most of the dog (or wolf) remains were postcranial elements; however, at least eight complete skulls were identified as dog based on their morphology. In addition, McCullough notes the recovery of sled-dog traction gear and dog whips, and thus emphasizing the importance of dog-sled teams to the early Thule occupants ((McCullough 1989):268–269).

Houses 1–4 are clustered together; House 4 is a large square-shaped structure identified as a "festival house" or *qargi*. A total of five dog remains were sampled for mtDNA across three houses: a tibia was sampled from the kitchen floor of House 2, a proximal femur from House 3, and two left humerii (one juvenile and one adult) from House 4 and the associated midden area. A single humerus was sampled from House 6, which was isolated at the north end of the site. House 8 is another *qargi*, and houses 9 and 10 are associated dwelling structures, both of which have distinct kitchen alcoves. From House 8, a mandibular tooth was sampled, and from House 10 a humerus. Teeth from two adult mandibles were sampled from House 11—a large, triple-roomed structure with a kitchen alcove. A single house ruin located on a relatively steep slope was designated as House 17; a distal femur was sampled. House 20 is isolated above the cluster of houses 8, 9, and 10, and a humerus was selected for sampling. Houses 21 and 22 were clustered together and also included distinct kitchen alcoves; a mandibular tooth was sampled from the former and a tibia from the latter. All of the faunal remains from this site are stored at the Canadian Museum of History (with the exception of midden materials from House 4 at the University of Calgary).

\subsubsection{Stormbugt I, Northeast Greenland}*
Extract IDs: TRF-01-01(ZMK 13w/1909-P60), TRF-01-02(ZMK 13w/1909-P61)

The site of Stormbugt I is lies at 76°46'N in northeastern Greenland on the Stormbugt Bay. Samples were collected by Bendix Thostrup in 1907 during the Danmarks Expedition for ethnographic research (Thostrup 1911). The site consists of tent rings, winter houses, traps, and house stores, and four human graves. The dog remains that were collected during the expedition are presumed to be of Thule origin (Thostrup 1911). DNA was extracted from two dog petrous bones from the Danmarks Expedition materials from Stormbugt I, inventoried in 1909, which are stored by the Natural History Museum of Denmark.

\subsubsection{Sukersit, Ammassalik District, East Greenland}*
Extract ID: TRF-01-03(ZMK 23/1932)

In the Ammassalik District, Sukersit represents one of the oldest and largest villages in the region (Therkel Mathiassen 1933). The site consists of nine house ruins, seven human graves, four tent rings, and several other structures (Therkel Mathiassen 1933). Dog bones were found within several houses, in addition to fragments of human remains (Therkel Mathiassen 1933). On the basis of house type Mathiassen suggests that the site witnessed two separate occupations (Therkel Mathiassen 1933). The earliest of these occupations taking place around the end of the 14th century CE(Therkel Mathiassen 1933). At the time of excavation in 1931, Mathiassen estimates that it had been at least 200 years since

the more recent occupation of the site (Therkel Mathiassen 1933). One petrous bone was sampled from Sukersit and extracted for this study.

\subsubsection{Talaguak, Kimmirut (Lake Harbour), Baffin Island, Nunavut}*
Extract ID: KEDQ-2:M1, KEDQ-2:P1

Excavations were undertaken at the Talaguak site (KeDq-2) near the village of Kimmirut (Lake Harbour) along Hudson Strait on southern Baffin Island under the direction of M.S. Maxwell and George Sabo III of Michigan State University in 1977–78 (Sabo and Jacobs 1980). Five Thule-style houses were excavated at this site (House 2, 5, 7, 9, 11); a juvenile mandible and an adult pelvis from the Lower Component of House 2 were sampled for mtDNA at the Canadian Museum of History. This house component was identified as early Thule (AD 1200–1300) based on the presence of, among other items, Thule Type 2 harpoon heads, and the architectural style of the house (Sabo and Jacobs 1980).

\subsubsection{Tikilik, Igloolik Island, Nunavut}*
Extract ID: NIHF-4:132C, NIHF-4:1C, NIHF-4:4PLM, NIHF-4:TDL

The Tikilik site (NiHf-4) is located on a peninsula jutting out along the southwestern coast of Igloolik Island and was excavated by Susan Rowley in 1993–94. Faunal remains from the Paleo-Inuit components of this site were analyzed (Murray 1996); however, none of the bone from the early Thule components had been identified or previously published. Using notes associated with the original excavation bags and unpublished documents at the Prince of Wales Northern Heritage Centre in Yellowknife, four *Canis* specimens were collected from the Thule (AD 1200–1400) upper peat layer components for mtDNA sampling: a palate/alveolus from Unit 13N/2E (NIHF-4:132C), a palate/alveolus from the floor of House 1 (NIHF-4:1C), a mandible from Feature 4 (Unit -1N/-3E; NIHF-4:4PLM), and a lumbar from Test Unit D (NIHF-4:TDL).

\subsubsection{Tiktalik, south Amundsen Gulf, Nunavut}*
Extract ID: NKRI-3:616M, NKRI-3:623M

Tiktalik (NkRi-3) is comprised of the remains of at least five early Thule winter houses on the southern coast of Amundsen Gulf, Northwest Territories. Excavation of House 5, the largest of the semi-subterranean house ruins, was undertaken by David Morrison in 1998; faunal remains from this site were analyzed by John Moody (Moody and Hodgetts 2013) and are stored at the Prince of Wales Northern Heritage Centre in Yellowknife. Two adult, right mandibles recovered from Level 7 of the oil-soaked "Kitchen" were sampled for mtDNA. Radiocarbon dates on caribou bone submitted by Morrison as well as the abundance of Sicco-style harpoon heads all point to this being an early Thule house dating to AD 1200–1300 (Moody and Hodgetts 2013). An additional caribou incisor from House 5 at the Canadian Museum of History was submitted by C. Darwent for radiocarbon dating with Arizona AMS (AA102350: 684 ± 49 BP, d13C -18.1‰), which produced a calibrated age of AD 1255–1400.

\subsubsection{Uivak Point 1, Okak Bay, Labrador}*

Extract ID: HJCL-9:01, HJCL-9:03, HJCL-9:04, HJCL-9:05, HJCL-9:07, HJCL-9:09, HJCL-9:10, HJCL-9:13, HJCL-9:14, HJCL-9:18, HJCL-9:20, HJCL-9:25, HJCL-9:27

The Uivak Point 1 site is located in the Okak Bay region of Labrador. According to Kaplan ((Kaplan 2012):28), "Uivaq was one of three eighteenth-century Inuit whaling settlements in Okak", dating to the late 1700s. A large communal Inuit sod house (House 7) along with an associated midden was partially excavated Susan Kaplan and James Wollett in 1993 and 1994. This house is estimated to have been an 18 x 12 meter structure that included an entrance passage, cold trap, paved house floor, and raised sleeping platform that ran the entire length of the back wall of the house. Moravian accounts and European manufactured materials indicate the house was most likely occupied in the late eighteenth century (Kaplan and Woollett 2000; Kaplan 2012). Over 52,000 animal bones were recovered and analyzed from the House 7 midden alone (Woollett 2003, 2007), which includes hundreds of dog remains. All samples for this study were recovered from various levels of the midden area adjacent to House 7, with the exception of HJCL-9:03, which is from a nearby test unit (Test Pit 5).

\subsubsection{Utkiavik, Utqiaġvik, North Slope, Alaska}*
Sample ID: NMNH258810

Utkiavik (49-BAR-002) is an Inupiaq site in Utqiaġvik (Barrow), Alaska. The area was first surveyed by the Smithsonian, led by James A. Ford in the 1930s (Ford 1959). The site consists of various anthropogenic mounds, which have been excavated as recently as the 1980s (Sheehan 1997). For this analysis, the specimen from Utkiavik consists of a single lower M1 used for GMM analysis brought to the Smithsonian after Ford's 1931 expedition to Alaska.

\subsection{*Siberian Archaeological Sample Descriptions}

\subsubsection{Ainu Creek 1, Urup Island, Kurils}*
Extract ID: KBP2569.045, KBP2569.170, KBP2569.179, KBP2569.283

The Ainu Creek 1 site is located on the south end of Urup Island, on the Sea of Okhotsk. The site consists of an unknown number of house pits at an elevation ~10 m asl (Ben Fitzhugh et al. 2016; B. Fitzhugh et al. 2007; {Fitzhugh et al. 2009). The midden deposit was tested by members of the Kuril Biocomplexity Project (KBP) in 2006 and 2007, and has materials dating to both the Epi-Jomon period and the Okhotsk period. Although the site has been heavily impacted by twentieth century military activities, the dog bones (two metapodials; scapula and maxillary fragments) sampled for mtDNA are derived from undisturbed Epi-Jomon deposits from Unit B2, Level 3, dated to ca. 780-230 BC (Ben Fitzhugh et al. 2016).

\subsubsection{Boisman II, Primorsky}*
Sample IDs: BOIS \#1, BOIS \#2, BOIS \#3, BOIS \#5, BOIS \#9, BOIS \#10
Extract IDs: AL2988, AL2989, AL2990, AL2991, AL2994, AL2995

Boisman II is located on Boisman Bay to the southwest of Vladivostok on the east coast of the Russian Federation in the Primorksii region. The site complex consists of several mounds of shell midden and

multiple human burials from the Siberian Neolithic Boisman culture (A. N. Popov, Chikisheva, and Shpakova 1997). The term Neolithic here primarily refers to the advent of the widespread use of pottery by forager populations, not agriculture (Alexander N. Popov, Tabarev, and Mikishin 2014). The site has been repeatedly radiocarbon dated, which suggests a maximum period of occupation between ~8500 and 5000 cal. BP (Alexander N. Popov, Tabarev, and Mikishin 2014). All of the dog remains at the site were found as isolated elements within midden sediments. Four of the six Boisman II dog remains were directly radiocarbon dated for this project (Table S5). Given the heavily reliance at the site on marine fish, mammals, and shellfish, the radiocarbon dates likely carry some marine reservoir bias, perhaps up to a few centuries.

\subsubsection{Bukhta Zakat, Makanrushi Island, Kurils}*
Extract ID: KBP3316.216, KBP3316.228, KBP3316.229

The Bukhta Zakat site is located on Bukhta Zakat (transl. "Sunset Bay") on the southwest end of Makanrushi Island. The site consists of 17 house pits situated on flat grassy plain (B. Fitzhugh et al. 2007). One test pit was excavated by KBP in 2008, exposing thick midden deposits dominated by cod bones dating to the Okhotsk period (ca. AD 1030-1190). An acetabulum, scapula blade fragment, and 5th metacarpal of a dog were all sampled for mtDNA.

\subsubsection{Cherniatino II, Primorsky}*
Sample ID: CHER \#14
Extract ID: AL2997

Cherniatino II is located ~80 km inland from Vladivostok near the village of Cherniatino in the Primorsky region of the Russian Federation. The site was occupied during the Bronze Age, Early Iron Age, and Medieval period. The dog sample analyzed for this study is from the Medieval Mokhe cultural deposits at the site, which contained remnants of dwellings, household waste pits, and faunal remains from wild and domestic species (Nikitin, Iu.G., Chzhun, S., Chzho, Li Ch. 2008; Nikitin and Chzhun 2009). The Mokhe were settled agricultural peoples. The dog specimen was directly dated for this study to 1294 $\pm$ 33 (Ua-51216), or 1290 to 1180 cal. BP.

\subsubsection{Ekven Cemetery, Chukchi Peninsula, Chukotka}*
Extract ID: 14.EC95.G329B.518, 15.EC95.G329B.651

The Ekven site is located near Cape Dezhnev on the Chukchi Peninsula that is renowned for its cemetery, which has produced exquisite ivory artifacts and other grave goods belonging to the Old Bering Sea culture. Numerous investigations have been undertaken at the site (Arutiunov, S. A., & Sergeev, D. A. 1990; Bronshtein and Dneprovsky 2002; Dneprovsky and Bronshtein 2002; Csonka 2014). In 1995, two dogs were found interred together, both intentionally placed in a human grave (Grave 359) after the initially buried individual's body was removed. Two grave goods (a walrus baculum and polar bear mandible) were present as well, with one of the dogs being an older male exhibiting cut marks on its occipital. A mandible fragment from one dog and a maxillary fragment from the other were submitted for DNA testing. One of the two dogs was radiocarbon dated to 1660$\pm$ $\pm$120 BP; however, the

mean δ13C value for the specimen was -13.62, which indicates that the dog had a predominantly a marine-based diet. Therefore, using a 100\% marine-diet calibration curve (Stuiver, Reimer, and Reimer 2019), the sample dates to AD 483–996, which is more aligned with Punuk, but still pre-Thule or Paleo-Inuit.

\subsubsection{Kaniskak, Chukchi Peninsula, Chukotka}*
Extract ID: 11.KG90.H1.658, 12.KG90.H1.747, 13.KG90.H1.708, 17.KM90.2/11.204, 18.KM90.2/11.971, 19.KM90.2/11.937, 20.KM90.2/11.1700, 21.KM90.2/11.938, 22.KM90.2/12.908, 23.KM90.2/12.907, 24.KM90.2/12.168, 25.KM90.2/10.922, 26.KM90.2/10.923, 27.KM90.2/9.931, 28.KM90.2/8.933, 29.KM90.2/7.1578, 32.KM90.2/4.266, 33.KM90.1/4.532, 34.KM89.1/3.247, 35.KM90.2/5.873, 36.KM90.2/5.272, 37.KM89.1/6.1566

Kaniskak is a multicomponent village site located on a bluff overlooking the coast near Cape Dezhnev on the Chukchi Peninsula, Chukotka, Russia(Gusev, Zagoroulko, and Porotov 1999). Investigations at the site were undertaken between 1989 and 1991. The site has a residential area with five semi-subterranean houses and an accompanying large stratified kitchen midden. The houses are generally associated with the Old Bering Sea and Thule cultures; however, the midden deposits span the entire sequence from 100 BC to AD 1800.

Paleo-Inuit samples: In House 1, the remains of two adult *Canis* ulnae were recovered from excavation of Level 4 and Level 5 and submitted for DNA testing (Extract ID: 12, 13); these date to AD 300–800 or the Old Bering Sea culture(Gusev, Zagoroulko, and Porotov 1999). A total of 19 adult *Canis* specimens were submitted from levels 3 through 12 of the stratified kitchen midden deposits. Three left mandible fragments from Level 12 (Extract ID: 22–24), and five maxillary fragments from Level 11 (Extract ID: 17–21) date from 100 BC to AD 400 or early Old Bering Sea levels(Gusev, Zagoroulko, and Porotov 1999). Two tibia from Level 10 (Extract ID: 25–26), an ulna from Level 9 (Extract ID: 27), two mandibles from Level 7 and Level 8 (Extract ID: 28–29), and a maxillary fragment from Level 6 (Extract ID: 37) date to AD 400–500 or late Old Bering Sea. Two dog atlas fragments were submitted from Level 5 (Extract ID: 35–36), and three maxillary fragments from Level 3 and Level 4 (Extract ID: 33–34); all five of these samples are from Birnirk levels dating to AD 900–1350(Gusev, Zagoroulko, and Porotov 1999).

Inuit samples: From House 1, one adult *Canis* ulna recovered from Level 2 (Extract ID: 11) dates to AD 1650–1850 or the late Thule period (214). The absence of historical artifacts suggests occupation was from earlier in the date range. Two mandibles were submitted for DNA testing from Level 2, or the Thule level of the midden dated to AD 1100–1500 (214).

\subsubsection{Paipelghak, Chukchi Peninsula, Chukotka}*
Extract ID: 01.P09.B1.1077, 02.P04.H1.128, 03.P04.H1.1024, 04.P04.H1.151, 05.P04.H1.1612, 06.P05.H2.200, 07.P05.H2.959, 08.P05.H2.961, 09.P05.H2.894, 10.P05.AD5.895

Paipelghak is a multicomponent village site located on the shore of the Chukchi Sea 90 km north of Cape Dezhnev on the Chukchi Peninsula(Dneprovsky 2006). Investigations at the site began in 2002 and continued until 2010 under the direction of Kurill Dneprovsky. The excavated sections of the site

contained a complex of five dwellings constructed of driftwood, stone slabs, and whale bones, with accompanying midden deposits(Dneprovsky 2006). It appears that there was a regular cycle of reusing components of old dwellings in newer structures,which has made the stratigraphy of the site highly complex. Based on radiocarbon dates and temporally sensitive artifacts, it appears the occupation of Paipelghak spanned from 300 BC to AD 1850(Dneprovsky 2006).

Paleo-Inuit samples: Four adult, left mandibles from Level 11 of House 1 (Extract ID: 2–5), and three tibiae from levels 7 and 8 of House 2 (Extract ID: 6–8) were submitted for DNA testing; these deposits date to AD 700–1250 or the Birnirk culture. Burial 1 represents the remains of a Birnirk dog that appears to have been a ritually buried with several grave goods (Dneprovsky 2006). It was so well-preserved that hide and fur were still intact; a P3 and P4 were submitted for testing (Extract ID: 1).

Inuit samples: Two adult, right tibiae were recovered from Area D, Level 5 (Extract ID: 9–10), which dates to AD 1600–1850. The absence of historical artifacts suggests it was likely occupied earlier in this date range.

\subsubsection{Pospelova 1, Primorsky}*
Sample ID: POS \#24
Extract ID: AL2992

Pospelova 1 is located on Russki Island within the city of Vladivostok on the east coast of the Russian Federation in the Primorksii region. It consists of a shell midden and a ritual platform with a human burial. The excavated materials from Pospelova 1 are from the Early Iron Age Iankov culture, thought to date from ~2800 to 2000 BP (Rakov et al. 2009), or ~2950 to 1970 cal. BP. The dog sample analyzed here was not dated, but remains of six other dogs at the site are directly dated to ~3020 and 2750 cal. BP. Given the use of marine resources at the site, the radiocarbon dates likely carry some marine reservoir bias, perhaps up to a few centuries.

\subsubsection{Rasshua 1, Rasshua Island, Kurils}*
Extract ID: KBP3674.045, KBP3695.003, KBP3700.074

The Rasshua 1 site is located on the southwest end of Rasshua Island, on the Sea of Okhotsk. The site is situated on a low terrace, and consists of over 100 pit features, many or all of them likely to be prehistoric house pits (B. Fitzhugh et al. 2007; Ben Fitzhugh et al. 2016). Two test pits were excavated by KBP in 2008, exposing thick midden deposits dating to the Okhotsk period (AD 980-1215). The dog bones (calcaneus, 4th metatarsal, femur) were recovered from Test Pits 1 and 1A and from levels 3 and 4.

\subsubsection{Russkii 1, Primorsky}*
Sample ID: RUS \#2
Extract ID: AL2998

Russkii 1 is located on Russki Island within the city of Vladivostok on the east coast of the Russian Federation in the Primorksii region. This shell midden site was sampled in 2010-11 and all archaeological

remains recovered are from the Early Iron Age Iankov culture, thought to date from ~2800 to 2000 BP (A. N. Popov and Lazin 2011), ~2950 to 1970 cal. BP. The midden contained remains of both wild and domestic fauna (pigs and cattle), as do other sites in the surrounding area from this general time period (Kuzmin 1997). The single dog specimen analyzed here was not radiocarbon dated, but six others from the site are directly dated, with ages ranging from ~3200 to 2750 cal. BP(A. N. Popov and Lazin 2011). Given the use of marine resources at the site, the radiocarbon dates likely carry some marine reservoir bias, perhaps up to a few centuries.

\subsubsection{Sagan Nuge III, Lake Baikal}*
Sample ID: 2012-003b
Extract ID: AL2645

Sagan-Nuge III is located on the Little Sea shoreline of Lake Baikal in southern Siberia northeast of the city of Irkutsk. This site consists of deflated stone constructions that were visible on the modern surface that were created during the Iron age (T. I. Nomokonova and Goriunova 2013). The constructions contained ceramics, iron tools, and other artifacts, along with remains of possible offerings of domestic (horse, cattle, sheep/goat, dog) and wild (Baikal seal, fox, hare, fish) animals. The dog sample analyzed here was directly dated for this project to 1057 $\pm$ 31(Ua-44173), which calibrates to 1054 to 926 cal. BP. Organisms consuming fauna from Lake Baikal carry varying levels of freshwater reservoir effects (T. Nomokonova et al. 2013). The stable isotope values for this dog (δ13C = -17.2; δ15N = 10.7) suggest its diet consisted of some lake foods and that the date is ~152 radiocarbon years too old (Schulting et al. 2014). Note that an earlier radiocarbon date (Beta-335122; 550 $\pm$ 30) on ungulate bone from a directly adjacent feature at the site suggests the site was used somewhat later, between 639 and 517 cal. BP (T. I. Nomokonova and Goriunova 2013).

\subsubsection{Verbljuzhka, Chukotka}*
Extract ID: 16.V89.C9.239

Excavations were undertaken at Verbljuzhka in the Chukotka region in 1989. The fragment of an adult *Canis* mandible recovered from Cut #9 at 110–120 cm below surface and dating to AD 400–700 was submitted for DNA testing.

\subsection{*Icelandic and Norse Greenlandic Archaeological Sample Descriptions}

\subsubsection{Alþingisreitur, Reykjavík}*
The Alþingisreitur site in downtown Reykjavík, Iceland was excavated in 2008-2009 and 2012 in advance of the construction of a new office building for the Icelandic Parliament (Garðarsdóttir 2010, 2011, 2013). During the excavation large quantities of animal bones spanning the entire settlement period of Iceland from the late 9th century until the 20th century were recovered (Pálsdóttir 2010, 2013). Over 21 000 bone fragments were recovered at the site but the analysis of the animal bones from the site has not been completed (Pálsdóttir 2010, 2013). All known dog bones from the site have been recorded in detail, a total of 315 elements were identified as coming from dogs or dog/arctic fox and 15 as coming from arctic fox. Dog bones were often found partially articulated at the site so the MNI of dogs from the site is significantly lower but that analysis has not been completed (Pálsdóttir 2010, 2013). Organic preservation

at the site was in general excellent and a large number of textile fragments, wood and leather were found during the excavation (Garðarsdóttir 2010). This good organic preservation can probably be traced to the partial waterlogging of the site which lies on what was a narrow stretch of land between the Reykjavík pond and the coastline. The Alþingisreitur site is divided into four main phases based on tephrochronology and radiocarbon dating but fine scale dating of the site using dateable artefacts has not taken place (Garðarsdóttir 2010, 2011, 2013). Ancient DNA analysis has been done on three walrus bones from Phases III and IV from the site. One of the walrus bones had no DNA and two of the bones had extremely low endogenous DNA content (Pálsdóttir, Boessenkool, and Star 2018). Many of the bones from the Alþingisreitur site have blue mineral deposits, most likely vivianite, which forms in iron rich soil where there is plenty of water. It has been suggested that vivianite can cause problems for DNA extraction (McGowan and Prangnell 2006) and indeed that does seem to be the case at Alþingisreitur where most aDNA analysis of animal bones has failed.

\subsubsection{Alþingisreitur, Reykjavík Modern}*
Extract ID: 2008-32-1343

Phase I of Alþingisreitur spans the time period from 1800 to the late 20th century. The archaeological remains in this phase were very poorly preserved due to construction, cellars and various utilities in the area. Remains of concrete walls and cellars as well as stone walls were found along with disturbed peat ash refuse layers. The total number of bones from this phase is only 567 and the NISP 347. In general bone preservation from this phase was poor due to the amount of building activity during this site. Bones from this phase rarely had signs of vivianite deposits. The species found in this Phase I include sheep/goat, horse, cattle, dog, cat, pig, seals, whales, rat, chicken, auks, ptarmigan and gadid fish (Pálsdóttir 2010). A total of 13 dog bones are recorded from this phase of the site (Pálsdóttir 2010) .

\subsubsection{Alþingisreitur, Reykjavík Historical}*
Extract ID: 2008-32-1616, 2008-32-621, 2009-32-1539, 2009-32-1629, 2009-32-1637, 2009-32-1760, 2009-32-1819, 2009-32-1838, 2009-32-3316

Phase II spans between 1550-1800 CE. A number of building remains belong to this phase, including a smithy from The New Enterprises (Innréttingarnar) but there was also a 350 m^2 refuse dump at the site during Phase II. The midden was mainly used for household refuse and large quantities of animal bones accumulated in the area. The total number of bones analysed from this phase is 7514 (TNF) with a NISP of 6043. The species found in this Phase II include sheep/goat, horse, cattle, dog, cat, pig, arctic fox, seals, whales, gulls, auks, ptarmigan, eider ducks, swan, goose and gadid fish (Pálsdóttir 2010). A total of 145 dog or dog/arctic fox bones are recorded from this phase of the site (Pálsdóttir 2010).

The one dog bone from Alþingisreitur that had any endogenous dog DNA was a complete skull, no. 2009-32-3316, from an adult animal that came from Phase II, context 358141. The context which the skull was found in was a peat ash waste deposit with wood fragments and a number of animal bones. The deposit is a result of a dumping event and part of a sequence of midden deposits covering the area. A cluster of five stones were placed in a circle at the southwest end of context 358141. The purpose for these stones is not clear, but the stones might have been acting as a stone packing to hold a post upright ((Garðarsdóttir

2010), p.227). The context most likely dates to after 1600 since it included two clay pipes (Ragnheiður Traustadóttir, pers. comm 14.01.2019)

\subsubsection{Alþingisreitur, Reykjavík Medieval}*
Extract ID: 2008-32-1184, 2008-32-1199, 2008-32-1208, 2008-32-2327, 2008-32-38, 2008-32-670, 2008-32-908, 2009-32-1556, 2009-32-1936, 2009-32-1948, 2009-32-2985

Phase III 1223-1550 was also largely a refuse area with large quantities of well-preserved animal bones (Garðarsdóttir 2010). The total number of bones from Phase III is 6399 (TNF) and the NISP is 3828. The species found in this Phase III include sheep/goat, horse, cattle, dog, cat, pig, seals, walrus, whale, arctic fox, gulls, auks, eider duck, goose and gadid fish (Pálsdóttir 2010). A total of 145 dog or dog/arctic fox bones are recorded from this phase of the site (Pálsdóttir 2010).

\subsubsection{Alþingisreitur, Reykjavík Norse/Viking}*
Extract ID: 2008-32-1185, 2008-32-1221, 2008-32-1253, 2008-32-1276, 2008-32-1321, 2008-32-843, 2009-32-3111, 2012-32-4144

Phase IV spans from the early settlement period of Iceland around 870 to 1223 CE. During this time period the site was a general activity area where ironworking and processing took place (Garðarsdóttir 2010, 2011, 2013). The total number of bones from Phase IV is 1037 (TNF) and the NISP is 552. The species found in this Phase IV include sheep/goat, horse, cattle, dog, cat, pig, seals, walruses, whales, great auk, gulls, auks, eider duck, eagle, falcon and gadid fish (Pálsdóttir 2010). A total of 10 dog or dog/arctic fox bones are recorded from this Phase IV of the site (Pálsdóttir 2010).

\subsubsection{Bessastaðir, Álftanes Medieval}*
Extract ID: 1988-213-374

The Bessastaðir site is located on the low lying Álftanes peninsula in Southwest Iceland. It has been inhabited since the settlement period in the late 9th century CE and is currently the residence of the president of Iceland. The Bessastaðir project was the first large scale rescue excavations in Iceland and took place from 1987 to 1996 during extensive renovation of the standing buildings at the site. Those buildings largely date to the mid-18th century and were constructed to serve as the residency for the Danish governor of Iceland (Ólafsson 2010). The archaeological site was much larger than initially anticipated and by 1996 4600 m^2 had been excavated of the estimated 11 000 m^2. The site stratigraphy was complicated with the many phases of construction from the 9th century until the mid-18th century as well as the organization of the renovations and little experience in Iceland with such a complicated large scale site (Amorosi et al. 1992; Ólafsson 2010, 2013). Over 16 000 bone fragments from the Late Medieval (1450-1500 CE) and Early Modern (1600-1849 CE) phases have been analyzed from Bessastaðir identifying cattle, sheep/goats, horses, dogs, and wild species (Amorosi 1996; Amorosi et al. 1992). The number of recorded dog bones from the site is 11 but the actual number of dog bones is likely much higher ((Amorosi 1996) p. 173 and unpublished data from AHP). Bones from the oldest phase of the site have not been analyzed and the size of the total size of the archaeofauna from Bessastaðir is unknown (Amorosi 1996; Amorosi et al. 1992; Ólafsson 2013, 2010). The publication of site reports

fromm Bessastaðir by the National Museum of Iceland is still underway but is set to be completed in 2019 (Ólafsson 2010).

\subsubsection{Garðar, Igaliku, Eastern Settlement, Greenland}*
Extract ID: E47-DOG-36, E47-DOG-505-4

Garðar was the seat of the Bishop in Norse Greenland (AD 1124–1378). The closest city to Igaliku is Qaqortoq in South Greenland. Archaeological excavations of these Medieval ruins were undertaken as early as the 19th century; however, more recent systematic fieldwork at Igaliku was directed by Orri Vésteinsson in 2012 (Vésteinsson 2014). From context "036" in the charcoal horizon of the Area A excavation block, a dog tibia and an astragalus were submitted by faunal analyst Konrad Smiarowski for DNA testing. Radiocarbon dates from the charcoal horizon suggest an occupation from the late 13th to early 15th century (Vésteinsson 2014).

\subsubsection{Gásir, Eyjafjörður, Iceland}*
Extract ID: GÁS-DOG-101, GÁS-DOG-2349, GÁS-DOG-2812, GÁS-DOG-2851, GÁS-DOG-655

Gásir remains date to the 14th century cultural contexts connected to a trading post. All materials result from the 2001-2006 archaeological investigations by the Institute of Archaeology, Iceland (Roberts 2009). The site is located on a coastal inlet of Eyjafjörður, NE Iceland, about 11 km north of Akureyri and 3 km northeast of the Möðruvellir farm estate. The Gásir archaeofaunal and artifact assemblage shows aspects of high status consumption of food and drink, via imported pottery used for wine or beer, and exchange, for example gyrfalcon bones, a sulfur processing pit, homespun sheep wool of varied quality used for likely both packing materials and export goods, as well as roving (unfinished strings of wool) were recovered from Gásir. Furthermore, there were large slabs of imported schist for whetstone making that might have originally functioned as ballast (Harrison 2009, 2014). Dog upper limb bones and a mandible from the Medieval Trading site, Area A, were tested for DNA.

\subsubsection{Hofstaðir, Mývatnssveit, Iceland}*
Extract ID: HST-DOG-5169

Hofstaðir is located in the Mývatn Lake Basin area of northern Iceland, and a midden deposit located just outside the churchyard of the Medieval Christian cemetery (AD 1280–1320) was excavated in 2011 (McGovern, Smiarowski, and Harrison 2013). Dog bones are generally rare in Icelandic archaeofauna; however, at least four dogs of various ages were recovered in the same deposit (McGovern et al. 2013). A tibia fragment from an adult dog was submitted for DNA testing.

\subsubsection{Oddsstaðir, Vesturland, Iceland}*
Extract ID: ODÖ-DOG-108

The Oddstaðir farm ruins are located on relatively flat, south facing pasture land at c. 150-160m asl, north of the river Hörgá and circa 20 km SW of Gásir. Materials discussed here are from the midden excavations as part of the Gásir Hinterlands Project in 2009. The Oddstaðir midden deposits provide a chronology of activities spanning several centuries, from the farm's likely establishment in the late 9th or early 10th century, through to farm abandonment in the late 14th century, around the same time when Gásir exchange ended. The archaeofauna suggests that Oddstaðir could have been an independent farm, at least at the time of its Settlement in the late 9th to early 10th century (Harrison 2011, 2014). One dog vertebra was sampled for DNA.

\subsubsection{Skútustaðir, Mývatnssveit, Iceland}*
Extract ID: SKU-DOG-1, SKU-DOG-2

Skútustaðir is a farm site located in Mývatnssveit, northern Iceland. Excavations undertaken in the 2011 field season involved expansion of the large Trench H midden (Hicks et al. 2013). A mandibular and a maxillary deciduous premolar of a juvenile dog were submitted for DNA testing from the early Modern deposits (1600–1700) from this midden.

\subsubsection{Tasiliqulooq, Aniaaq Fjord, Eastern Settlement}*
Extract ID: E171-DOG-11

A dog humerus from Tasiliqulooq (E171), a Norse farm settlement in Aniaaq Fjord in the Qaqortoq municipality of south Greenland, was submitted for DNA testing.

\subsubsection{Tatsip Ataa, Igaliku Fjord, Eastern Settlement}*
Extract ID: E172-DOG-03, E172-DOG-07, E172-DOG-12

The Norse farm site of Tatsip Ataa (E172) is located on the eastern shore of Igaliku Fjord in the Qaqortoq municipality of south Greenland. Excavation of a Norse midden area was undertaken under the direction of Konrad Smiarowski in 2009–10 (Smiarowski 2012). Part of the midden was in use from around AD 1000 to 1300. A sacrum of a small-sized dog (DOG-07) and a sacrum (DOG-03) and tibia (DOG-12) of a large-sized dog from the midden area were submitted for DNA testing.

\subsection{*Museum and Ethnographic Sample Descriptions}

\subsubsection{Natural History Museum of Denmark}*
CN2587, CN3294, CN2040, CN2041

Four tanned historical museum hides were sampled for this study. Two of the samples originate from East Greenland, Scoresbysund in 1928 and Angmagssalik in 1934, respectively. Two additional specimens from The Fifth Thule Expedition 1921-24, collected in eastern Canada, one in winter 1922/23 in Iglulik and one in 1923 on "Danske Øen".

\subsubsection{National Museum of Denmark}*

Extract IDs: *TRF-02-14(P.32.6), TRF-02-16(P.1659), TRF-02-17(Lu.798a), TRF-02-18(Lu.798b), TRF-02-19(Lu.788), TRF-02-20(Lu.788), TRF-02-21(K.1-1a), TRF-02-27(P.6402), TRF-02-28(P.6396), TRF-02-29(P.6397), TRF-02-30(P.6399), TRF-02-31(K.6411a), TRF-02-32(P26.2), TRF-02-33(P.229a), TRF-02-34(P.231a), TRF-02-35(P.1358), TRF-02-36(Lu.930), TRF-02-37(L.5083), TRF-02-38(L19.3), TRF-02-39(L.2106), TRF-02-40(Lu.979), TRF-02-41(L.18.194), TRF-02-42(Lu.805), TRF-02-43(Lu.804), TRF-02-44(Lu.808), TRF-02-45(Lu.800), TRF-02-46(Ld.8a), TRF-02-47(L.9357), TRF-02-63(Kd.13), TRF-02-66(L.2103), TRF-02-67(K.3-13), TRF-02-68(P.229e), 31369(Ld.103), 31588_a(L.2097), 31588_b(L.2097), 32179(L.5082), 30491(Lc.186c), 31867(L.4364), 32135(L.5039), 32987(L.9344), 35561(L19.4a-b), 67941(L.7827)*

A total of 42 specimens were sampled from ethnographic collections housed at the National Museum of Denmark in Copenhagen. Three samples derived from Siberian clothing (Table S1 for specimen sampled from garments). The remaining 39 samples derive from Inuit clothing in Greenland, Canada, and Alaska. These samples were taken from fur clothing which was collected on ethnographic expeditions or donated to the National Museum. Macroscopic analysis was done to identify the types of fur used to make the garments, which includes headgear, parkas, trousers, footwear, or mittens. The samples included in this study from these materials originated from the American Arctic and was identified to belong to either dogs or wolves.

\subsubsection{Greenland National Museum and Archives}*
Extract IDs: TRF-01-04, TRF-01-05, TRF-01-06, TRF-01-07, TRF-01-30, TRF-01-31, TRF-01-35, TRF-01-36, TRF-01-41, TRF-01-42, TRF-01-43, TRF-01-46, TRF-01-49, TRF-01-52, TRF-01-53, TRF-01-54, TRF-01-55, TRF-01-56, TRF-01-57, TRF-01-58, TRF-01-59, 19451(K.613), 19619(K.907), 19741(K.1162), 36960(P32.3), 37351(P30.9), 37351_a(P30.9), 67733(Kc.108a), 89375(K.1-3), 89594(K.2-11), 90710(Kc.99), 91085(Kd.10), 36635(P.1658), 37233(P31.1), 66272(Ic.191), 70111(Pd.2a)

A total of 36 historical and ethnographic samples from the Greenland National Museum and Archives, housed at the Natural History Museum of Denmark were also included in this study. These samples were collected on various expeditions to Greenland, some in the form of surface finds and others were collected from living individuals during these expeditions.

\subsubsection{Natural History Museum, University of Oslo, Norway}*
Extract ID: GL1(NHMO-DMA-55366/1-P), GL2(NHMO-DMA-55369/1-P)

Two Greenland dog hides from dogs used under Amundsen's South Pole expedition 1911-1913, the dogs are supposedly of Greenland origin.

\subsubsection{Museum of Cultural History, University of Oslo, Norway}*
Sample ID: Sort, Grå

Two stuffed dogs brought to Norway with Amundsen after his exploration of the Northwest Passage (1903–1906). The dogs supposedly origin from the modern day Gjoa Haven area.

\subsubsection{The Ski Museum, Oslo, Norway}*
Sample ID: Obersten

A famous lead sled dog used under Amundsen's South Pole expedition 1911-1913, it is recorded the dog originated from Greenland.

\subsubsection{Zoological Institute of the Russian Academy of Sciences}*
Sample ID: StP_2848, StP_3437, StP_5811, StP_20033, StP_20039, StP_9290, StP_9292, StP_9299, StP_14732, StP-6122,
Extract ID: AL3004, AL3007:3437, AL3023:5811, AL3031:20033, AL3032:20039, AL3052:9290, AL3053:9292, AL3054:9299, AL3080:14732

Nine specimens were sampled for this study from the ethnographic collections at the Zoological Institute of the Russian Academy of Sciences. During the expedition of Russian scientist N.A. Grebnickii in 1884 specimens were collected from Native Kamchatkan dogs (2848, 3437) and Native sled dogs from the Bering Island's (9290, 9292, 9299). A Native Chukotka dog specimen (14732) collected during the expedition of Russian scientist K. Chapskii in 1928. During the expedition of Russian scientist N. Gondatti in 1896 a Native Chukotka dog specimen (5811) was collected. Native Chukotka dogs (20033, 20039) were also collected during the expedition of Russian scientist Dr. Grinevickii in 1891.

\subsubsection{American Museum of Natural History, New York}*
Sample ID: AMHN_30475, AMHN_30476, AMHN_30477, AMHN_441_21934, AMHN_67191, AMHN_67212, AMHN_67229, AMHN_67245, AMHN_67179, AMHN_67187, AMHN_67190, AMHN_67196, AMHN_67203, AMHN_67204, AMHN_67207, AMHN_67220, AMHN_67227, AMHN_67228, AMHN_67230, AMHN_67233, AMHN_67235, AMHN_67243, AMHN_67253, AMHN_67168, AMHN_67182, AMHN_67192, AMHN_67194, AMHN_67195, AMHN_67198, AMHN_67201, AMHN_67210, AMHN_67213, AMHN_67216, AMHN_70955, AMHN_1907_155, AMHN_67199, AMHN_67215, AMHN_67223, AMHN_67232, AMHN_67251, AMHN_30442, AMHN_30450, AMHN_67169, AMHN_67197, AMHN_67211, AMHN_67214, AMHN_67222, AMHN_67224, AMHN_70942, AMHN_70943, AMHN_70944, AMHN_70945, AMHN_70946, AMHN_70947, AMHN_70948, AMHN_70949, AMHN_70950, AMHN_70952, AMHN_70953, AMHN_70954, AMHN_70956, AMHN_70958, AMHN_30439, AMHN_30440, AMHN_30441, AMHN_30474, AMHN_30478, AMHN_30482, AMHN_67202, AMHN_67205, AMHN_67208, AMHN_67219, AMHN_67234, AMHN_67240, AMHN_67242, AMHN_67248, AMHN_67255, AMHN_67256, AMHN_67239, AMHN_67244

Pleistocene North American wolf specimens from the Alaska area held at the American Museum of Natural History were used for GMM analysis, including 64 lower M1s, 35 mandibles and 1 crania from 81 individual specimens.

\subsubsection{American Museum of Natural History, New York}*
Extract ID: M-14045, M-14445

Robert E. Peary's Expeditions to the North Pole resulted in a multitude of natural history specimens collected for the American Museum of Natural History in New York. Included in this collection are more than a dozen skulls and complete skeletons of dogs used by Peary's expedition as sled dogs. A mixture of male and female adult and old adult dogs many of which had healed and unhealed cranial wounds were collected by Peary's team in 1896 and 1897. Teeth from two adult males were sampled for mtDNA.

\subsubsection{University of Alaska Museum of the North}*
Hooper Bay Village
Extract ID: UA1-1950-1789, UA1-1950-2093, UA1-1950-2094, UA1-1950-2110, UA1-1950-2111, UA1-1950-2112, UA1-1950-2113

The village of Hooper Bay is located on the Bering Sea coast in southwest Alaska on the north side of the bay. The oldest existing portion of the still-occupied village is situated along the banks of a slough where Wendell Oswalt (Oswalt 1952) noted evidence of prehistoric houses and a midden, and where he chose to focus archaeological excavations during the summer of of 1951. Contemporary buildings covered much of the prehistoric site and therefore Oswalt excavated two blocks, one on either end of the landform. These excavations were slowed by permafrost and by the end of the field season one block reached a depth of 1.5 meters and the other a depth of 2.5 meters, and the base of the cultural deposit was not found in either. Based on tree-ring dates from three spruce driftwood samples recovered during excavation, Oswalt hypothesized a date of A.D. 1600 for the lowest levels of the midden excavation with shallower levels dating to the late A.D. 1800s. Seven complete dog craniums from this site were sampled for DNA and all of them were collected from the surface and likely date to the late Historic or Modern time periods.

\subsubsection{Point Hope, Old Village}*
Extract ID: UA5072-1, UA5072-2, UA5072-3, UA5072-5, UA5072-7

Twelve dog specimens (7 skulls and 5 mandibles) were collected by Otto Geist at Point Hope in 1961. The primary documentation that exists for this collection indicates these remains were collected from the old village site excavation, which may refer to archaeological work conducted during airfield construction at the village in 1961. Five of the seven skulls included in this collection from the University of Alaska Museum of the North were sampled for mtDNA and likely date to the late Historic or Modern period.

\subsubsection{School of Anthropology, University of Arizona}*
Anaktuvuk Pass, Alaska
Extract ID: U.Arizona-1, U.Arizona-2, U.Arizona-3, U.Arizona-4, U.Arizona-5, U.Arizona-6

A total of six upper second molar samples were submitted from six separate crania collected by Lewis Binford (Binford 1978) during his Nunamiut Ethnoarchaeological research in the Anaktuvuk Pass region of northern Alaska between 1969 and 1972. These crania, which are housed at the University of Arizona, were surface collected on the tundra from unknown contexts, and thus were somewhat weathered and many of the teeth had been lost post-mortem; several individuals display joint trauma of the upper neck and/or occipital region. It is estimated that the animals died sometime in the mid-20th century. Based on

the physiological conditions of the crania, which includes broken or worn teeth, some or all of the individuals appear to have been used as sled dogs by local Nunamiut hunters.

\subsubsection{Modern Greenland Sledge Dog Samples}*
Extract ID: GSD_19.5, GSD_2.6, GSD_51602, GSD_51603, GSD_GS16, GSD_GS31, GSD_Mums, GSD_Pondus, GSD_Q11, GSD_QSON

10 modern samples from tissue or DNA extract biobanks at Greenland Institute of Natural Resources, Arctic Research Centre - Aarhus University or Department of Veterinary and Animal Sciences - University of Copenhagen. There are 2 samples from each of 5 modern Greenland cities or villages, specifically, Qaanaaq, Ilulissat, Aasiaat, Sisimiut and Tasiilaq.

\subsubsection{Bern Natural History Museum}*
Sample ID: NMBE_1051369, NMBE_1051371, NMBE_1051372, NMBE_1051373, NMBE_1051377, NMBE_1051386, NMBE_1051393, NMBE_1055699, NMBE_1055702, NMBE_1055705, NMBE_1058274, NMBE_1058384, NMBE_1059182, NMBE_1059211

Modern breed Greenland Dogs were collected for GMM analysis from the collections at the Natural History Museum, Bern Switzerland. These specimens all come from local Greenland dog breeders based in Switzerland between 1972–2004.

\subsubsection{National Museum of Natural History, Washington DC}*
Sample ID: NMNH_147705, NMNH_257445, NMNH_181244, NMNH_203895, NMNH_175669, NMNH_175670, NMNH_175671, NMNH_282817, NMNH_282818, NMNH_282819, NMNH_283586, NMNH_291012, NMNH_507338, NMNH_242704, NMNH_A06256, NMNH_A09000, NMNH_A09001, NMNH_A09003, NMNH-134131, NMNH_134781, NMNH-008710-A07154, NMNH_243573, NMNH_243574, NMNH_243576, NMNH_243577, NMNH_212609, NMNH-257446, NMNH_147470, NMNH_147472, NMNH_168440, NMNH_127550, NMNH_130897, NMNH_180282, NMNH_180283, NMNH_302628, NMNH_302629, NMNH_243324, NMNH_243325, NMNH_A23137, NMNH_A23140, NMNH-A23138, NMNHA23138, NMNH_134496, NMNH_134497, NMNH_274485, NMNH_274486, NMNH_146361, NMNH_212610, NMNH_214477, NMNH_214480, NMNH_214481, NMNH_214482, NMNH_A16750, NMNH_291008, NMNH_291010, NMNH-291009, NMNH_177370, NMNH177370, NMNH_243327, NMNH_243329, NMNH_243329, NMNH_244206, NMNH_244207

Modern wolf specimens from the collections at the Smithsonian National Museum of Natural History used for GMM analysis.

\subsubsection{British Natural History Museum, London}*
Sample ID: BNHM_1979_5444

Ethnographic sample from Baffin Island, Nunavut

\subsubsection{Qaanaaq, Thule District , Greenland}*
Sample ID: NMNH_275071

Qaanaaq, also called Thule or New Thule is a modern village in Northwestern Greenland. The specimen used in this study is a mandible for GMM analysis, and was collected by A.J. Duvall and C.O. Handley during an expedition sponsored by the Smithsonian in 1946.

\subsubsection{Harvard Museum of Comparative Zoology}*
Sample ID: HMCZ_B_9070, HMCZ_7409, HMCZ_7410, HMCZ_7422, HMCZ_7482, HMCZ_7411

Modern wolf specimens from the collections at the Harvard museum of comparative zoology used for GMM analysis.

\subsubsection{Naturalis Biodiversity Center Leiden}*
Sample ID: SNRM_595129, SNRM_582376

Modern wolf specimen from Naturalis Biodiversity Center, Leiden.

\subsubsection{University of Alaska Museum of the North}*
Cape Dyer, North Slope
Extract ID: UA5231

A single dog skull was collected by William Pruitt on July 20, 1960 from the ground surface at Cape Dyer, east of the Kipalog Creek mouth north of Pt. Hope, Alaska. No other contextual information exists for this specimen.

\subsubsection{Cape Nome, Extinct Village, Alaska}*
Extract ID: UA2001-74-1, UA2001-74-2, UA2001-74-3

Three dog skulls were collected by Otto Geist at Cape Nome in 1946. Primary documentation notes the specific locality as “Extinct Eskimo village,” but no other contextual details are provided. All three of these specimens were sampled for mtDNA. The remains likely date to the late Historical or Modern period.

\subsubsection{Deering, Kotzebue Sound, Alaska}*
Extract ID: UA2001-73-1, UA2001-73-2, UA2001-73-3

Three dog skulls, one associated with a disarticulated mandible, were collected by an unknown individual at an unknown point in time (although likely between the 1930s and 1960s). Primary documentation states that these remains were collected at Deering, but no other contextual information exists for this mtDNA sample. The remains likely date to the late Historical or Modern period.

\subsubsection{Meade River, North Slope, Alaska}*

Extract ID: UA4667, UA4667-1

In 1959 and 1960 Otto Geist collected six dog skulls, one dog mandible, and one dog femur from the Meade River, North Slope, Alaska. The primary documentation for these specimens does not provide any additional contextual information. Two of these dog skulls were sampled for DNA. The remains likely date to the late Historical or Modern period.

\subsubsection{Oliktok Point, Lower Colville River, Alaska}*
Extract ID: UA1-1952-1159, UA1-1952-1160

Two dog mandibles were collected by William Irving during archaeological reconnaissance of the lower Colville River and the Colville Delta in 1952 (Irving 1953). Specific details about the context of these two specimens is lacking, but they are likely from historic or protohistoric camps near Oliktok Point. Both mandibles were sample for DNA.

\subsubsection{Point Barrow, North Slope, Alaska}*
Extract ID: UA5225

A single dog skull was collected by Otto Geist from Point Barrow in 1960 and is likely modern. Primary documentation for this specimen lists "Birnirk excavation" in parenthesis as the specific location, but no other contextual information exists; it was sampled for mtDNA. The cranium could date from Birnirk through the Thule/Historic periods.

\subsubsection{Point Spencer, Port Clarence, Norton Sound, Alaska}*
Extract ID: UA2001-72-2

A single dog skull and mandible were collected by an unknown individual at Point Spencer, Alaska. No date is given for when these specimens were collected, but there is a remark in the primary documentation saying the "skull is recent prehistoric." The skull from this collection was sampled for DNA but the mandible was not.

\subsubsection{Wales, Seward Peninsula, Alaska}*
Extract ID: UA594-M-5824

A single dog skull and unarticulated mandible was collected in Wales, Alaska by Charles Lucier prior to December of 1952, which is the date it was accessioned. Little information exists for this specimen sampled for mtDNA, but it was likely collected from the surface and dates to the historic period.

\subsubsection{Wainright Lagoon, North Slope, Alaska}*
Extract ID: UA1-1959-proj247

A single dog mandible collected by Otto Geist at Wainwright Lagoon in 1959 was sample for this project. No other contextual information exists for this specimen sampled for mtDNA, but it was likely collected from the surface and represents the historic time period.

\subsubsection{Walakpa, North Slope, Alaska}*
Extract ID: WAL-B10-J16, WAL-B10-J19A, WAL-B10-J19B, WAL-B10-J19C, WAL-B10-J19D, WAL-B10-J19E, WAL-B10-J20, WAL-B10-K17, WAL-B10-K18, WAL-B10-K19, WAL-B9-J19, WAL-B9-J20, WAL-B9-K15, WAL-B9-K19

The Walakpa site is located near the modern village of Utqiaġvik (Barrow), in the North Slope region of northern Alaska. Excavation of this multi-component site was undertaken by Dennis Stanford in 1968 and 1969 (Stanford 1976) and revealed historic, late Thule, early Thule, Birnirk, and some possible ASTt deposits. Birnirk deposits were superposed in levels 8, 9, and 10. Radiocarbon dating from this site and typological cross-dating with other dated Birnirk sites in the area indicate the Birnirk occupation here was from AD 650–1000. Although dominated by small seals, and to a lesser extent caribou, a total of ts 296 Canis remains were recovered from across all Thule and Birnirk levels of the site; nearly 70\% of these were recovered from levels 9 and 10 of Area B. We concentrated our sample search in these lower Birnirk levels, plus dog crania from these levels were boxed together. Faunal remains from excavations at Walakpa are stored at the Maxwell Museum at the University of New Mexico. Tooth roots from of 14 different crania/mandibles, some of which still had dried or burned tissue adhering, were sampled for mtDNA from Area B, levels 8–9, units J-16, J-19, J-20, K-15, K-17, K-18, K-19.

\subsubsection{University of California, Davis Zooarchaeology Laboratory}*
Inglefield Land, Northwest Greenland
Extract ID: UCDZL-1080, UCDZL-1081, UCDZL-1082

Three dog crania were collected by C. Darwent during the Inglefield Land Archaeology Project in 2009. Two were surface collected adjacent to houses originally built at Innuarfissuaq, Marshall Bay, by Inughuit assistants to Danish archaeologist Eric Holtved in the 1930s; however, the houses continued to be reused by hunters in the region until the 1990s. The third cranium was collected at Qaqaitsut, Paris Fjord, and likely dates to the mid-1980s when four to five Inughuit families, on the initiative of the community of Qaanaaq, built houses and endeavored to "live on the land" (LeMoine and Darwent 2010) .

SI Figure Captions

Fig. S1: MapDamage plots showing general patterns of deamination seen in samples included in the study with transition frequency plotted against position in read. Deaminated bases appear most frequently at the ends of reads, the frequency of the appearance of deaminated bases increases between younger and older samples. A) Plot of historical sample collected ca. 135 years before present. B) Plot of Thule Inuit sample, ca. 600 years old. C) Plot of Siberian Holocene sample, ca. 6,000 years old.
Fig. S2: Maximum likelihood tree generated with RAXML with 1,000 bootstraps from samples with a minimum mean of 10x read depth of 95\% of sites in the mitochondrial genome with the exclusion of samples in basal positions of the subclades to ensure robusticity of node support.
Fig. S3: Bayesian skyline plot of whole dataset, plot displates Ne against time (years before present). The black line represents the average Ne and the 95\% HPD is depicted with the purple area. Bayesian skyline plot of all samples with a minimum mean of 10x read depth, generated with BEAST2.4.

Fig. S4: Maximum likelihood tree generated with RAXML with 1,000 bootstraps from samples with a minimum mean of 3x read depth of 80\% of positions in the mitochondrial genome. Arrows indicating A-clade subclades and X-clade. Three specimen (coyote outgroup and two Chinese wolves) were excluded from figure for better visualisation, remaining samples were coloured according to cultural affiliation. Dog clades B-E are collapsed into black clades with labels in red.

Fig. S5: Maximum likelihood tree generated with RAXML with 1,000 bootstraps from samples with a minimum mean of 3x read depth of 80\% of positions in the mitochondrial genome. Arrows indicating A-clade subclades and X-clade. The outgroup and Chinese wolves are included in figure.

Fig. S6: Maximum likelihood tree generated with RAXML with 1,000 bootstraps from samples with a minimum mean of 10x read depth of 95\% of sites in the mitochondrial genome. A-clade subclades and X-clade are indicated with arrows.

Fig. S7: Bayesian mitochondrial phylogeny constructed with BEAST2.4 for A-clade with dated nodes from samples with a minimum mean of 10x read depth. The subclades of the A-clade have been coloured according with Fig. 2, purple for clade A2b, blue for clade A2a, red for clade A1b, and yellow for clade A1a.

Fig. S8: Bayesian mitochondrial phylogeny constructed with BEAST2.4 for whole dataset with dated nodes from samples with a minimum mean of 10x read depth. The subclades of the A-clade have been coloured according with Fig. 2, purple for clade A2b, blue for clade A2a, red for clade A1b, and yellow for clade A1a.

Fig. S9: Bayesian skyline plots corresponding to Inuit and Historical/Modern sample sets from the North American Arctic. Each plot displates Ne against from 15,000 years ago to present. The black line represents the average Ne and the 95\% HPD is depicted with the purple area. A) North American Arctic dogs from 300 years ago until 50 years before present. B) North American Arctic dogs from 300 years ago until present. C) Samples of only North American dogs from Inuit contexts. D) Combined dataset of Inuit dogs until modern dogs in the North American Arctic.

Fig. S10: Nuclear ancestry analyses of AL2797 (Thule sample from Nunalleq). A-B. Principal component analysis of nuclear SNP data with (A) and without (B) coyotes genomes. This data set comprises 49 canids genomes (see Table X for accession numbers and sample provenance), including an outgroup (Andean fox) 2 coyotes (C_MidW, C_Cal), 10 modern Arctic, pre-contact dogs and CTVT genomes (D_AHusky91, D_Green, D_Husky, D_Husky89, D_Mal68, AL3194, AL3223, C_399T, C_79T, C_24T), 9 modern East Asian dogs (D_China8, D_Tibet3, D_Tibet4, D_TMastif4, D_TMastif5, D_China9, D_Dingo, D_Viet21, D_Viet59) and 17 Western dogs (D_India168, D_India60, D_SLaika, D_Na89, D_Basenji, D_Na8, D_Qatar27, D_Qatar5, D_NGDG, D_Peru, D_Port71, D_Mex, D_Port61, D_Leb85, D_Leb79, D_GerShep6, D_GerShep3), 6 modern Eurasian wolves (W_Iran, W_India, W_Spa, W_Port, W_Mongo, W_Altai), 1 ancient Eurasian wolf (TAI) and 3 modern American wolves (W_Mex1, W_Yellow1, W_Yellow2). C. Neighbour joining tree tree depicting the relationship between these genomes D. D-statistics of the form D(AndeanFox, AL2797, Pop3, Pop4) where Pop3 was fixed as either Western dogs, Asian dogs, Arctic dogs, or CTVT (coloured box plot) and Pop4 represents the genome on the y-axis.

Fig. S11: Estimated F_{ST} and statistical significance were computed using absolute haplogroup frequencies (i.e., sample sizes were accounted for) in Arlequin 3.5.1.3. F_{ST} ranges from 0 to 1 with 0 reflecting no difference in haplogroup frequencies and 1 reflecting the greatest possible difference. Statistical significance was denoted by asterisks (*** $P < 0.001$, *$P < 0.05$).

Fig. S12: A) A median joining network of historical and modern haplotypes from North American Arctic dogs constructed using Network v5.0.1.0., illustrating that subclades A2a and A2b are more deeply rooted than subclades A1a or A1b; rho estimates indicate the average number of mutations accumulating since divergence from the root haplotype (too few samples from A2b were present in historical dogs to estimate rho). B) The relatively recent derivation of A1a and A1b haplotypes (past ~200 years) relative to A2a haplotypes with ages based on calibration to the North American A2a subclade assuming it began accumulating the observed mutations at least 1,000 years ago (rho minus 1 SD = 1000 years), coinciding with the Inuit expansion.
Fig. S13: Maps showing the sampling location and overall sample sizes for the three elements analysed using GMM methods.
Fig. S14: (top) Images of the canid crania showing the placement of the 30 permanent 3D landmarks. 2D landmark configurations for the (middle) the lower M1 showing permanent landmarks (red) and 2 curves of sliding semi-landmarks (grey) and (bottom) mandible showing 15 permanent landmarks (red).
Fig. S15 Two first axes of Principal component analyses of the Crania (top), lower M1 (middle), and mandible (bottom).
Fig. S16 Two first axes of Canonical Variate Analyses of the Crania (top), lower M1 (middle), and mandible (bottom).
Fig. S17: Bone collagen δ^{13}C and δ^{15}N data for the canids from the Nunalleq Site. The wolf is labelled (AL2797), and shows a diet with less marine protein than the dogs from the site.

Captions for Databases S1 - S7

Table S1: Data table listing all specimens used in for DNA analysis, including group assignment, type of element analysed, and sample collection location. (Excel)
Table S2: Data table listing all published mitochondrial sequences used as a reference panel. (Excel)
Table S3: Data table listing all specimens used for GMM analysis, including group assignment, type of element analysed, and sample collection location. Sheets two to four include the raw coordinates used for the GMM analyses. (Excel)
Table S4: Results of carbon and nitrogen stable isotope analysis, including the 7 new samples generated for this analysis, as well as those from (E. McManus-Fry et al. 2016). (Excel)
Table S5: Radiocarbon dates generated for this project from Deering, Seward Peninsula, Kotzebue Sound, Alaska; Silumiut site (KkJg-2), northwestern Hudson Bay, Nunavut and Boisman II, Primorsky Siberia. (Excel)
Table S6: Nuclear genomes used for the analysis of AL2797 (excel)

SI References

Adams, D. C., M. L. Collyer, and A. Kaliontzopoulou. 2018. “Geomorph: Software for Geometric Morphometric Analyses.” *R Package Version 3.0.6.* https://cran.r-project.org/package=geomorph.

Alix, Claire, Owen K. Mason, Nancy H. Bigelow, Shelby L. Anderson, Jeffrey Rasic, and John F. Hoffecker. 2015. “Archéologie Du Cap Espenberg Ou La Question Du Birnirk et de L’origine Du Thule Dans Le Nord-Ouest de l’Alaska.” *Nouvelles Orientations Dans Les Peches. Serie de Notes de Synthese Sur Les Questions de Developpement*, no. 141: 13–19.

Allen, G. 1920. *Dogs of the American Aborigines*. Vol. 63. Bulletin of the Museum of Comparative Zoology at Harvard College.

Ambrose, Stanley H. 1990. “Preparation and Characterization of Bone and Tooth Collagen for Isotopic Analysis.” *Journal of Archaeological Science* 17 (4): 431–51.
Amorosi, T. 1996. “Icelandic Zooarchaeology: New Data Applied to Issues of Historical Ecology, Paleoeconomy and Global Change.” The City University of New York, New York.
Amorosi, T., P. C. Buckland, Ólafsson G., J. P. Sadler, and P. Skidmore. 1992. “Site Status and the Palaeoecological Record. Glasgow:” In *Norse and Later Settlement and Subsistence in the North Atlantic.*, edited by C. D. Morris and D. J. Rackham. University of Glasgow, Dept. of Archaeology.
Anderson, Douglas D. 1988. *Onion Portage: An Archaeological Site on the Kobuk River, Northwestern Alaska*. University of Alaska Press.
Arendt, Beatrix. 2013. “The Return to Hopedale: Excavations at Anniowaktook Island, Hopedale, Labrador.” *Canadian Journal of Archaeology* 37 (-): 302–30.
Arnold, Charles D. 1979. “Possible Evidence of Domestic Dog in a Paleoeskimo Context.” *Arctic* 32 (3): 263–65.
———. 1980. “A Paleoeskimo Occupation on Southern Banks Island, NWT.” *Arctic*, 400–426.
———. 1994. “Investigation on Richards Island.” In *Bridges Across Time: The NOGAP Archaeology Project*, edited by Jean-Luc Pilon, 85–93. Canadian Archaeological Association Occasional Paper 2.
———. 2010. “Archaeological Investigations near Fish Lake, Southwestern Banks Island.” *Unpublished Report on File at the Prince of Wales Northern Heritage Centre, Yellowknife.*
Arutiunov, S. A., & Sergeev, D. A. 1990. “Chapter 1.” In *Issues of the Ethnic History of the Bering Sea.*, edited by Arutiunov, S. A., & Sergeev, D. A., 50–61.
Balkwill, Darlene, and Anne Rick. 1994. “Siglit Subsistence: Preliminary Report on Faunal Remains from a Large Midden at the Gupuk Site (NiTs-1), Mackenzie Delta, NWT.” *Bridges Across Time: The NOGAP Archaeology Project*, 95–116.
Bandi, Hans Georg, and Reto Blumer. 2002. “Investigations by Swiss Archaeologists on St. Lawrence Island. InD. E. Dumond and RL Bland.” *Archaeology in the Bering Strait Region: Research on Two Continents*, 25–60.
Baylac, Michel, and Martin Friess. 2005. “Fourier Descriptors , Procrustes Superimposition , and Data Dimensionality : An Example of Cranial Human Populations.” *Modern Morphometrics in Physical Anthropology*, 145–65.
Betts, Matthew W. 2008. *Subsistence and Culture in the Western Canadian Arctic: A Multicontextual Approach*. Canadian Mus of Civilization.
Betts, M. W., M. Burchell, and B. R. Schöne. 2017. “An Economic History of the Maritime Woodland Period in Port Joli Harbour, Nova Scotia” 10 (sp10): 18–41.
Binford, L. R. 1978. *Nunamiut: Ethnoarchaeology*. New York: Academic Press.
Botigue, Laura, Shiya Song, Amelie Scheu, Shyamalika Gopalan, Amanda Pendleton, Matthew Oetjens, Angela Taravella, et al. 2016. “Ancient European Dog Genomes Reveal Continuity since the Early Neolithic.” *bioRxiv*. https://doi.org/10.1101/068189.
Bowers, Peter M., and Catherine M. Williams. 2009. “The Archaeology of Deering, Alaska: Final Report on the Deering Village Safe Water Archaeological Program.”
Bowers, P. M., and M. Moss. 2001. “The North Point Wet Site and the Subsistence Importance of Pacific Cod on the Northern Northwest Coast.” In *People and Wildlife in Northern North America: Essays in Honor of R. Dale Guthrie*, edited by S. Craig Gerlach and M. S. Murray, 159–77. 944. Oxford: BAR International Series.
Bowers, P. M., C. M. Williams, R. C. Betts, O. K. Mason, R. T. Gould, and M. L. Moss. 2001. “The North Point Site: Archaeological Investigations of a Prehistoric Wet Site at Port Houghton, Alaska.” *Report Prepared for the USDA Forest Service, Tongass National Forest, Sitka* Prepared by Northern Land Use Research, Inc., Fairbanks.
Britton, Kate, Rick Knecht, Olaf Nehlich, Charlotta Hillerdal, Richard S. Davis, and Michael Richards. 2013. “Maritime Adaptations and Dietary Variation in Prehistoric Western Alaska: Stable Isotope Analysis of Permafrost-Preserved Human Hair.” *American Journal of Physical Anthropology* 151 (3): 448–61.

Bronshtein, M. M., and K. A. Dneprovsky. 2002. "The Northeastern Chukchi Peninsula during the Birnirk and Early Punuk Periods." In *Archaeology in the Bering Strait Region: Research on Two Continents*, edited by D. E. Dumon and R. L. Bland, 153–65. University of Oregon Anthropological Papers. University of Oregon.

Brown, Sarah K., Christyann M. Darwent, and Benjamin N. Sacks. 2013. "Ancient DNA Evidence for Genetic Continuity in Arctic Dogs." *Journal of Archaeological Science* 40 (2): 1279–88.

Brown, S. K., C. M. Darwent, E. J. Wictum, and B. N. Sacks. 2015. "Using Multiple Markers to Elucidate the Ancient, Historical and Modern Relationships among North American Arctic Dog Breeds." *Heredity* 115 (6): 488–95.

Brown, T. A., D. E. Nelson, J. S. Vogel, and J. R. Southon. 1988. "Improved Collagen Extraction by Modified Longin Method." *Radiocarbon* 30 (2): 171–77.

Buonasera, Tammy Y., Andrew H. Tremayne, Christyann M. Darwent, Jelmer W. Eerkens, and Owen K. Mason. 2015. "Lipid Biomarkers and Compound Specific δ13C Analysis Indicate Early Development of a Dual-Economic System for the Arctic Small Tool Tradition in Northern Alaska." *Journal of Archaeological Science* 61 (September): 129–38.

Carlson, Eric S., Howard L. Smith, Peter M. Bowers, Sarah McGowan, and Holly J. McKinney. 2013. "Kotzebue Shore Avenue Reconstruction Project: Construction Monitoring Cultural Resource Project." *Report Prepared for the Alaska Interstate Construction, LLC, Anchorage, Alaska, by Northern Land Use Research, Inc., Fairbanks, Alaska.*

Carøe, Christian, Shyam Gopalakrishnan, Lasse Vinner, Sarah S. T. Mak, Mikkel Holger S. Sinding, José A. Samaniego, Nathan Wales, Thomas Sicheritz-Pontén, and M. Thomas P. Gilbert. 2018. "Single-Tube Library Preparation for Degraded DNA." Edited by Susan Johnston. *Methods in Ecology and Evolution / British Ecological Society* 9 (2): 410–19.

Carter, Wilbert. 1966. "Archaeological Survey of Eskimo, or Earlier, Material in the Vicinity of Point Barrow, Alaska." *Final Report to the Office of Naval Research and Arctic Institute of North America.*

Clark, Brenda. 1980. "The Lake Site (KkHh-2), Southampton Island, NWT and Its Position in Sadlermiut Prehistory." *Canadian Journal of Archaeology/Journal Canadien d'Archéologie*, 53–81.

Cleland, Charles Edward, and W. G. Haag. 1973. "Appendix I: Notes on the Dog Skull from Nanook Component 2." *Maxwell, MS Archaeology of the Lake Harbour District, Baffin Island. Mercury Series, Archaeological Survey of Canada Paper* 6: 353–56.

Collins, H. B. 1951. "Dr H. B. Collins' Archaeological Excavations at Frobisher Bay, Baffin Island, and at Resolute Bay, Cornwallis Island, 1948 and 1949." *The Polar Record* 6 (42): 255–255.

Collins, M. J., and P. Galley. 1998. "Towards an Optimal Method of Archaeological Collagen Extraction: The Influence of pH and Grinding." *Ancient Biomolecules* 2 (2/3): 209–23.

Cooper, A., and H. N. Poinar. 2000. "Ancient DNA: Do It Right or Not at All." *Science* 289 (5482): 1139.

Cooper, H. Kory, Owen K. Mason, Victor Mair, John F. Hoffecker, and Robert J. Speakman. 2016. "Evidence of Eurasian Metal Alloys on the Alaskan Coast in Prehistory." *Journal of Archaeological Science* 74 (October): 176–83.

Csonka, Yvon. 2014. *The Ekven Settlement: Eskimo Beginnings on the Asian Shore of Bering Strait*. Archaeopress.

Dabney, Jesse, Michael Knapp, Isabelle Glocke, Marie-Theres Gansauge, Antje Weihmann, Birgit Nickel, Cristina Valdiosera, et al. 2013. "Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave Bear Reconstructed from Ultrashort DNA Fragments." *Proceedings of the National Academy of Sciences of the United States of America* 110 (39): 15758–63.

Damgaard, Peter B., Ashot Margaryan, Hannes Schroeder, Ludovic Orlando, Eske Willerslev, and Morten E. Allentoft. 2015. "Improving Access to Endogenous DNA in Ancient Bones and Teeth." *Scientific Reports* 5 (June): 11184.

Darwent, Christyann M. 2006. "Reassessing the Old Whaling Locality at Cape Krusenstern, Alaska." In *Dynamics of Northern Societies. Proceedings of the SILA/NABO Conference on Arctic and North*

Atlantic Archaeology, Copenhagen, May 10th–14th, 2004, edited by Jette Arneborg and Bjarne Grønnow, 10:95–102. Studies in Archaeology and History. Copenhagen.: PNM, Publications from the National Museum.

Darwent, Christyann Marie. 2002. "High Arctic Paleoeskimo Fauna: Temporal Changes and Regional Differences (Canada, Greenland)." https://elibrary.ru/item.asp?id=6690866.

Darwent, Christyann M., and John Darwent. 2016. "The Enigmatic Choris and Old Whaling 'Cultures' of the Western Arctic." *The Oxford Handbook of the Prehistoric Arctic. Oxford University Press, New York*, 371–94.

Darwent, John, and Christyann M. Darwent. 2005. "Occupational History of the Old Whaling Site at Cape Krusenstern, Alaska." *Alaska Journal of Anthropology* 3 (2): 135–54.

Darwent, John, Owen K. Mason, John F. Hoffecker, and Christyann M. Darwent. 2013. "1,000 Years of House Change at Cape Espenberg, Alaska: A Case Study in Horizontal Stratigraphy." *American Antiquity* 78 (3): 433–55.

Dneprovsky, K. 2006. "A Late Birnirk House at Paipelghak in Northern Chukotka: A Preliminary Report Based on the Excavations from 2002-2004." *Alaska J. Anthropol.* 4: 34–53.

Dneprovsky, K., and M. Bronshtein. 2002. "Ekven H-18: A Birnirk and Early Punuk-Period Site in Chukotka. Archaeology in the Bering Strait Region: Research on Two Continents." *University of Oregon Anthropological Papers,* 59: 166–206.

Duggan, Ana T., Alison J. T. Harris, Stephanie Marciniak, Ingeborg Marshall, Melanie Kuch, Andrew Kitchen, Gabriel Renaud, et al. 2017. "Genetic Discontinuity between the Maritime Archaic and Beothuk Populations in Newfoundland, Canada." *Current Biology: CB*, 1–8.

Enk, Jacob M., Alison M. Devault, Melanie Kuch, Yusuf E. Murgha, Jean-Marie Rouillard, and Hendrik N. Poinar. 2014. "Ancient Whole Genome Enrichment Using Baits Built from Modern DNA." *Molecular Biology and Evolution* 31 (5): 1292–94.

Ersmark, Erik, Ludovic Orlando, Edson Sandoval-Castellanos, Ian Barnes, Ross Barnett, Anthony Stuart, Adrian Lister, and Love Dalén. 2015. "Population Demography and Genetic Diversity in the Pleistocene Cave Lion." https://doi.org/10.5334/oq.aa.

Evin, Allowen, Thomas Cucchi, Andrea Cardini, Una Strand Vidarsdottir, Greger Larson, and Keith Dobney. 2013. "The Long and Winding Road: Identifying Pig Domestication through Molar Size and Shape." *Journal of Archaeological Science* 40 (1): 735–43.

Evin, Allowen, Thibaud Souter, Ardern Hulme-Beaman, Carly Ameen, Richard Allen, Pietro Viacava, Greger Larson, Thomas Cucchi, and Keith Dobney. 2016. "The Use of Close-Range Photogrammetry in Zooarchaeology: Creating Accurate 3D Models of Wolf Crania to Study Dog Domestication." *Journal of Archaeological Science: Reports* 9: 87–93.

Fitzhugh, Ben, Erik Gjesfjeld, William Brown, Mark J. Hudson, and Jennie D. Shaw. 2016. "Resilience and the Population History of the Kuril Islands, Northwest Pacific: A Study in Complex Human Ecodynamics." *Quaternary International: The Journal of the International Union for Quaternary Research* 419 (October): 165–93.

Fitzhugh, B., M. Etnier, B. MacInnes, S. C. Phillips, and J. Taylor. 2007. "Report of Archaeological Field Research in 2006, Including Geological Descriptions of Archaeological Locales (Unpublished Project Report for Kuril Biocomplexity Project)." University of Washington (tDAR id: 376122); https://doi.org/10.6067/XCV8PK0FCF.

{Fitzhugh, B., M. Etnier, S. C. Phillips, and E. }. Gjesfjeld. 2009. "Report of Archaeological Field Research in 2008, Including Geological Descriptions of Archaeological Locales (Unpublished Project Report for Kuril Biocomplexity Project)." *University of Washington* tDAR id: 376123. https://doi.org/10.6067/XCV8F18XZ9.

Fitzhugh, William W. 2016. "Solving the 'Eskimo Problem': Henry Bascom Collins and Arctic Archaeology." In *Early Inuit Studies, Themes and Transitions, 1850s-1980s.*, edited by Igor Krupnik. Washington, D.C.: Smithsonian Institution Scholarly Press.

Flegontov, Pavel, Nefize Ezgi Altinisik, Piya Changmai, Nadin Rohland, Swapan Mallick, Deborah A. Bolnick, Francesca Candilio, et al. 2017. "Paleo-Eskimo Genetic Legacy across North America."

bioRxiv. https://doi.org/10.1101/203018.
Forbes, Véronique, Kate Britton, and Rick Knecht. 2015. "Preliminary Archaeoentomological Analyses of Permafrost-Preserved Cultural Layers from the Pre-Contact Yup'ik Eskimo Site of Nunalleq , Alaska: Implications, Potential and Methodological Considerations." *Environmental Archaeology* 20 (2): 158–67.
Ford, J. A. 1959. "Eskimo Prehistory in the Vicinity of Point Barrow, Alaska." *Anthropological Papers of the American Museum of Natural History* 47 (1): 1–272.
Frantz, Laurent A. F., Victoria E. Mullin, Maud Pionnier-Capitan, Ophélie Lebrasseur, Morgane Ollivier, Angela Perri, Anna Linderholm, et al. 2016. "Genomic and Archaeological Evidence Suggest a Dual Origin of Domestic Dogs." *Science* 352 (6290): 1228–31.
Friesen, T. M. A. X. 2009. "PAN-ARCTIC POPULATION MOVEMENTS The Early Paleo-Inuit and Thule Inuit Migrations." In *The Oxford Handbook of the Prehistoric Arctic*, edited by Max Friesen and Owen Mason.
Friesen, T. Max. 2015. "On the Naming of Arctic Archaeological Traditions: The Case for Paleo-Inuit." *Arctic* 68 (3). http://arctic.journalhosting.ucalgary.ca/arctic/index.php/arctic/article/download/4504/4622.
Friesen, T. Max, and Charles D. Arnold. 1995. "Zooarchaeology of a Focal Resource: Dietary Importance of Beluga Whales to the Precontact Mackenzie Inuit." *Arctic* 48 (1): 22–30.
Friesen, T. Max, and Matthew Betts. 2006. "Archaeofaunas and Architecture: Zooarchaeological Variability within an Inuit Semi-Subterranean House, Arctic Canada." In *Integrating Zooarchaeology*, edited by Mark Maltby, 65–76. Oxford: Oxbow Press.
Garðarsdóttir, V. B. 2010. "Fornleifauppgröftur á Alþingisreitnum 2008-2010." *Reykjavík: Alþingi Og Framkvæmdasýsla Ríkisins.*
———. 2011. "Alþingisreiturinn: Upphaf Landnáms í Reykjavík." *Árbók Hins íslenzka Fornleifafélags, 2011, 5–43.*
———. 2013. "Alþingisreitur 2012." *Reykjavík.*
Geist, Otto William, and Froelich G. Rainey. 1936. "Archaeological Excavations at Kukulik, St. Lawrence Island, Alaska." *Government Printing Office, Washington, DC*.
Giddings, J. 1973. "Louis 1964 The Archaeology of Cape Denbigh." *Brown UniversityPress, Providence*.
Giddings, J. L. 1957. "Round Houses in the Western Arctic." *American Antiquity* 23 (2Part1): 121–35.
———. 1960. "The Archeology of Bering Strait." *Current Anthropology* 1 (2): 121.
Giddings, J. Louis. 1952. *The Arctic Woodland Culture of the Kobuk River*. Museum Monographs,. Philadelphia: University Museum, University of Pennsylvania,.
Giddings, J. Louis, and Douglas D. Anderson. 1986. "Beach Ridge Archaeology of Cape Krusenstern: Eskimo and Pre-Eskimo Settlements around Kotzebue Sound Alaska." *Publications in Archaeology 20. National Park Service* U.S. Department of the Interior, Washington, D.C.
Gilbert, M. Thomas P., Hans-Jürgen Bandelt, Michael Hofreiter, and Ian Barnes. 2005. "Assessing Ancient DNA Studies." *Trends in Ecology & Evolution* 20 (10): 541–44.
Gilbert, M. Thomas P., Lynn P. Tomsho, Snjezana Rendulic, Michael Packard, Daniela I. Drautz, Andrei Sher, Alexei Tikhonov, et al. 2007. "Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts." *Science* 317 (5846): 1927–30.
Glob, P. V. 1935. *Eskimo Settlements in Kempe Fjord and King Oskar Fjord, with Zoological Appendix by M. Degerbøl*. Vol. 102 No. 2. Meddelelser Om Grønland . Copenhagen.: C.A. Reitzel.
Goodall, C. R. 1995. "Procrustes Methods in the Statistical Analysis of Shape Revisited." *Current Issues in Statistical Shape Analysis*, 18–33.
Gotfredsen, Anne Birgitte. 2010. "Faunal Remains from the Wollaston Forland – Clavering Ø Region, Northeast Greenland – Thule Culture Subsistence in a High Arctic Polynya and Ice-Edge Habitat." *Geografisk Tidsskrift-Danish Journal of Geography* 110 (2): 175–200.
Gotfredsen, Anne Birgitte, and Tinna Møbjerg. 2010. "Nipisat-a Saqqaq Culture Site in Sisimiut, Central West Greenland (Vol. 331)." http://www.oapen.org/search?identifier=342365.
Grønnow, Bjarne. 1994. "Qeqertasussuk-the Archaeology of a Frozen Saqqaq Site in Disko Bugt, West

Greenland." *Threads of Arctic Prehistory: Papers in Honour of William E. Taylor Jr. Ottawa, Canada: Canadian Museum of Civilization*, 197–238.
———. 2017. *The Frozen Saqqaq Sites of Disko Bay, West Greenland. Qeqertasussuk and Qajaa (2400 – 900 BC)*. Monographs on Greenland. Museum Tusculanum Press.
Gronnow, B., and J. F. Jensen. 2009. *The Northernmost Ruins of the Globe*.
Gullov, Hans Christian. 1997. *From Middle Ages to Colonial Times*. Museum Tusculanum Press.
Gusev, Sergey V., Andrey V. Zagoroulko, and Aleksey V. Porotov. 1999. "Sea Mammal Hunters of Chukotka, Bering Strait: Recent Archaeological Results and Problems." *World Archaeology* 30 (3): 354–69.
Harrison, R. 2009. "The Gásir Area A Archaeofauna: An Update of the Results from the Faunal Analysis of the High Medieval Trading Site in Eyjafjörður, N Iceland."
———. 2011. "Myrkárdalur in Hörgárdalur, N. Iceland: Brief Summary of the 2008/2009 Archaeofauna." *NORSEC/HERC Zooarchaeology Laboratory Report* 57.
———. 2014. *Connecting the Land to the Sea at Gásir. Long-Term Human Ecodynamics in the North Atlantic: An Archaeological Study*. Vol. 117–36. Lanham, Maryland, : Lexington Publishers.
Hays, J. M., J. D. Reuther, C. Wooley, J. S. Rogers, R. C. Bowman, J. Baxter-McIntosh, and M. M. Proue. 2012. "Life on the River: Community Archaeology at SLT-094, Middle Kuskokwim River, Alaska." *Report Prepared by Northern Land Use Research, Inc., Fairbanks, and Chumis Cultural Resource Services, Anchorage.*
Heizer, Robert Fleming. 1956. *Archaeology of the Uyak Site, Kodiak Island, Alaska*. Vol. 17. University of California Press.
Helmer, James W. 1991. "The Palaeo-Eskimo Prehistory of the North Devon Lowlands." *Arctic* 44 (4): 301–17.
Hicks, Megan, Frank Feeley Adolf Friðriksson, George Hambrecht, Lilja Pálsdottir, Garðar Guðmundsson, and Magnus Á. Sigurgeirsson. 2013. "Midden Excavations at Skútustaðir N. Iceland, 2011. Report FS508-08274." *Fornleifastofnun Íslands Institute of Archaeology, Reykjavík.*
Hill, Erica. 2018. "Humans , Birds and Burial Practices at Ipiutak , Alaska : Perspectivism in the Western Arctic Humans , Birds and Burial Practices at Ipiutak , Alaska : Perspectivism in the Western Arctic." *Environmental Archaeology* 0 (0): 1–15.
Hodgetts, Lisa M., and Edward J. H. Eastaugh. 2017. "The Role of Magnetometry in Managing Arctic Archaeological Sites in the Face of Climate Change." *Advances in Archaeological Practice* 5 (2): 110–24.
Houlette, Chris. 2009. "Lost in the Collection: Reconsidering the Meat Cache 35 Assemblage and the Question of a Thule Occupation at Kukulik." *Alaska Journal of Anthropology* 7 (2): 101–20.
Hrdlicka, Ales. 1928. "Explorations in Alaska." *46th Annual Report. Bureau of American Ethnography*.
———. 1930. "Anthropological Survey in Alaska." *46th Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution,1928-1929. Government Printing Office, Washington, D.C.*, 19–347.
Imes, Donna L., Elizabeth J. Wictum, Marc W. Allard, and Benjamin N. Sacks. 2012. "Identification of Single Nucleotide Polymorphisms within the mtDNA Genome of the Domestic Dog to Discriminate Individuals with Common HVI Haplotypes." *Forensic Science International. Genetics* 6 (5): 630–39.
Irving, W. N. 1953. *Evidence of Early Tundra Cultures in Northern Alaska*. University of Alaska.
Jensen, Anne M. 2009. "Nuvuk, Point Barrow, Alaska: The Thule Cemetery and Ipiutak Occupation." Edited by Richard S. Davis. Ann Arbor, United States: Bryn Mawr College. https://search.proquest.com/docview/304833059.
———. 2016. "Archaeology of the Late Western Thule/Iñupiat in North Alaska (a.d. 1300–1750)." In *The Oxford Handbook of the Prehistoric Arctic*, edited by Max Friesen And. Oxford University Press.
Johansen, Trine B. 2012. "A Zoorchaeological and Ethnographic Investigation of Subsistence Change through Time at Iita, Northwest Greenland." Ph.D. Dissertation, University of California, Davis.

Johnson, Donald Mcintosh. 1933. “Observations on the Eskimo Remains on the East Coast of Greenland Between 72 and 75 North Latitude.” *Meddelelser Om Grønland* 92 (6).
Jónsson, Hákon, Aurélien Ginolhac, Mikkel Schubert, Philip L. F. Johnson, and Ludovic Orlando. 2013. “mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters.” *Bioinformatics* 29 (13): 1682–84.
Kaplan, Susan A. 1980. “Neo-Eskimo Occupations of the Northern Labrador Coast.” *Arctic* 33 (3): 646–58.
———. 2012. “Labrador Inuit Ingenuity and Resourcefulness: Adapting to a Complex Environmental, Social, and Spiritual Environment.” In *Settlement, Subsistence, and Change among the Labrador Inuit: The Nunatsiavummiut Experience*, edited by David C. Natcher, Lawrence Felt, and Andrea Procter, 15–42. University of Manitoba Press, Winnipeg.
Kaplan, Susan A., and James M. Woollett. 2000. “Challenges and Choices: Exploring the Interplay between Climate, History, and Culture on Canada’s Labrador Coast.” *Arctic, Antarctic, and Alpine Research* 32 (3): 351–59.
Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80.
Kemp, Brian M., Cara Monroe, and David Glenn Smith. 2006. “Repeat Silica Extraction: A Simple Technique for the Removal of PCR Inhibitors from DNA Extracts.” *Journal of Archaeological Science* 33 (12): 1680–89.
Kim, K. S., S. E. Lee, H. W. Jeong, and J. H. Ha. 1998. “The Complete Nucleotide Sequence of the Domestic Dog (Canis Familiaris) Mitochondrial Genome.” *Molecular Phylogenetics and Evolution* 10 (2): 210–20.
Klinken, G. J. van. 1999. “Bone Collagen Quality Indicators for Paleodietary and Radiocarbon Measurements.” *Journal of Archaeological Science* 26 (6): 687–96.
Kuzmin, Yaroslav V. 1997. “Vertebrate Animal Remains from Prehistoric and Medieval Settlements in Primorye (Russian Far East).” *International Journal of Osteoarchaeology* 7 (2): 172–80.
Laguna, Frederica de. 1936. “An Archaeological Reconnaissance of the Middle and Lower Yukon Valley, Alaska 2.” *American Antiquity* 2 (1): 6–12.
Larsen, Helge. 2001. *Deering: A Men’s House from Seward Peninsula, Alaska*. Edited by Martin Appelt. Department of Ethnography, the National Museum of Denmark.
Larsen, Helge, and Thorvald Sørensen. 1934. *Dødemandsbugten: An Eskimo Settlement on Clavering Island; Treaarsexpeditionen Til Christian Den X’s Land 1931-34*. Reitzel.
Ledger, Paul M., Véronique Forbes, Edouard Masson-Maclean, Charlotta Hillerdal, W. Derek Hamilton, Ellen McManus-Fry, Ana Jorge, Kate Britton, and Richard A. Knecht. 2018. “THREE GENERATIONS UNDER ONE ROOF? BAYESIAN MODELING OF RADIOCARBON DATA FROM NUNALLEQ, YUKON-KUSKOKWIM DELTA, ALASKA.” *American Antiquity* 83 (3): 505–24.
Ledger, Paul M., Véronique Forbes, Edouard Masson-maclean, and Richard A. Knecht. 2016. “Dating and Digging Stratified Archaeology in Circumpolar North America : A View from Nunalleq , Southwestern Alaska” 69 (4): 378–90.
Lee, Esther J., D. Andrew Merriwether, Alexei K. Kasparov, Pavel A. Nikolskiy, Marina V. Sotnikova, Elena Yu Pavlova, and Vladimir V. Pitulko. 2015. “Ancient DNA Analysis of the Oldest Canid Species from the Siberian Arctic and Genetic Contribution to the Domestic Dog.” *PloS One* 10 (5): 1–13.
LeMoine, Genevieve M., and Christyann M. Darwent. 2010. “Zooarchaeological Analysis of a Late Dorset and an Early Thule Dwelling at Cape Grinnell, Northwest Greenland.” *Geografisk Tidsskrift – Danish Journal of Geography* 100 (2): 315–36.
Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo

Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.
Lindblad-Toh, Kerstin, Claire M. Wade, Tarjei S. Mikkelsen, Elinor K. Karlsson, David B. Jaffe, Michael Kamal, Michele Clamp, et al. 2005. “Genome Sequence, Comparative Analysis and Haplotype Structure of the Domestic Dog.” *Nature* 438 (7069): 803–19.
Lobdell, John E. 1980. “Prehistoric Human Populations and Resource Utilization in Kachemak Bay, Gulf of Alaska.” Unpublished PhD Dissertation, University of Tennessee, Knoxville.
Longin, R. 1971. “New Method of Collagen Extraction for Radiocarbon Dating.” *Nature* 230 (5291): 241–42.
Loring, Stephen, and Beatrix Arendt. 2009. “‘… They Gave Hebron, The City of Refuge…’(Joshua 21: 13): An Archaeological Reconnaissance at Hebron, Labrador.” *Journal of the North Atlantic* 2 (sp1): 33–56.
Mary-Rouselière, Guy. 1976. “The Paleoeskimo in Northern Baffin Island.” In *Eastern Arctic Prehistory: Paleoeskimo Problems,* edited by Moreau S. Maxwell, 31:40–57. Memoirs of the Society for American Archaeology. Washington, D.C.: Society for American Archaeology.
———. 1979. “The Thule Culture on North Baffin Island: Early Thule Characteristics and the Survival of the Thule Tradition.” In *Thule Eskimo Culture: An Anthropological Retrospective*, edited by Alan P. McCartney, 54–75. Archaeological Survey of Canada Paper 88, National Museum of Man Mercury Series. Ottawa: National Museums of Canada.
Mason, Owen K. 2016. “Thule Origins in the Old Bering Sea Culture: The Interrelationship of Punuk and Birnirk Cultures.” In *The Oxford Handbook of the Prehistoric Arctic*, edited by T. Max Friesen and Owen K. Mason, 489–512. Oxford: Oxford University Press.
Mathiassen, Therkel. 1927a. *Archaeology of the Central Eskimos: II. The Thule Culture and Its Position within the Eskimo Culture*. Vol. IV. Report of the Fifth Thule Expedition 1921-24. Copenhagen: Gyldendalske.
———. 1927b. “Archaeology of the Central Eskimos: The Thule Culture and Its Position within the Eskimo Culture. Report of the Fifth Thule Expedition 1921--24.” Nordisk Forlag Copenhagen.
———. 1930a. “An Old Eskimo Culture in West Greenland: Report of an Archeological Expedition to Upernivik.” *Geographical Review* 20 (4): 605–14.
———. 1930b. *Inugsuk: A Mediaeval Eskimo Settlement in Upernivik District, West Greenland*. Luno.
———. 1933. *Prehistory of the Angmagssalik Eskimos*. CA Reitzel.
———. 1958. *The Sermermiut Excavations: 1955*. CA Reitzel.
Mathiassen, T., and E. Holtved. 1936. “The Eskimo Archaeology of Julianehaab District: With a Brief Summary of the Prehistory of the Greenlanders.” *Reitzel*.
Max Friesen, T., and Charles D. Arnold. 2008. “The Timing of the Thule Migration: New Dates from the Western Canadian Arctic.” *American Antiquity* 73 (3): 527–38.
Maxwell, Moreau S. 1973. *Archaeology of the Lake Harbour District, Baffin Island*. Archaeological Survey of Canada, National Museum of Man, National Museums of ….
McCartney, Allen P. 1977. *Thule Eskimo Prehistory along Northwestern Hudson Bay*. National Museum of Man Mercury Series. Archaeological Survey of Canada Paper 70,. Ottawa: National Museums of Canada,.
McCullough, Karen Margrethe. 1989. *The Ruin Islanders: Thule Culture Pioneers in the Eastern High Arctic*. Archaeological Survey of Canada. Canadian Museum of Civilization.
McGhee, Robert. 1975. “Beluga Hunters: An Archaeological Reconstruction of the History and Culture of the Mackenzie Delta Kittegaryumiut.” *Newfoundland Social and Economic Studies 13. Memorial University of Newfoundland, St. John’s.*
———. 1979. “The Paleoeskimo Occupations at Port Refuge, High Arctic Canada.” *Musée National de l’Homme. Collection Mercure. Commission Archéologique Du Canada. Publications d'Archéologie. Dossier Ottawa*, no. 92: 1–132.
McGovern, Thomas Howatt, Konrad Smiarowski, and Ramona Harrison. 2013. “Hard Times at Hofstaðir?: An Archaeofauna circa 1300 AD from Hofstaðir in M\`yvatnssveit, N Iceland.”

https://rafhladan.is/bitstream/handle/10802/7465/HERC-NORSEC_Report_60.pdf?sequence=1.
McGowan, Glenys, and Jonathan Prangnell. 2006. “The Significance of Vivianite in Archaeological Settings.” *Geoarchaeology* 21 (1): 93–111.
McManus-Fry, Ellen, Rick Knecht, Keith Dobney, Michael P. Richards, and Kate Britton. 2016. “Dog-Human Dietary Relationships in Yup’ik Western Alaska: The Stable Isotope and Zooarchaeological Evidence from Pre-Contact Nunalleq.” *Journal of Archaeological Science: Reports*. https://doi.org/10.1016/j.jasrep.2016.04.007.
McManus-Fry, Ellen Teresa. 2015. “Pre-Contact Ecology , Subsistence and Diet on the Yukon-Kuskokwim Delta An Integrated Ecosystem Approach.” University of Aberdeen.
Meldgaard, J. 1983. “Qaja, En Kokkenmodding I Dybfrost.” *Copenhagen: Feltrapport Fra Arbejdsmarken I Grönland, Saertryk Af National Museet, Arbejdsmark*.
Meldgaard, Jørgen. 1960. “Prehistoric Culture Sequences in the Eastern Arctic as Elucidated by Stratified Sites at Igloolik.” In *Men and Cultures*, edited by Anthony F. C. Wallace and Anthony F. C. Wallace. Philadelphia: University of Pennsylvania Press.
Meldgaard, Morten. 2004. *"Ancient Harp Seal Hunters of Disko Bay. Subsistence and Settlement at the Saqqaq Culture Site Qeqertasussuk (2400-1400BC), West Greenland*. Vol. 330. : Monographs on Greenland | Meddelelser Om Grønland. Museum Tusculanum Press.
Méreuze, Rémi. 2015. “La Construction de La Maison 33 Du Cap Espenberg, Nord-Ouest de l’Alaska, Au Xviiie Siècle.” *Nouvelles Orientations Dans Les Peches. Serie de Notes de Synthese Sur Les Questions de Developpement*, no. 141 (September): 19–25.
Meyer, Matthias, and Martin Kircher. 2010. “Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing.” *Cold Spring Harbor Protocols* 2010 (6): db.prot5448.
Møhl, Jeppe. 1979. “AN ESKIMO SETTLEMENT REPRESENTATIVE OF THE THULE CULTURE IN WEST GREENLAND.” In *Thule Eskimo Culture*, 380–94. An Anthropological Retrospective. University of Ottawa Press.
———. 1986. “Dog Remains from a Paleoeskimo Settlement in West Greenland.” *Arctic Anthropology* 23 (1/2): 81–89.
Monteiro, L. R. 1999. “Multivariate Regression Models and Geometric Morphometrics: The Search for Causal Factors in the Analysis of Shape.” *Systematic Biology* 48 (1): 192–99.
Moody, John F., and Lisa M. Hodgetts. 2013. “Subsistence Practices of Pioneering Thule–Inuit: A Faunal Analysis of Tiktalik.” *Arctic Anthropology* 50 (2): 4–24.
Morey, Darcy F. 2010. *Dogs: Domestication and the Development of a Social Bond*. Cambridge University Press.
Morey, Darcy F., and Kim Aaris-Sørensen. 2002. “Paleoeskimo Dogs of the Eastern Arctic.” *Arctic* 55 (1): 44–56.
Morgan, E. D., L. Titus, R. J. I. Small, and Corony Edwards. 1983. “The Composition of Fatty Materials from a Thule Eskimo Site on Herschel Island.” *Arctic* 36 (4): 356–60.
Morrison, David. 1990. *Iglulualumiut Prehistory: The Lost Inuit of Franklin Bay*. Archaeological Survey of Canada Mercury Series Paper 116. Hull, Quebec.: Canadian Museum of Civilization.
———. 2000. “Inuvialuit Fishing and the Gutchiak Site.” *Arctic Anthropology* 37 (1): 1–42.
Morrison, David A. 1983. *Thule Culture in Western Coronation Gulf, N.W.T*. Vol. Archaeological Surey of Canada Paper 116,. National Museum of Man Mercury Series. . Ottawa: National Museums of Canada.
Moss, Madonna. 2004. *Archaeological Investigation of Cape Addington Rockshelter: Human Occupation of the Rugged Seacoast on the Outer Prince of Wales Archipelago, Alaska*. University of Oregon Anthropological Paper No. 63. University of Oregon, Department of Anthropology.
Moss, Madonna, Justin M. Hays, Peter Bowers, and Douglas R. Reger. 2016. *The Archaeology of Coffman Cove: 5500 Years of Settlement in the Heart of Southeast Alaska*. University of Oregon Museum of Natural and Cultural History and the Department of Anthropology, University of Oregon.
Murray, Maribeth S. 1996. “Economic Change in the Palaeoeskimo Prehistory of the Foxe Basin, NWT.”

https://macsphere.mcmaster.ca/handle/11375/12973.
Murray, Maribeth Suzanne. 2005. “Prehistoric Use of Ringed Seals: A Zooarchaeological Study from Arctic Canada.” *Environmental Archaeology* 10 (1): 19–38.
Nash, R. J. 1976. “Cultural Systems and Culture Change in the Central Arctic.” *Memoirs of the Society for American Archaeology,* 31: 150–55.
Nash, Ronald J. 1972. “Dorset Culture in Northeastern Manitoba, Canada.” *Arctic Anthropology* 9 (1): 10–16.
Natcher, David C., and Andrea H. Procter. 2012. *Settlement, Subsistence, and Change Among the Labrador Inuit: The Nunatsiavummiut Experience*. Univ. of Manitoba Press.
Nikitin, Iu G., and S. Chzhun. 2009. *Arkheologicheskie Issledovaniia Na Poselenii Cherniatino 2 v Primor’e v 2008 Gody*. Vol. 1-2. Tedzhon, DVGTU, IIAiE DVO RAN, Vladivostok.
Nikitin, Iu.G., Chzhun, S., Chzho, Li Ch. 2008. *Arkheologicheskie Issledovaniia Na Poselenii Cherniatino 2 v Primor’e v 2007 Gody*. Vol. 1-2. Chunnam Vuekyn, DVGTU, IIAiE DVO RAN, Vladivostok.
Ní Leathlobhair, Máire, Angela R. Perri, Evan K. Irving-Pease, Kelsey E. Witt, Anna Linderholm, James Haile, Ophelie Lebrasseur, et al. 2018. “The Evolutionary History of Dogs in the Americas.” *Science*, no. 361: 81–85.
Nomokonova, Tatiana, Robert J. Losey, Ol’ga I. Goriunova, and Andrzej W. Weber. 2013. “A Freshwater Old Carbon Offset in Lake Baikal, Siberia and Problems with the Radiocarbon Dating of Archaeological Sediments: Evidence from the Sagan-Zaba II Site.” *Quaternary International: The Journal of the International Union for Quaternary Research* 290-291 (March): 110–25.
Nomokonova, T. Iu, and O. I. Goriunova. 2013. “Kosti Zhivotnykh Iz Kammenykh Konstruktsii Bukhty Sagan-Nuge Na Malom More Oz. Baikal [Animal Bones from Stone Features of Sagan-Nuge Cove in the Little Sea Region of Lake Baikal].” *Bulletin of the Irkutsk State University. Geoarchaeology, Ethnology, and Anthropology* Series 1(2): 44–61.
Norman, Lauren E. Y., T. Max Friesen, Claire Alix, Michael J. E. O’Rourke, and Owen K. Mason. 2017. “An Early Inupiaq Occupation: Observations on a Thule House From Cape Espenberg, Alaska.” *Open Archaeology* 3 (1): 57.
Ólafsson, G. 2010. “Bessastaðarannsókn 1987: Aðdragandi Og Upphaf - Uppgraftarsvæði 1-11 (Skýrslur Þjóðminjasafns Íslands No. 2010/1).” *Reykjavík: Þjóðminjasafn Íslands.*
———. 2013. “Bessastaðarannsókn II: Kirkjugarður Og Miðaldaminjar, Uppgraftarsvæði 12-15 (Skýrslur Þjóðminjasafns Íslands No. 2013/2).” *Reykjavík: Þjóðminjasafn Íslands.*
Oswalt, Wendell. 1952. “The Archaeology of Hooper Bay Village, Alaska.” *Anthropological Papers of the University of Alaska* 1 (1): 47–91.
Pálsdóttir, A. H. 2010. “Dýrabein Frá Alþingisreit: Greining á Dýrabeinum Frá Svæðum A, B Og C (Skýrslur Íslenskra Fornleifarannsókna Ehf. No. 2010–1).” *Reykjavík: Íslenskar Fornleifarannsóknir Ehf.*
———. 2013. “Dýrabeinin Frá Alþingisreit IV. Fasi (871-1226): Uppgröftur 2008-2012 (Skýrslur Íslenskra Fornleifarannsókna Ehf. No. 2013–1).” *Reykjavík: Íslenskar Fornleifarannsóknir Ehf.*
Pálsdóttir, A. H., S. Boessenkool, and B. Star. 2018. “Ancient DNA Sampling Report: Walrus Bones from Alþingisreitur (Technical Report No. 100).” *Reykjavík: Landbúnaðarháskóli Íslands.*
Pang, Jun-Feng, Cornelya Kluetsch, Xiao-Ju Zou, Ai-Bing Zhang, Li-Yang Luo, Helen Angleby, Arman Ardalan, et al. 2009. “mtDNA Data Indicate a Single Origin for Dogs South of Yangtze River, Less than 16,300 Years Ago, from Numerous Wolves.” *Molecular Biology and Evolution* 26 (12): 2849–64.
Paradis, E., J. Claude, and K. Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2): 289–90.
Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2): 289–90.
Park, Robert W. 1987. “Dog Remains from Devon Island, NWT: Archaeological and Osteological Evidence for Domestic Dog Use in the Thule Culture.” *Arctic*, 184–90.

———. 1993. “The Dorset-Thule Succession in Arctic North America : Assessing Claims for Culture Contact Author (S): Robert W . Park Published by : Society for American Archaeology Stable URL : http://www.jstor.org/stable/281966 Http://www.jstor.org/page/info/about/p” 58 (2): 203–34.

Patterson, Nick, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. 2012. “Ancient Admixture in Human History.” *Genetics* 192 (3): 1065–93.

Patterson, Nick, Alkes L. Price, and David Reich. 2006. “Population Structure and Eigenanalysis.” *PLoS Genetics* 2 (12): e190.

Popov, Alexander N., Andrei V. Tabarev, and Yuri A. Mikishin. 2014. “Neolithization and Ancient Landscapes in Southern Primorye, Russian Far East.” *Journal of World Prehistory* 27 (3-4): 247–61.

Popov, A. N., N. F. Chikisheva, and E. G. Shpakova. 1997. “Boismanskaya Arheologicheskaya Kultura v Yuzhnom Primorye.(Boisman Archaeological Culture in Southern Primorye). Novosibirsk: Institute of Archaeology and Ethnography.”

Popov, A. N., and B. V. Lazin. 2011. “Arkheologicheskie Issledovaniia Na Ostrove Russkom v G. Vladivostoke v 2010-2011 Godakh.” In *Drevnosti Po Obe Storony Velikogo Okeana,* edited by D. L. Brodianskii, 118–26. DVGTU, Vladivostok.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American Journal of Human Genetics* 81 (3): 559–75.

Raghavan, M., M. DeGiorgio, A. Albrechtsen, I. Moltke, P. Skoglund, T. S. Korneliussen, B. Gronnow, et al. 2014. “The Genetic Prehistory of the New World Arctic.” *Science* 345 (6200): 1255832–1255832.

Rainey, Froelich G. 1941. “Eskimo Prehistory: The Orvik Site on the Punuk Islands.” *Anthropological Papers of the American Museum of Natural History*.

Rakov, V. A., A. N. Popov, L. E. Vasil’eva, Iu V. Zavertanova, and Iu A. }. Mikishin. 2009. “Fauna Pribrezhnoi Zony Proliva Bosfor-Vostochnyi Perioda Zheleznogo Veka (po Materialan Spasatel’nykh Raskopok Pamiatnikov Nazimova-1 I Pospelova-1 v G. Vladivostoke).” In *Ot Mongolii Do Primor’ia I Sakhalina*, edited by D. L. Brodianskii, 162–212. DVGTU, Vladivostok.

Rankin, L. 2014. “Indian Harbour, Norman’s Island and Double Mer Point, Labrador.” *Provincial Archaeology Office 2013 Archaeology Review St. John’s, NL*.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reger, D. R., Madonna L. Moss, Peter M. Bowers, and Justin M. Hays. 2017. *Recovery of Archaeological Data from the Ferry Terminal Site (49-PET-556), Coffman Cove, Alaska*. Preliminary Report Submitted to the USDA Forest Service, December 2017. Fairbanks, AK: Northern Land Use Research.

Reimer, Paula J., Edouard Bard, Alex Bayliss, J. Warren Beck, Paul G. Blackwell, Christopher Bronk Ramsey, Caitlin E. Buck, et al. 2013. “IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years Cal BP.” *Radiocarbon* 55 (4): 1869–87.

Renouf, M. A. P. 1993. “Palaeoeskimo Seal Hunters at Port Au Choix, Northwestern Newfoundland.” *Newfoundland and Labrador Studies*. https://journals.lib.unb.ca/index.php/NFLDS/article/viewFile/941/1294.

Rick, Anne Meachem. 1980. “Non-Cetacean Vertebrate Remains from Two Thule Winter Houses on Somerset Island, NWT.” *Canadian Journal of Archaeology/Journal Canadien d’Archéologie*, 99–117.

Riewe, R., Suluk, L., & Brandon, L. 1989. “Inuit Land Use and Occupancy in Northern Manitoba.” *Northern Review* 3/4. http://journals.sfu.ca/nr/index.php/nr/article/view/458.

Roberts, H. M. 2009. “Gásir Post Excavation Reports-.” *Fornleifastofnun Íslands, Reykjavík.* 1 (FS423-010712,).

Rohland, Nadin, and Michael Hofreiter. 2007. “Ancient DNA Extraction from Bones and Teeth.” *Nature Protocols* 2 (7): 1756–62.

Rohland, Nadin, Heike Siedel, and Michael Hofreiter. 2010. “A Rapid Column-Based Ancient DNA Extraction Method for Increased Sample Throughput.” *Molecular Ecology Resources* 10 (4): 677–83.
Rohlf, F. J. 2004. “TpsDig, Version 1.4.” *Http://life.bio.sunysb.edu/morph/*. https://ci.nii.ac.jp/naid/20001496718/.
Rohlf, F. James, and Dennis Slice. 1990. “Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks.” *Systematic Biology* 39 (1): 40–59.
Ross, W. Gillies. 1977. “Whaling and the Decline of Native Populations [Corrected Title: Intoduced Disease and Eskimo Mortality during the Whaling Period: The Extinction of the Saglermiut in Hudson Bay, 1902].” *Arctic Anthropology* 14 (2): 1–8.
Ryan, Karen, and Janet Young. 2013. “Identification of a Probable Aarnguaq in a Sadlermiut Grave from Native Point, Southampton Island, Nunavut, Canada.” *Arctic Anthropology* 50 (1): 20–48.
Sabo, George, and John D. Jacobs. 1980. “Aspects of Thule Culture Adaptations in Southern Baffin Island.” *Arctic* 33 (3): 487–504.
Sampson, P. D., F. L. Bookstein, F. H. Sheenan, and E. L. Bolson. 1996. “Eigenshape Analysis of Left Ventricular Outlines from Contrast Ventriculograms.” In *Advances in Morphometrics*, edited by L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylar, and D. E. Slice, 211–33. NATO ASI Series A: Life Sciences. New York: Plenum.
Savolainen, Peter, Thomas Leitner, Alan N. Wilton, Elizabeth Matisoo-Smith, and Joakim Lundeberg. 2004. “A Detailed Picture of the Origin of the Australian Dingo, Obtained from the Study of Mitochondrial DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (33): 12387–90.
Savolainen, Peter, Ya-Ping Zhang, Jing Luo, Joakim Lundeberg, and Thomas Leitner. 2002. “Genetic Evidence for an East Asian Origin of Domestic Dogs.” *Science* 298 (5598): 1610–13.
Schlager, S. 2017. “Morpho and Rvcg – Shape Analysis in R.” In *Statistical Shape and Deformation Analysis*, edited by G. Zheng, S. Li, and G. Szekely, 217–56.
Schledermann, Peter. 1990. *Crossroads to Greenland: 3000 Years of Prehistory in the Eastern High Arctic*. Arctic Institute of North America, the University of Calgary.
Schledermann, Peter, and Karen M. McCullough. 2003. “Late Thule Culture Developments on the Central East Coast of Ellesmere Island.” *Sila – The Greenland Research Centre at the National Museum of Denmark and Danish Polar Center, Copenhagen.*
Schulting, Rick J., Christopher Bronk Ramsey, Vladimir I. Bazaliiskii, Olga I. Goriunova, and Andrzej Weber. 2014. “Freshwater Reservoir Offsets Investigated Through Paired Human-Faunal 14C Dating and Stable Carbon and Nitrogen Isotope Analysis at Lake Baikal, Siberia.” *Radiocarbon* 56 (3): 991–1008.
Sheehan, Glenn W. 1997. *In the Belly of the Whale: Trade and War in Eskimo Society*. Vol. 6. Alaska Anthropological Association.
Shirar, Scott. 2011. “Late Holocene Chronology of the Noatak and Kobuk Rivers.” *Alaska Journal of Anthropology*, 1.
Sloan, Anna C. 2014. “Spirituality and the Seamstress: Birds in Ipiutak and Western Thule Lifeways at Deering, Alaska.” *Arctic Anthropology* 51 (2): 35–59.
Smiarowski, Konrad. 2012. “E172 Tatsip Ataa Midden Excavation 2009 & 2010 Preliminary Excavation Report.” *http://www.nabohome.org/publications/ipy/E172ReportDraft3KS3-20-12.pdf*.
Sørensen, Mikkel, and Hans Christian Gulløv. 2012. “The Prehistory of Inuit in Northeast Greenland.” *Arctic Anthropology* 49 (1): 88–104.
Staab, M. L. 1979. “Analysis of Faunal Material Recovered from a Thule Eskimo Site on the Island of Silumiut, N.W.T., Canada.” In *The Thule Eskimo Culture: An Anthropological Retrospective.*, edited by A. P. McCartney, 349–79. Mercury Series. Archaeological Survey of Canada Paper 88. Ottawa: National Museum of Man.
Stamatakis, Alexandros. 2006. “RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models.” *Bioinformatics* 22 (21): 2688–90.

———. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–13.
Steffian, Amy F. 1992. “Fifty Years after Hrdlicka: Further Excavation of the Uyak Site, Kodiak Island, Alaska.” *Contributions to the Anthropology of Southcentral and Southwestern Alaska*, 141–64.
Stenton, Douglas R. 1987. “Recent Archaeological Investigations in Frobisher Bay, Baffin Island, NWT.” *Canadian Journal of Archaeology/Journal Canadien d’Archéologie*, 13–48.
Stuiver, M., P. J. Reimer, and R. W. Reimer. 2019. “CALIB 7.1.” *Http://calib.org*.
Sutherland, Patricia D. 1996. “Continuity and Change in the Paleo-Eskimo Prehistory of Northern Ellesmere Island.” *The Paleo-Eskimo Cultures of Greenland: New Perspectives in Greenlandic Archaeology. Publication*, no. 1: 271–94.
Tackney, Justin, Anne M. Jensen, Caroline Kisielinski, and Dennis H. O’Rourke. 2019. “Molecular Analysis of an Ancient Thule Population at Nuvuk, Point Barrow, Alaska.” *American Journal of Physical Anthropology* 168 (2): 303–17.
Thai, Quan Ke, Dung Anh Chung, and Hoang-Dung Tran. 2017. “Canis mtDNA HV1 Database: A Web-Based Tool for Collecting and Surveying Canis mtDNA HV1 Haplotype in Public Database.” *BMC Genetics* 18 (1): 60.
Thalmann, O., B. Shapiro, P. Cui, V. J. Schuenemann, S. K. Sawyer, D. L. Greenfield, M. B. Germonpré, et al. 2013. “Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs.” *Science* 342 (6160): 871–74.
Thostrup, Christian Bendix. 1911. *Ethnographic Description of the Eskimo Settlements and Stone Remains in North-East Greenland*. Vol. 4. CA Reitzel.
Tremayne, Andrew H., Christyann M. Darwent, John Darwent, Kelly A. Eldridge, and Jeffrey T. Rasic. 2018. “Iyatayet Revisited: A Report on Renewed Investigations of a Stratified Middle-to-Late Holocene Coastal Campsite in Norton Sound, Alaska.” *Arctic Anthropology* 55 (1): 1–23.
VanStone, James W., and Charles V. Lucier. 1974. “An Early Archaeological Example of Tattooing from Northwestern Alaska.” *Fieldiana. Anthropology* 66 (1): 1–9.
Vésteinsson, Orri. 2014. “Archaeological Investigations in Igaliku: Excavations in the Meadow, 2012–2013. Report FS517-12151.” *Fornleifastofnun Íslands Institute of Archaeology, Reykjavík.*
Woollett, James M. 2003. “An Historical Ecology of Labrador Inuit Culture Change.” City University of New York.
———. 2007. “Labrador Inuit Subsistence in the Context of Environmental Change: An Initial Landscape History Perspective.” *American Anthropologist* 109 (1): 69–84.
Workman, Karen W. 1975. “Chugachik Island (SEL 033): Test Excavations at a Middle Kachemak Tradition Site in Kachemak Bay, Alaska, July 1974.” *Prepared for Alaska Division of Parks.*
———. 1978. “The Second Season at Chugachik Island (SEL 033), Kachemak Bay, Alaska.” In *Paper Presented at the 5th Annual Meeting of the Alaska Anthropological Association, March 17-18, Anchorage.*
Workman, William B. 1998. “Archaeology of the Southern Kenai Peninsula.” *Arctic Anthropology* 35 (1): 146–59.
Workman, William B., Lobdell John E., and Karen W. Workman. 1980. “Recent Archeological Work in Kachemak Bay, Gulf of Alaska.” *Arctic* 33 (3): 385–99.
Workman, William B., and Peter Zollars. 2002. “The Dispersal of the Arctic Small Tool Tradition into Southern Alaska: Dates and Data from the Kenai Peninsula, Southcentral Alaska.” *Anthropological Papers of the University of Alaska New Series* 2 (1): 39–49.
Yorga, Brian W. D. 1980. “Washout: A Western Thule Site on Herschel Island, Yukon Territory.” *Musée National de l’Homme. Collection Mercure. Commission Archéologique Du Canada. Publications d'Archéologie. Dossier Ottawa*, no. 98: 1–201.
Zelditch, Miriam Leah, Donald L. Swiderski, H. David Sheets, and William L. Fink. 2004. *Geometric Morphometrics for Biologists: A Primer*. Elsevier.
Zollars, Peter. 1982. “Chugachik Island Project Report, 1982.” *Prepared for the Otto Geist Fund Committee, University of Alaska Museum.*

Bibliography

1. Kämmerer U, Germeyer A, Stengel S, Kapp M, Denner J. Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. J Reprod Immunol. 2011;91:1–8. doi:10.1016/j.jri.2011.06.102.

2. Beyer U, Moll-Rocek J, Moll UM, Dobbelstein M. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. Proc Natl Acad Sci U S A. 2011;108:3624–9. doi:10.1073/pnas.1016201108.

3. Bénit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J Virol. 1999;73:3301–8. https://www.ncbi.nlm.nih.gov/pubmed/10074184.

4. Smit AF. Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res. 1993;21:1863–72. doi:10.1093/nar/21.8.1863.

5. López-Sánchez P, Costas JC, Naveira HF. Paleogenomic record of the extinction of human endogenous retrovirus ERV9. J Virol. 2005;79:6997–7004. doi:10.1128/JVI.79.11.6997-7004.2005.

6. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology. 2011;8:90. doi:10.1186/1742-4690-8-90.

7. Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. J Virol. 2014;88:9529–37. doi:10.1128/JVI.00919-14.

8. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, et al. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res. 2006;16:1548–56. doi:10.1101/gr.5565706.

9. Volkman HE, Stetson DB. The enemy within: endogenous retroelements and autoimmune disease. Nat Immunol. 2014;15:415–22. doi:10.1038/ni.2872.

10. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351:1083–7. doi:10.1126/science.aad5497.

11. Zeng M, Hu Z, Shi X, Li X, Zhan X, Li X-D, et al. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. Science. 2014;346:1486–92. doi:10.1126/science.346.6216.1486.

12. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, et al.

Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522:221–5. doi:10.1038/nature14308.

13. Sutkowski N, Chen G, Calderon G, Huber BT. Epstein-Barr virus latent membrane protein LMP-2A is sufficient for transactivation of the human endogenous retrovirus HERV-K18 superantigen. J Virol. 2004;78:7852–60. doi:10.1128/JVI.78.14.7852-7860.2004.

14. Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN. The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. J Immunol. 2006;176:7636–44. https://www.ncbi.nlm.nih.gov/pubmed/16751411.

15. Bhardwaj N, Maldarelli F, Mellors J, Coffin JM. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses in vivo but not to increased virion production. J Virol. 2014;88:11108–20. doi:10.1128/JVI.01623-14.

16. Büscher K, Hahn S, Hofmann M, Trefzer U, Ozel M, Sterry W, et al. Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. Melanoma Res. 2006;16:223–34. doi:10.1097/01.cmr.0000215031.07941.ca.

17. Bieda K, Hoffmann A, Boller K. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. J Gen Virol. 2001;82 Pt 3:591–6. doi:10.1099/0022-1317-82-3-591.

18. Hahn S, Ugurel S, Hanschmann K-M, Strobel H, Tondera C, Schadendorf D, et al. Serological response to human endogenous retrovirus K in melanoma patients correlates with survival probability. AIDS Res Hum Retroviruses. 2008;24:717–23. doi:10.1089/aid.2007.0286.

19. Chen T, Meng Z, Gan Y, Wang X, Xu F, Gu Y, et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating β-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. Leukemia. 2013;27:1469–78. doi:10.1038/leu.2013.8.

20. Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, et al. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int J Cancer. 2007;120:81–90. doi:10.1002/ijc.22256.

21. Ono M, Kawakami M, Ushikubo H. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J Virol. 1987;61:2059–62. https://www.ncbi.nlm.nih.gov/pubmed/2883329.

22. Holmes VA, Wallace JMW, Gilmore WS, McFaul P, Alexander HD. Plasma levels of the immunomodulatory cytokine interleukin-10 during normal human pregnancy: a longitudinal study. Cytokine. 2003;21:265–9. https://www.ncbi.nlm.nih.gov/pubmed/12823999.

23. Dummer W, Becker JC, Schwaaf A, Leverkus M, Moll T, Bröcker EB. Elevated serum levels of interleukin-10 in patients with metastatic malignant melanoma. Melanoma Res. 1995;5:67–8. https://www.ncbi.nlm.nih.gov/pubmed/7734958.

24. Hummel J, Kämmerer U, Müller N, Avota E, Schneider-Schaulies S. Human endogenous retrovirus envelope proteins target dendritic cells to suppress T-cell activation. Eur J Immunol. 2015;45:1748–59. doi:10.1002/eji.201445366.

25. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med. 2010;16:571–9, 1p following 579. doi:10.1038/nm.2129.

26. Denne M, Sauter M, Armbruester V, Licht JD, Roemer K, Mueller-Lantzsch N. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. J Virol. 2007;81:5607–16. doi:10.1128/JVI.02771-06.

27. Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, et al. Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. J Virol. 2004;78:10310–9. doi:10.1128/JVI.78.19.10310-10319.2004.

28. Galli UM, Sauter M, Lecher B, Maurer S, Herbst H, Roemer K, et al. Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. Oncogene. 2005;24:3223–8. doi:10.1038/sj.onc.1208543.

29. Fischer S, Echeverría N, Moratorio G, Landoni AI, Dighiero G, Cristina J, et al. Human endogenous retrovirus np9 gene is over expressed in chronic lymphocytic leukemia patients. Leuk Res Rep. 2014;3:70–2. doi:10.1016/j.lrr.2014.06.005.

30. Fuchs NV, Loewer S, Daley GQ, Izsvák Z, Löwer J, Löwer R. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 2013;10:115. doi:10.1186/1742-4690-10-115.

31. Herve CA, Lugli EB, Brand A, Griffiths DJ, Venables PJW. Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. Clin Exp Immunol. 2002;128:75–82. https://www.ncbi.nlm.nih.gov/pubmed/11982593.

32. Ehlhardt S, Seifert M, Schneider J, Ojak A, Zang KD, Mehraein Y. Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia. J Rheumatol. 2006;33:16–23. https://www.ncbi.nlm.nih.gov/pubmed/16395745.

33. Sicat J, Sutkowski N, Huber BT. Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. J Rheumatol. 2005;32:1821–31. https://www.ncbi.nlm.nih.gov/pubmed/16142883.

34. Krzysztalowska-Wawrzyniak M, Ostanek M, Clark J, Binczak-Kuleta A, Ostanek L, Kaczmarczyk M, et al. The distribution of human endogenous retrovirus K-113 in health and autoimmune diseases in Poland. Rheumatology . 2011;50:1310–4. doi:10.1093/rheumatology/ker022.

35. Ariza M-E, Williams MV. A human endogenous retrovirus K dUTPase triggers a TH1, TH17 cytokine response: does it have a role in psoriasis? J Invest Dermatol. 2011;131:2419–27. doi:10.1038/jid.2011.217.

36. Douville R, Liu J, Rothstein J, Nath A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. Ann Neurol. 2011;69:141–51. doi:10.1002/ana.22149.

37. Ou SH, Wu F, Harrich D, García-Martínez LF, Gaynor RB. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. J Virol. 1995;69:3584–96. https://www.ncbi.nlm.nih.gov/pubmed/7745706.

38. Li W, Lee M-H, Henderson L, Tyagi R, Bachani M, Steiner J, et al. Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med. 2015;7:307ra153. doi:10.1126/scitranslmed.aac8201.

39. Antony JM, van Marle G, Opii W, Allan Butterfield D, Mallet F, Yong VW, et al. Human endogenous retrovirus glycoprotein–mediated induction of redox reactants causes oligodendrocyte death and demyelination. Nature Neuroscience. 2004;7:1088–95. doi:10.1038/nn1319.

40. Dolei A, Serra C, Mameli G, Pugliatti M, Sechi G, Cirotto MC, et al. Multiple sclerosis-associated retrovirus (MSRV) in Sardinian MS patients. Neurology. 2002;58:471–3. https://www.ncbi.nlm.nih.gov/pubmed/11839854.

41. Christensen T, Petersen T, Thiel S, Brudek T, Ellermann-Eriksen S, Møller-Larsen A. Gene–environment interactions in multiple sclerosis: Innate and adaptive immune responses to human endogenous retrovirus and herpesvirus antigens and the lectin complement activation pathway. Journal of Neuroimmunology. 2007;183:175–88. doi:10.1016/j.jneuroim.2006.09.014.

42. Brudek T, Christensen T, Hansen HJ, Bobecka J, Møller-Larsen A. Simultaneous presence of endogenous retrovirus and herpes virus antigens has profound effect on cell-mediated immune responses: implications for multiple sclerosis. AIDS Res Hum Retroviruses. 2004;20:415–23. doi:10.1089/088922204323048168.

43. Tai AK, O'Reilly EJ, Alroy KA, Simon KC, Munger KL, Huber BT, et al. Human endogenous retrovirus-K18 Env as a risk factor in multiple sclerosis. Mult Scler. 2008;14:1175–80. doi:10.1177/1352458508094641.

44. Karlsson H, Bachmann S, Schröder J, McArthur J, Torrey EF, Yolken RH. Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci U S A. 2001;98:4634–9.

doi:10.1073/pnas.061021998.

45. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7. doi:10.1038/nature13595.

46. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016;530:177–83. doi:10.1038/nature16549.

47. Karlsson H, Schröder J, Bachmann S, Bottmer C, Yolken RH. HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. Molecular Psychiatry. 2004;9:12–3. doi:10.1038/sj.mp.4001439.

48. Perron H, Hamdani N, Faucard R, Lajnef M, Jamain S, Daban-Huard C, et al. Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. Transl Psychiatry. 2012;2:e201. doi:10.1038/tp.2012.125.

49. Perron H, Mekaoui L, Bernard C, Veas F, Stefas I, Leboyer M. Endogenous retrovirus type W GAG and envelope protein antigenemia in serum of schizophrenic patients. Biol Psychiatry. 2008;64:1019–23. doi:10.1016/j.biopsych.2008.06.028.

50. Suntsova M, Gogvadze EV, Salozhin S, Gaifullin N, Eroshkin F, Dmitriev SE, et al. Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. Proc Natl Acad Sci U S A. 2013;110:19472–7. doi:10.1073/pnas.1318172110.

51. Kempf L, Nicodemus KK, Kolachana B, Vakkalanka R, Verchinski BA, Egan MF, et al. Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. PLoS Genet. 2008;4:e1000252. doi:10.1371/journal.pgen.1000252.

52. Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. J Neurosci. 2012;32:1884–97. doi:10.1523/JNEUROSCI.3136-11.2012.

53. Löwer R, Löwer J, Kurth R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci U S A. 1996;93:5177–84. https://www.ncbi.nlm.nih.gov/pubmed/8643549.

54. Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, et al. A novel gene from the human endogenous retrovirus K expressed in transformed cells. Clin Cancer Res. 2002;8:1800–7. https://www.ncbi.nlm.nih.gov/pubmed/12060620.

55. Magin C, Löwer R, Löwer J. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. J Virol. 1999;73:9496–507. https://www.ncbi.nlm.nih.gov/pubmed/10516058.

56. Heyne K, Kölsch K, Bruand M, Kremmer E, Grässer FA, Mayer J, et al. Np9, a

cellular protein of retroviral ancestry restricted to human, chimpanzee and gorilla, binds and regulates ubiquitin ligase MDM2. Cell Cycle. 2015;14:2619–33. doi:10.1080/15384101.2015.1064565.

57. Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014;516:405–9. doi:10.1038/nature13804.

58. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320:1344–9. doi:10.1126/science.1158441.

59. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402. doi:10.1146/annurev.genom.9.081307.164359.

60. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78. doi:10.1038/nprot.2012.016.

61. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods. 2008;5:613–9. doi:10.1038/nmeth.1223.

62. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 2010;26:493–500. doi:10.1093/bioinformatics/btp692.

63. Love MI, Hogenesch JB, Irizarry RA. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat Biotechnol. 2016;34:1287–91. doi:10.1038/nbt.3682.

64. Capobianco E. RNA-Seq Data: A Complexity Journey. Comput Struct Biotechnol J. 2014;11:123–30. doi:10.1016/j.csbj.2014.09.004.

65. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013;20:1131–9. doi:10.1038/nsmb.2660.

66. Mayer J, Ehlhardt S, Seifert M, Sauter M, Müller-Lantzsch N, Mehraein Y, et al. Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity. Virology. 2004;322:190–8. doi:10.1016/j.virol.2004.01.023.

67. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9. doi:10.1093/bioinformatics/bts199.

68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. doi:10.1038/nmeth.1923.

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.

70. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. doi:10.1101/gr.107524.110.

71. Team RC, Others. R: A language and environment for statistical computing. 2013. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.5851&rep=rep1&type=pdf.

72. Hayward WS, Neel BG, Astrin SM. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature. 1981;290:475–80. https://www.ncbi.nlm.nih.gov/pubmed/6261142.

73. Jenkins NA, Copeland NG, Taylor BA, Lee BK. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature. 1981;293:370–4. https://www.ncbi.nlm.nih.gov/pubmed/6268990.

74. Scheel C, Weinberg RA. Cancer stem cells and epithelial–mesenchymal transition: Concepts and molecular links. Semin Cancer Biol. 2012;22:396–403. doi:10.1016/j.semcancer.2012.04.001.

75. Golan M, Hizi A, Resau JH, Yaal-Hahoshen N, Reichman H, Keydar I, et al. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. Neoplasia. 2008;10:521–33. https://www.ncbi.nlm.nih.gov/pubmed/18516289.

76. Boller K, Schönfeld K, Lischer S, Fischer N, Hoffmann A, Kurth R, et al. Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. J Gen Virol. 2008;89 Pt 2:567–72. doi:10.1099/vir.0.83534-0.

77. Ruprecht K, Ferreira H, Flockerzi A, Wahl S, Sauter M, Mayer J, et al. Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. J Virol. 2008;82:10008–16. doi:10.1128/JVI.01016-08.

78. Boublikova L, Buchler T, Stary J, Abrahamova J, Trka J. Molecular biology of testicular germ cell tumors: unique features awaiting clinical application. Crit Rev Oncol Hematol. 2014;89:366–85. doi:10.1016/j.critrevonc.2013.10.001.

79. Bhardwaj N, Montesion M, Roy F, Coffin JM. Differential expression of HERV-K (HML-2) proviruses in cells and virions of the teratocarcinoma cell line Tera-1. Viruses. 2015;7:939–68. doi:10.3390/v7030939.

80. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-

Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell. 2015;162:961–73. doi:10.1016/j.cell.2015.07.056.

81. Gross H, Barth S, Pfuhl T, Willnecker V, Spurk A, Gurtsevitch V, et al. The NP9 protein encoded by the human endogenous retrovirus HERV-K(HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2). Int J Cancer. 2011;129:1105–15. doi:10.1002/ijc.25760.

82. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. Annu Rev Genomics Hum Genet. 2006;7:149–73. doi:10.1146/annurev.genom.7.080505.115700.

83. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: extinction and survival. Retrovirology. 2015;12:8. doi:10.1186/s12977-015-0136-x.

84. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997;13:335–40. https://www.ncbi.nlm.nih.gov/pubmed/9260521.

85. Fuentes DR, Swigut T, Wysocka J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife. 2018;7. doi:10.7554/eLife.35989.

86. Jordan CT, Guzman ML, Noble M. Cancer stem cells. N Engl J Med. 2006;355:1253–61. doi:10.1056/NEJMra061808.

87. Lin AT, Santander SG, Marchi E, Karamitros T, Katzourakis A, Magiorkinis G. Targeting endogenous retrovirus gene transcription in human cancers. bioRxiv. 2018. doi:10.1101/449686.

88. Ben-David U, Benvenisty N. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer. 2011;11:268–77. doi:10.1038/nrc3034.

89. Team RC, Others. R: A language and environment for statistical computing. 2016. ftp://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf.

90. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100:3983–8. doi:10.1073/pnas.0530291100.

91. May CD, Sphyris N, Evans KW, Werden SJ, Guo W, Mani SA. Epithelial-mesenchymal transition and cancer stem cells: a dangerously dynamic duo in breast cancer progression. Breast Cancer Res. 2011;13:202. doi:10.1186/bcr2789.

92. Al-Hajj M, Clarke MF. Self-renewal and solid tumor stem cells. Oncogene. 2004;23:7274–82. doi:10.1038/sj.onc.1207947.

93. Ishii H, Saitoh M, Sakamoto K, Kondo T, Katoh R. Epithelial Splicing Regulatory Protein 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via

different mechanisms. Journal of Biological. 2014. http://www.jbc.org/content/early/2014/08/20/jbc.M114.589432.short.

94. Ranganathan P, Weaver KL, Capobianco AJ. Notch signalling in solid tumours: a little bit of everything but not all the time. Nat Rev Cancer. 2011;11:338–51. doi:10.1038/nrc3035.

95. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu M-F, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst. 2008;100:672–9. doi:10.1093/jnci/djn123.

96. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, Brennan KR, et al. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res. 2010;70:709–18. doi:10.1158/0008-5472.CAN-09-1681.

97. Pesce M, Schöler HR. Oct-4: control of totipotency and germline determination. Mol Reprod Dev. 2000;55:452–7. doi:3.0.CO;2-S">10.1002/(SICI)1098-2795(200004)55:4<452::AID-MRD14>3.0.CO;2-S.

98. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. 1998;95:379–91. https://www.ncbi.nlm.nih.gov/pubmed/9814708.

99. Shankar SR, Bahirvani AG, Rao VK, Bharathy N, Ow JR, Taneja R. G9a, a multipotent regulator of gene expression. Epigenetics. 2013;8:16–22. doi:10.4161/epi.23331.

100. Brock A, Krause S, Ingber DE. Control of cancer formation by intrinsic genetic noise and microenvironmental cues. Nat Rev Cancer. 2015;15:499–509. doi:10.1038/nrc3959.

101. Radisky DC, LaBarge MA. Epithelial-mesenchymal transition and the stem cell phenotype. Cell stem cell. 2008;2:511–2. doi:10.1016/j.stem.2008.05.007.

102. Nieto MA. Context-specific roles of EMT programmes in cancer cell dissemination. Nat Cell Biol. 2017;19:416–8. doi:10.1038/ncb3520.

103. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86. doi:10.1038/nrg.2016.139.

104. Wang-Johanning F, Li M, Esteva FJ, Hess KR, Yin B, Rycaj K, et al. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. International journal of cancer. 2014;134:587–95. https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.28389.

105. Zhao J, Rycaj K, Geng S, Li M, Plummer JB, Yin B, et al. Expression of Human Endogenous Retrovirus Type K Envelope Protein is a Novel Candidate Prognostic Marker for Human Breast Cancer. Genes Cancer. 2011;2:914–22. doi:10.1177/1947601911431841.

106. Wang-Johanning F, Rycaj K, Plummer JB, Li M, Yin B, Frerich K, et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. J Natl Cancer Inst. 2012;104:189–210. doi:10.1093/jnci/djr540.

107. Jin X, Xu X-E, Jiang Y-Z, Liu Y-R, Sun W, Guo Y-J, et al. The endogenous retrovirus-derived long noncoding RNA TROJAN promotes triple-negative breast cancer progression via ZMYND8 degradation. Sci Adv. 2019;5:eaat9820. doi:10.1126/sciadv.aat9820.

108. Montesion M, Bhardwaj N, Williams ZH, Kuperwasser C, Coffin JM. Mechanisms of HERV-K (HML-2) Transcription during Human Mammary Epithelial Cell Transformation. J Virol. 2018;92. doi:10.1128/JVI.01258-17.

109. Montesion M, Williams ZH, Subramanian RP, Kuperwasser C, Coffin JM. Promoter expression of HERV-K (HML-2) provirus-derived sequences is related to LTR sequence variation and polymorphic transcription factor binding sites. Retrovirology. 2018;15:57. doi:10.1186/s12977-018-0441-2.

110. Siwaszek A, Ukleja M, Dziembowski A. Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems. RNA Biol. 2014;11:1122–36. doi:10.4161/rna.34406.

111. Kapranov P, St Laurent G, Raz T, Ozsolak F, Patrick Reynolds C, Sorensen PHB, et al. Erratum to: The majority of total nuclear-encoded non-ribosomal RNA in a human cell is "dark matter" un-annotated RNA. BMC Biology. 2011;9. doi:10.1186/1741-7007-9-86.

112. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol. 2011;18:1435–40. doi:10.1038/nsmb.2143.

113. Matteucci C, Balestrieri E, Argaw-Denboba A, Sinibaldi-Vallebona P. Human endogenous retroviruses role in cancer cell stemness. Semin Cancer Biol. 2018;53:17–30. doi:10.1016/j.semcancer.2018.10.001.

114. Bannert N, Hofmann H, Block A, Hohn O. HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors. Front Microbiol. 2018;9:178. doi:10.3389/fmicb.2018.00178.

115. Kahlert UD, Joseph JV, Kruyt FAE. EMT- and MET-related processes in nonepithelial tumors: importance for disease progression, prognosis, and therapeutic opportunities. Mol Oncol. 2017;11:860–77. doi:10.1002/1878-0261.12085.

116. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev Cancer. 2018;18:128–34. doi:10.1038/nrc.2017.118.

117. Tran HD, Luitel K, Kim M, Zhang K, Longmore GD, Tran DD. Transient SNAIL1 expression is necessary for metastatic competence in breast cancer. Cancer

Res. 2014;74:6330–40. doi:10.1158/0008-5472.CAN-14-0923.

118. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature. 2015;527:525–30. doi:10.1038/nature16064.

119. Heidmann O, Béguin A, Paternina J, Berthier R, Deloger M, Bawa O, et al. HEMO, an ancestral endogenous retroviral envelope protein shed in the blood of pregnant women and expressed in pluripotent stem cells and tumors. Proc Natl Acad Sci U S A. 2017;114:E6642–51. doi:10.1073/pnas.1702204114.

120. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012;13:283–96. doi:10.1038/nrg3199.

121. Jern P, Coffin JM. Effects of Retroviruses on Host Genome Function. Annual Review of Genetics. 2008;42:709–32. doi:10.1146/annurev.genet.42.110807.091501.

122. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009;41:563–71. doi:10.1038/ng.368.

123. Conley AB, Piriyapongsa J, Jordan IK. Retroviral promoters in the human genome. Bioinformatics. 2008;24:1563–7. doi:10.1093/bioinformatics/btn243.

124. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nature Reviews Genetics. 2013;14:880–93. doi:10.1038/nrg3594.

125. Rous P. A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS. J Exp Med. 1911;13:397–411. doi:10.1084/jem.13.4.397.

126. Gerdes P, Richardson SR, Mager DL, Faulkner GJ. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biol. 2016;17:100. doi:10.1186/s13059-016-0965-5.

127. Ellison CE, Bachtrog D. Dosage compensation via transposable element mediated rewiring of a regulatory network. Science. 2013;342:846–50. doi:10.1126/science.1239552.

128. Chang Y, Cesarman E, Pessin M, Lee F, Culpepper J, Knowles D, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science. 1994;266:1865–9. doi:10.1126/science.7997879.

129. van der Kuyl AC. HIV infection and HERV expression: a review. Retrovirology. 2012;9:6. doi:10.1186/1742-4690-9-6.

130. Dai L, Del Valle L, Miley W, Whitby D, Ochoa AC, Flemington EK, et al. Transactivation of human endogenous retrovirus K (HERV-K) by KSHV promotes Kaposi's sarcoma development. Oncogene. 2018;37:4534–45. doi:10.1038/s41388-018-0282-4.

131. Chen J, Foroozesh M, Qin Z. Transactivation of human endogenous retroviruses by tumor viruses and their functions in virus-associated malignancies. Oncogenesis. 2019;8:6. doi:10.1038/s41389-018-0114-y.

132. Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, et al. KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. EMBO Rep. 2018;19. doi:10.15252/embr.201745000.

133. Kassiotis G, Stoye JP. Making a virtue of necessity: the pleiotropic role of human endogenous retroviruses in cancer. Philos Trans R Soc Lond B Biol Sci. 2017;372. doi:10.1098/rstb.2016.0277.

134. Felsenstein J. Statistical Inference of Phylogenies. Journal of the Royal Statistical Society. Series A (General). 1983;146:246. doi:10.2307/2981654.

135. Duchêne DA, Duchêne S, Holmes EC, Ho SYW. Evaluating the Adequacy of Molecular Clock Models Using Posterior Predictive Simulations. Mol Biol Evol. 2015;32:2986–95. doi:10.1093/molbev/msv154.

136. Zuckerkandl E, Pauling L. Molecular Disease, Evolution, and Genic Heterogeneity. 1962. https://books.google.com/books/about/Molecular_Disease_Evolution_and_Genic_He.html?hl=&id=wufgtAEACAAJ.

137. Zuckerkandl E, Pauling L. Evolutionary Divergence and Convergence in Proteins. Evolving Genes and Proteins. 1965;:97–166. doi:10.1016/b978-1-4832-2734-4.50017-6.

138. Dickerson RE. The structure of cytochromec and the rates of molecular evolution. Journal of Molecular Evolution. 1971;1:26–45. doi:10.1007/bf01659392.

139. Doolittle RF, Feng DF, Tsang S, Cho G, Little E. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science. 1996;271:470–7. https://www.ncbi.nlm.nih.gov/pubmed/8560259.

140. Runnegar B. A molecular-clock date for the origin of the animal phyla. Lethaia. 1982;15:199–205. doi:10.1111/j.1502-3931.1982.tb00645.x.

141. Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, et al. Genome-scale rates of evolutionary change in bacteria. Microb Genom. 2016;2:e000094. doi:10.1099/mgen.0.000094.

142. Luo A, Ho SYW. The molecular clock and evolutionary timescales. Biochem Soc Trans. 2018;46:1183–90. doi:10.1042/BST20180186.

143. Kumar S. Molecular clocks: four decades of evolution. Nat Rev Genet. 2005;6:654–62. doi:10.1038/nrg1659.

144. Kimura M. Evolutionary Rate at the Molecular Level. Nature. 1968;217:624–6. doi:10.1038/217624a0.

145. Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science. 1986;232:980–2. https://www.ncbi.nlm.nih.gov/pubmed/2939560.

146. Kimura M. Molecular evolutionary clock and the neutral theory. J Mol Evol. 1987;26:24–33. https://www.ncbi.nlm.nih.gov/pubmed/3125335.

147. Ho SYW. The changing face of the molecular evolutionary clock. Trends Ecol Evol. 2014;29:496–503. doi:10.1016/j.tree.2014.07.004.

148. Wilson Sayres MA, Venditti C, Pagel M, Makova KD. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. Evolution. 2011;65:2800–15. doi:10.1111/j.1558-5646.2011.01337.x.

149. Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, et al. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. Nat Commun. 2017;8:303. doi:10.1038/s41467-017-00323-y.

150. Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science. 2016;352:1228–31. doi:10.1126/science.aaf3161.

151. Alexander M, Ho SYW, Molak M, Barnett R, Carlborg Ö, Dorshorst B, et al. Mitogenomic analysis of a 50-generation chicken pedigree reveals a rapid rate of mitochondrial evolution and evidence for paternal mtDNA inheritance. Biol Lett. 2015;11. doi:10.1098/rsbl.2015.0561.

152. Ho S, Larson G. Molecular clocks: when timesare a-changin'. Trends in Genetics. 2006;22:79–83. doi:10.1016/j.tig.2005.11.006.

153. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4:e88. doi:10.1371/journal.pbio.0040088.

154. Welch JJ, Bromham L. Molecular dating when rates vary. Trends Ecol Evol. 2005;20:320–7. doi:10.1016/j.tree.2005.02.007.

155. Gaut B, Yang L, Takuno S, Eguiarte LE. The Patterns and Causes of Variation in Plant Nucleotide Substitution Rates. Annual Review of Ecology, Evolution, and Systematics. 2011;42:245–66. doi:10.1146/annurev-ecolsys-102710-145119.

156. Heath TA, Holder MT, Huelsenbeck JP. A dirichlet process prior for estimating lineage-specific substitution rates. Mol Biol Evol. 2012;29:939–55. doi:10.1093/molbev/msr255.

157. Ho SYW, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 2005;22:1561–8. doi:10.1093/molbev/msi145.

158. Ho SYW, Duchêne S, Molak M, Shapiro B. Time-dependent estimates of molecular evolutionary rates: evidence and causes. Mol Ecol. 2015;24:6007–12. https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13450.

159. Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, et al. Time-dependent rates of molecular evolution. Molecular Ecology. 2011;20:3087–101. doi:10.1111/j.1365-294x.2011.05178.x.

160. Molak M, Ho SYW. Prolonged decay of molecular rate estimates for metazoan mitochondrial DNA. PeerJ. 2015;3:e821. doi:10.7717/peerj.821.

161. Subramanian S, Denver DR, Millar CD, Heupink T, Aschrafi A, Emslie SD, et al. High mitogenomic evolutionary rates and time dependency. Trends Genet. 2009;25:482–6. doi:10.1016/j.tig.2009.09.005.

162. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr Biol. 2015;25:1395–400. doi:10.1016/j.cub.2015.04.007.

163. Membrebe JV, Suchard MA, Rambaut A, Baele G, Lemey P. Bayesian inference of evolutionary histories under time-dependent substitution rates. Mol Biol Evol. 2019. doi:10.1093/molbev/msz094.

164. Papadopoulou A, Anastasiou I, Vogler AP. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Mol Biol Evol. 2010;27:1659–72. doi:10.1093/molbev/msq051.

165. Ho SYW, Lo N. The insect molecular clock. Australian Journal of Entomology. 2013;52:101–5. doi:10.1111/aen.12018.

166. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, et al. Age of cichlids: new dates for ancient lake fish radiations. Mol Biol Evol. 2007;24:1269–82. doi:10.1093/molbev/msm050.

167. Burridge CP, Craw D, Fletcher D, Waters JM. Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol Biol Evol. 2008;25:624–33. doi:10.1093/molbev/msm271.

168. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013;45:1176–82. doi:10.1038/ng.2744.

169. Aiewsakun P, Katzourakis A. Endogenous viruses: Connecting recent and ancient viral evolution. Virology. 2015;479-480:26–37. doi:10.1016/j.virol.2015.02.011.

170. Aiewsakun P, Katzourakis A. Time dependency of foamy virus evolutionary rate estimates. BMC Evol Biol. 2015;15:119. doi:10.1186/s12862-015-0408-z.

171. Li Y, Carroll DS, Gardner SN, Walsh MC, Vitalis EA, Damon IK. On the origin of smallpox: correlating variola phylogenics with historical smallpox records. Proc

Natl Acad Sci U S A. 2007;104:15787–92. doi:10.1073/pnas.0609268104.

172. Gibbs AJ, Fargette D, García-Arenal F, Gibbs MJ. Time--the emerging dimension of plant virus studies. J Gen Virol. 2010;91 Pt 1:13–22. doi:10.1099/vir.0.015925-0.

173. Duchêne S, Holmes EC, Ho SYW. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci. 2014;281. doi:10.1098/rspb.2014.0732.

174. Duchêne S, Ho SYW, Holmes EC. Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. BMC Evol Biol. 2015;15:36. doi:10.1186/s12862-015-0312-6.

175. Nowak MD, Smith AB, Simpson C, Zwickl DJ. A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses. PLoS One. 2013;8:e66245. doi:10.1371/journal.pone.0066245.

176. Lee MSY, Ho SYW. Molecular clocks. Current Biology. 2016;26:R399–402. doi:10.1016/j.cub.2016.03.071.

177. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. Trends in Ecology & Evolution. 2003;18:481–8. doi:10.1016/s0169-5347(03)00216-7.

178. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. Trends Ecol Evol. 2015;30:306–13. doi:10.1016/j.tree.2015.03.009.

179. Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J, et al. Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Nature. 2017;544:180–4. doi:10.1038/nature21416.

180. Fellows Yates JA, Drucker DG, Reiter E, Heumos S, Welker F, Münzel SC, et al. Central European Woolly Mammoth Population Dynamics: Insights from Late Pleistocene Mitochondrial Genomes. Sci Rep. 2017;7:17714. doi:10.1038/s41598-017-17723-1.

181. O'Hanlon SJ, Rieux A, Farrer RA, Rosa GM, Waldman B, Bataille A, et al. Recent Asian origin of chytrid fungi causing global amphibian declines. Science. 2018;360:621–7. doi:10.1126/science.aar1965.

182. Rieux A, Balloux F. Inferences from tip-calibrated phylogenies: a review and a practical guide. Molecular Ecology. 2016;25:1911–24. doi:10.1111/mec.13586.

183. Arcila D, Alexander Pyron R, Tyler JC, Ortí G, Betancur-R R. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). Mol Phylogenet Evol. 2015;82 Pt A:131–45. doi:10.1016/j.ympev.2014.10.011.

184. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian inference of

phylogeny and its impact on evolutionary biology. Science. 2001;294:2310–4. doi:10.1126/science.1065889.

185. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nat Rev Genet. 2012;13:303–14. doi:10.1038/nrg3186.

186. Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst Biol. 2011;60:150–60. doi:10.1093/sysbio/syq085.

187. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol. 2013;30:239–43. doi:10.1093/molbev/mss243.

188. Dutheil JY, Galtier N, Romiguier J, Douzery EJP, Ranwez V, Boussau B. Efficient Selection of Branch-Specific Models of Sequence Evolution. Molecular Biology and Evolution. 2012;29:1861–74. doi:10.1093/molbev/mss059.

189. Sullivan J, Joyce P. Model Selection in Phylogenetics. Annual Review of Ecology, Evolution, and Systematics. 2005;36:445–66. doi:10.1146/annurev.ecolsys.36.102003.152633.

190. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 2002;161:1307–20. https://www.ncbi.nlm.nih.gov/pubmed/12136032.

191. Duchêne S, Lanfear R, Ho SYW. The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol Phylogenet Evol. 2014;78:277–89. doi:10.1016/j.ympev.2014.05.032.

192. Paradis E. Molecular dating of phylogenies by likelihood methods: A comparison of models and a new information criterion. Molecular Phylogenetics and Evolution. 2013;67:436–44. doi:10.1016/j.ympev.2013.02.008.

193. Sanderson MJ. A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy. Molecular Biology and Evolution. 1997;14:1218–31. doi:10.1093/oxfordjournals.molbev.a025731.

194. Ho SYW, Phillips MJ. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol. 2009;58:367–80. doi:10.1093/sysbio/syp035.

195. Leonardi M, Librado P, Der Sarkissian C, Schubert M, Alfarhan AH, Alquraishi SA, et al. Evolutionary Patterns and Processes: Lessons from Ancient DNA. Syst Biol. 2017;66:e1–29. doi:10.1093/sysbio/syw059.

196. Darwin C. The variation of animals and plants under domestication / By Charles Darwin. 1887. doi:10.5962/bhl.title.2090.

197. Dobney K, Larson G. Genetics and animal domestication: new windows on an

elusive process. Journal of Zoology. 2006;0:060222013030001 – ??? doi:10.1111/j.1469-7998.2006.00042.x.

198. Driscoll CA, Macdonald DW, O'Brien SJ. From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci U S A. 2009;106 Suppl 1:9971–8. doi:10.1073/pnas.0901586106.

199. Jensen P. Behavior Genetics and the Domestication of Animals. Annual Review of Animal Biosciences. 2014;2:85–104. doi:10.1146/annurev-animal-022513-114135.

200. MacHugh DE, Larson G, Orlando L. Taming the Past: Ancient DNA and the Study of Animal Domestication. Annual Review of Animal Biosciences. 2017;5:329–51. doi:10.1146/annurev-animal-022516-022747.

201. Wilkins AS, Wrangham RW, Tecumseh Fitch W. The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. Genetics. 2014;197:795–808. doi:10.1534/genetics.114.165423.

202. Hemmer H. Domestication: The Decline of Environmental Appreciation. Cambridge University Press; 1990. https://books.google.com/books/about/Domestication.html?hl=&id=HTZRRwjwP3AC.

203. F.Bath GF, Bath GFF. A Natural History of Domesticated Mammals. African Zoology. 2001;36:113–20. doi:10.1080/15627020.2001.11657122.

204. Larson G, Burger J. A population genetics view of animal domestication. Trends Genet. 2013;29:197–205. doi:10.1016/j.tig.2013.01.003.

205. Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, et al. Current perspectives and the future of domestication studies. Proc Natl Acad Sci U S A. 2014;111:6139–46. doi:10.1073/pnas.1323964111.

206. Zeder MA. Documenting Domestication: New Genetic and Archaeological Paradigms. Univ of California Press; 2006. https://books.google.com/books/about/Documenting_Domestication.html?hl=&id=EaVTxjrbIFQC.

207. MacNeish RS. The Origins of Agriculture and Settled Life. 1992. https://books.google.com/books/about/The_Origins_of_Agriculture_and_Settled_L.html?hl=&id=Q7N7QgAACAAJ.

208. Sauer CO. Agricultural Origins and Dispersals: The Domestication of Animals and Foodstuffs. 1969. https://books.google.com/books/about/Agricultural_Origins_and_Dispersals.html?hl=&id=0DG0kDq493UC.

209. Fuller DQ. An Emerging Paradigm Shift in the Origins of Agriculture. General Anthropology. 2010;17:1–12. doi:10.1111/j.1939-3466.2010.00010.x.

210. Willcox G. The distribution, natural habitats and availability of wild cereals in

relation to their domestication in the Near East: multiple events, multiple centres. Vegetation History and Archaeobotany. 2005;14:534–41. doi:10.1007/s00334-005-0075-x.

211. Larson G. Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. Science. 2005;307:1618–21. doi:10.1126/science.1106927.

212. Ottoni C, Flink LG, Evin A, Geörg C, De Cupere B, Van Neer W, et al. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. Mol Biol Evol. 2013;30:824–32. doi:10.1093/molbev/mss261.

213. Cucchi T, Hulme-Beaman A, Yuan J, Dobney K. Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. Journal of Archaeological Science. 2011;38:11–22. doi:10.1016/j.jas.2010.07.024.

214. Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO. African pastoralism: genetic imprints of origins and migrations. Science. 2002;296:336–9. doi:10.1126/science.1069878.

215. Pitt D, Sevane N, Nicolazzi EL, MacHugh DE, Park SDE, Colli L, et al. Domestication of cattle: Two or three events? Evolutionary Applications. 2019;12:123–36. doi:10.1111/eva.12674.

216. Luikart G, Gielly L, Excoffier L, Vigne JD, Bouvet J, Taberlet P. Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc Natl Acad Sci U S A. 2001;98:5927–32. doi:10.1073/pnas.091591198.

217. Daly KG, Maisano Delser P, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science. 2018;361:85–8. doi:10.1126/science.aas9411.

218. Pedrosa S, Uzun M, Arranz J-J, Gutiérrez-Gil B, San Primitivo F, Bayón Y. Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. Proc Biol Sci. 2005;272:2211–7. doi:10.1098/rspb.2005.3204.

219. Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, et al. Revealing the history of sheep domestication using retrovirus integrations. Science. 2009;324:532–6. doi:10.1126/science.1170587.

220. Vilà C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, Liden K, et al. Widespread origins of domestic horse lineages. Science. 2001;291:474–7. doi:10.1126/science.291.5503.474.

221. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell. 2019. doi:10.1016/j.cell.2019.03.049.

222. Larson G, Fuller DQ. The Evolution of Animal Domestication. Annual Review of Ecology, Evolution, and Systematics. 2014;45:115–36. doi:10.1146/annurev-

ecolsys-110512-135813.

223. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci U S A. 2012;109:8878–83. doi:10.1073/pnas.1203005109.

224. Willcox G, Stordeur D. Large-scale cereal processing before domestication during the tenth millennium cal BC in northern Syria. Antiquity. 2012;86:99–114. doi:10.1017/s0003598x00062487.

225. Vigne J-D, -D. Vigne J, Briois F, Zazzo A, Willcox G, Cucchi T, et al. First wave of cultivators spread to Cyprus at least 10,600 y ago. Proceedings of the National Academy of Sciences. 2012;109:8445–9. doi:10.1073/pnas.1201693109.

226. Ervynck A, Dobney K, Hongo H, Meadow R. Born Free ? New Evidence for the Status of Sus scrofa at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). Paléorient. 2001;27:47–73. doi:10.3406/paleo.2001.4731.

227. Vigne J-D, Carrère I, Briois F, Guilaine J. The Early Process of Mammal Domestication in the Near East. Current Anthropology. 2011;52:S255–71. doi:10.1086/659306.

228. Zhang H, Paijmans JLA, Chang F, Wu X, Chen G, Lei C, et al. Morphological and genetic evidence for early Holocene cattle management in northeastern China. Nat Commun. 2013;4:2755. doi:10.1038/ncomms3755.

229. Fuller DQ. Agricultural Origins and Frontiers in South Asia: A Working Synthesis. Journal of World Prehistory. 2007;20:127–127. doi:10.1007/s10963-006-9007-7.

230. Bjørnstad G, Flagstad Ø, Hufthammer AK, Røed KH. Ancient DNA reveals a major genetic change during the transition from hunting economy to reindeer husbandry in northern Scandinavia. Journal of Archaeological Science. 2012;39:102–8. doi:10.1016/j.jas.2011.09.006.

231. Goñalons GLM. Camelids in ancient Andean societies: A review of the zooarchaeological evidence. Quaternary International. 2008;185:59–68. doi:10.1016/j.quaint.2007.05.022.

232. Thornton EK, Emery KF, Steadman DW, Speller C, Matheny R, Yang D. Earliest Mexican Turkeys (Meleagris gallopavo) in the Maya Region: implications for pre-Hispanic animal trade and the timing of turkey domestication. PLoS One. 2012;7:e42630. doi:10.1371/journal.pone.0042630.

233. Zeder MA. The Domestication of Animals. Journal of Anthropological Research. 2012;68:161–90. doi:10.3998/jar.0521004.0068.201.

234. Marshall F, Weissbrod L. Domestication Processes and Morphological Change. Current Anthropology. 2011;52:S397–413. doi:10.1086/658389.

235. Anthony DW. The Horse, the Wheel, and Language: How Bronze-Age Riders

from the Eurasian Steppes Shaped the Modern World. Princeton University Press; 2010. https://market.android.com/details?id=book-0FDqf415wqgC.

236. Peters J, den Driesch A von. The two-humped camel (Camelus bactrianus): new light on its distribution, management and medical treatment in the past. Journal of Zoology. 1997;242:651–79. doi:10.1111/j.1469-7998.1997.tb05819.x.

237. Boivin N, Fuller DQ. Shell Middens, Ships and Seeds: Exploring Coastal Subsistence, Maritime Trade and the Dispersal of Domesticates in and Around the Ancient Arabian Peninsula. Journal of World Prehistory. 2009;22:113–80. doi:10.1007/s10963-009-9018-2.

238. Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, Barnett R, Campos PF, Cappellini E, et al. Ancient genomics. Philos Trans R Soc Lond B Biol Sci. 2015;370:20130387. doi:10.1098/rstb.2013.0387.

239. Parks M, Subramanian S, Baroni C, Salvatore MC, Zhang G, Millar CD, et al. Ancient population genomics and the study of evolution. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014;370:20130381–20130381. doi:10.1098/rstb.2013.0381.

240. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A. 2013;110:15758–63. doi:10.1073/pnas.1314445110.

241. Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. Nature. 2015;522:167–72. doi:10.1038/nature14507.

242. Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS, et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Mol Ecol Resour. 2016;16:459–69. doi:10.1111/1755-0998.12470.

243. Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am J Hum Genet. 2013;93:852–64. doi:10.1016/j.ajhg.2013.10.002.

244. Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard J-M, Poinar HN. Ancient whole genome enrichment using baits built from modern DNA. Mol Biol Evol. 2014;31:1292–4. doi:10.1093/molbev/msu074.

245. Enk J, Rouillard J-M, Poinar H. Quantitative PCR as a predictor of aligned ancient DNA read counts following targeted enrichment. Biotechniques. 2013;55:300–9. doi:10.2144/000114114.

246. Castellano S, Parra G, Sánchez-Quinto FA, Racimo F, Kuhlwilm M, Kircher M, et al. Patterns of coding variation in the complete exomes of three Neandertals. Proc

Natl Acad Sci U S A. 2014;111:6666–71. doi:10.1073/pnas.1405138111.

247. Gilbert MTP, Wilson AS, Bunce M, Hansen AJ, Willerslev E, Shapiro B, et al. Ancient mitochondrial DNA from hair. Current Biology. 2004;14:R463–4. doi:10.1016/j.cub.2004.06.008.

248. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications. 2014;5. doi:10.1038/ncomms6257.

249. Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 2014;506:225–9. doi:10.1038/nature13025.

250. Adler CJ, Haak W, Donlon D, Cooper A. Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science. 2011;38:956–64. doi:10.1016/j.jas.2010.11.010.

251. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. Sci Rep. 2015;5:11184. doi:10.1038/srep11184.

252. Slon V, Hopfe C, Weiß CL, Mafessoni F, de la Rasilla M, Lalueza-Fox C, et al. Neandertal and Denisovan DNA from Pleistocene sediments. Science. 2017;356:605–8. doi:10.1126/science.aam9695.

253. Seersholm FV, Pedersen MW, Søe MJ, Shokry H, Mak SST, Ruter A, et al. DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. Nat Commun. 2016;7:13389. doi:10.1038/ncomms13389.

254. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012;338:222–6. doi:10.1126/science.1224344.

255. Gansauge M-T, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols. 2013;8:737–48. doi:10.1038/nprot.2013.038.

256. Gansauge M-T, Meyer M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Research. 2014;24:1543–9. doi:10.1101/gr.174201.114.

257. Green RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, et al. The Neandertal genome and ancient DNA authenticity. EMBO J. 2009;28:2494–502. doi:10.1038/emboj.2009.222.

258. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, Willerslev E, et al. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 2012;13:178. doi:10.1186/1471-2164-13-178.

259. Kerpedjiev P, Frellsen J, Lindgreen S, Krogh A. Adaptable probabilistic

mapping of short reads using position specific scoring matrices. BMC Bioinformatics. 2014;15:100. doi:10.1186/1471-2105-15-100.

260. Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat Protoc. 2014;9:1056–82. doi:10.1038/nprot.2014.063.

261. Rubin C-J, -J. Rubin C, -J. Megens H, Barrio AM, Maqbool K, Sayyab S, et al. Strong signatures of selection in the domestic pig genome. Proceedings of the National Academy of Sciences. 2012;109:19529–36. doi:10.1073/pnas.1217149109.

262. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature. 2003;425:832–6. doi:10.1038/nature02064.

263. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464:587–91. doi:10.1038/nature08832.

264. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, et al. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proc Natl Acad Sci U S A. 2008;105:17312–7. doi:10.1073/pnas.0806569105.

265. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, et al. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc Natl Acad Sci U S A. 2007;104:15276–81. doi:10.1073/pnas.0703411104.

266. Girdland Flink L, Allen R, Barnett R, Malmström H, Peters J, Eriksson J, et al. Establishing the validity of domestication genes using DNA from ancient chickens. Proc Natl Acad Sci U S A. 2014;111:6184–9. doi:10.1073/pnas.1308939110.

267. Loog L, Thomas MG, Barnett R, Allen R, Sykes N, Paxinos PD, et al. Inferring Allele Frequency Trajectories from Ancient DNA Indicates That Selection on a Chicken Gene Coincided with Changes in Medieval Husbandry Practices. Mol Biol Evol. 2017;34:1981–90. doi:10.1093/molbev/msx142.

268. Marciniak S, Perry GH. Harnessing ancient genomes to study the history of human adaptation. Nature Reviews Genetics. 2017;18:659–74. doi:10.1038/nrg.2017.65.

269. Allaby RG, Ware RL, Kistler L. A re□evaluation of the domestication bottleneck from archaeogenomic evidence. Evolutionary Applications. 2019;12:29–37. doi:10.1111/eva.12680.

270. Ramakrishnan U, Hadly EA. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. Mol Ecol. 2009;18:1310–30. doi:10.1111/j.1365-294X.2009.04092.x.

271. Mourier T, Ho SYW, Gilbert MTP, Willerslev E, Orlando L. Statistical

guidelines for detecting past population shifts using ancient DNA. Mol Biol Evol. 2012;29:2241–51. doi:10.1093/molbev/mss094.

272. Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol. 2004;13:729–44. https://www.ncbi.nlm.nih.gov/pubmed/15012752.

273. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics. 2000;154:1785–91. https://www.ncbi.nlm.nih.gov/pubmed/10747069.

274. Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. Genetic evidence for an East Asian origin of domestic dogs. Science. 2002;298:1610–3. doi:10.1126/science.1073906.

275. Jansen T, Forster P, Levine MA, Oelke H, Hurles M, Renfrew C, et al. Mitochondrial DNA and the origins of the domestic horse. Proceedings of the National Academy of Sciences. 2002;99:10905–10. doi:10.1073/pnas.152330099.

276. Hiendleder S, Kaupe B, Wassmuth R, Janke A. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proc Biol Sci. 2002;269:893–904. doi:10.1098/rspb.2002.1975.

277. Liu Y-P, Wu G-S, Yao Y-G, Miao Y-W, Luikart G, Baig M, et al. Multiple maternal origins of chickens: out of the Asian jungles. Mol Phylogenet Evol. 2006;38:12–9. doi:10.1016/j.ympev.2005.09.014.

278. Beja-Pereira A, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, et al. African origins of the domestic donkey. Science. 2004;304:1781. doi:10.1126/science.1096008.

279. Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, et al. The Near Eastern origin of cat domestication. Science. 2007;317:519–23. doi:10.1126/science.1139518.

280. Kadwell M, Fernandez M, Stanley HF, Baldi R, Wheeler JC, Rosadio R, et al. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proc Biol Sci. 2001;268:2575–84. doi:10.1098/rspb.2001.1774.

281. Guo S, Savolainen P, Su J, Zhang Q, Qi D, Zhou J, et al. Origin of mitochondrial DNA diversity of domestic yaks. BMC Evol Biol. 2006;6:73. doi:10.1186/1471-2148-6-73.

282. Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. Serial SimCoal: a population genetics model for data from multiple populations and points in time. Bioinformatics. 2005;21:1733–4. doi:10.1093/bioinformatics/bti154.

283. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, et al. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science. 2013;342:871–4. doi:10.1126/science.1243650.

284. Ní Leathlobhair M, Perri AR, Irving-Pease EK, Witt KE, Linderholm A, Haile J, et al. The evolutionary history of dogs in the Americas. Science. 2018;361:81–5. doi:10.1126/science.aao4776.

285. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet. 2015;47:1141–8. doi:10.1038/ng.3394.

286. Heintzman PD, Zazula GD, MacPhee RDE, Scott E, Cahill JA, McHorse BK, et al. A new genus of horse from Pleistocene North America. Elife. 2017;6. doi:10.7554/eLife.29944.

287. de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science. 2018;360. doi:10.1126/science.aar7711.

288. Gerbault P, Allaby RG, Boivin N, Rudzinski A, Grimaldi IM, Pires JC, et al. Storytelling and story testing in domestication. Proc Natl Acad Sci U S A. 2014;111:6159–64. doi:10.1073/pnas.1400425111.

289. Flad RK, Yuan J袁靖, Li S李水城. Zooarcheological evidence for animal domestication in northwest China. Developments in Quaternary Sciences. 2007;:167–203. doi:10.1016/s1571-0866(07)09012-4.

290. Jing Y, Flad RK. Pig domestication in ancient China. Antiquity. 2002;76:724–32. doi:10.1017/s0003598x00091171.

291. Frantz L, Meijaard E, Gongora J, Haile J, Groenen MAM, Larson G. The Evolution of Suidae. Annu Rev Anim Biosci. 2016;4:61–85. doi:10.1146/annurev-animal-021815-111155.

292. Fang M, Andersson L. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proc Biol Sci. 2006;273:1803–10. doi:10.1098/rspb.2006.3514.

293. Kim K-I, Lee J-H, Li K, Zhang Y-P, Lee S-S, Gongora J, et al. Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. Anim Genet. 2002;33:19–25. doi:10.1046/j.1365-2052.2002.00784.x.

294. Fang M, Hu X, Jiang T, Braunschweig M, Hu L, Du Z, et al. The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. Anim Genet. 2005;36:7–13. doi:10.1111/j.1365-2052.2004.01234.x.

295. Megens H-J, Crooijmans RPMA, San Cristobal M, Hui X, Li N, Groenen MAM. Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. Genet Sel Evol. 2008;40:103–28. doi:10.1051/gse:2007039.

296. Ai H, Huang L, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. PLoS One. 2013;8:e56001. doi:10.1371/journal.pone.0056001.

297. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14:R107. doi:10.1186/gb-2013-14-9-r107.

298. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491:393–8. doi:10.1038/nature11622.

299. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet. 2015;47:217–25. doi:10.1038/ng.3199.

300. Ramírez O, Burgos-Paz W, Casas E, Ballester M, Bianco E, Olalde I, et al. Genome data from a sixteenth century pig illuminate modern breed relationships. Heredity . 2015;114:175–84. doi:10.1038/hdy.2014.81.

301. Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, et al. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. Proc Natl Acad Sci U S A. 2010;107:7686–91. doi:10.1073/pnas.0912264107.

302. Zeder MA, Hesse B. The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. Science. 2000;287:2254–7. https://www.ncbi.nlm.nih.gov/pubmed/10731145.

303. Zeder MA. The Origins of Agriculture in the Near East. Current Anthropology. 2011;52:S221–35. doi:10.1086/659307.

304. Horwitz LK, Tchernov E, Ducos P, Becker C, Den Driesch AV, Martin L, et al. Animal domestication in the Southern Levant. Paléorient. 1999;25:63–80. doi:10.3406/paleo.1999.4687.

305. Arbuckle BS, Atici L. Initial diversity in sheep and goat management in Neolithic south-western Asia. Levant. 2013;45:219–35. doi:10.1179/0075891413z.00000000026.

306. Fuller DQ, Willcox G, Allaby RG. Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. World Archaeology. 2011;43:628–52. doi:10.1080/00438243.2011.624747.

307. Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, et al. Early Neolithic genomes from the eastern Fertile Crescent. Science. 2016;353:499–503. doi:10.1126/science.aaf7943.

308. Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. Genomic insights into the origin of farming in the ancient Near East. Nature. 2016;536:419–24. doi:10.1038/nature19310.

309. Llorente MG, Connell S, Jones ER, Merrett D, Jeon J, Eriksson A, et al. The genetics of an early Neolithic pastoralist from the Zagros, Iran. doi:10.1101/059568.

310. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, Zaibert V, et al. The earliest horse harnessing and milking. Science. 2009;323:1332–5. doi:10.1126/science.1168594.

311. Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, Perego UA, et al. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc Natl Acad Sci U S A. 2012;109:2449–54. doi:10.1073/pnas.1111637109.

312. Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, Reissmann M, et al. Origin and history of mitochondrial DNA lineages in domestic horses. PLoS One. 2010;5:e15311. doi:10.1371/journal.pone.0015311.

313. Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, Ludwig A, et al. Discovery of lost diversity of paternal horse lineages using ancient DNA. Nat Commun. 2011;2:450. doi:10.1038/ncomms1447.

314. Der Sarkissian C, Sarkissian CD, Ermini L, Schubert M, Yang MA, Librado P, et al. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Current Biology. 2015;25:2577–83. doi:10.1016/j.cub.2015.08.032.

315. Boyd L, Houpt KA. Przewalski's Horse: The History and Biology of an Endangered Species. SUNY Press; 1994. https://books.google.com/books/about/Przewalski_s_Horse.html?hl=&id=Bvg99Qq54nUC.

316. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, Schubert M, et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science. 2018;360:111–4. doi:10.1126/science.aao3297.

317. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci U S A. 2014;111:E5661–9. doi:10.1073/pnas.1416991111.

318. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al. Ancient genomic changes associated with domestication of the horse. Science. 2017;356:442–5. doi:10.1126/science.aam5298.

319. Shannon LM, Boyko RH, Castelhano M, Corey E, Hayward JJ, McLean C, et al. Genetic structure in village dogs reveals a Central Asian domestication origin. Proc Natl Acad Sci U S A. 2015;112:13639–44. doi:10.1073/pnas.1516215112.

320. Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, Liu Y-H, et al. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. 2016;26:21–33. doi:10.1038/cr.2015.147.

321. Skoglund P, Ersmark E, Palkopoulou E, Dalén L. Ancient wolf genome reveals

an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr Biol. 2015;25:1515–9. doi:10.1016/j.cub.2015.04.019.

322. Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, Stiller M, et al. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. Journal of Archaeological Science. 2009;36:473–90. doi:10.1016/j.jas.2008.09.033.

323. Perri A. A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. Journal of Archaeological Science. 2016;68:1–4. doi:10.1016/j.jas.2016.02.003.

324. Sablin MV, Khlopachev GA. The Earliest Ice Age Dogs: Evidence from Eliseevichi 1. Current Anthropology. 2002;43:795–9. doi:10.1086/344372.

325. Drake AG, Coquerelle M, Colombeau G. 3D morphometric analysis of fossil canid skulls contradicts the suggested domestication of dogs during the late Paleolithic. Sci Rep. 2015;5:8299. doi:10.1038/srep08299.

326. Goebel T, Waters MR, O'Rourke DH. The late Pleistocene dispersal of modern humans in the Americas. Science. 2008;319:1497–502. doi:10.1126/science.1153569.

327. Jakobsson M, Pearce C, Cronin TM, Backman J, Anderson LG, Barrientos N, et al. Post-glacial flooding of the Bering Land Bridge dated to 11 cal ka BP based on new geophysical and sediment records. Climate of the Past. 2017;13:991–1005. doi:10.5194/cp-13-991-2017.

328. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature. 2018;553:203–7. doi:10.1038/nature25173.

329. Brown SK, Darwent CM, Wictum EJ, Sacks BN. Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. Heredity . 2015;115:488–95. doi:10.1038/hdy.2015.49.

330. Derr M. A Dog's History of America: How Our Best Friend Explored, Conquered, and Settled a Continent. Macmillan; 2004. https://books.google.com/books/about/A_Dog_s_History_of_America.html?hl=&id=9Xa6q-mT5xwC.

331. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013;499:74–8. doi:10.1038/nature12323.

332. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505:43–9. doi:10.1038/nature12886.

333. Slon V, Mafessoni F, Vernot B, de Filippo C, Grote S, Viola B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature. 2018;561:113–6. doi:10.1038/s41586-018-0455-x.

334. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, et al. Rise and fall of the Beringian steppe bison. Science. 2004;306:1561–5. doi:10.1126/science.1101074.

335. Heintzman PD, Froese D, Ives JW, Soares AER, Zazula GD, Letts B, et al. Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. Proc Natl Acad Sci U S A. 2016;113:8057–63. doi:10.1073/pnas.1601077113.

336. Gilbert MTP, Drautz DI, Lesk AM, Ho SYW, Qi J, Ratan A, et al. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc Natl Acad Sci U S A. 2008;105:8327–32. doi:10.1073/pnas.0802315105.

337. Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, et al. Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. Curr Biol. 2008;18:1320–6. doi:10.1016/j.cub.2008.07.061.

338. Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, et al. The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. Elife. 2013;2:e00731. doi:10.7554/eLife.00731.

339. Taubenberger JK, Reid AH, Janczewski TA, Fanning TG. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. Philos Trans R Soc Lond B Biol Sci. 2001;356:1829–39. doi:10.1098/rstb.2001.1020.

340. Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife. 2018;7. doi:10.7554/eLife.36666.

341. Mühlemann B, Jones TC, Damgaard P de B, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018;557:418–23. doi:10.1038/s41586-018-0097-z.

342. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011;478:506–10. doi:10.1038/nature10549.

343. Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. Elife. 2016;5:e12994. doi:10.7554/eLife.12994.

344. Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán de Heredia J, et al. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host Microbe. 2016;19:874–81. doi:10.1016/j.chom.2016.05.012.

345. Harbeck M, Seifert L, Hänsch S, Wagner DM, Birdsell D, Parise KL, et al. Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals

Insights into Justinianic Plague. PLoS Pathogens. 2013;9:e1003349. doi:10.1371/journal.ppat.1003349.

346. Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nature Ecology & Evolution. 2018;2:520–8. doi:10.1038/s41559-017-0446-6.

347. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514:494–7. doi:10.1038/nature13591.

348. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science. 2013;341:179–83. doi:10.1126/science.1238286.

349. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC Genomics. 2014;15:270. doi:10.1186/1471-2164-15-270.

350. Ho SYW, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. Molecular Ecology. 2014;23:5947–65. doi:10.1111/mec.12953.

351. Duchêne S, Duchêne D, Holmes EC, Ho SYW. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. Mol Biol Evol. 2015;32:1895–906. doi:10.1093/molbev/msv056.

352. Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, et al. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol Evol. 2016;7:80–9. doi:10.1111/2041-210X.12466.

353. Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, Castaños P, et al. Coat color variation at the beginning of horse domestication. Science. 2009;324:485. doi:10.1126/science.1172750.

354. Wilson Sayres MA, Venditti C, Pagel M, Makova KD. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. Evolution. 2011;65:2800–15. doi:10.1111/j.1558-5646.2011.01337.x.

355. Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet. 2003;72:659–70. doi:10.1086/368264.

356. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22:160–74. doi:10.1007/bf02101694.

357. Brown WM, George M, Wilson AC. Rapid evolution of animal mitochondrial

DNA. Proceedings of the National Academy of Sciences. 1979;76:1967–71. doi:10.1073/pnas.76.4.1967.

358. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 2009;84:740–59. doi:10.1016/j.ajhg.2009.05.001.

359. Williamson S, Orive ME. The genealogy of a sequence subject to purifying selection at multiple sites. Mol Biol Evol. 2002;19:1376–84. doi:10.1093/oxfordjournals.molbev.a004199.

360. Tong KJ, Duchêne DA, Duchêne S, Geoghegan JL, Ho SYW. A comparison of methods for estimating substitution rates from ancient DNA sequence data. BMC Evol Biol. 2018;18:70. doi:10.1186/s12862-018-1192-3.

361. Emerson BC. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. Syst Biol. 2007;56:337–45. doi:10.1080/10635150701258795.

362. Bandelt HJ. Time dependency of molecular rate estimates: tempest in a teacup. Heredity . 2008;100:1–2. https://www.nature.com/articles/6801054.

363. Emerson BC, Alvarado-Serrano DF, Hickerson MJ. Model misspecification confounds the estimation of rates and exaggerates their time dependency. Molecular ecology. 2015;24:6013–20. doi:10.1111/mec.13451.

364. Emerson BC, Hickerson MJ. Lack of support for the time-dependent molecular evolution hypothesis. Mol Ecol. 2015;24:702–9. https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13070.

365. Navascués M, Emerson BC. Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. Mol Ecol. 2009;18:4390–7. doi:10.1111/j.1365-294X.2009.04333.x.

366. Debruyne R, Poinar HN. Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? Syst Biol. 2009;58:348–60. doi:10.1093/sysbio/syp028.

367. Wertheim JO, Fourment M, Kosakovsky Pond SL. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. Mol Biol Evol. 2012;29:451–6. doi:10.1093/molbev/msr266.

368. Duchêne D, Duchêne S, Ho SYW. Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. Mol Ecol Resour. 2015;15:785–94. doi:10.1111/1755-0998.12352.

369. Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, et al. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol Evol. 2016;7:80–9. doi:10.1111/2041-210x.12466.

370. Aiewsakun P, Katzourakis A. Time-Dependent Rate Phenomenon in Viruses. J Virol. 2016;90:7184–95. doi:10.1128/JVI.00593-16.

371. Pitulko VV, Kasparov AK. Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. Journal of Archaeological Science: Reports. 2017;13 Supplement C:491–515. doi:10.1016/j.jasrep.2017.04.003.

372. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. doi:10.1186/1471-2148-7-214.

373. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15:e1006650. doi:10.1371/journal.pcbi.1006650.

374. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. doi:10.1093/ve/vex042.

375. Mourad R, Chevennet F, Dunn DT, Fearnhill E, Delpech V, Asboe D, et al. A phylotype-based analysis highlights the role of drug-naive HIV-positive individuals in the transmission of antiretroviral resistance in the UK. AIDS. 2015;29:1917–25. doi:10.1097/QAD.0000000000000768.

376. To T-H, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. Syst Biol. 2016;65:82–97. doi:10.1093/sysbio/syv068.

377. Volz EM, Frost SDW. Scalable relaxed clock phylogenetic dating. Virus Evol. 2017;3. doi:10.1093/ve/vex025.

378. Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K. Estimating divergence times in large phylogenetic trees. Syst Biol. 2007;56:741–52. doi:10.1080/10635150701613783.

379. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A. 2012;109:19333–8. doi:10.1073/pnas.1213199109.

380. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. Trends Ecol Evol. 2003;18:481–8. doi:10.1016/S0169-5347(03)00216-7.

381. Drummon A, OLIVERGPYBUS, Rambaut A. Inference of Viral Evolutionary Rates from Molecular Sequences. Advances in Parasitology Volume 54. 2003;:331–58. doi:10.1016/s0065-308x(03)54008-8.

382. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol. 2010;27:2038–51. doi:10.1093/molbev/msq088.

383. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:vew007. doi:10.1093/ve/vew007.

384. Frantz LAF, Haile J, Lin AT, Scheu A, Geörg C, Benecke N, et al. Ancient pigs

reveal a near-complete genomic turnover following their introduction to Europe. Proc Natl Acad Sci U S A. 2019. doi:10.1073/pnas.1901169116.

385. Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, Maisano Delser P, et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 2019;365:173–6. doi:10.1126/science.aav1002.

386. Froese D, Stiller M, Heintzman PD, Reyes AV, Zazula GD, Soares AER, et al. Fossil and genomic evidence constrains the timing of bison arrival in North America. Proc Natl Acad Sci U S A. 2017;114:3457–62. doi:10.1073/pnas.1620754114.

387. Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. Syst Biol. 2007;56:515–22. doi:10.1080/10635150701435401.

388. Hedrick PW. Conservation genetics and North American bison (Bison bison). J Hered. 2009;100:411–20. doi:10.1093/jhered/esp024.

389. Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, et al. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. Proc Natl Acad Sci U S A. 2014;111:4826–31. doi:10.1073/pnas.1320412111.

390. David Mech L. Use of Erroneous Wolf Generation Time in Assessments of Domestic Dog and Human Evolution. 2017. https://digitalcommons.unl.edu/usgsnpwrc/303/. Accessed 14 Aug 2019.

391. Rochus CM, Johansson AM. Estimation of genetic diversity in Gute sheep: pedigree and microsatellite analyses of an ancient Swedish breed. Hereditas. 2017;154:4. doi:10.1186/s41065-017-0026-4.

392. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30:3276–8. doi:10.1093/bioinformatics/btu531.

393. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52. doi:10.1093/oxfordjournals.molbev.a026334.

394. Ramsden C, Holmes EC, Charleston MA. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol. 2009;26:143–53. doi:10.1093/molbev/msn234.

395. Duffy S, Holmes EC. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. J Gen Virol. 2009;90 Pt 6:1539–47. doi:10.1099/vir.0.009266-0.

396. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3. doi:10.1093/bioinformatics/btu033.

397. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and

evolutionary analyses in R. Bioinformatics. 2018;1:3. doi:10.1093/bioinformatics/bty633.

398. Neher RA, Bedford T. nextflu: real-time tracking of seasonal influenza virus evolution in humans. Bioinformatics. 2015;31:3546–8. doi:10.1093/bioinformatics/btv381.

399. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10:e1003537. doi:10.1371/journal.pcbi.1003537.

400. Molak M, Suchard MA, Ho SYW, Beilman DW, Shapiro B. Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into Bayesian phylogenetic analyses of ancient DNA. Mol Ecol Resour. 2015;15:81–6. doi:10.1111/1755-0998.12295.

401. Rambaut A, Drummond AJ, Xie D, Baele G. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;10. https://academic.oup.com/sysbio/advance-article-pdf/doi/10.1093/sysbio/syy032/25008458/syy032.pdf.

402. Elson JL, Turnbull DM, Howell N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet. 2004;74:229–38. doi:10.1086/381505.

403. Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, et al. The Role of Selection in the Evolution of Human Mitochondrial Genomes. Genetics. 2006;172:373–87. doi:10.1534/genetics.105.043901.

404. Dos Reis M, Yang Z. Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons? Genetics. 2013;195:195–204. doi:10.1534/genetics.113.152025.

405. Gignoux CR, Henn BM, Mountain JL. Rapid, global demographic expansions after the origins of agriculture. Proc Natl Acad Sci U S A. 2011;108:6044–9. doi:10.1073/pnas.0914274108.

406. Ho SYW, Kolokotronis S-O, Allaby RG. Elevated substitution rates estimated from ancient DNA sequences. Biol Lett. 2007;3:702–5. doi:10.1098/rsbl.2007.0377.

407. Rambaut A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics. 2000;16:395–9. doi:10.1093/bioinformatics/16.4.395.

408. Molak M, Lorenzen ED, Shapiro B, Ho SYW. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. Mol Biol Evol. 2013;30:253–62. doi:10.1093/molbev/mss232.

409. Ho SYW, Larson G. Molecular clocks: when times are a-changin'. Trends Genet. 2006;22:79–83. doi:10.1016/j.tig.2005.11.006.

410. Ho SYW, Heupink TH, Rambaut A, Shapiro B. Bayesian estimation of sequence damage in ancient DNA. Mol Biol Evol. 2007;24:1416–22. doi:10.1093/molbev/msm062.

411. Rambaut A, Ho SYW, Drummond AJ, Shapiro B. Accommodating the effect of ancient DNA damage on inferences of demographic histories. Mol Biol Evol. 2009;26:245–8. doi:10.1093/molbev/msn256.

412. Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, et al. True single-molecule DNA sequencing of a pleistocene horse bone. Genome Res. 2011;21:1705–19. doi:10.1101/gr.122747.111.

413. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011;27:2153–5. doi:10.1093/bioinformatics/btr347.

414. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences. 2014;111:2229–34. doi:10.1073/pnas.1318934111.

415. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A. 2007;104:14616–21. doi:10.1073/pnas.0704665104.

416. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013;29:1682–4. doi:10.1093/bioinformatics/btt193.

417. Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. Nature Reviews Genetics. 2015;16:395–408. doi:10.1038/nrg3935.

418. Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, et al. A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. Current Biology. 2010;20:231–6. doi:10.1016/j.cub.2009.11.068.

419. Dabney J, Meyer M, Paabo S. Ancient DNA Damage. Cold Spring Harbor Perspectives in Biology. 2013;5:a012567–a012567. doi:10.1101/cshperspect.a012567.

420. Prufer K, Meyer M. Comment on “Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans.” Science. 2015;347:835–835. doi:10.1126/science.1260617.

421. Haldane JBS. Suggestions as to quantitative measurement of rates of evolution. Evolution. 1949;3:51–6. doi:10.1111/j.1558-5646.1949.tb00004.x.

422. Purugganan MD, Fuller DQ. Archaeological data reveal slow rates of evolution during plant domestication. Evolution. 2011;65:171–83. doi:10.1111/j.1558-5646.2010.01093.x.

423. Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, et al.

Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. Proc Natl Acad Sci U S A. 2014;111:6147–52. doi:10.1073/pnas.1308937110.

424. Doust AN, Lukens L, Olsen KM, Mauro-Herrera M, Meyer A, Rogers K. Beyond the single gene: How epistasis and gene-by-environment effects influence crop domestication. Proc Natl Acad Sci U S A. 2014;111:6178–83. doi:10.1073/pnas.1308940110.

425. Geiger M, Sánchez-Villagra MR. Similar rates of morphological evolution in domesticated and wild pigs and dogs. Front Zool. 2018;15:23. doi:10.1186/s12983-018-0265-x.

426. Miao Y-W, Peng M-S, Wu G-S, Ouyang Y-N, Yang Z-Y, Yu N, et al. Chicken domestication: an updated perspective based on mitochondrial genomes. Heredity . 2013;110:277–82. doi:10.1038/hdy.2012.83.

427. Austin RM, Sholts SB, Williams L, Kistler L, Hofman CA. Opinion: To curate the molecular past, museums need a carefully considered set of best practices. Proceedings of the National Academy of Sciences. 2019;116:1471–4. doi:10.1073/pnas.1822038116.

428. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci. 2012;279:4724–33. doi:10.1098/rspb.2012.1745.

429. Burger J, Hummel S, Herrmann B, Henke W. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. Electrophoresis. 1999;20:1722–8. doi:3.3.co;2-w">10.1002/(sici)1522-2683(19990101)20:8<1722::aid-elps1722>3.3.co;2-w.

430. Koch PL, Barnosky AD. Late Quaternary Extinctions: State of the Debate. Annual Review of Ecology, Evolution, and Systematics. 2006;37:215–50. doi:10.1146/annurev.ecolsys.34.011802.132415.

431. Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, et al. Species-specific responses of Late Quaternary megafauna to climate and humans. Nature. 2011;479:359–64. doi:10.1038/nature10574.

432. Nieves-Colón MA, Ozga AT, Pestle WJ, Cucina A, Tiesler V, Stanton TW, et al. Comparison of two ancient DNA extraction protocols for skeletal remains from tropical environments. Am J Phys Anthropol. 2018;166:824–36. doi:10.1002/ajpa.23472.

433. Rohland N, Glocke I, Aximu-Petri A, Meyer M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat Protoc. 2018;13:2447–61. doi:10.1038/s41596-018-0050-5.

434. Andam CP, Worby CJ, Chang Q, Campana MG. Microbial Genomics of Ancient Plagues and Outbreaks. Trends in Microbiology. 2016;24:978–90.

doi:10.1016/j.tim.2016.08.004.

435. Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. Scientific Reports. 2015;4. doi:10.1038/srep04003.

436. Smith O, Clapham AJ, Rose P, Liu Y, Wang J, Allaby RG. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. Sci Rep. 2014;4:5559. doi:10.1038/srep05559.

437. Fordyce SL, Ávila-Arcos MC, Rasmussen M, Cappellini E, Romero-Navarro JA, Wales N, et al. Deep sequencing of RNA from ancient maize kernels. PLoS One. 2013;8:e50961. doi:10.1371/journal.pone.0050961.

438. Smith O, Palmer SA, Clapham AJ, Rose P, Liu Y, Wang J, et al. Small RNA Activity in Archeological Barley Shows Novel Germination Inhibition in Response to Environment. Mol Biol Evol. 2017;34:2555–62. doi:10.1093/molbev/msx175.

439. Keller A, Kreis S, Leidinger P, Maixner F, Ludwig N, Backes C, et al. miRNAs in ancient tissue specimens of the Tyrolean Iceman. Molecular Biology and Evolution. 2016;:msw291. doi:10.1093/molbev/msw291.

440. Smith O, Dunshea G, Sinding M-HS, Fedorov S, Germonpre M, Bocherens H, et al. Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. PLoS Biol. 2019;17:e3000166. doi:10.1371/journal.pbio.3000166.

441. Fordyce SL, Kampmann M-L, van Doorn NL, Gilbert MTP. Long-term RNA persistence in postmortem contexts. Investig Genet. 2013;4:7. doi:10.1186/2041-2223-4-7.

442. Halfon P, Khiri H, Gerolami V, Bourliere M, Feryn JM, Reynier P, et al. Impact of various handling and storage conditions on quantitative detection of hepatitis C virus RNA. J Hepatol. 1996;25:307–11. doi:10.1016/s0168-8278(96)80116-4.

443. Forster JL, Harkin VB, Graham DA, McCullough SJ. The effect of sample type, temperature and RNAlater™ on the stability of avian influenza virus RNA. J Virol Methods. 2008;149:190–4. doi:10.1016/j.jviromet.2007.12.020.

444. Prendergast ME, Lipson M, Sawchuk EA, Olalde I, Ogola CA, Rohland N, et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. Science. 2019;365. doi:10.1126/science.aaw6275.

445. Brace S, Diekmann Y, Booth TJ, van Dorp L, Faltyskova Z, Rohland N, et al. Ancient genomes indicate population replacement in Early Neolithic Britain. Nat Ecol Evol. 2019;3:765–71. doi:10.1038/s41559-019-0871-9.

446. Delsuc F, Kuch M, Gibb GC, Karpinski E, Hackenberger D, Szpak P, et al. Ancient Mitogenomes Reveal the Evolutionary History and Biogeography of Sloths. Curr Biol. 2019;29:2031–42.e6. doi:10.1016/j.cub.2019.05.043.

447. Brüniche-Olsen A, Jones ME, Burridge CP, Murchison EP, Holland BR, Austin JJ. Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. Journal of Biogeography. 2018;45:963–76. doi:10.1111/jbi.13214.

448. Woods R, Turvey ST, Brace S, MacPhee RDE, Barnes I. Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative radiation. Proceedings of the National Academy of Sciences. 2018;115:12769–74. doi:10.1073/pnas.1808603115.

449. Louchart A, Bastian F, Baptista M, Guarino-Vignon P, Hume JP, Jacot-des-Combes C, et al. Ancient DNA reveals the origins, colonization histories, and evolutionary pathways of two recently extinct species of giant scops owl from Mauritius and Rodrigues Islands (Mascarene Islands, south-western Indian Ocean). Journal of Biogeography. 2018;45:2678–89. doi:10.1111/jbi.13450.

450. Oswald JA, Allen JM, Witt KE, Folk RA, Albury NA, Steadman DW, et al. Ancient DNA from a 2,500-year-old Caribbean fossil places an extinct bird (Caracara creightoni) in a phylogenetic context. Molecular Phylogenetics and Evolution. 2019;140:106576. doi:10.1016/j.ympev.2019.106576.

451. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, et al. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. Science. 2014;343:1243727. doi:10.1126/science.1243727.

452. Blanquart F, Grabowski MK, Herbeck J, Nalugoda F, Serwadda D, Eller MA, et al. A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. Elife. 2016;5. doi:10.7554/eLife.20492.

453. Larsen BB, Cole KL, Worobey M. Ancient DNA provides evidence of 27,000-year-old papillomavirus infection and long-term codivergence with rodents. Virus Evol. 2018;4:vey014. doi:10.1093/ve/vey014.

454. Mühlemann B, Margaryan A, Damgaard P de B, Allentoft ME, Vinner L, Hansen AJ, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc Natl Acad Sci U S A. 2018;115:7557–62. doi:10.1073/pnas.1804921115.

455. Simmonds P, Aiewsakun P, Katzourakis A. Prisoners of war — host adaptation and its constraints on virus evolution. Nature Reviews Microbiology. 2019;17:321–8. doi:10.1038/s41579-018-0120-2.

456. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. Initial genetic characterization of the 1918 “Spanish” influenza virus. Science. 1997;275:1793–6. doi:10.1126/science.275.5307.1793.

457. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455:661–4. doi:10.1038/nature07390.

458. Gryseels S, Watts TD, Kabongo J-MM, Larsen BB, Lemey P, Muyembe-

Tamfum J-J, et al. A near-full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. bioRxiv. 2019;:687863. https://www.biorxiv.org/content/10.1101/687863v1.abstract.

459. Donczo B, Guttman A. Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: The Holy Grail for molecular diagnostics. J Pharm Biomed Anal. 2018;155:125–34. doi:10.1016/j.jpba.2018.03.065.

460. POTTER, CW. Chronicle of influenza pandemics. Textbook of Influenza. 1998;:3–17. https://ci.nii.ac.jp/naid/10030269896/. Accessed 5 Oct 2019.

461. Pyle GF. The Diffusion of Influenza: Patterns and Paradigms. Rowman & Littlefield; 1986. https://play.google.com/store/books/details?id=7w-H11vMykgC.

462. Beveridge WI. The chronicle of influenza epidemics. Hist Philos Life Sci. 1991;13:223–34. https://www.ncbi.nlm.nih.gov/pubmed/1724803.

463. Potter CW. A history of influenza. J Appl Microbiol. 2001;91:572–9. doi:10.1046/j.1365-2672.2001.01492.x.

464. Horman WSJ, Nguyen THO, Kedzierska K, Bean AGD, Layton DS. The Drivers of Pathology in Zoonotic Avian Influenza: The Interplay Between Host and Pathogen. Front Immunol. 2018;9:1812. doi:10.3389/fimmu.2018.01812.

465. Yoon S-W, Webby RJ, Webster RG. Evolution and Ecology of Influenza A Viruses. Influenza Pathogenesis and Control - Volume I. 2014;:359–75. doi:10.1007/82_2014_396.

466. Webster RG, Govorkova EA. Continuing challenges in influenza. Ann N Y Acad Sci. 2014;1323:115–39. doi:10.1111/nyas.12462.

467. Reperant LA, Kuiken T, Osterhaus ADME. Influenza viruses: from birds to humans. Hum Vaccin Immunother. 2012;8:7–16. doi:10.4161/hv.8.1.18672.

468. Ma W, Kahn RE, Richt JA. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. J Mol Genet Med. 2008;3:158–66. https://www.ncbi.nlm.nih.gov/pubmed/19565018.

469. Khiabanian H, Trifonov V, Rabadan R. Reassortment patterns in Swine influenza viruses. PLoS Curr. 2009;1:RRN1008. doi:10.1371/currents.RRN1008.

470. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459:1122–5. doi:10.1038/nature08182.

471. Crisci E, Mussá T, Fraile L, Montoya M. Review: Influenza virus in pigs. Molecular Immunology. 2013;55:200–11. doi:10.1016/j.molimm.2013.02.008.

472. Ma W, Lager KM, Vincent AL, Janke BH, Gramer MR, Richt JA. The role of swine in the generation of novel influenza viruses. Zoonoses Public Health. 2009;56:326–37. doi:10.1111/j.1863-2378.2008.01217.x.

473. Nelson MI, Wentworth DE, Culhane MR, Vincent AL, Viboud C, LaPointe MP, et al. Introductions and evolution of human-origin seasonal influenza a viruses in multinational swine populations. J Virol. 2014;88:10110–9. doi:10.1128/JVI.01080-14.

474. Parrish CR, Murcia PR, Holmes EC. Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. J Virol. 2015;89:2990–4. doi:10.1128/JVI.03146-14.

475. PERRONCITO, E. Epizoozia tifoide nei gallinacei. Annali Accad Agri Torino. 1878;:87–126. https://ci.nii.ac.jp/naid/10029625637/. Accessed 5 Oct 2019.

476. Lupiani B, Reddy SM. The history of avian influenza. Comp Immunol Microbiol Infect Dis. 2009;32:311–23. doi:10.1016/j.cimid.2008.01.004.

477. Fanning TG, Slemons RD, Reid AH, Janczewski TA, Dean J, Taubenberger JK. 1917 avian influenza virus sequences suggest that the 1918 pandemic virus did not acquire its hemagglutinin directly from birds. J Virol. 2002;76:7860–2. doi:10.1128/jvi.76.15.7860-7862.2002.

478. DeWitte SN. Archaeological Evidence of Epidemics Can Inform Future Epidemics. Annu Rev Anthropol. 2016;45:63–77. doi:10.1146/annurev-anthro-102215-095929.

479. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011;334:89–94. doi:10.1126/science.1209202.

480. Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. Am J Hum Genet. 2016;98:22–33. doi:10.1016/j.ajhg.2015.11.015.

481. Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova J-L, et al. Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am J Hum Genet. 2016;98:5–21. doi:10.1016/j.ajhg.2015.11.014.

482. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell. 2016;167:643–56.e17. doi:10.1016/j.cell.2016.09.024.

483. Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 2012;91:265–74. doi:10.1016/j.ajhg.2012.06.015.

484. Enard D, Petrov DA. Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. Cell. 2018;175:360–71.e13. doi:10.1016/j.cell.2018.08.034.

485. O'Brien TR, Prokunina-Olsson L, Donnelly RP. IFN-λ4: the paradoxical new member of the interferon lambda family. J Interferon Cytokine Res. 2014;34:829–38.

doi:10.1089/jir.2013.0136.

486. Key FM, Peter B, Dennis MY, Huerta-Sánchez E, Tang W, Prokunina-Olsson L, et al. Selection on a variant associated with improved viral clearance drives local, adaptive pseudogenization of interferon lambda 4 (IFNL4). PLoS Genet. 2014;10:e1004681. doi:10.1371/journal.pgen.1004681.

487. Terczyńska-Dyla E, Bibert S, Duong FHT, Krol I, Jørgensen S, Collinet E, et al. Reduced IFNλ4 activity is associated with improved HCV clearance and reduced expression of interferon-stimulated genes. Nat Commun. 2014;5:5699. doi:10.1038/ncomms6699.

488. Pybus OG, Gray RR. Virology: The virus whose family expanded. Nature. 2013;498:310–1. doi:10.1038/498310b.

489. Pybus OG, Thézé J. Hepacivirus cross-species transmission and the origins of the hepatitis C virus. Curr Opin Virol. 2016;16:1–7. doi:10.1016/j.coviro.2015.10.002.

490. Kenney AD, Dowdle JA, Bozzacco L, McMichael TM, St Gelais C, Panfil AR, et al. Human Genetic Determinants of Viral Diseases. Annu Rev Genet. 2017;51:241–63. doi:10.1146/annurev-genet-120116-023425.

491. Sonderup MW, Afihene M, Ally R, Apica B, Awuku Y, Cunha L, et al. Hepatitis C in sub-Saharan Africa: the current status and recommendations for achieving elimination by 2030. Lancet Gastroenterol Hepatol. 2017;2:910–9. doi:10.1016/S2468-1253(17)30249-2.

492. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312:282–4. https://www.ncbi.nlm.nih.gov/pubmed/6504142.

493. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc Natl Acad Sci U S A. 2014;111:18655–60. doi:10.1073/pnas.1412627111.

494. Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J-L, et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature. 2014;505:403–6. doi:10.1038/nature12788.

495. Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. Nat Rev Genet. 2015;16:395–408. doi:10.1038/nrg3935.

496. Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, et al. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science. 2014;344:523–7. doi:10.1126/science.1250368.

497. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. 2014;24:454–66. doi:10.1101/gr.163592.113.

498. Seguin-Orlando A, Gamba C, Sarkissian CD, Ermini L, Louvel G, Boulygina E, et al. Pros and cons of methylation-based enrichment methods for ancient DNA. Sci Rep. 2015;5:11826. doi:10.1038/srep11826.

499. Orlando L, Willerslev E. Evolution. An epigenetic window into the past? Science. 2014;345:511–2. doi:10.1126/science.1256515.

500. Paijmans JLA, Gilbert MTP, Hofreiter M. Mitogenomic analyses from ancient DNA. Mol Phylogenet Evol. 2013;69:404–16. doi:10.1016/j.ympev.2012.06.002.

501. Woodhams M. Can deleterious mutations explain the time dependency of molecular rate estimates? Mol Biol Evol. 2006;23:2271–3. doi:10.1093/molbev/msl107.

502. Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? Heredity. 2016;116:362–71. doi:10.1038/hdy.2015.104.

503. Santos C, Montiel R, Sierra B, Bettencourt C, Fernandez E, Alvarez L, et al. Understanding Differences Between Phylogenetic and Pedigree-Derived mtDNA Mutation Rate: A Model Using Families from the Azores Islands (Portugal). Molecular Biology and Evolution. 2005;22:1490–505. doi:10.1093/molbev/msi141.

504. Carvajal RD, Schwartz GK, Tezel T, Marr B, Francis JH, Nathan PD. Metastatic disease from uveal melanoma: treatment options and future prospects. Br J Ophthalmol. 2017;101:38–44. doi:10.1136/bjophthalmol-2016-309034.

505. Chattopadhyay C, Kim DW, Gombos DS, Oba J, Qin Y, Williams MD, et al. Uveal melanoma: From diagnosis to treatment and the science in between. Cancer. 2016;122:2299–312. doi:10.1002/cncr.29727.

506. Oliva M, Rullan AJ, Piulats JM. Uveal melanoma as a target for immune-therapy. Ann Transl Med. 2016;4:172. doi:10.21037/atm.2016.05.04.

507. Caminal JM, Ribes J, Clèries R, Ibáñez N, Arias L, Piulats JM, et al. Relative survival of patients with uveal melanoma managed in a single center. Melanoma Research. 2012;22:271–7. doi:10.1097/cmr.0b013e328353ef30.

508. Van Raamsdonk CD, Bezrookove V, Green G, Bauer J, Gaugler L, O'Brien JM, et al. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. Nature. 2009;457:599–602. doi:10.1038/nature07586.

509. Van Raamsdonk CD, Griewank KG, Crosby MB, Garrido MC, Vemula S, Wiesner T, et al. Mutations in GNA11 in uveal melanoma. N Engl J Med. 2010;363:2191–9. doi:10.1056/NEJMoa1000584.