Palaeogenomics of Animal Domestication:

Computational modelling of ancestry, demography and selection

Evan Kim Irving-Pease
St Hugh's College
University of Oxford

Thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2019

Declaration

I hereby declare that this thesis has never been submitted, either in the same or a different form, to this or any other university, for a degree or any other qualification. All the work presented in this thesis is the result of my own work unless otherwise stated in the author contributions. The work was carried out while I was registered as a graduate student at the Research Laboratory for Archaeology and the History of Art, University of Oxford, under the supervision of Prof. Greger Larson and Dr. Laurent Frantz.

Evan Kim Irving-Pease

Oxford, October 2019

Palaeogenomics of Animal Domestication: Computational modelling of ancestry, demography and selection

Evan Kim Irving-Pease
St Hugh's College
University of Oxford

Thesis submitted for the degree of *Doctor of Philosophy*
Trinity Term 2019

Abstract

The study of animal domestication is in the midst of a genomic revolution, as technological advances dramatically increase the availability of DNA sequence data from ancient and modern domestic populations. For the first time, palaeogenomic datasets are being produced of sufficient size and scope to directly observe anthropological and evolutionary processes through time; from the emergence of domestic animals in the Neolithic, to the formation of modern breeds in the Victorian era.

This thesis applies computational modelling of ancient and modern DNA to infer ancestry, demography, and selection in domestic cattle (*Bos taurus*), horses (*Equus ferus caballus*), and dogs (*Canis familiaris*).

Chapter 1 reviews the substantial contribution palaeogenomics has made to the study of animal domestication, and details the latest theoretical and empirical advances, alongside individual profiles of sixteen major domestic species.

Chapter 2 reconstructs the allele frequency trajectories of genetic variants linked to quantitative traits in genome-wide associations studies (GWAS) of cattle and horses. Using a Bayesian modelling approach, and a dataset of more than 350 ancient genomes spread across the last 10,000 years, the age of the allele under selection for thousands of GWAS variants is inferred, along with the selection coefficients for hundreds of polygenic traits.

Chapter 3 analyses the first ancient genome-wide DNA from pre-contact North American dogs. Admixture analyses show that the earliest American dogs were not independently domesticated, and originated from a population of arctic dogs in Eastern Siberia. These dogs likely accompanied humans during the peopling of the continent, and rapidly diversified after their arrival, until their sudden replacement following the arrival of European colonists. Remarkably, their most closely related living relatives are now an 8,000-year-old contagious cancer clone known as Canine Transmissible Venereal Tumour (CTVT).

Chapter 4 presents an ancestry analysis of a rare breed of humpless, dwarf shorthorn cattle from the remote island of Socotra, in the Arabian Sea. Admixture analyses show that contrary to published hypotheses, these cattle are predominantly Eurasian taurine (*Bos taurus*) in ancestry, with only minimal African taurine heritage, and preliminary results indicate possible admixture with a basal clade of zebu cattle (*Bos indicus*).

Table of Contents

ABSTRACT IV

TABLE OF CONTENTS V

TABLE OF FIGURES VIII

TABLE OF TABLES XIII

ACKNOWLEDGEMENTS XIV

INTRODUCTION 1

1 PALAEOGENOMICS OF ANIMAL DOMESTICATION 3

1.1 STATEMENT OF AUTHORSHIP 3
1.2 AUTHORS AND AFFILIATIONS 3
1.3 ABSTRACT 4
1.4 INTRODUCTION 4
1.5 SEQUENCING ANCIENT DNA 5
1.6 PATHWAYS TO ANIMAL DOMESTICATION 8
1.6.1 THE COMMENSAL PATHWAY 9
1.6.2 THE PREY PATHWAY 10
1.6.3 THE DIRECTED PATHWAY 10
1.7 WHEN, WHERE AND WHICH PATHWAY 11
1.7.1 COMMENSAL DOMESTICATES 12
1.7.2 PREY DOMESTICATES 17
1.7.3 DIRECTED DOMESTICATES 23
1.8 THE BIOLOGICAL ARCHITECTURE OF DOMESTICATION 29
1.8.1 THEORIES AND EXPERIMENTS 29
1.8.2 GENETIC CHANGES DURING DOMESTICATION 32
1.8.3 TEMPORAL PATTERN OF GENETIC AND MORPHOLOGICAL CHANGES 34
1.8.4 GENES AS DOMESTIC MARKERS 37
1.8.5 INTROGRESSION IN DOMESTICATION 39
1.9 FUTURE PERSPECTIVES 41
1.9.1 ANCIENT EPIGENOMES 41
1.9.2 TECHNICAL ADVANCES 42
1.9.3 NOVEL SUBSTRATES FOR ADNA 43
1.10 CONCLUSION 44
1.11 REFERENCES 46
1.12 ACKNOWLEDGMENTS AND NOTES 65
1.13 PERMISSION FROM CO-AUTHORS 66

2 SELECTION TRAJECTORIES OF GENETIC VARIANTS UNDERLYING DOMESTIC ANIMAL TRAITS 68

2.1 STATEMENT OF AUTHORSHIP 68
2.2 AUTHORS AND AFFILIATIONS 68
2.3 ABSTRACT 69
2.4 INTRODUCTION 69
2.5 METHODS AND MATERIALS 71
2.5.1 COMPUTATIONAL PIPELINE 71
2.5.2 REFERENCE GENOMES 72

2.5.3 Modern Samples 72
2.5.4 Demographic modelling 76
2.5.5 Ancient Samples 78
2.5.6 Combined dataset 81
2.5.7 Genome-Wide Association Studies (GWAS) 81
2.5.8 Ensembl genes and variants 82
2.5.9 SNP array metadata 82
2.5.10 Selection analysis 82
2.6 Results 90
2.6.1 Demographic modelling 90
2.6.2 Selection analysis 93
2.6.3 Joint posterior distributions 104
2.7 Discussion 107
2.7.1 SNP polarisation 107
2.7.2 Sample binning 109
2.7.3 MCMC modelling 109
2.7.4 GWAS associations 112
2.7.5 Modelling assumptions 114
2.8 Conclusions 115
2.9 References 117
2.10 Supplementary Materials 127
2.10.1 Sample accessions 127
2.10.2 MCMC convergence metrics 144
2.10.3 Supplementary References 166
2.11 Acknowledgments and Notes 170
2.12 Permission from Co-authors 171

3 THE EVOLUTIONARY HISTORY OF DOGS IN THE AMERICAS 172

3.1 Statement of Authorship 172
3.2 Authors and Affiliations 173
3.3 Abstract 176
3.4 Main Text 176
3.5 References 185
3.6 Supplementary Materials 198
3.6.1 Nuclear ancestry analyses 198
3.7 Acknowledgments and Notes 238
3.8 Permission from Co-authors 240

4 GENETIC ANALYSIS OF THE DWARF SHORTHORN CATTLE OF SOCOTRA 251

4.1 Statement of Authorship 251
4.2 Authors and Affiliations 251
4.3 Abstract 252
4.4 Introduction 252
4.4.1 Cattle domestication 252
4.4.2 African cattle 254
4.4.3 Arabian cattle 255
4.4.4 Socotra 256
4.5 Materials and methods 257
4.5.1 Sample collection and sequencing 257
4.5.2 Computational Pipeline 258
4.5.3 Alignment and variant calling 258
4.5.4 Neighbour-joining tree 260

4.5.5 Principal component analysis 260
4.5.6 Model-based clustering 261
4.5.7 Admixture graphs 261
4.6 Results 263
4.6.1 Principal component analysis 263
4.6.2 Neighbour-joining tree 265
4.6.3 Model-based clustering 267
4.6.4 Admixture graphs 271
4.7 Discussion 279
4.8 References 282
4.9 Acknowledgments and Notes 288
4.10 Permission from Co-authors 288

5 CONCLUSIONS 289

5.1 CharacterING modern patterns of diversity 289
5.2 Introgression and population turnover 290
5.3 Ancient DNA time-series 291
5.4 Future perspectives 292
5.5 References 296

6 APPENDIX 301

A. Rabbits and the Specious Origins of Domestication 301
B. Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events 305
C. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion 313
D. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe 322

Table of Figures

Figure 1.1 Geographical/chronological time frame of domestication and potential pathways for major domestic animals. The timelines are in Ky (1,000 years) increment. Adapted after (Larson and Fuller 2014)..... 9

Figure 1.2 Multiple traits, commonly referred to as 'domestication syndrome' and their occurrence in different mammalian species. Adapted after (Sánchez-Villagra et al. 2016).....30

Figure 1.3 Pie charts representing the allele frequency of variants at *TSHR* and *BCDO2* (affecting skin colour) in ancient chickens. This figure demonstrates that the variants at those genes, which are thought to influence traits in modern domestic chickens, were not found at high frequency in ancient chickens. Their rise in frequency is thought to be associated with breeding during the middle age, but not with the domestication process (Loog et al. 2017). Adapted after (Girdland Flink et al. 2014). 35

Figure 1.4 Haplotype network for the MC1R gene coding region. This figure demonstrates the existence of three lineages of black pigs, in Hawaii, Europe and East Asia. Adapted after (Linderholm et al. 2016). 38

Figure 2.1 Schematic of the first three sequential epoch ðaði models 77

Figure 2.4 Flow chart summary of the pipeline to perform the selection analyses 83

Figure 2.5 Selection trajectory for 'Milk fat yield' (rs378261830) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.....99

Figure 2.6 Selection trajectory for 'Milk fat percentage' and 'Milk protein percentage' (rs137567750) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size. 100

Figure 2.7 Selection trajectory for 'Age at puberty' and 'Scrotal circumference' (rs134839836) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size. 101

Figure 2.8 Selection trajectory for 'Withers height' (rs68689852) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.....102

Figure 2.9 Selection trajectory for 'Number of progressively motile sperm' (rs68628642) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.....103

Figure 2.10 Selection trajectory for 'Withers height' (rs68619452) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.....104

Figure 2.11 Joint posterior distribution of the selection coefficient $\boldsymbol{s1}$ for Exterior, Health, Meat and Carcass, and Reproduction traits in cattle 105

Figure 2.12 Joint posterior distribution of the selection coefficient $\boldsymbol{s1}$ for Milk and Production traits in cattle 106

Figure 2.13 Joint posterior distribution of the selection coefficient $\boldsymbol{s1}$ for Behaviour, Exterior, Growth, Health, Performance and Reproduction traits in horses.....107

Figure 3.1 Sample location and ancestry of pre-contact dogs. a. A map depicting the location and age of the archaeological remains analysed in this study. Each dot represents a single sample, and multiple

samples per archaeological site are grouped in boxes. Sites mentioned in the text are labelled. b. A tip calibrated Bayesian mitochondrial phylogenetic tree of dogs, within haplogroup A. This analysis was conducted with 71 novel ancient mitogenomes together with 145 publicly available mitogenomes from both modern and ancient canids (Driggers, 2015) (Figure S3.6). Red branches represent modern dogs. Blue horizontal bars on nodes represent 95% High Density Posterior age. The grey shaded area represents the time frame during which people entered the Americas (Graf and Buvit, 2017; Moreno-Mayar et al., 2018; Raghavan et al., 2015) c. A neighbour-joining tree built with whole genomes (Driggers, 2015). d An admixture graph constructed with TreeMix (based on transversions; Supplementary Material) depicting the relationship between PCD (including the Port Du Choix [AL3194] and Weyanoke Old Town [AL3223] samples) and other dog and wolf populations. We only used Greenland dogs and Malamute (American Arctic dogs) for this analysis as these are the least admixed with Western Eurasian dogs (Driggers, 2015).....178

Figure 3.2 Legacy of pre-contact dogs in modern American dogs. a. A map showing the locations of dog populations obtained from (Shannon et al., 2015) and their degree of relatedness (D-statistics) with the ~4ky old Port au Choix dog (AL3194; see (Driggers, 2015) and Figure S3.14). Higher values (in red) represent closer relatedness. b. A map depicting the multiple introductions of dogs into the Americas. 183

Figure 4.1 Principal Components Analysis (PC1 vs. PC2) of 798 cattle samples (from 70 populations) based on 736,071 SNPs ascertained using the BovineHD SNP array. All Socotran and Dhofari samples were projected. 263

Figure 4.2 Principal Components Analysis (PC1 vs. PC2) of 1,361 cattle samples (from 59 populations) based on 52,870 SNPs ascertained using the BovineSNP50 SNP array. All Socotran and Dhofari samples were projected. 264

Figure 4.3 Neighbour Joining (NJ) tree built from Identity By State (IBS) matrix of 247 cattle samples (including 15 populations) based on 736,071 SNPs ascertained using the BovineHD SNP array. 266

Figure 4.4 Cross-validation error for ADMIXTURE analysis, K between 1 and 22. K = 20 has the lowest cross-validation error. 268

Figure 4.5 ADMIXTURE clusters, for K = 3, 12 and 20, from 788 cattle samples (including 67 populations) based on 236,614 SNPs ascertained using the BovineHD SNP array and pruned for LD. Population tiles are shows with a fixed width, regardless of sample size. 269

Figure 4.6 Interpolated map of ADMIXTURE clusters for K = 6, from 798 cattle samples (including 70 populations) based on 236,614 SNPs ascertained using the BovineHD SNP array and pruned for LD. Dots represent sampling locations for each population. SOC and OMN populations shown as square dots.....270

Figure 4.7 Admixture graph with no migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array). 272

Figure 4.8 Admixture graph with one migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array). 273

Figure 4.9 Admixture graph with two migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array). 274

Figure 4.10 Admixture graph with three migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array). 275

Figure 4.11 Admixture graph, inferred using qpGraph, with no f4 outliers (i.e., $|Z| < 3$), based on SNPs from the BovineHD array, showing two independent zebu introgressions into African taurine.....277

Figure 4.12 Admixture graph, inferred using qpGraph, with no f4 outliers (i.e., $|Z| < 3$), based on SNPs from the BovineHD array, showing two consecutive zebu introgressions into African taurine. 278

Table of Supplementary Figures

FIGURE S3.1 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. 208

FIGURE S3.2 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 MTDNA ANALYSES. LACK OF 5' DAMAGE IN SOME LIBRARIES IS DUE TO LIBRARY PREPARATION PROTOCOL (SEE ANCIENT DNA ILLINOIS SECTION). 209

FIGURE S3.3 MAXIMUM LIKELIHOOD TREE BASED ON MTDNA DATA. THE FOUR MAJOR DOG HAPLOGROUPS ARE INDICATED: A (RED; INCLUDES ALL BUT ONE PRE-CONTACT DOGS), B (PURPLE), C (YELLOW), D (GREEN). BLUE TIP LABEL REPRESENTS NEWLY SEQUENCED SAMPLES (THIS STUDY). DARK BLUE HIGHLIGHTED CLADE REPRESENTS AMERICAN DOGS (MONOPHYLETIC, BOOTSTRAP SUPPORT VALUE=87). LIGHT BLUE HIGHLIGHTED CLADES (CGG1-10) REPRESENT ZHOKHOV ISLAND SAMPLES (~9KYA SLED DOGS FROM EASTERN SIBERIA; SEE TABLE S3.1). CGG10-11 (OUTSIDE OF THE ZHOKHOV / PRE-CONTACT CLADE) ARE MORE RECENT SLED DOGS FROM SIBERIA (~1.5KYA; TABLE S3.1). NODE LABELS INDICATE BOOTSTRAP REPLICATES....213

FIGURE S3.4 MAXIMUM LIKELIHOOD TREE BASED ON 605BP OF THE CONTROL REGION. ALL SAMPLES STARTING WITH PREFIX HQ WERE OBTAINED FROM (*7*) WHILE ALL SAMPLES STARTING WITH PREFIX AY WERE OBTAINED FROM (LEONARD ET AL., 2002). THE FOUR MAJOR DOG HAPLOGROUPS ARE INDICATED WITH DIFFERENT BRANCH COLOURS: A (RED; INCLUDES ALL BUT ONE PRE-CONTACT DOGS), B (PURPLE), C (YELLOW), D (GREEN). ALL PRE-COLUMBIAN DOGS FROM THIS STUDY ARE HIGHLIGHTED IN LIGHT BLUE. NODE LABELS INDICATE BOOTSTRAP REPLICATES. 214

FIGURE S3.5 MAXIMUM LIKELIHOOD TREE BASED ON MTDNA DATA INCLUDING DATA FROM (STRAKOVA ET AL., 2016). RED BRANCHES REPRESENT ANCIENT PRE-CONTACT DOGS. NODE LABELS INDICATE BOOTSTRAP REPLICATES. 215

FIGURE S3.6 BAYESIAN TREE (BEAST) OF MTDNA DATA. RED, PURPLE, AND GREEN CIRCLE REPRESENT NODES WITH >0.9, >0.7 AND >0.5 POSTERIOR PROBABILITY RESPECTIVELY. BLUE BAR REPRESENTS CONFIDENCE INTERVAL OF DIVERGENCE TIME (SCALED IN YEAR BEFORE PRESENT). 216

FIGURE S3.7 PRINCIPAL COMPONENTS ANALYSIS (PC1 VERSUS PC2) OF 57 CANID SAMPLES (INCLUDING WOLVES AND COYOTES) BASED ON 2,063,129 SNPS ASCERTAINED USING THE GENOME-WIDE DATA-SET. ALL PRE-CONTACT DOG SAMPLES WERE PROJECTED. 217

FIGURE S3.8 PRINCIPAL COMPONENTS ANALYSIS (PC1 VERSUS PC2) OF 44 DOG SAMPLES (EXCLUDING WOLVES AND COYOTES) BASED ON 2,063,129 SNPS ASCERTAINED USING THE GENOME-WIDE DATA-SET. ALL PRE-CONTACT DOG SAMPLES WERE PROJECTED. 218

FIGURE S3.9 PRINCIPAL COMPONENTS ANALYSIS (PC1 VERSUS PC2) OF 44 DOG SAMPLES (EXCLUDING WOLVES AND COYOTES) BASED ON 2,063,129 SNPS ASCERTAINED USING THE GENOME-WIDE DATA-SET. ALL PRE-CONTACT DOG AND CTVT SAMPLES WERE PROJECTED. 219

FIGURE S3.10 NEIGHBOUR JOINING (NJ) TREE BASED ON IDENTITY BY STATE (IBS). THIS FIGURE IS THE SAME AS IN FIGURE 1C. CONFIRMS THAT CTVT IS MORE CLOSELY RELATED TO PRE-CONTACT DOGS THAN ANY OTHER DOG POPULATION. CONFIRMS THAT PCD FORM A MONOPHYLETIC CLADE. 220

FIGURE S3.11 NEIGHBOUR JOINING (NJ) TREE BASED ON IDENTITY BY STATE (IBS). SAME AS FIGURE S3.3 BUT WITHOUT EAST ASIAN DOGS THAT ARE ADMIXED WITH EUROPEAN DOGS I.E. EXCLUDING ALL EAST ASIAN DOGS EXCEPT VIETNAMESE. PCD, ARCTIC DOGS AND CTVT FOUNDER NOW APPEAR MORE CLOSELY RELATED TO WESTERN DOGS. 221

FIGURE S3.12 BAYESIAN TREE BASED ON ~26K TRANSVERSIONS. CONFIRMS THAT CTVT AND PCD ARE MONOPHYLETIC WITH HIGH SUPPORT AND SUPPORTS THE BASAL PLACEMENT OF THE CTVT/PCD CLADE. SUPPORT VALUES REPRESENT POSTERIOR PROBABILITY. 222

FIGURE S3.13 SHARED GENETIC DRIFT MEASURED BY f_3(OUTGROUP; Y, X) WHERE Y IS EITHER PORT AU CHOIX DOG (AL3194), WEYANOKE OLD TOWN DOG (AL3223), ALASKAN MALAMUTE (D_MAL68), GREENLAND SLEDGE DOG (D_GREEN) AND X REPRESENTS MODERN DOG POPULATIONS. ERROR BARS REPRESENT 1 SE....223

FIGURE S3.14 BOX PLOT REPRESENTING A. D-STATISTICS AND B. SIGNIFICANCE OF D-STATISTICS (Z) FOR EVERY COMBINATION OF D(OUTGROUP, AL3194[PORT AU CHOIX], POP3, POP4), WHERE POP3 IS FIXED AND POP4 REPRESENTS ANY OTHER GENOME. POSITIVE VALUES SUPPORT A CLOSE RELATIONSHIP BETWEEN POP3 AND PCD WHILE NEGATIVE

values imply PCD are closer to other dog populations. If Pop3 is not admixed with PCD, we expect -4<Z<4 (x-axis). 224

Figure S3.15 Box plot representing a. D-statistics and b. significance of D-statistics (Z) for every combination of D(Outgroup, AL3223[Weyanoke Old town], Pop3, Pop4), where Pop3 is fixed and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and PCD while negative values imply PCD are closer to other dog populations. If Pop3 is not admixed with PCD, we expect -4<Z<4 (x-axis). 225

Figure S3.16 Box plot representing a. D-statistics and b. significance of D-statistics (Z) for every combination of D(Outgroup, Coyote, Pop3, Pop4), where Pop3 is fixed (x-axis) and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and coyotes while negative values imply coyotes are closer to other dog populations. If Pop3 is not admixed with coyotes, we expect -4<Z<4 (x-axis). 226

Figure S3.17 Box plot representing a. D-statistics b. significance of D-statistics (Z) for every combination of D(Outgroup, north American wolf, Pop3, Pop4), where Pop3 is fixed and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and north American wolves while negative values imply NA wolves are closer to other dog populations. If Pop3 is not admixed with north American wolves, we expect -4<Z<4 (x-axis). 227

Figure S3.18 Schematic representation of the assumed phylogeny for the f4 ratio test used to estimate pre-contact ancestry into modern North American dogs. Alpha represents the degree of ancestry from pre-contact dogs. 228

Figure S3.19 Cross validation (CV) values for ADMIXTURE analysis of SNP array data. 229

Figure S3.20 ADMIXTURE results based on SNP array data for K=4, 10 and 15 (see Table S3.9 for population codes). 230

Figure S3.21 Admixture graph without migration edge and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup. 231

Figure S3.22 Admixture graph with a single migration edge and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). We see evidence for admixture from coyotes (COY) into the pre-contact dog lineage (PCD/CTVT), consistent with Figure S3.16. *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup. 232

Figure S3.23 Admixture graph with two migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup. 233

Figure S3.24 Admixture graph with three migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup. 234

FIGURE S3.25 QPGRAPH MODEL WITH ADMIXTURE FRACTIONS. *WESTERN EURASIAN DOGS* - PORTUGUESE VILLAGE DOGS (DEU), *EAST ASIAN DOGS* - VIETNAMESE VILLAGE DOGS (DVN), *PRE-CONTACT DOGS* (PCD), INCLUDING BOTH PORT AU CHOIX (AL3194) AND WEYANOKE OLD TOWN (AL3223), *ARCTIC DOGS* - MALAMUTE (DMA), *CTVT* - (79T AND 24T), *EURASIAN WOLVES* (WEU) FROM SPAIN AND PORTUGAL, *NORTH AMERICAN WOLVES* (WAM) FROM YELLOWSTONE, *COYOTES* (COY) AS AN OUTGROUP. 235

FIGURE S3.26 ADMIXTURE GRAPH AND MATRIX OF RESIDUALS, EXPRESSED AS THE NUMBER OF STANDARD ERRORS, INFERRED USING TREEMIX FOR THE SAME POPULATION AS IN FIGURE S3.25. *WESTERN EURASIAN DOGS* - PORTUGUESE VILLAGE DOGS (DEU), *EAST ASIAN DOGS* - VIETNAMESE VILLAGE DOGS (DVN), *PRE-CONTACT DOGS* (PCD), INCLUDING BOTH PORT AU CHOIX (AL3194) AND WEYANOKE OLD TOWN (AL3223), *ARCTIC DOGS* - MALAMUTE (DMA), *CTVT* - (79T AND 24T), *EURASIAN WOLVES* (WEU) FROM SPAIN AND PORTUGAL, *NORTH AMERICAN WOLVES* (WAM) FROM YELLOWSTONE, *COYOTES* (COY) AS AN OUTGROUP. 236

FIGURE S3.27 A. CTVT MUTATION SPECTRUM. 1,925,779 TUMOUR-ONLY MUTATIONS IN CTVT ARE DISPLAYED BY MUTATION TYPE (IN PYRIMIDINE CONTEXT) WITH IMMEDIATE 5' AND 3' CONTEXT. EACH OF THE 96 MUTATION CLASSES IS DISPLAYED ON THE HORIZONTAL AXIS. MUTATION PROPORTIONS ARE DISPLAYED RELATIVE TO CANFAM3.1 B. FRACTION OF CTVT TUMOUR-ONLY MUTATIONS ATTRIBUTABLE TO COSMIC SIGNATURES 1, 5, 7, AND THE DOG GERMLINE SIGNATURE, AS ESTIMATED USING SIGFIT. C. RECONSTRUCTION OF CTVT TUMOUR-ONLY SPECTRUM USING COSMIC SIGNATURES 1,5 AND 7 AND THE DOG GERMLINE SIGNATURE. 237

Table of Tables

Table 2.1 Genome assemblies used for each species 72
Table 2.2 Summary of modern reference panels 72
Table 2.3 Summary of polarised biallelic SNPs 76
Table 2.4 Summary of published and unpublished ancient samples 78
Table 2.5 Summary of combined modern and ancient datasets 81
Table 2.6 Summary of unique GWAS associations, variants, and traits 81
Table 2.7 Ensembl release versions, dates, and assemblies 82
Table 2.8 Summary of commercial SNP arrays 82
Table 2.10 Summary of cattle temporal bins 85
Table 2.11 Summary of horse temporal bins 86
Table 2.12 Mutation rates and generation times for each species 87
Table 2.13 ∂a∂i results for cattle using the folded SFS 91
Table 2.14 Cattle demography with 4 epochs 91
Table 2.15 ∂a∂i results for horses using the folded SFS 92
Table 2.16 Horse demography with 5 epochs 92
Table 2.17 Frequency of putatively mispolarised SNPs for cattle 93
Table 2.18 Frequency of putatively mispolarised SNPs for horses 94
Table 2.19 Summary of ESS metrics 95
Table 2.20 Summary of MPSRF metrics 96
Table 2.21 Ten largest maximum *a posteriori* selected traits for cattle 97
Table 2.22 Ten largest maximum *a posteriori* selected traits for horses 98
Table 4.1 Summary of modern reference panels 258
Table 4.2 Sequencing depth and number of intersecting SNPs per array 259
Table 4.3 Summary of combined modern datasets 260

Table of Supplementary Tables

Table S2.1 Modern cattle samples 127
Table S2.2 Modern horse samples 128
Table S2.4 Ancient cattle samples 129
Table S2.5 Ancient horse samples 133
Table S2.6 MCMC convergence metrics for cattle GWAS SNPs 144
Table S2.7 MCMC convergence metrics for horse GWAS SNPs 158

Acknowledgements

Over the course of the last three years, many people have assisted my doctoral research and provided help in the preparation of this thesis. First and foremost, thanks go to my supervisors, Greger Larson and Laurent Frantz, for guiding this work to completion.

From the first day I approached Greger about doing a PhD, his overwhelming enthusiasm for research and unflagging support has been a driving force in everything that followed. Laurent's deep and detailed knowledge of computational biology has been an invaluable resource over the years, and I am very appreciative of all the time he spent answering questions and sharing ideas. I would also like to thank Joshua G. Schraiber for all his assistance, and for his continued support for my research despite leaving academia.

As a dry lab specialist, I am deeply indebted to the many people working untold hours in the wet lab; drilling bones, performing extractions, building libraries, and sequencing the samples. Thanks especially to Victoria E. Mullin, Kevin G. Daly, James Haile, Anna Linderholm, Antoine Fages, Daniel G. Bradley and Ludovic Orlando for providing the ancient DNA sequence data used in this thesis. None of this work would have been possible without your generosity.

Many thanks are also due to all the wonderful members of the PalaeoBARN lab, past and present. Especially to Alexandra Jamieson, Anna Linderholm, Antony Dimopoulos, Hannah Ryan, James Haile and Ophélie Lebrasseur for making this such a supportive environment, and to Joel M. Alves and Jesse Dabney for all their sage advice and our lively lunchtime debates about science, philosophy, and politics.

I am also grateful to the Clarendon Fund and St Hugh's College for my Clarendon Scholarship, and to the John Fell Fund for their financial support of my research.

Lastly, I am forever indebted to Hayley Mountford for her patience and support, for putting up with my endless science talk at the dinner table, and for being so understanding about the pressures of PhD life. You're the best!

Introduction

The study of animal domestication is a multidisciplinary endeavour, encompassing research approaches and perspectives from archaeology, history, genetics, and ecology. These diverse approaches have endeavoured to address core questions of when, where, and how many times animal domestications took place, and to elucidate the underlying genetic architecture of animal domestication. More recently, the study of animal domestication has been transformed by the revolution in modern and ancient genome sequencing, and the nascent field of palaeogenomics.

Chapter 1 presents a review of the application of palaeogenomics to animal domestication. This encompasses theoretical and empirical advances in our understanding of the domestication continuum, detailed through a discussion of the different pathways to domestication and individual case studies for dogs (*Canis familiaris*), pigs (*Sus scrofa domesticus*), cats (*Felis catus*), chickens (*Gallus gallus domesticus*), goats (*Capra hircus*), sheep (*Ovis aries*), cattle (*Bos taurus* and *B. indicus*), llama (*Lama glama*), alpaca (*Vicugna pacos*), horses (*Equus ferus caballus*), rabbits (*Oryctolagus cuniculus*), camels (*Camelus dromedarius* and *C. bactrianus*), silkworms (*Bombyx mori*) and honey bees (*Apis mellifera*).

The growing availability of large genome-wide datasets for major domestic taxa have facilitated the mapping of complex traits to genetic variants through genome-wide association studies (GWAS). Chapter 2 presents an ancient DNA time-series analysis of selection for quantitative traits in taurine cattle and horses. Using a Bayesian modelling approach, the allele frequency trajectories of thousands of GWAS variants are reconstructed, and the selection coefficients for hundreds of polygenic traits inferred. Detailed technical recommendations are also made for improvements to the computational modelling software necessary to carry this research forward.

Chapter 3 presents an analysis of the first whole-genome sequence data from ancient North American dogs. We show that domestic dogs likely accompanied humans during the peopling of the Americas, originating from a source population of Arctic dogs in

Eastern Siberia. Following the arrival of European colonists, pre-contact dogs were quickly replaced by European dogs, and have left almost no genetic legacy in modern populations. Surprisingly, the most closely related living organism to these pre-contact dogs is not a dog, but is instead an 8,000-year-old contagious cancer clone, known as Canine Transmissible Venereal Tumour (CTVT).

Chapter 4 presents an analysis of modern genome-wide sequence data from a rare breed of humpless, dwarf shorthorn cattle found on the remote island of Socotra, in the Arabian Sea. Analysis shows that these regionally incongruous cattle come from a single geographically dispersed population, found on Socotra and in the Dhofar Mountains, Oman. In comparison to a global reference panel of >1,000 cattle, results show that this population is unlike any other cattle sampled in the region. Admixture analyses reveal that they are predominantly Eurasian taurine in ancestry (*Bos taurus*), and preliminary evidence suggests possible admixture with a basal clade of zebu cattle (*Bos indicus*), previously unseen in modern populations.

The thesis presents a varied set of approaches to computational modelling of ancestry, demography and selection in domestic cattle, horses, and dogs. The field of palaeogenomics is fast moving and dynamic, and the research presented here spans the breadth of technical analyses and species domains within the field.

1 Palaeogenomics of Animal Domestication

This chapter was published on 11 Oct 2018 in: Lindqvist, C., Rajora, O.P. (Eds.), *Paleogenomics*, Population Genomics. Springer, Cham, pp. 1–48.
https://doi.org/10.1007/13836_2018_55

1.1 Statement of Authorship

Manuscript

I wrote the following sections of the manuscript:

1.3 Abstract

1.4 Introduction

1.5 Sequencing Ancient DNA

1.6 Pathways to Animal Domestication

1.7.1.3 Cats

1.7.2.1 Goats

1.7.2.2 Sheep

1.7.2.3 Cattle

1.7.3.1 Horses

1.7.3.2 Rabbits

1.8 The Biological Architecture of Domestication

1.9 Future Perspectives

1.10 Conclusion

I edited the manuscript with L.A.F.F., with input from all other authors.

1.2 Authors and Affiliations

Authors: Evan K. Irving-Pease[1], Hannah Ryan[1], Alexandra Jamieson[1], Evangelos A. Dimopoulos[1], Greger Larson[1], Laurent A. F. Frantz[1,2]

Affiliations:

[1] *The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK.*

[2] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK.*

1.3 Abstract

Starting with dogs, over 15,000 years ago, the domestication of animals has been central in the development of modern societies. Because of its importance for a range of disciplines—including archaeology, biology, and the humanities—domestication has been studied extensively. This chapter reviews how the field of palaeogenomics, has, and will continue to, revolutionise our understanding of animal domestication. We discuss how the recovery of ancient DNA from archaeological remains is allowing researchers to overcome inherent shortcomings arising from the analysis of modern DNA alone. In particular, we show how DNA, extracted from ancient substrates, has proven to be a crucial source of information to reconstruct the geographic and temporal origin of domestic species. We also discuss how ancient DNA is being used by geneticists and archaeologists to directly observe evolutionary changes linked to artificial and natural selection to generate a richer understanding of this fascinating process.

1.4 Introduction

The domestication of plants and animals was one of the most significant transformations in human history. Domestication was central to the emergence of settled agricultural communities (Larson et al. 2014). The advent of farming and pastoralism, during the Neolithic transition, lead to massive social, economic, religious, and demographic changes (Zeder 2012a). It supported vastly increased human population sizes (Bocquet-Appel 2011), and laid the foundation for the development of complex civilisations (Larson and Burger 2013). Ultimately these changes transformed the biosphere and ushered in the age of the Anthropocene (Smith and Zeder 2013).

The study of animal domestication is a broad endeavour, which draws in expertise from archaeology, genetics, ecology, and the physical sciences (Zeder et al. 2006; Vigne 2011; Larson et al. 2012, 2014; Zeder 2016; MacHugh et al. 2017). This multidisciplinary approach has provided the power to address the critical questions of when, where, and how animal domestications took place (Larson et al. 2014), as well as to help elucidate the biological basis for animal domestication (Jensen 2014). More recently, the study of animal domestication has been transformed by the revolution in modern and ancient

genome sequencing (Larson and Burger 2013; Larson and Bradley 2014; Gerbault et al. 2014).

This chapter will review how palaeogenomics has informed, and will continue to inform, our understanding of animal domestication. We will discuss how palaeogenomic approaches applied to domestic species have been used to resolve their geographic and temporal origin, track human migration, and to understand how animal genomes have been shaped by changes in human culture and technology.

1.5 Sequencing Ancient DNA

Early ancient DNA (aDNA) studies were constrained by the high cost and low yield of the sequencing technology which was available at the time. The first aDNA study, which recovered DNA from an extinct quagga (Higuchi et al. 1984), used molecular cloning to amplify target DNA molecules, by ligating them into plasmids and replicating them within bacteria (Maniatis et al. 1982). This approach was rapidly superseded by the discovery of the polymerase chain reaction (PCR) (Saiki et al. 1985; Mullis and Faloona 1987), which allowed researchers to efficiently amplify predetermined genomic loci, for sequencing using the Sanger chain-terminating method (Sanger et al. 1977). *In vitro* amplification (PCR) also had its limitations as it required *a priori* knowledge of the loci being targeted, which restricted analyses to species and genes which had already been sequenced in modern populations. PCR targets the intended locus using a pair of primers (forward and reverse) which flank the target region. As aDNA is highly fragmented—mostly less than 100 base pairs (bp) (Sawyer et al. 2012)—the loci targeted by the PCR primers needs to be shorter than the average length of endogenous molecules in an ancient sample, or the experiment might fail.

These early PCR based studies focused primarily on the recovery of a single gene locus from the mitochondrial DNA (mtDNA). The most commonly targeted regions were highly variable loci, such as cytochrome *b* and the mtDNA control region, which were used extensively for resolving molecular phylogenies (Irwin et al. 1991; Meyer 1994). Unlike the nuclear genome, which has only two copies in each cell, there can be many

thousands of copies of the mitochondrial genome in each cell (Reynier et al. 2001). This greater relative abundance of mtDNA improves the likelihood of retrieving any particular locus via PCR amplification. Whilst mtDNA is easier to recover, its information content is more limited than nuclear autosomal DNA. Autosomal DNA is inherited equally from both parents, in contrast to mtDNA which is uniparentally inherited, along the maternal line only. Consequently, mtDNA may not reflect the broader evolutionary history of the species as a whole (reviewed in Ballard and Whitlock 2004). Discrepancies between mtDNA and nuclear DNA analysis can be particularly acute when there are sex-biased processes, population replacement, or gene-flow occurring at the population level, such as those documented in horses (Vilà et al. 2001; Lippold et al. 2011b), pigs (Frantz et al. 2013b) and cattle (Hanotte et al. 2002).

The advent of high-throughput, or "next generation sequencing" (NGS) platforms in the mid-2000s (Margulies et al. 2005; Bentley et al. 2008) dramatically reduced the cost of sequencing and massively increased the volume of throughput (reviewed in Goodwin et al. 2016). For palaeogenomics, NGS technology was instrumental in the sequencing of the first ancient whole genomes, beginning with a ~40,000-year-old woolly mammoth (*Mammuthus primigenius*) (Miller et al. 2008), and shortly followed by a similarly aged Neanderthal (*Homo neanderthalensis*) (Green et al. 2010). In the years since then, ancient whole genomes have been published for several non-human mammalian taxa; including the horse (Orlando et al. 2013; Schubert et al. 2014; Librado et al. 2015), Przewalski's horse (Der Sarkissian et al. 2015; Gaunitz et al. 2018), quagga (Jónsson et al. 2014), auroch (Park et al. 2015), mammoth (Palkopoulou et al. 2015; Lynch et al. 2015), wolf (Skoglund et al. 2015), dog (Frantz et al. 2016b; Botigué et al. 2017; Ní Leathlobhair et al. 2018) and goat (Daly et al. 2018). Whilst ancient whole genomes have yet to be published for domestic pigs, cattle, sheep or chicken, sequences for these taxa will likely be forthcoming in the near future.

Sequencing ancient genomes, however, even with NGS technologies remains challenging—the primary constraint being the poor preservation of endogenous aDNA in sub-fossil remains recovered from archaeological sites. It is not uncommon for the endogenous DNA fraction of an NGS sequencing run to be below 1% (Carpenter et al.

2013). This problem is particularly acute in geographic regions with warm climates (Hofreiter et al. 2015), where most domestic animals originated. Many factors contribute to the degradation of ancient DNA; including time, temperature, humidity, soil pH, and microbial action. Despite decades of research, however, the decay kinetics of DNA degradation are still not well understood (Allentoft et al. 2012). In practice, the heterogeneity of DNA degradation makes preservation infeasible to accurately predict.

Recent studies have shown that aDNA preservation is also highly variable across different archaeological samples—awareness of which has led to dramatic improvements in aDNA recovery by focusing research on samples with higher endogenous yields. The petrous portion of the temporal bone can contain up to 183 times the concentration of endogenous DNA found in less dense bone (Gamba et al. 2014; Pinhasi et al. 2015). Tooth cementum has also been shown to contain comparably high levels of endogenous DNA content (Adler et al. 2011; Higgins et al. 2013; Damgaard et al. 2015). In experiments comparing petrous bones and tooth cementum, recovered from corresponding skeletons, the petrous bone was found to contain higher endogenous yields in only one tested assemblage, with the majority showing no systematic difference in yield (Hansen et al. 2017). As teeth are often over-represented in archaeological assemblages (Lam et al. 1999), they are an ideal target for aDNA recovery. In addition, teeth are great markers of domestication in multiple species, including pigs (Evin et al. 2013), horses (Cucchi et al. 2017) and dogs (Ameen et al. 2017).

Even with these strong constraints, genome-wide datasets have recently been published for early Neolithic farmers from sites across the Near East; including Anatolia, the Levant, and Zagros Mountains (Broushaki et al. 2016; Gallego-Llorente et al. 2016; Lazaridis et al. 2016; Kılınç et al. 2016). Comparable sequences for domestic animals from the region have so far been limited to goats (Daly et al. 2018). Given the great importance of the Near East as a centre for domestication, it is likely that more genome-wide sequences from ancient domestic animals will be forthcoming in the near future. Recovery of nuclear aDNA will be crucial to our understanding of the underlying process of domestication.

1.6 Pathways to Animal Domestication

When, where and how animals were domesticated are central questions to our understanding of human civilization. The current consensus among archaeologists and geneticists is that most domestic animals originated in a small number of "core" zones, from whence they were dispersed across the globe (Larson and Fuller 2014). As such, animal domestication is thought to be a rare process. Ancient DNA has been key to establish (as well as to challenge) our perception of the geographical and temporal origin of many species, and to test the idea that domestication is a rare phenomenon.

The idea that domestication is rare is also based on our current theoretical perspective that depicts domestication as a non-linear, diffuse, and long-term process that requires specific conditions to occur (Conolly et al. 2011; Vigne et al. 2011). The complexity and nuance of these processes have informed the development of two new theoretical models of animal domestication, by Vigne (2011) and Zeder (2012b), which have cast off the anthropocentrism of many previous models.

Vigne's (2011) model described animal domestication as the ultimate phase of intensification in the relationship between animal and human populations. This multi-stage model proposes a continuum of intensification, progressing through phases of (i) anthropophily; (ii) commensalism; (iii) control in the wild; (iv) control of captive animals; (v) extensive breeding; (vi) intensive breeding; and ultimately (vii) pet keeping (Vigne 2011). Not all domestic animals, however, progressed through each of these stages. By focusing on the shared phases of intensification between different groups of domestic taxa, Zeder's (2012b) model has proposed three main pathways to domestication. This model describes animal domestication as a mutualistic process, with progressive intensification of animal human relationships, however, it further distinguishes between three distinct evolutionary trajectories a (i) commensal pathway; a (ii) prey pathway; and a (ii) directed pathway (Zeder 2012b) (Figure 1.1).

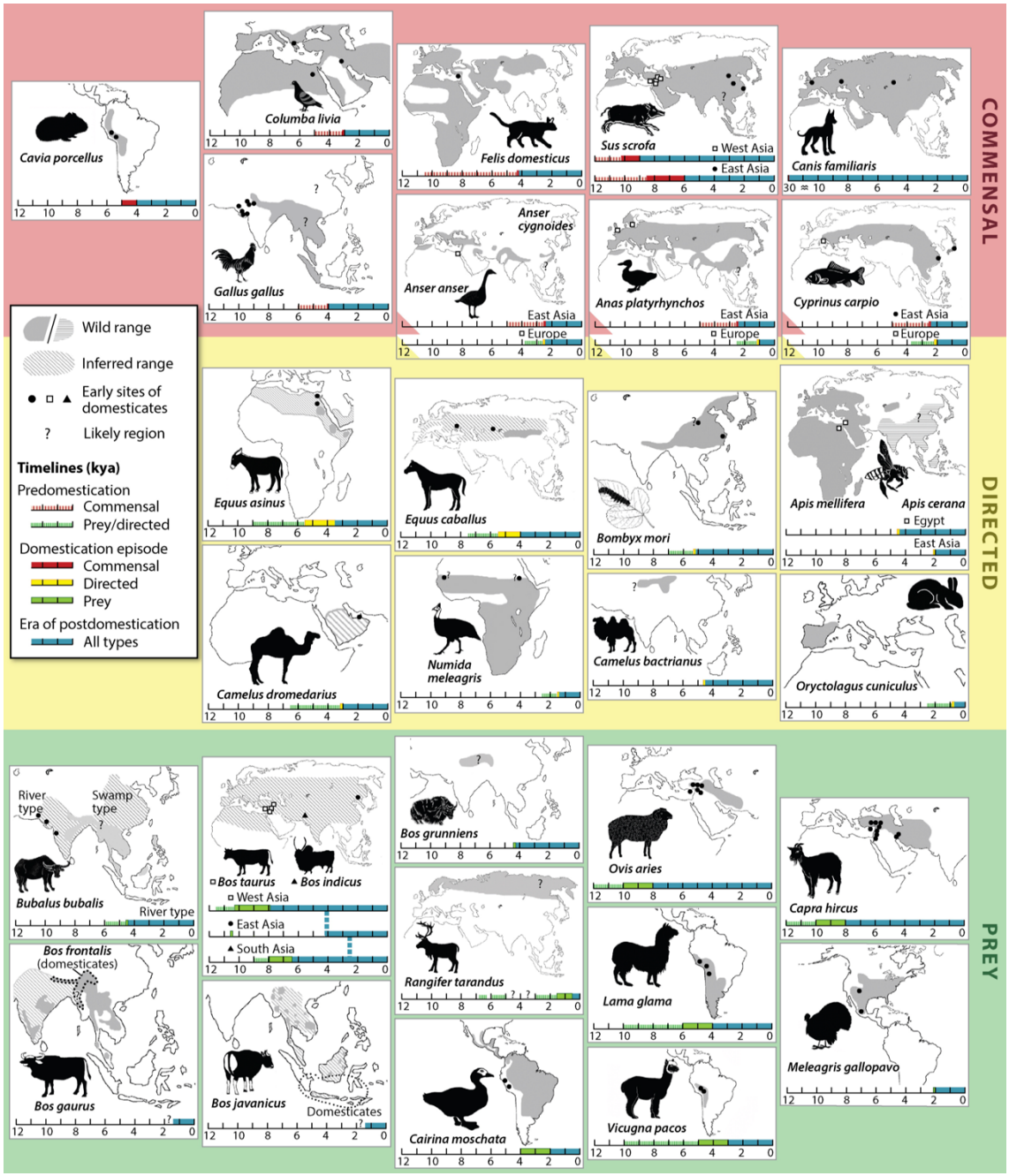


Figure 1.1 Geographical/chronological time frame of domestication and potential pathways for major domestic animals. The timelines are in Ky (1,000 years) increment. Adapted after (Larson and Fuller 2014).

1.6.1 The Commensal Pathway

Under the commensal pathway, wild animals were firstly attracted to, and then entangled by, elements of the human constructed niche (Zeder 2012b). The attraction occurred as wild animals were drawn to food sources available on the margins of human occupation—such as refuse scavenging (e.g. wolves and wild boars), food stores (e.g.

mice, or chickens) or increased prey availability (e.g. cats). These commensal animals would have been subjected to subtle selection favouring individuals who were more adapted to exploit the human niche. Over time this human animal relationship intensified, ultimately leading to a full domestic partnership. This commensal pathway implies no intentionality or forethought on the part of the human partners during most of the process, but rather describes a slowly evolving beneficial relationship (Zeder 2012b).

1.6.2 The Prey Pathway

Under the prey pathway, wild animals were firstly exploited for their meat and hides before demographic pressures lead humans to take an ever-greater role in herd management (Zeder 2012b). Where hunting pressures may have changed the size and composition of prey herds, humans responded by adjusting their hunting strategies to maintain sufficient prey availability—such as preferential targeting of young males (Zeder 2006). Over time, these hunting strategies developed progressively through more advanced systems of herd management, captive breeding, through to directed breeding for favourable behavioural and phenotypic traits. In this way, the early stages of the prey pathway can be seen as just as unintentional as the commensal pathway. In contrast, however, the latter stages of the prey pathway are characterised by an intensification of human intervention, in an attempt to maintain supply of a diminishing resource (Zeder 2012b).

1.6.3 The Directed Pathway

Under the directed pathway, humans leveraged their prior experience with domestic animals, and their emergent understanding of directed breeding, to capture wild animals and intentionally bring them under increasing levels of human control (Zeder 2012b). The directed pathway describes the route taken for almost all recently domesticated taxa—particularly the exponential increase in aquatic species—but was of much lesser importance in the distant past. A recent meta-analysis found that 97% of all aquatic domesticates have been domesticated in the past hundred years, including

more than 100 species in the preceding decade alone (Duarte et al. 2007). This recent prevalence of the directed pathway, coupled with modern intensive breeding practices, has been formative in the minds of many researchers, and obscured a clearer understanding of early animal domestications. The idea of the directed pathway as the preeminent mode of domestication is typified by the theories of Galton (1865) and Clutton-Brock (1994), among many others, in which domestication is seen as the logical outcome of the intentional taming of wild animals.

1.7 When, Where and Which Pathway

The domestication of animals began more than 15,000 years ago, with the domestication of the grey wolf (*Canis lupus*) by nomadic hunter gatherers (Larson et al. 2012). It was not until much later (beginning around 11,000 years ago) that people in the Near East intensified their relationships with wild populations of sheep, goat, auroch and boar, such that incipient domestication processes began to emerge (Conolly et al. 2011). By 10,000 years ago, these four elements of the so-called 'Neolithic package' had spread extensively throughout Southwest Asia and the eastern Mediterranean (Vigne 2008). Despite the later ubiquity of these domesticates across the region, detailed zooarchaeological studies have revealed the complex non-linear nature of these domestication processes; complete with ebbs and flows in tempo in response to local environment and conditions (Conolly et al. 2011; Vigne et al. 2011).

In the following section, we will briefly review what is known about the domestication of a range of key mammalian, avian and insect species—with a particular focus on palaeogenomic contributions to our understanding of these domestications. The species profiles are grouped by the pathways they each took to domestication, to highlight shared elements of the underlying process.

1.7.1 Commensal Domesticates

1.7.1.1 Dogs

The first animal likely to have followed the commensal pathway to domestication was the grey wolf (*Canis lupus*) (reviewed in Thalmann and Perri 2018). It has been theorised that wolves who were naturally less wary of people would have been drawn to human encampments to scavenge refuse left by hunters (Thalmann and Perri 2018). Where, when and how many times wolves were domesticated remains a contentious issue, due to the sparsity of evidence and conflicting interpretations of both the archaeology and genetics (Germonpré et al. 2009; Larson and Bradley 2014; Skoglund et al. 2015; Frantz et al. 2016b; Botigué et al. 2017; Ní Leathlobhair et al. 2018).

The earliest widely accepted archaeological dog remains date to about 15,000 years ago (Thalmann and Perri 2018). Earlier canids remains, dating back to over 30,000 years ago (Germonpré et al. 2009), were recently described as dogs but their status (as dogs or wolves) remains highly controversial (Perri 2016). Palaeogenomic data has provided additional information about the potential time frame for dog domestication. In particular, analyses of genome-wide data from an ancient Siberian wolf (Skoglund et al. 2015) and an ancient Irish dog (Frantz et al. 2016b) together with radiocarbon dates have provided the means to estimate a reliable mutation rate for canids and to obtain an estimate of the divergence time between extant wolves and dogs of 20,000-40,000 years ago. This timing, which represents an upper bound for dog domestication, needs to be interpreted with caution as the ancestor of dogs may have become extinct (Thalmann et al. 2013; Freedman et al. 2014; Frantz et al. 2016b). This would mean this time instead represents the time of divergence between extant wolves and the ancestor of dogs, rather than the time at which dogs were domesticated.

Over the years, genomic (including palaeogenomic) studies have provided conflicting information about the geographical origin of dogs, with papers suggesting that dogs originated in East Asia (Pang et al. 2009; Wang et al. 2015), Central Asia (Shannon et al. 2015), the Middle East (vonHoldt et al. 2010) and Europe (Thalmann et al. 2013). Additional genome-wide palaeogenomic studies, however, have provided novel clues

on the geographical origin of dogs. For example, studies based on ancient genomes from European dogs have suggested that modern Western Eurasian populations (including Africa, Europe and Middle East) were most likely imported from Asia, over 7,000 years ago (Frantz et al. 2016b; Botigué et al. 2017). Based on additional archaeological data and multiple ancient mtDNA sequences, Frantz et al. (2016b) suggested that populations that inhabited Europe and the Middle East, prior to the arrival of dogs from East Asia, had been domesticated independently. This hypothesis, which implies that the dogs that were present prior to the arrival of East Asian dogs are now extinct, remains to be tested.

1.7.1.2 Pigs

Although they were hunted like other ungulate species (sheep, goat, cattle etc.), the omnivorous lifestyle of pigs provided them with the ability to consume human waste, suggesting that they were potentially domesticated via a commensal pathway (Larson and Fuller 2014). Interestingly, pigs are the only animals for which we have unequivocal, genetic and archaeological evidence for two independent domestication processes, from two different subspecies of *Sus scrofa*, in China and Anatolia respectively (Larson et al. 2005). Ancient DNA studies have played a key role in unravelling a complex domestication history marked by frequent population replacements.

The western Eurasian domestic pigs were most likely first domesticated in Anatolia, over 10,000 years ago, as suggested by zooarchaeological evidence of selection and culling from long-term occupation sites such as the Çayönü Tepesi (Hongo and Meadow 1998; Ervynck et al. 2001). Ancient DNA evidence suggests that they were then transported, from the Near East into Europe as part of the Neolithic package (Larson et al. 2007a), around 9,000 years ago (Conolly et al. 2011). Evidence for such an early human mediated dispersal of pigs, from the Near-East into Europe, however, is absent from modern DNA sequences (Larson et al. 2005). Lack of Near Eastern ancestry in modern domestic breeds is most likely the result of a population turnover resulting from long-term gene-flow between European wild boars and domestic pigs (Frantz et al. 2015), a process that likely started as soon as pigs were introduced in Europe (Larson et al. 2007a).

Further ancient DNA evidence suggest that European domestic pigs, lacking Near Eastern ancestry, were later introduced back into the Near East (Anatolia), during the Iron Age where they replaced pigs with Near Eastern ancestry (Ottoni et al. 2013). More recently, Chinese pigs, which were domesticated from a highly divergent subspecies (Frantz et al. 2013b, 2016a) were imported in Europe to improve production traits during the industrial revolution (White 2011; Bosse et al. 2014a). This process dramatically affected the genetic (Bosse et al. 2014b) and phenotypic make-up (Bosse et al. 2014a) of European populations.

In East Asia, the first unequivocal evidence of pig domestication dates back to ~8,600 years ago at the site of Jiahu near the Yellow River (China) (Cucchi et al. 2011). Similar to the process seen in Europe, ancient mtDNA evidence suggests that East Asian domestic pigs were transported from their domestication centre to Island South East Asia, Papua and Polynesia were they were later replaced by pigs of European decent (Larson et al. 2010; Linderholm et al. 2016). During their human mediated dispersal throughout Island South East Asia, domestic pigs encountered a high diversity of wild *suid* species (and subspecies) which readily interbreed with domestic pigs (Frantz et al. 2013b, 2014; Ai et al. 2015). Future ancient nuclear DNA will be able to assess whether deliberate interbreeding with wild stock may have allowed for adaptation of domestic pigs to the wide range of habitat they encountered in Europe, Asia, and Polynesia (Frantz et al. 2016a).

1.7.1.3 Cats

Cats (*Felis catus*) also became domesticated via the commensal pathway, however, despite their worldwide popularity relatively little is known about the origins of the domestic cat (reviewed in Geigl and Grange 2018). The archaeological and genetic evidence points to both the Near East and Egypt as important regions for the domestication of the cat (Vigne et al. 2004; Driscoll et al. 2007; Ottoni et al. 2017). The wild progenitor of the domestic cat (*Felis catus*) is the Near Eastern wildcat (*Felis silvestris lybica*), which has a natural range spanning North Africa and the Near East (Driscoll et al. 2007). Archaeological remains of wildcats in the Near East point to a long

history of commensal relationship with early farming communities, where they are thought to have predated on invasive rodent populations (Vigne et al. 2004, 2012). This relationship persisted for thousands of years before the appearance of any classic domestication traits—such as reduction in overall body size and the emergence of novel coat colours (Vigne et al. 2016). The dispersal of domestic cats around the world was aided by their role on ships and trade vessels as protection against rodents. This is reflected in their patterns of dispersal, which mirror major trade routes (Lipinski et al. 2008; Ottoni et al. 2017).

A worldwide phylogenetic study of modern cats, using a fragment of the mitochondrial genome (mtDNA) and microsatellite markers, has shown that the Near Eastern wildcat (*F. s. lybica*) is more closely related to the domestic cat than other subspecies of wildcat (Driscoll et al. 2007). Based on the current distribution of wildcats, the authors concluded that cats were most likely domesticated in the Near East. Of the 979 analysed samples, they identified 15 wildcats from Israel, the United Arab Emirates, Bahrain, and Saudi Arabia with mtDNA and microsatellite markers consistent with those found in modern domestic cats (Driscoll et al. 2007). Further research, using microsatellite markers to analyse the phylogeographical structure of modern domestic cats, also found support for a Mediterranean basin origin for their dispersal (Lipinski et al. 2008).

A recent ancient mtDNA study of 209 archaeological cat remains has shown that mitochondrial lineages from both the Near East and Egypt contributed to worldwide domestic cat populations at different times (Ottoni et al. 2017). Their analysis showed that domestic cats are drawn from five deeply divergent mtDNA subclades (IV-A to IV-E) of *F. s. lybica*, and that the relative proportions of domestic cat haplogroups has shifted over time. The IV-A and IV-B subclades were identified as originating in the Near East and represent the first wave of domestic cats which spread across the Old World. The IV-C subclade originated in Egypt and was found in the majority of Egyptian cat mummies. Despite a supposed ban on the export of Egyptian cats (Zeuner 1963), the Egyptian subclade increased in frequency outside Egypt, such that during the 1st millennium AD in Western Anatolia, it had expanded to twice the frequency of the local Near Eastern subclade (Ottoni et al. 2017). The authors speculated that the cause of this

increase might have been due to more desirable behavioural characteristics of the Egyptian cats.

The same study also looked at the history of the tabby coat trait, one of the most widely used markers for identifying domestic cats (Ottoni et al. 2017). Their analysis found that the coat-colour variant responsible for the derived blotched tabby marking only reached high frequency after the Middle Ages, around the time that cat pelts were being traded for clothing. Coupled with the relatively small changes in overall size of domestic cats, this suggests that directed breeding of cats for morphological novelty was a very late phenomenon (Ottoni et al. 2017).

1.7.1.4 Chickens

Genetic data from modern domestic chickens and wild junglefowl have established that the red junglefowl (*Gallus gallus*) is the primary wild ancestor of the domestic chicken (Liu et al. 2006; Miao et al. 2013). Studies of nuclear genetic data have demonstrated, however, that the yellow skin allele found in domestic chickens was not inherited from red junglefowl, but was instead inherited from the grey junglefowl (*Gallus sonneratii*) demonstrating that the genome of modern domestic chickens combines elements of at least two junglefowl species (Eriksson et al. 2008).

An initial review of the archaeological evidence argued that chicken domestication had begun by the 3rd millennium BC, since the first robust evidence for poultry farming has been recovered in the Indus Valley ~2,600-1,900 BC, before chickens were then translocated to the Near East, Africa and Europe during the 1st millennium BC (Zeuner 1963). Based on an analysis of osteological evidence that attested to the presence of chickens during the middle Neolithic (~6,000-4,000 BC) in the Yellow River basin, West and Zhou (West and Zhou 1988) concluded that chickens were domesticated in the Southeast Asian native range of red junglefowl prior to 6,000 BC before being dispersed westwards along a northern route through Central and Western Eurasia. Two subsequent studies (Berke 1995; Peters 1997) questioned whether the Chinese specimens actually belonged to domestic chickens since they possessed morphological

features typical of other galliform birds including pheasants. Despite these critiques, a mid-Holocene origin of domestic chickens has been frequently claimed in the literature.

A recent ancient DNA analysis of galliform bone specimens from early and middle Neolithic sites in the Yellow River basin reinforced the claims for an early domestication of chickens (Xiang et al. 2014). This study suggested that red junglefowl dispersed naturally to Northern China following the Younger Dryas where they were then domesticated during the early Neolithic. This assertion has since been questioned. An independent morphological re-evaluation of galliform bones from northern Chinese Neolithic sites concluded that the bones in question belonged primarily to the common pheasant (*Phasianus colchicus*) (Peters et al. 2016; Eda et al. 2016). In addition, several lines of evidence including an assessment of associated wild mammalian faunas and high resolution climate and precipitation records from temperate Holocene East Asia suggested that the (sub-)tropical forest habitat conducive to thermophilic red junglefowl did not extend into Northern China during the mid-Holocene climatic optimum (Peters et al. 2016). Lastly, multiple studies of modern domestic chickens have suggested that red junglefowl from peninsular Southeast Asia are the likely initial population from which domestic chickens were derived, and a recent genetic study of complete mitochondrial genomes has cast doubt on the likelihood that chickens were domesticated in Northern China (Huang et al. 2018). Future archaeological and genomic studies of modern and ancient chickens are necessary to reveal not only the spatial and temporal pattern of chicken domestication, but also the process which led to the close association between chickens and people.

1.7.2 Prey Domesticates

1.7.2.1 Goats

Goats (*Capra hircus*), along with sheep (*Ovis aries*) and cattle (*Bos taurus*), all followed the prey pathway to domestication in the Fertile Crescent region of the Near East (Zeder 2012b). Archaeological and genetic evidence has established that goats were domesticated from the bezoar ibex (*Capra aegagrus*), a species of wild goat inhabiting

the mountainous region spanning Southwestern Turkey, to central Afghanistan and southern Pakistan (Zeder and Hesse 2000; Naderi et al. 2008).

Detailed zooarchaeological studies of wild goat assemblages have allowed researchers to reconstruct the age and sex-specific harvest profiles employed by hunters prior to domestication. These harvest profiles reveal incipient herd management strategies, in which hunters transitioned from targeting of prime age males, which maximised short-term meat return, towards selective culling of sub-adult males and older adult females, to promote growth in herd sizes (Zeder and Hesse 2000; Zeder 2006, 2008). These management strategies of wild ranging goats gradually intensified from herding towards a fully domestic relationship, and domestic phenotypes appear in archaeological goat assemblages around 10,500 years ago at multiple sites across Southeast Anatolia, the Zagros mountains and Cyprus (Conolly et al. 2011).

Domestic goats were subsequently brought into Europe as part of the Neolithic Package, however, unlike with pigs and cattle, there were no extant wild populations for the incoming domestic population to admix with (Scheu et al. 2012). Studies of modern mitochondrial DNA in domestic goat populations have revealed unusually high levels of genetic diversity coupled with low levels of geographical structuring (Luikart et al. 2001; Naderi et al. 2007, 2008). This diversity has been attributed to population structure in the region of domestication, followed by extensive trade and transport of domestic goats. Modern goat populations comprise six maternal haplogroups (A, B, C, D, F and G), with most domestic goats belonging to haplogroup A (Naderi et al. 2007). The first ancient DNA study of goats established that haplogroups A and C were both present in the Early Neolithic in France, with moderately high genetic diversity, a result that the authors interpreted as potential evidence for two independent domestications with subsequent gene flow between populations (Fernández et al. 2006).

Recently, the diverse origins of domestic goats were further investigated in the first genome-wide study of ancient caprids (Daly et al. 2018). The authors selectively targeted petrous bones to retrieve genome-wide data from 51 ancient goats and used mtDNA capture to retrieve complete mtDNA genomes for 83 ancient goats. Their

analyses of nuclear genomes provided evidence for variable proportions of ancestry shared between pre-domestic wild goats and early domestic goat populations, which suggested local recruitment of divergent wild populations during domestication (Daly et al. 2018). This was mirrored by the mtDNA data, which showed that multiple highly divergent haplogroups were involved in the domestication process and have differentially contributed to the genetic make-up of modern populations. This study also revealed that, in contrast to modern populations (Naderi et al. 2008), mtDNA haplogroups were highly structured in ancient populations (Daly et al. 2018). Interestingly, the collapse in haplogroup structure happened relatively early in their evolutionary history (~7,000 years ago), when haplogroup A replaced most others to become the dominant haplogroup across the region (Daly et al. 2018).

1.7.2.2 Sheep

Sheep (*Ovis aries*) also followed a prey pathway to domestication in the Fertile Crescent, around 10,500 years ago, with the Asiatic mouflon (*Ovis orientalis*) as the most likely wild progenitor (Conolly et al. 2011). Both the urial (*Ovis vignei*) and the argali (*Ovis ammon*) have also been suggested as potential ancestors, however, no mitochondrial lineages from either species have been observed in domestic sheep populations (Meadows et al. 2011). The European mouflon (*Ovis aries musimon*), is a feral descendent of a primitive domestic population (Bruford and Townsend 2006), and a recent genome-wide analysis revealed widespread bidirectional admixture between European mouflon and modern domestic sheep (Barbato et al. 2017).

Domestic sheep populations comprise five maternal haplogroups (A, B, C, D and E), with most modern sheep belonging to haplogroups A, B and C (Meadows et al. 2011). Two major Y-chromosome patrilines have also been identified, showing limited geographic structure (Meadows and Kijas 2009). Similar to the pattern seen in domestic goats, the maternal haplogroups diverged long before domestication, suggesting that multiple divergent lineages were involved in the domestication process (Pedrosa et al. 2005; Meadows et al. 2011). The relative abundance of these haplogroups have changed over time, with haplogroups A and B dominating the initial expansion into Europe, followed

by haplogroup C around 3,000 years ago (Tapio et al. 2006). This first wave of domestic sheep, bred primarily for meat production, were replaced by a second wave of domestic stock carrying improved production traits for wool and milk (Chessa et al. 2009; Demars et al. 2017).

Recently, a genome-wide study of selection in modern sheep and goats found 90 selective sweep regions which segregated between domestic and wild populations of *Capra* and *Ovis* (Alberto et al. 2018). A gene ontology enrichment analysis (reviewed in Huang et al. 2009) identified significant enrichment for genes involved in nervous system, immunity and productivity traits (Alberto et al. 2018). Interestingly, this analysis identified only 20 regions under selection which were common to both *Capra* and *Ovis*, suggesting that convergent phenotypes in goats and sheep were primarily established by selection on non-homologous gene regions.

1.7.2.3 Cattle

Cattle (*Bos taurus* and *Bos indicus*) also followed the prey pathway to domestication, however, there is ongoing uncertainty about how many times cattle were domesticated (Loftus et al. 1994; Troy et al. 2001; Hanotte et al. 2002; Beja-Pereira et al. 2006; Chen et al. 2010; Pitt et al. 2018). Large genome-wide studies of modern domestic cattle have shown that they form three deeply divergent groups: (i) Eurasian and (ii) African taurine cattle (*Bos taurus*), and (iii) Asian indicine cattle, or zebu (*Bos indicus*) (Gibbs et al. 2009; Decker et al. 2014).

The earliest cattle domestication occurred in the Fertile Crescent, approximately 10,500–10,000 years ago, where Eurasian taurine cattle were domesticated from wild Eurasian aurochs (*Bos primigenius*) (Hanotte et al. 2002; Helmer et al. 2005; Hongo et al. 2009; Conolly et al. 2011). The domestication of Asian indicine cattle occurred in South Asia, approximately 8,000–7,500 years ago, and was the product of either an independent domestication process or admixture between domestic taurine cattle and Asian aurochs (*Bos primigenius namadicus*) (Meadow 1983; Loftus et al. 1994; Chen et al. 2010; Larson and Burger 2013). Current archaeological and genetic evidence is

consistent with an independent domestication process, however, without ancient genome-wide data admixture between Asian aurochs and domestic taurine cattle cannot be ruled out as a potential cause of indicine cattle domestication (Larson and Burger 2013). Uncertainty around a hypothesised independent domestication of African aurochs (*Bos primigenius africanus*) (Bradley et al. 1996; Hanotte et al. 2002; Wendorf and Schild 2005; Stock and Gifford-Gonzalez 2013), in the Western Desert of Egypt, has been largely resolved following reanalysis of the archaeological material (Brass 2018), and explicit model based testing of the genetic data (Pitt et al. 2018), which found no evidence for an independent African domestication.

A recent study published the first whole-genome sequence of an extinct Eurasian auroch (*Bos primigenius*), recovered from a 6,750-year-old British specimen (Park et al. 2015). Analysis of the genome-wide data revealed localised nuclear gene flow into the ancestors of British and Irish taurine cattle, contrary to previous mtDNA studies, which found no evidence of introgression (Edwards et al. 2007). Model based testing of ancient genetic data suggest that the matrilineal founding population of taurine cattle may have been as low as just 80 individuals (Bollongino et al. 2012; Scheu et al. 2015). As taurine cattle migrated from the Near East into Europe, their mtDNA genetic diversity decreased along the axis of migration, and intercontinental migration continued up until ~7,000 years ago (Scheu et al. 2015). When whole-genome sequences of early domestic cattle become available, we will be able to better resolve the role of introgression between Eurasian domestic cattle and wild aurochs.

Within Asia, the evolutionary history of the *Bos* genus is characterised by reticulate admixture between domestic cattle populations and other *Bos* species (Wu et al. 2018). East Asian cattle populations show a mosaic of ancestry components, including an ancestral East Asian taurine component, a later Eurasian taurine component, and a deeply divergent Chinese indicine component (Chen et al. 2018). Cattle populations from Tibet also show signs of adaptive introgression of yak (*Bos grunniens*) genes, in the response-to-hypoxia pathway, likely supporting an adaptation to high altitude (Chen et al. 2018; Wu et al. 2018)—similar to the adaptive introgression from Denisovans into

Tibetans (Huerta-Sánchez et al. 2014), and Tibetan wolves into Tibetan mastiffs (Miao et al. 2017).

1.7.2.4 New World Camelids

In South America, llamas (*Lama glama*) and alpacas (*Vicugna pacos*) likely also followed a prey pathway. Archaeological evidence suggests the domestication of llamas and alpacas from their potential wild progenitors, vicuñas (*Vicugna vicugna*) and guanacos (*Lama guanicoe*), began ~6,000 years ago (Diaz-Lameiro 2016) within their overlapping native ranges in the mountainous regions of Bolivia, Chile and Peru and the central Andes Mountains (Barreta et al. 2013). There are two current hypotheses for how the domestication of these two species took place. The first is that both llamas and alpacas are domesticated forms of guanacos. Alternatively, alpacas may be a domesticated form of vicuñas while llamas were derived from guanacos.

Both these hypotheses have support from genetic data. Ancient mitochondrial DNA sequenced from llama and alpaca remains from pre-Columbian South American sites (Cerro Narrio, Ecuador and Iwawi, Bolivia) demonstrated that the ancient alpacas and llamas clustered together within a well-supported monophyletic group more closely related to guanacos than to vicuñas, thus suggesting that both species were domesticated from guanacos in the Northern South American Andes (Diaz-Lameiro 2016). The second hypothesis is supported by a study using modern nuclear data, which suggested that alpacas and llamas are more closely related to vicuñas and guanacos, respectively (Kadwell et al. 2001; Wheeler et al. 2006). Though this second study based upon a larger number of nuclear and mitochondrial loci has more weight, the large observed differences between the wild species may be partly due to a strong bottlenecking in the recent past. For instance, the vicuña population in the 1960s had a population size of only 2,000 across South America (Barreta et al. 2013) and guanaco populations have been small over the past century. The biases associated with these recent demographic shifts may have had an effect on the interpretation of these datasets. Understanding the origins and domestication history of these two species will be much more clearly understood through the generation and interpretation of ancient

nuclear DNA datasets derived from archaeological material across the spatiotemporal range of the wild and domestic species.

1.7.3 Directed Domesticates

1.7.3.1 Horses

The earliest suggested case of an animal following the directed pathway to domestication is that of the horse (*Equus ferus caballus*) (Zeder 2012b), which may have been domesticated to assist steppe pastoralists in hunting wild horses (Levine 1999; Olsen 2006a). Identifying horse domestication in the archaeological record is difficult because many of the classic markers of domestication show no discernible variation between early wild and domestic populations—e.g. morphological changes (Eisenmann and Mashkour 2005) and mortality profiles (Olsen 2006b, a). The earliest evidence for horse domestication (reviewed in Orlando 2018) comes from Central Asia, around ~5,500 years ago, where skeletal pathologies indicate horses were bridled and probably ridden, and stable isotope analysis of lipid residues in pottery indicate processing of mare's milk (Outram et al. 2009).

Modern horse populations comprise 18 major maternal haplogroups (A–R), 17 of which are found in domestic horses and one of which (haplogroup F) is found only in Przewalski's horses (*Equus ferus przewalskii*) (Achilli et al. 2012). This high number of mtDNA haplogroups, which diverged long before the start of domestication, has been interpreted as evidence of extensive restocking of wild mares during the domestication process (Vilà et al. 2001; Lippold et al. 2011b). In contrast, modern Y-chromosome patrilines have an extreme lack of diversity (Lindgren et al. 2004), likely caused by a strong bottleneck in male horses. The timing of this bottleneck is not clear, however, as aDNA studies have revealed that ancient domestic horses had greater Y-chromosome diversity than modern horses (Lippold et al. 2011a; Librado et al. 2017). The recent publication of the first complete assembly of the horse Y-chromosome should assist in future aDNA studies of male-biased processes in horse domestication (Janečka et al. 2018).

A recent genome-wide aDNA study of 14 ancient domestic horses, has also challenged the traditional view that the high rate of deleterious mutations found in modern horses can be attributed to a male population bottleneck during domestication (Librado et al. 2017). The “cost of domestication” hypothesis (reviewed in Moyers et al. 2018) argues that the process of domestication leads to increased levels of deleterious mutations in domestic animals—principally via population bottlenecks and strong artificial selection. In the case of horses, however, aDNA has revealed that ancient domestic horses had high rates of genetic diversity, and an analysis of the fitness consequences of that diversity found that the mutational load of ancient horses was less than that of both modern horses and pre-domestic horses (Librado et al. 2017). This implies that current levels of deleterious mutations are most likely a product of subsequent breeding practices, rather than a consequence of the domestication process itself.

Przewalski's horses are often described as the only extant wild horses (e.g. Der Sarkissian et al. 2015), after they were rescued from extinction in the wild following a captive breeding program involving 12 wild-caught individuals (Volf et al. 1991). A recent genome-wide aDNA study of ancient domestic and Przewalski's horses, from the domestication centre in Central Asia, however, showed that Przewalski's horses are not truly wild, but are instead the feral descendants of the first domestic horses (Gaunitz et al. 2018). This study revealed that it was the ancestors of modern Przewalski's horses which were first domesticated ~5,500 years ago, and that by ~4,000 years ago there had been a nearly complete genetic turnover among domestic horses, coinciding with the dramatic population expansion associated with the Yamnaya culture during the Early Bronze Age (Allentoft et al. 2015; Gaunitz et al. 2018). The exact timing of this turnover, and the geographic origin of the population, which gave rise to all modern domestic horses, remains unknown. Whilst there is still much to discover about the evolutionary history of horses, this study highlights the incredible insights that palaeogenomics can bring to our understanding of the history of domestication.

1.7.3.2 Rabbits

The European rabbit (*Oryctolagus cuniculus*) is often reported to have been domesticated via the directed pathway. In the most widely cited historical account, rabbits were supposedly domesticated by Catholic monks in France, circa AD 600, when they were granted a dispensation to eat foetal rabbits during Lent (Zeuner 1963; Clutton-Brock 1981). The practice of eating *laurices*—newborn or foetal rabbits—goes back to at least the 1st century AD, when Pliny the Elder describes the Spanish delicacy of cutting foetal rabbits from the belly of their mother and eating them whole and uneviscerated (*Naturalis Historia*, 8.55). It follows, that by granting permission to consume *laurices* during the many fasting days of the medieval calendar, French monks were suddenly motivated to move the breeding of rabbits above ground to obtain a reliable supply of newborn rabbits. First put forward by (Nachtsheim 1936) (1936), this account has its origins in a widely-miscited text from the late 6th century by St Gregory of Tours (Gregory 1969). Through successive retellings, the account became incrementally embellished, such that consumption of *laurices* became especially popular amongst the monks during Lent (Nachtsheim 1936), then permitted by the Church because they were not considered meat (Zeuner 1963), and ultimately that the dispensation was granted by Pope Gregory the Great (Carneiro et al. 2011), an unrelated contemporary of St Gregory of Tours. In fact, there is no evidence that eating *laurices* was ever commonplace nor that they were not considered meat, and the timing and nature of rabbit domestication remains unknown (Irving-Pease et al. 2018).

Despite this, European rabbits have a well-resolved geographic origin, in Southwest France, and the presence of an extant wild progenitor makes it comparatively easy to obtain modern genomic samples from which to model the process of selection during domestication (Carneiro et al. 2011, 2014, 2015). A recent study compared genome-wide data from six breeds of domestic rabbits and wild rabbits from across their native range, to scan for segregating signatures of selection (Carneiro et al. 2014). The authors found more than 100 selective sweep regions distinct to domestic rabbits, and a gene ontology enrichment analysis identified significant enrichment for genes involved in brain and neuronal development (Carneiro et al. 2014). Interestingly, the authors found

very few fixed derived alleles in the domestic breeds, suggesting that domestication was achieved via changes in allele frequencies at hundreds of loci, each with low effect size. When ancient genome-wide data becomes available for European rabbits it should be possible to test the timing of selection at these loci, to better elucidate the process of rabbit domestication.

1.7.3.3 Old World Camels

The progenitor of modern Old World Camels reached Eurasia ~3 million years ago (Gaulthiers-Pilters & Dagg 1981, Koehler 1981, Peters 1997). By the middle Pleistocene, Old World Camels ranged from China and Mongolia over Central Asia, to the Arabian Peninsula, including parts of North Africa and Eastern Europe (Koehler 1981, Titov 2008). By the end of the Pleistocene, the range of wild camelids had contracted dramatically (Gaulthiers-Pilters & Dagg 1981, Kozhamkulova 1986, Titov 2008) and several wild camel species became extinct leaving only the species *Camelus ferus*.

The distribution of the small extant wild population is restricted to China and Mongolia (Bannikov 1976, Hare 1997, Reading et al. 1999, Mix et al. 1997, Mix et al. 2002), though the domestic form, *Camelus bactrianus*, has spread throughout Central Asia and is now found from North-East China, Mongolia, South-Russia, and Central Asia. In Asia Minor, its distribution overlaps with that of the dromedary. The one-humped dromedary camel, probably once found as a wild animal throughout the Arabian region but known with certainty only in the domestic or feral state, is now widespread in the hot deserts of Northern Africa and Arabia (Walker 1964).

Archaeological records show evidence for a relationship between people and the Bactrian camel ~5,000 years ago (Bulliet 1975, Benecke 1994) and the earliest records of camel bones come from sites Turkmenistan and Iran (Kuzmina 2008). Given the presence of camel bones in Bronze Age strata from sites in Iran and southern Turkmenistan, it has been hypothesized that the inhabitants of the Iranian Plateau and the Kopet-Dagh-foothills area played a major role in the domestication of the two-humped camel (Benecke 1994).

Due to their use as pack animals, the modern populations of dromedary camels do not possess significant phylogeographic structure (Almathen et al. 2016). A recent study of dromedary camels (Almathen et al. 2016) successfully recovered DNA from ancient dromedary remains. The authors concluded that the founders of the modern domestic dromedary camels were likely a population of wild camels present in the southeastern corner of the Arabian Peninsula, and that domestic populations were routinely hybridised with wild individuals with novel mtDNA haplotypes.

1.7.3.4 Insects

Two domesticated insect species likely followed the directed pathway: silkworms (*Bombyx mori*) and honey bees (*Apis mellifera).* People probably began selectively breeding moths for silk production ~5,000 years ago. (Bisch-Knaden et al. 2014). The extreme changes in morphology and their reliance on humans for survival and reproduction has led to the recognition of the domestic form as a unique species, *B. mori.* Recent genetic analyses of complete mitochondrial sequences from different geographic regions (Li et al. 2010) and a mixture of mitochondrial and nuclear loci (Sun et al. 2012) now suggests that silkworm domestication began in China, in line with fossil, historical and archaeological lines of evidence (Sun et al. 2012).

Though there is a clear genetic distinction between wild and domestic silkworm lineages, *B. mori* retain ~83% of the genetic variance of its wild relatives. Xia et al. (2009) interpreted this observation as evidence for a short domestication process with a large starting population. Yang et al. (2014) used coalescence simulations and the approximate Bayesian computation (ABC) on 29 nuclear loci to suggest that domestication began ~7,500 years ago with a subsequent bottleneck ~4,000 years ago. Though the genetic architecture of domestication remains uncertain, several studies have identified genes and phenotypes that have been selected during domestication, including loci related to the olfactory system (Xiang et al. 2013), orphan genes (Sun et al. 2015) (reviewed in Tautz and Domazet-Lošo 2011) and epigenetic changes (Xiang et al. 2013).

The genus *Apis* has 10 distinct species, 9 of which are confined to Asia which suggests that the domesticated species, *A. mellifera*, also originated in Asia. This is supported by the fact that the closest species to *A. mellifera*, *Apis cerana,* is found in western and central Asia. Unlike domestic silkworms however, there is a range of subspecies of domesticated honey bee and these are phenotypically distinct in different geographic regions. Because these species are adapted to their environment of origin, the basis for this phenotypic variation is largely unknown (Wallberg et al. 2014). These sub-species fall into four categories supported by morphometric and genetic studies: A are a subspecies found throughout Africa, M from western and northern Europe, C eastern Europe and O includes species from Turkey and the middle East (Han et al. 2012). Despite the parsimonious explanation, an early paper using 1,136 nuclear SNPs suggested Africa as the origin of *A. mellifera* due to distance trees rooting in the African clade (Whitfield et al. 2006). More recent studies have questioned this conclusion. One study demonstrated that some of the analysed sub-species were actually recent hybrids, and by removing these species from the analyses the root of phylogenetic trees did not fall unequivocally into the A clade (Han et al. 2012). A similarly ambiguous conclusion was drawn when trees were built using 8.3 million SNPs (Wallberg et al. 2014). As a result, Asia remains the most likely origin of *A. mellifera*.

Harpur et al. (2012) found that honey bees exhibit unusually high levels of genetic diversity, as domestic bees are more genetically diverse than wild populations in Europe. This high level of diversity is believed to be maintained by the crossing of queens from diverse locations to produce more diverse hives. De la Rúa et al. (2013) pointed out that backcrossing with the local populations may be reducing the overall variation in the global honey bee population. Interbreeding between wild and domestics may be reducing the number of individuals with local adaptations that may be advantageous in a changing environment.

Relative to domestic mammal species, it is far more difficult to identify domestic insects in the archaeological record. As a result, investigations into the early process of

domestication will have to rely upon genetic and morphological insights derived from museum specimens of silkworm and honey bees (e.g. Cridland et al. 2018).

1.8 The Biological Architecture of Domestication

Given its importance for our understanding of evolution, domestication has also been extensively studied by experimental biologists and geneticists. These studies have focused on characterising the nature of the specific biological changes underlying the differences between domestic and wild species, as well as the interspecific similarities among domestic animals (known as the "domestication syndrome;" Figure 1.2). Palaeogenomics has an enormous potential to address many questions regarding the biological underpinning of domestication by, for example, providing time-series data that can help detect artificial selection in the genome. Here we review how studies have, and will continue to, leverage the power of palaeogenomics to answer fundamental questions in domestication.

1.8.1 Theories and Experiments

The evolutionary basis of animal domestication is one of the most enduring questions in evolutionary biology. Shortly after Charles Darwin (1859) published the theory of evolution by natural selection, he turned his attention to the study of domestication. Darwin's (1868) seminal work on the topic, 'The Variation of Animals and Plants under Domestication' examined in extensive detail the remarkable phenotypic similarity shown by a diverse range of domestic animals. Darwin's observations on the role of selection during domestication distinguished between two phases of artificial selection; termed 'unconscious' and 'methodical.' Darwin argued that the initial phase of domestication would have involved people unknowingly selecting for domestication traits by, for example, choosing the more productive cattle to breed and the less productive to eat (Darwin 1868). Over time, these unconscious selective pressures formed the many regional landraces of animals. Subsequently, he theorised, people began practicing conscious or methodical selection, in which animals were bred with a specific phenotypic outcome in mind—a view largely informed by the animal husbandry

practices of the 19th century (Marshall et al. 2014). This perspective on animal domestication placed central focus on the role of human intent in the development of domestication traits, and reproductive isolation from wild populations to preserve them.

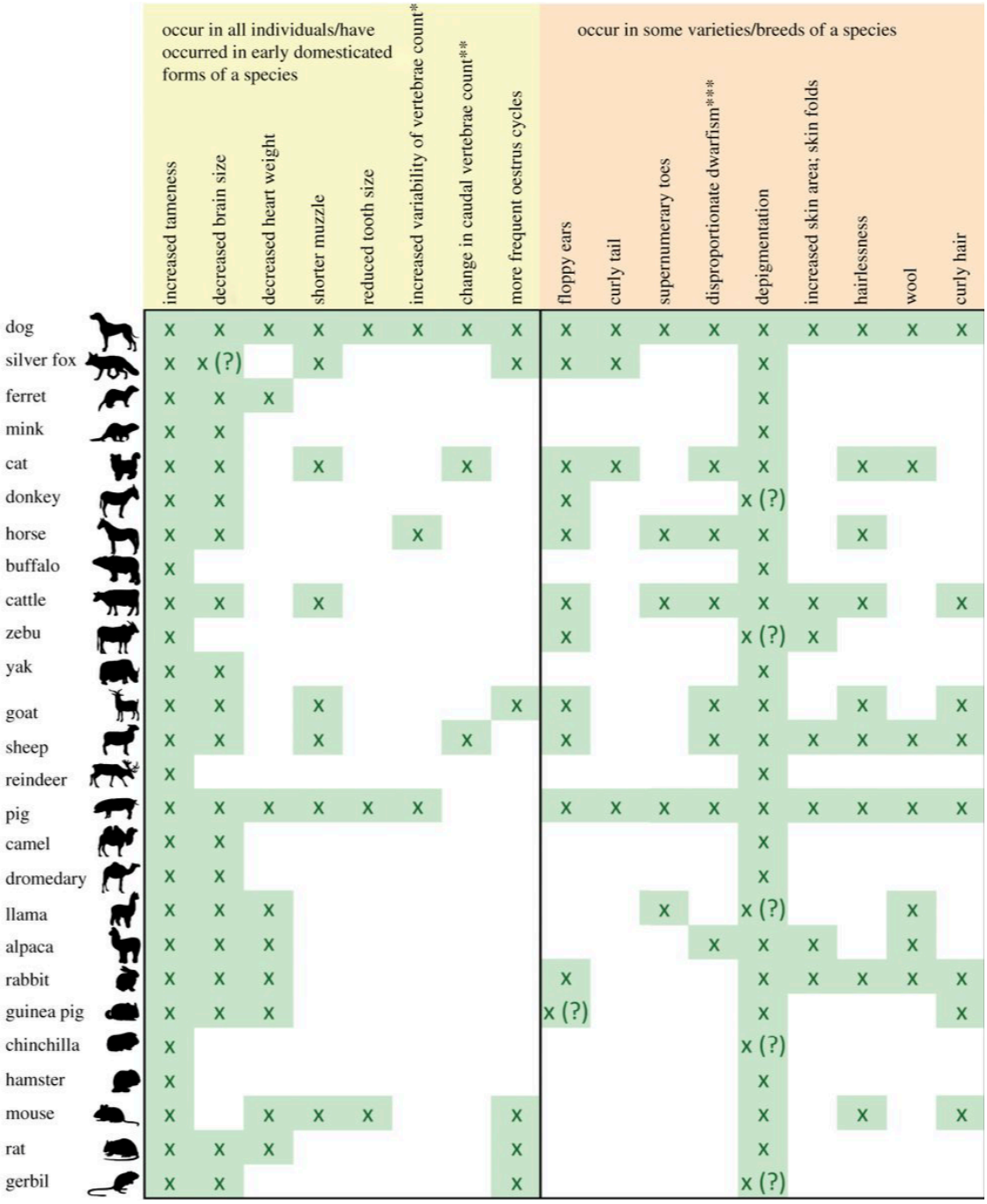


Figure 1.2 Multiple traits, commonly referred to as 'domestication syndrome' and their occurrence in different mammalian species. Adapted after (Sánchez-Villagra et al. 2016)

These ideas were further developed by Francis Galton (1865, 1883), based on ethnographic observation of pet keeping in hunter-gatherer communities. Galton argued that the domestication of animals was a direct consequence of the human desire to capture and tame wild animals. All animals would be exposed to this process, but only those with a natural predisposition towards domestication would be permanently tamed. These anthropocentric views of domestication proved very influential, placing human intent at the heart of many contemporary definitions of domestication (Bökönyi 1989; Ducos 1989; Clutton-Brock 1994).

Darwin's (1868) study of domestication identified a series of behavioural, physiological, and morphological traits shared by domestic animals, but not by their wild progenitors. These shared traits subsequently became known as the "domestication syndrome" (Hammer 1984). Amongst domestic animals, these traits are now considered to include increased docility and tameness, reduction in body mass and brain size, novel coat colours and patterns, altered tails and floppy ears, smaller teeth and shorter snouts, prolonged physical and behavioural neoteny, more frequent and non-seasonal reproductive cycles, as well as changes in hormonal and neurotransmitter expression (Darwin 1868; Hammer 1984; Wilkins et al. 2014). The prevalence of these traits amongst domestic animals, including birds, fish, and mammals suggest that domestic animals respond similarly to artificial selection. The subsequent domestication syndrome (Figure 1.2), has been hypothesised to result from a disruption in developmental process of the neural crest (Wilkins et al. 2014; Sánchez-Villagra et al. 2016).

Experimental studies of animal domestication have played a key role in our understanding of how the domestication syndrome develops. The earliest experiments involved domestication of the brown rat (*Rattus norvegicus*) (King and Donaldson 1929; Castle 1947), however, the most informative experiments involved the silver fox—a melanistic form of the red fox (*Vulpes vulpes*). Beginning in 1959 at the Institute of Cytology and Genetics in Novosibirsk, Dmitri Belyaev established an experimental breeding program which selectively bred silver foxes, brown rats, and European mink (*Mustela lutreola*) for tameness (Belyaev 1969; Trut et al. 2004, 2009). Captive

silver foxes were sourced from fur farms, where they had been selectively bred for their unique coat pigmentation (Belyaev 1969). Their level of aggression towards humans was tested by attempting to hand feed, stroke or handle the foxes, and those which exhibited the least aggressive response were chosen for subsequent breeding (Trut et al. 2004). The selective pressures in each generation were very high, with only 3% of males and 8–10% of females permitted to breed (Trut et al. 2004). Within 30 generations, almost half of the experimental foxes had behavioural relationships with humans that were analogous to domestic dogs. Interestingly, they also exhibited classic symptoms of the domestication syndrome—changes in coat colour and snout length, floppy ears, and altered developmental timing (Trut et al. 2004).

Whole-genome sequences for tame, aggressive, and conventional foxes raised under these experimental conditions have recently been published (Kukekova et al. 2018). Analysis of this data identified more than 100 regions showing signatures of selection in one or more of the experimental populations, and the *SorCS1* gene was identified as a strong candidate gene for tame behaviour.

1.8.2 Genetic changes during domestication

Many researchers have investigated the genetic basis for the phenotypic and behavioural changes seen in the domestication syndrome (Dobney and Larson 2006; Trut et al. 2009; Albert et al. 2009; Driscoll et al. 2009; Axelsson et al. 2013; Jensen 2014; Wilkins et al. 2014; Carneiro et al. 2014). With regard to plant domestication, good progress has been made in identifying genes linked to domestication and crop improvement (reviewed in Doebley et al. 2006; Olsen and Wendel 2013), however, the identification of similar genes linked to animal domestication has been more elusive. Increasingly, research has suggested that the phenotypic diversity found in domestic animal populations is based on complex genetic architectures involving hundreds of genes and regulatory regions, each with small effect sizes (Larson et al. 2014; Wilkins et al. 2014; Carneiro et al. 2014).

Evidence drawn from across the range of domestic taxa, and phenotypic traits, suggest complex pleiotropic, polygenic, and epistatic effects (Reissmann and Ludwig 2013; Wilkins et al. 2014; Wright 2015). For example, pleiotropy—in which single genes affect multiple discrete phenotypic traits—has been putatively identified in behavioural, morphological, life-history and sexual ornament traits in domestic chickens (*Gallus gallus*) (Wright et al. 2010; Johnsson et al. 2012). Polygenic traits—in which single phenotypic traits are controlled by multiple genes—is most clearly evident in pigmentation traits for hair, skin, and eyes, where more than 125 causal genes have been identified in domestic mice (*Mus musculus*) (Bennett and Lamoreux 2003). Epistasis—in which the expression of a genetic variant is dependent on the effect of one or more variants in modifier regions (Cordell 2002)—has been putatively identified in more than a dozen epistatic pairs effecting tameness, flight and startle responses, body weight and other traits, in experimentally domesticated brown rats (Albert et al. 2009). Among domesticated crops, where the architecture of domestication traits is better understood, epistasis is thought to play a key role in phenotypic expression (reviewed in Doust et al. 2014).

The identification of genes involved in animal domestication and their mapping to complex traits has been achieved via two main approaches: (i) quantitative trait loci (QTL) mapping (reviewed in Mackay et al. 2009); and (ii) genome-wide association studies (GWAS) (reviewed in McCarthy et al. 2008). Both techniques have been critical in identifying candidate genes associated with traits that differentiate domestic populations (Goddard and Hayes 2009). This work has been aided by the development of online databases, cataloguing known gene associations. The Animal QTLdb now contains more than 57,000 trait mappings (Hu et al. 2007, 2016), and the Online Mendelian Inheritance in Animals (OMIA) database (Nicholas 2003; Lenffer et al. 2006) catalogues thousands of monogenic traits in domestic animals. QTL mapping and GWAS studies, however, often focus on traits that are important for food production rather than traits that differentiate wild and domestic populations.

Population genomic studies that focus on identifying the signatures of selection in genome-wide sequence data from wild and domestic populations have allowed for

more candidate genes involved in domestication to be identified. This approach has recently been used to identify putative selection in polygenic loci involved in brain and neuronal development traits in domestic rabbits (*Oryctolagus cuniculus*) (Carneiro et al. 2014), and digestion and nervous system development traits in dogs (Axelsson et al. 2013). Genome-wide sequencing data has also been used to test the hypothesis of gene-loss as a driver of rapid evolutionary change (Olson 1999), which has been discounted as an important process in the domestication of dogs (Freedman et al. 2016), chickens (Rubin et al. 2010), pigs (Rubin et al. 2012) and rabbits (Carneiro et al. 2014).

1.8.3 Temporal pattern of genetic and morphological changes

Identifying the genetic basis of animal domestication based solely on modern DNA, however, can be problematic (Larson and Burger 2013). In order to Identify the genetic basis of traits that are associated with early stages of the domestication process it is necessary to dissociate these from changes that happened during later stage of the process (Vigne 2011). This can be problematic as domestic animals bear little direct resemblance to their early forbears, due to thousands of years of artificial selection, divergent environmental conditions, and introgression with populations unrelated to the initial domestication. More recently, this has been further complicated by intensive breeding practices which have made reconstructing the early stages of domestication much harder (Larson and Burger 2013).

Recently, a genome-wide selection scan identified a putative domestication locus in the thyroid stimulating hormone receptor (*TSHR*) gene in domestic chickens (Rubin et al. 2010). Thyroid hormone metabolism has previously been suggested as a key factor in animal domestication (Crockford 2002; Dobney and Larson 2006), and the *TSHR* gene has been shown to play an important role in metabolic regulation and control of seasonal reproduction in birds (Nakao et al. 2008) and mammals (Hanon et al. 2008). Single nucleotide polymorphisms (SNPs) from the *TSHR* sweep region, including a candidate causal missense mutation, were genotyped in hundreds of domestic chickens, from dozens of geographically dispersed populations. The missense mutation was found to be almost completely fixed in the domestic population, with an allele frequency of

0.987 (Rubin et al. 2010). The same SNPs were typed in more than fifty red junglefowl (*Gallus gallus*)—thought to be the primary wild ancestor of the domestic chicken (Eriksson et al. 2008). The missense mutation was found with an allele frequency of 0.35, which the authors attributed to introgression from domestic chickens into zoo populations of red junglefowl (Rubin et al. 2010).

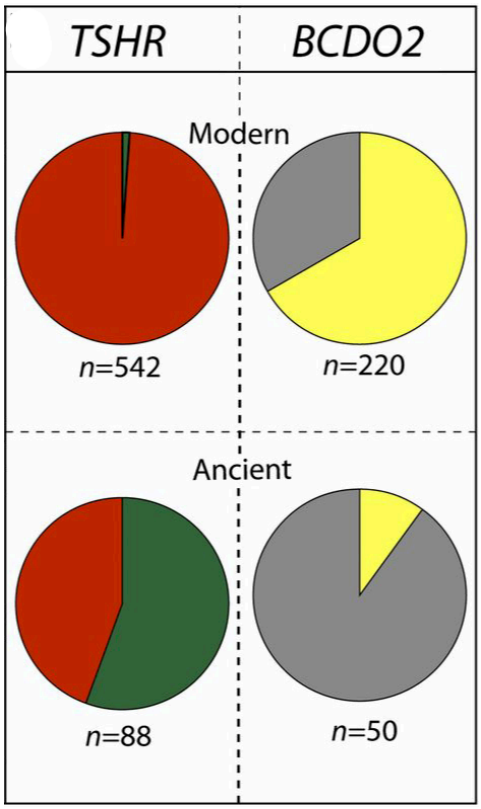


Figure 1.3 Pie charts representing the allele frequency of variants at *TSHR* and *BCDO2* (affecting skin colour) in ancient chickens. This figure demonstrates that the variants at those genes, which are thought to influence traits in modern domestic chickens, were not found at high frequency in ancient chickens. Their rise in frequency is thought to be associated with breeding during the middle age, but not with the domestication process (Loog et al. 2017). Adapted after (Girdland Flink et al. 2014).

The identification of the *TSHR* gene as a domestication locus relies on the assumption that selective pressure on the allele that is now almost fixed in domestic chickens took place during the early stage of the domestication process. This assumption was directly tested when another research group recovered aDNA from 80 domestic chickens, from a dozen sites across Europe, with a temporal range of approximately 2,000 years (Girdland Flink et al. 2014). The authors were able to genotype the SNP from the *TSHR*

sweep region in 44 ancient samples. The missense mutation was found with an allele frequency of just 0.432, and only 18% of the samples were homozygous for the derived allele (Girdland Flink et al. 2014). This analysis clearly demonstrated that the fixation of the *TSHR* missense mutation was associated with later trait improvements rather than the initial domestication process (Figure 1.3).

In a subsequent study, selection on the *TSHR* locus was revisited, with an expanded ancient DNA dataset, and the application of a novel Bayesian statistical framework for modelling the strength of selection over time (Loog et al. 2017). The authors concluded that selection on the derived allele began around AD 920, coinciding with Medieval religious dietary reforms, which may have increased demand for both chicken and eggs (Loog et al. 2017). These findings are supported by zooarchaeological assemblages from England and Germany, spanning the medieval period, which show an increase in both the overall frequency of chickens and the relative proportion of adult hens—interpreted as sign of increased egg production (Serjeantson 2006; Sykes 2007; Holmes 2014). Functional genetic investigation of the pleiotropic effects of the *TSHR* derived allele in chickens has shown that it is associated with increased egg production (Karlsson et al. 2016), decreased aggression and less fearful behaviours (Karlsson et al. 2015)—consistent with artificial selection for intensified egg production during the Medieval period. Evidently, *TSHR* has played an important role in the evolutionary history of domestic chickens, however, its identification as a domestication locus is erroneous and it can better be described as an improvement trait.

Similar cases, involving misidentified domestication genes, have been reported in domestic dogs and wheat. In the latter, a derived allele, fixed in modern populations of wheat, was identified as a putative domestication locus in the *NAM-B1* gene (Uauy et al. 2006). Ancient DNA recovered from herbarium seeds, however, established that the ancestral allele was still commonly found in cultivated populations as recently as 150 years ago (Asplund et al. 2010). Similarly, a recent genome-wide study of dogs, demonstrated that most modern populations harbour a high number of copies of the amylase alpha 2B (*AMY2B*) gene (Axelsson et al. 2013). This high copy number is almost fixed in modern dogs (Freedman et al. 2014) and allows them to better process starch

(Axelsson et al. 2013). Ancient DNA studies, however, showed that these genetic variations only started to occur following the onset of farming, more than 7,000 years after dogs were domesticated (Arendt et al. 2016; Ollivier et al. 2016). More recent aDNA analysis further suggests that selection on *AMY2B* copy-number variation did not begin until well after the advent of agriculture (Botigué et al. 2017).

These examples clearly demonstrate the importance of ancient DNA in verifying the timing of selection during the domestication process, and the pitfalls inherent in inference based solely on modern DNA. As the number of aDNA studies increases, the geographic range and temporal resolution of these datasets will allow ever more detailed studies to investigate which loci were under selection during early phases of domestication.

1.8.4 Genes as Domestic Markers

Genetic markers can potentially be used to evaluate whether animal remains belong to a wild or domestic individual, however, the use of genetics is controversial due to the disputed importance of genetic changes during the early phases of domestication (Zeder 2012a; Vigne 2015). These controversies stem from a general lack of consensus regarding the definition of domestication, particularly one which unifies both plants and animals. This lack of clear definition has recently been identified as one of the key challenges in domestication research (Zeder 2015). There are, however, some clear examples of genetic (and phenotypic) changes that are highly diagnostic of the domestication status of an animal. For example, multiple non-synonymous (protein changing) mutations have been found in the melanocortin 1 receptor (*MC1R*) gene of pigs which leads to a black coat colour, or black spotted colour, and loss of their wild-type camouflage coat pattern (Fang et al. 2009). At least three independent mutations, resulting in similar phenotypes exists in pigs, one in European pigs, one in East Asian pigs and one in Hawaiian feral pigs (introduced during the Polynesian expansion; Figure 1.4) (Linderholm et al. 2016). In modern European domestic pigs, this dominant allele, which leads to loss of camouflage, is found at very high frequency, while it is almost absent

from wild populations (Koutsogiannouli et al. 2010; Frantz et al. 2013a). This suggests a strong negative selection in wild boars.

This European dominant black allele was recently found in four ~6,500 years old pig remains from the site of Ertebølle (Mesolithic of northern Germany) (Krause-Kyora et al. 2013). These animals also had a mtDNA haplogroup originating in Near Eastern domestic populations, and geometric morphometric (GMM) analysis revealed they had molars with domestic shape characteristics and pathologies (Krause-Kyora et al. 2013). Their domestic status, however, conflicted with the cultural context in which they were found—Mesolithic hunter-gatherer rather than Neolithic farmers.

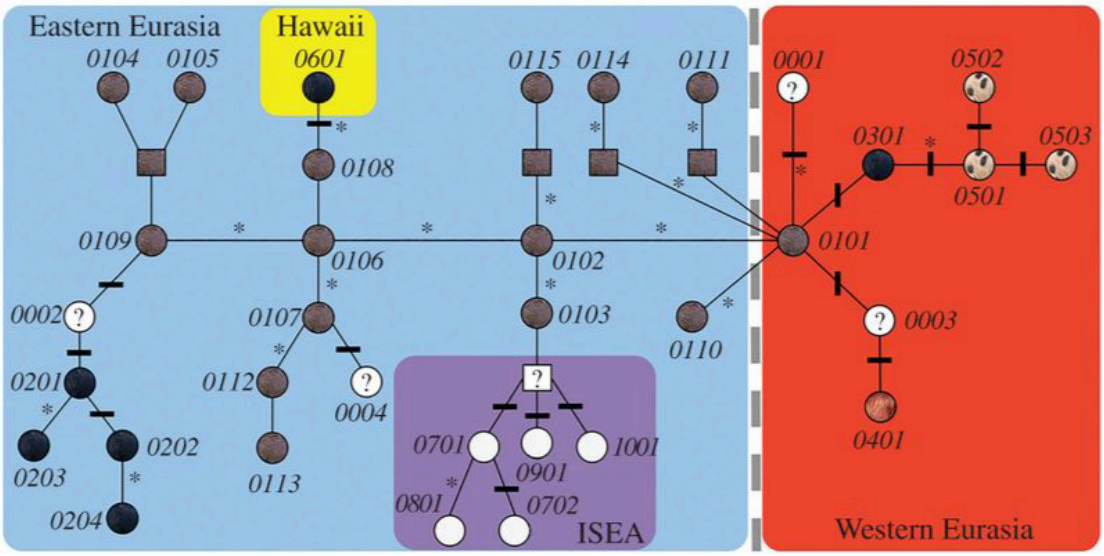


Figure 1.4 Haplotype network for the MC1R gene coding region. This figure demonstrates the existence of three lineages of black pigs, in Hawaii, Europe and East Asia. Adapted after (Linderholm et al. 2016).

This sparked a controversy and led to several published replies (Evin et al. 2014; Rowley-Conwy and Zeder 2014a, b). The principle critique centred on the lack of evidence that humans at Ertebølle had a special relationship with these animals, distinct from that of wild boar (Rowley-Conwy and Zeder 2014a). The authors argued that domestication involves more than the phenotypic expression of genetic traits and requires a mutualistic relationship between the domestic and the domesticator. Therefore, even unambiguously domestic animals—with the complete set of behavioural and

phenotypic traits—identified in this undifferentiated context would shed no light on the process of domestication, or the adoption of agriculture in the region (Rowley-Conwy and Zeder 2014b). All together this highlights the fact that biological markers, even highly discriminative as those described above, cannot on their own provide the sole basis for a definition of domestication.

1.8.5 Introgression in Domestication

Animal domestication is often thought to be defined, not solely by genetic and phenotypic characteristics, but also by population processes such as a strong bottlenecks, reproductive isolation from wild populations, and directed breeding (Marshall et al. 2014). More recently, modern and ancient genomic datasets have revealed that these conditions were much less common than previously thought—revealing complex and varied patterns of introgression between wild and domestic pigs (Frantz et al. 2015), goats (Daly et al. 2018), cattle (Park et al. 2015), horses (Schubert et al. 2014), dromedary camels (Almathen et al. 2016), cats (Ottoni et al. 2017), and many other species (Marshall et al. 2014). As early farmers spread outwards from the major centres of domestication, the domestic animals that accompanied them frequently interbred with wild populations encountered along their routes of dispersal. Successive waves of gene flow over thousands of years have resulted in modern genomes which are complex palimpsests, containing traces of many different ancestral populations. For researchers that use DNA to re-trace the temporal and geographic origin of domestic populations, introgression can be a double-edged sword. Patterns of admixture have been useful in untangling routes of animal dispersal and human migration (Larson et al. 2007b), but they have also led to misleading interpretations, based on limited mitochondrial datasets, for multiple independent domestications of cattle (Hanotte et al. 2002), pigs (Larson et al. 2005), goats (Luikart et al. 2001), sheep (Pedrosa et al. 2005) and horses (Vilà et al. 2001).

The specific patterns of introgression vary between different species, depending on the way domestic populations were managed, and the variety of wild populations which were encountered. For example, widespread introgression in European pigs has been

attributed to loose herd management practices; in which free ranging domestic pigs interbred freely with neighbouring wild boar populations, whose offspring were adopted into the loosely managed herds (Ottoni et al. 2013; Frantz et al. 2015). These patterns of introgression are highly asymmetric in pigs, with wild boars receiving little to no gene flow from domestic populations (Frantz et al. 2015). In general, the directionality of admixture is biased towards gene flow from local populations into migrant groups, especially with increasing distance from the source of the migration (Currat et al. 2008). Notable exceptions do occur, however, such as the *K*-locus variant introgressed from dogs into wolves (Schweizer et al. 2018) and *MITF* gene variants introgressed from cattle into yaks (Wu et al. 2018).

African cattle, which early genetic evidence suggested might have been independently domesticated (Hanotte et al. 2002), are now better explained by introgression between Near Eastern domestic cattle (*Bos taurus*), wild North African aurochs (*Bos primigenius africanus*) and successive waves of Asian domestic zebus (*Bos indicus*) (Mwai et al. 2015; Brass 2018; Pitt et al. 2018). In domestic chickens, the now ubiquitous yellow leg trait was acquired via introgression from the wild grey junglefowl (*Gallus sonneratii*) (Eriksson et al. 2008). For some species, introgression with wild populations continues to be an active process—particularly among reindeer (*Rangifer tarandus*) (Røed et al. 2008) and honey bees (*Apis mellifera*) (Harpur et al. 2012), which exhibit very high levels of genetic diversity.

An important, but limited, approach to investigating these complex histories is to use large genome-wide datasets to characterise the patterns of diversity and admixture seen in modern domestic populations—like cattle (Gibbs et al. 2009; Decker et al. 2014), sheep (Kijas et al. 2012), goats (Wang et al. 2016; Brito et al. 2017), pigs (Ai et al. 2013), horses (McCue et al. 2012; Petersen et al. 2013; Schaefer et al. 2017), chickens (Muir et al. 2008; Stainton et al. 2017), dogs (vonHoldt et al. 2010; Shannon et al. 2015) and mice (Yang et al. 2011; Staubach et al. 2012). These large modern datasets benefit from the relative ease of sampling and low cost of data generation, compared to aDNA. The recent development of novel computational methods using phased haplotypes (Lawson et al. 2012; Hellenthal et al. 2014) have increased the precision with which the timing,

direction and fraction of admixture can be resolved in these high quality modern datasets.

The inferences which can be made from modern DNA alone, however, are limited by the use of modern genetic variation as a proxy for ancestral populations. Modern DNA can be blind to population replacement (e.g. Haak et al. 2015; Gaunitz et al. 2018), because the extirpated populations make little contribution to modern variation. Nor can modern DNA effectively detect or measure admixture from unsampled extinct species (e.g. Prüfer et al. 2014; Park et al. 2015), because the ancestral states of both species are unknown—although statistical methods have been developed to infer admixture from unsampled archaic populations (e.g. Plagnol and Wall 2006; Vernot and Akey 2014). The best approaches are those that combine both ancient and modern DNA with explicit testable models of evolutionarily processes (Gerbault et al. 2014). The recent development of novel Bayesian techniques for modelling serially sampled DNA hold particular promise to reveal important insights into the evolutionary process of domestication (Schraiber et al. 2016; Ferrer-Admetlla et al. 2016; Loog et al. 2017).

1.9 Future perspectives

1.9.1 Ancient Epigenomes

The role of epigenetics in the domestication process, and in regulating domestic phenotypes, is a promising area of new research. For example, researchers working on the experimental domestication of the silver fox have suggested that observed differences in hormonal expression, associated with the domestication syndrome, may be linked to epigenetic modifications (Trut et al. 2009). A recent study comparing methylation patterns between dogs and wolves found 68 significantly differentially methylated sites across the two species, which included sites linked to the *GABRB1* and *SLC17A8* neurotransmitter genes, associated with a range of cognitive functions (Janowitz Koch et al. 2016). The role of epigenetics in a wide range of livestock phenotypes has also recently garnered a lot of attention (Feeney et al. 2014; Ibeagha-Awemu and Zhao 2015; Triantaphyllopoulos et al. 2016).

As our understanding of epigenetics improves, the ability to retrieve epigenetic information from ancient DNA will become increasingly important (reviewed in Hanghøj and Orlando 2018). Technical advances have recently made the recovery of ancient methylation maps possible (Briggs et al. 2010), which has resulted in the publication of the first genome-wide methylation maps for an ancient human (Pedersen et al. 2014), a Neandertal and a Denisovan (Gokhman et al. 2014). Specialist computational tools for performing these analyses have also recently become available (Hanghøj et al. 2016). Presently, an equivalent ancient genome-wide methylation map has yet to be produced for domestic animals, however, as the number of ancient whole genome sequences increase it is only a matter of time before these become available.

1.9.2 Technical Advances

On the technical front, palaeogenomics has benefited greatly from the development of increasingly cheaper and higher throughput sequencing platforms. As development of these machines continues apace, we can expect the cost of DNA sequencing to continue to reduce. In some experimental designs the limiting factor is no longer the cost of sequencing, but the costs of reagents and skilled labour for sample preparation (Rohland and Reich 2012). Protocols and laboratory equipment for automated library preparation, using liquid handling robots, are already available (Farias-Hesson et al. 2010; Lundin et al. 2010), and such approaches will likely become more commonplace in the future. As the cost of sequencing and sample preparation continues to drop, the number of samples and range of taxa which can be sequenced will increase concomitantly. Domestic animals are well represented in many archaeological sites, providing the potential for aDNA studies with fine-grained transects through time.

As palaeogenomics studies scale up, increasingly sophisticated population genetic models will be necessary to interpret the process of animal domestication (Gerbault et al. 2014). Current methods for inferring patterns of admixture will need to be extended and improved to deal with more complicated models and larger datasets. Model-based clustering techniques, like *STRUCTURE* (Pritchard et al. 2000) and *ADMIXTURE* (Alexander et al. 2009), are very popular but widely over-interpreted (Lawson et al.

2018). Graph fitting approaches, like *TreeMix* (Pickrell and Pritchard 2012) and *MixMapper* (Lipson et al. 2013), are useful for inferring models of admixture, but lack a formal statistical test of fit (Patterson et al. 2012). Formal models of admixture can be tested with *f*-statistics (Reich et al. 2009; Patterson et al. 2012) and *D*-statistics (Green et al. 2010; Durand et al. 2011), but these methods cannot resolve complex admixture topologies. Haplotype based methods (Lawson et al. 2012; Hellenthal et al. 2014) work well on high quality data but are not suitable for low-coverage ancient data. Bayesian techniques, like *admixturegraph* (Leppälä et al. 2017), can test goodness of fit between models using Bayes factors, but computing these factors is computationally expensive, making automated model exploration very slow. As datasets continue to increase in size, the main constraint on genome analysis will be scaling computation to contend with the growth in sequence data (Muir et al. 2016).

1.9.3 Novel Substrates for aDNA

Palaeogenomics is branching out into the recovery of aDNA from a range of novel substrates (reviewed in Green and Speller 2017). For example, the recent demonstration that aDNA can be successfully retrieved from historic parchments has opened up a whole new avenue for the study of domestic animals (Teasdale et al. 2015). Large numbers of historical parchments exist in archival and private collections across Europe. These parchments represent an exceptionally well dated source of aDNA for reconstructing the evolutionary history of regional landraces of sheep, goat, and cattle (Teasdale et al. 2015). Ancient coprolites from domestic animals have also recently been shown to be a suitable substrate for the recovery of aDNA. Using a combination of microscopy and aDNA sequencing, a recent study of domestic dog coprolites was able to establish the major diet components of ancient Polynesian dogs (Wood et al. 2016). Additionally, ancient latrines have been shown to contain retrievable quantities of parasite aDNA, the host specificity of which can be used to infer the presence of domestic animal species (Søe et al. 2018).

Calcified dental plaque, known as dental calculus, has also recently been established as an important new substrate for aDNA recovery (Adler et al. 2013; Warinner et al. 2014,

2015; Weyrich et al. 2015). Archaeological studies of dental calculus in domestic animals have a long history; the earliest of which used light microscopy to study phytoliths trapped in dental calculus from cattle, sheep, and horse teeth (Armitage 1975). Other early studies identified a broad range of organic substances in dental calculus (Dobney and Brothwell 1986), and developed a system for quantifying dental calculus in human, cattle and sheep teeth (Dobney and Brothwell 1987). More recently, palaeogenomic studies of dental calculus have focused on changes in human health and diet. For example, a recent study used aDNA from dental calculus to establish that Mesolithic foragers in the Balkans were consuming domesticated plant foods (Cristiani et al. 2016). As palaeogenomics broadens its focus away from human centred studies, similar studies of animal diet and oral health will no doubt be applied to domestic taxa and their wild progenitors.

Environmental and sediment DNA are also showing strong potential for reconstructing the movement of domestic animals and their environmental impacts. A recent study used DNA metabarcoding of Alpine lake sediments to build a high-resolution picture of agricultural land use since the Neolithic (Giguet-Covex et al. 2014). The authors were able to identify ancient sediment DNA from cattle, goats, sheep, horses, and chickens, and to correlate their abundance with changes in plant cover and erosion. The potential of environmental DNA, however, is moderated by the risk of vertical DNA movement through sediment stratigraphy. For example, one study identified sheep DNA in a New Zealand cave site from layers which pre-dated European contact, demonstrating that DNA leaching can be problematic under some soil conditions (Haile et al. 2007). The inability to directly date environmental DNA from sediments which lack macrofossils is also a significant concern, and has caused some to question the identification of the earliest domestic wheat in Britain; from an 8,000 year-old layer of a sediment core (Smith et al. 2015a, b; Bennett 2015).

1.10 Conclusion

The future of palaeogenomics and its application to the study of animal domestication looks bright. Ten years ago, the retrieval of a single gene locus from few ancient samples

was cause for celebration. Now, studies involving genome-wide data from dozens (Haak et al. 2015; Fu et al. 2016; Lazaridis et al. 2016) or even hundreds (Mathieson et al. 2015; Lipson et al. 2017) of ancient samples is increasingly commonplace. So far, large palaeogenomic studies have favoured retrieval of ancient human DNA, but similarly sized studies of domestic animals are certainly on the horizon. As our understanding of aDNA preservation (Hansen et al. 2017) and decay kinetics (Kistler et al. 2017) improves, more informed choice of skeletal elements and sampling locations will also permit the retrieval of aDNA from older time depths and warmer climates. We anticipate that the trend will be towards larger studies with many more samples, and much older and finer temporal resolution.

1.11 References

Achilli A, Olivieri A, Soares P, et al (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc Natl Acad Sci U S A 109:2449–2454. doi: 10.1073/pnas.1111637109

Adler CJ, Dobney K, Weyrich LS, et al (2013) Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 45:ng.2536. doi: 10.1038/ng.2536

Adler CJ, Haak W, Donlon D, Cooper A (2011) Survival and recovery of DNA from ancient teeth and bones. J Archaeol Sci 38:956–964. doi: 10.1016/j.jas.2010.11.010

Ai H, Fang X, Yang B, et al (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet 47:217–225. doi: 10.1038/ng.3199

Ai H, Huang L, Ren J (2013) Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. PLoS One 8:e56001. doi: 10.1371/journal.pone.0056001

Albert FW, Carlborg Ö, Plyusnina I, et al (2009) Genetic Architecture of Tameness in a Rat Model of Animal Domestication. Genetics 182:541–554. doi: 10.1534/genetics.109.102186

Alberto FJ, Boyer F, Orozco-terWengel P, et al (2018) Convergent genomic signatures of domestication in sheep and goats. Nat Commun 9:813. doi: 10.1038/s41467-018-03206-y

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. doi: 10.1101/gr.094052.109

Allentoft ME, Collins M, Harker D, et al (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci 279:4724–4733. doi: 10.1098/rspb.2012.1745

Allentoft ME, Sikora M, Sjögren K-G, et al (2015) Population genomics of Bronze Age Eurasia. Nature 522:167–172. doi: 10.1038/nature14507

Almathen F, Charruau P, Mohandesan E, et al (2016) Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. Proc Natl Acad Sci U S A 113:6707–6712. doi: 10.1073/pnas.1519508113

Ameen C, Hulme-Beaman A, Evin A, et al (2017) A landmark-based approach for assessing the reliability of mandibular tooth crowding as a marker of dog domestication. J Archaeol Sci 85:41–50. doi: 10.1016/j.jas.2017.06.014

Arendt M, Cairns KM, Ballard JWO, et al (2016) Diet adaptation in dog reflects spread of prehistoric agriculture. Heredity 117:301–306. doi: 10.1038/hdy.2016.48

Armitage PL (1975) The extraction and identification of opal phytoliths from the teeth of ungulates. J Archaeol Sci 2:187–197. doi: 10.1016/0305-4403(75)90056-4

Asplund L, Hagenblad J, Leino MW (2010) Re-evaluating the history of the wheat domestication gene NAM-B1 using historical plant material. J Archaeol Sci 37:2303–2307. doi: 10.1016/j.jas.2010.04.003

Axelsson E, Ratnakumar A, Arendt M-L, et al (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360–364. doi: 10.1038/nature11837

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729–744. doi: 10.1046/j.1365-294X.2003.02063.x

Barbato M, Hailer F, Orozco-terWengel P, et al (2017) Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Sci Rep 7:7623. doi:

10.1038/s41598-017-07382-7

Barreta J, Gutiérrez-Gil B, Iñiguez V, et al (2013) Analysis of mitochondrial DNA in Bolivian llama, alpaca and vicuna populations: A contribution to the phylogeny of the South American camelids. Anim Genet 44:158–168. doi: 10.1111/j.1365-2052.2012.02376.x

Beja-Pereira A, Caramelli D, Lalueza-Fox C, et al (2006) The origin of European cattle: Evidence from modern and ancient DNA. Proc Natl Acad Sci U S A 103:8113–8118. doi: 10.1073/pnas.0509210103

Belyaev DK (1969) Domestication of animals. Science Journal 5:47–52

Bennett DC, Lamoreux ML (2003) The Color Loci of Mice – A Genetic Century. Pigment Cell Res 16:333–344. doi: 10.1034/j.1600-0749.2003.00067.x

Bennett KD (2015) Comment on "Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago." Science 349:247–247. doi: 10.1126/science.aab1886

Bentley DR, Balasubramanian S, Swerdlow HP, et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59. doi: 10.1038/nature07517

Berke H (1995) Der Mensch und seine Haustiere Die Geschichte einer jahrtausendealten Beziehung

Bisch-Knaden S, Daimon T, Shimada T, et al (2014) Anatomical and functional analysis of domestication effects on the olfactory system of the silkmoth Bombyx mori. Proceedings of the Royal Society of London B: Biological Sciences 281:20132582. doi: 10.1098/rspb.2013.2582

Bocquet-Appel J-P (2011) When the World's Population Took Off: The Springboard of the Neolithic Demographic Transition. Science 333:560–561. doi: 10.1126/science.1208880

Bollongino R, Burger J, Powell A, et al (2012) Modern Taurine Cattle Descended from Small Number of Near-Eastern Founders. Mol Biol Evol 29:2101–2104. doi: 10.1093/molbev/mss092

Bosse M, Megens H-J, Frantz LAF, et al (2014a) Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat Commun 5:4392. doi: 10.1038/ncomms5392

Bosse M, Megens H-J, Madsen O, et al (2014b) Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. Mol Ecol 23:4089–4102. doi: 10.1111/mec.12807

Botigué LR, Song S, Scheu A, et al (2017) Ancient European dog genomes reveal continuity since the Early Neolithic. Nat Commun 8:16082. doi: 10.1038/ncomms16082

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. Proc Natl Acad Sci U S A 93:5131–5135. doi: 10.1073/pnas.93.10.5131

Brass M (2018) Early North African Cattle Domestication and Its Ecological Setting: A Reassessment. Journal of World Prehistory 31:81–115. doi: 10.1007/s10963-017-9112-9

Briggs AW, Stenzel U, Meyer M, et al (2010) Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res 38:e87. doi: 10.1093/nar/gkp1163

Brito LF, Kijas JW, Ventura RV, et al (2017) Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genomics 18:229. doi: 10.1186/s12864-017-3610-0

Broushaki F, Thomas MG, Link V, et al (2016) Early Neolithic genomes from the eastern Fertile Crescent. Science aaf7943. doi: 10.1126/science.aaf7943

Bruford MW, Townsend SJ (2006) Mitochondrial DNA diversity in modern sheep: Implications for domestication. In: Documenting Domestication: New Genetic and Archaeological Paradigms. University of California Press, pp 306–316

Bökönyi S (1989) Definitions of animal domestication. In: Clutton-Brock J (ed) The Walking larder: patterns of domestication, pastoralism, and predation. Unwin Hyman, London, pp 22–27

Carneiro M, Afonso S, Geraldes A, et al (2011) The Genetic Structure of Domestic Rabbits. Mol Biol Evol 28:1801–1816. doi: 10.1093/molbev/msr003

Carneiro M, Piorno V, Rubin C-J, et al (2015) Candidate genes underlying heritable differences in reproductive seasonality between wild and domestic rabbits. Anim Genet 46:418–425. doi: 10.1111/age.12299

Carneiro M, Rubin C-J, Di Palma F, et al (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345:1074–1079. doi: 10.1126/science.1253714

Carpenter ML, Buenrostro JD, Valdiosera C, et al (2013) Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. Am J Hum Genet 93:852–864. doi: 10.1016/j.ajhg.2013.10.002

Castle WE (1947) The domestication of the rat. Proc Natl Acad Sci U S A 33:109–117. doi: 10.1073/pnas.33.5.109

Chen N, Cai Y, Chen Q, et al (2018) Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun 9:2337. doi: 10.1038/s41467-018-04737-0

Chen S, Lin B-Z, Baig M, et al (2010) Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic. Mol Biol Evol 27:1–6. doi: 10.1093/molbev/msp213

Chessa B, Pereira F, Arnaud F, et al (2009) Revealing the history of sheep domestication using retrovirus integrations. Science 324:532–536. doi: 10.1126/science.1170587

Clutton-Brock J (1994) The unnatural world: behavioural aspects of humans and animals in the process of domestication. In: Manning A, Serpell J (eds) Animals and human society: Changing perspectives. Routledge, pp 23–35

Clutton-Brock J (1981) Domesticated animals from early times. British Museum (Natural History) and William Heinemann Ltd.

Conolly J, Colledge S, Dobney K, et al (2011) 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. doi: 10.1016/j.jas.2010.10.008

Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 11:2463–2468. doi: 10.1093/hmg/11.20.2463

Cridland JM, Ramirez SR, Dean CA, et al (2018) Genome Sequencing of Museum Specimens Reveals Rapid Changes in the Genetic Composition of Honey Bees in California. Genome Biol Evol 10:458–472. doi: 10.1093/gbe/evy007

Cristiani E, Radini A, Edinborough M, Borić D (2016) Dental calculus reveals Mesolithic foragers in the Balkans consumed domesticated plant foods. Proc Natl Acad Sci U S A 113:10298–10303. doi: 10.1073/pnas.1603477113

Crockford SJ (2002) Animal domestication and heterochronic speciation. In: Minugh-Purvis N, McNamara KJ (eds) Human Evolution Through Developmental Change. JHU Press, pp 122–153

Cucchi T, Hulme-Beaman A, Yuan J, Dobney K (2011) Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. J Archaeol Sci 38:11–22. doi: 10.1016/j.jas.2010.07.024

Cucchi T, Mohaseb A, Peigné S, et al (2017) Detecting taxonomic and phylogenetic signals in equid cheek teeth: towards new palaeontological and archaeological proxies. R Soc Open Sci 4:160997. doi: 10.1098/rsos.160997

Currat M, Ruedi M, Petit RJ, et al (2008) The Hidden Side of Invasions: Massive Introgression by Local Genes. Evolution 62:1908–1920. doi: 10.1111/j.1558-5646.2008.00413.x

Daly KG, Delser PM, Mullin VE, et al (2018) Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science 361:85–88. doi: 10.1126/science.aas9411

Damgaard PB, Margaryan A, Schroeder H, et al (2015) Improving access to endogenous DNA in ancient bones and teeth. Sci Rep 5:11184. doi: 10.1038/srep11184

Darwin C (1859) On the origin of species by means of natural selection. John Murray, Albemarle Street, London:

Darwin C (1868) The variation of animals and plants under domestication. O. Judd

Decker JE, McKay SD, Rolf MM, et al (2014) Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet 10:e1004254. doi: 10.1371/journal.pgen.1004254

De la Rúa P, Jaffé R, Muñoz I, et al (2013) Conserving genetic diversity in the honeybee: Comments on Harpur et al.(2012). Mol Ecol 22:3208–3210. doi: 10.1111/mec.12333

Demars J, Cano M, Drouilhet L, et al (2017) Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. Mol Biol Evol 34:1722–1729. doi: 10.1093/molbev/msx114

Der Sarkissian C, Ermini L, Schubert M, et al (2015) Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Curr Biol 25:2577–2583. doi: 10.1016/j.cub.2015.08.032

Diaz-Lameiro AM (2016) Evolutionary origins and domestication of South American camelids, the alpaca (Vicugna pacos) and the llama (Lama glama) explained through molecular DNA methods. State University of New York at Binghamton

Dobney K, Brothwell D (1986) Dental calculus: its relevance to ancient diet and oral ecology. In: Cruwys E, Foley R (eds) Teeth and anthropology. BAR, Oxford, England

Dobney K, Brothwell D (1987) A method for evaluating the amount of dental calculus on teeth from archaeological sites. J Archaeol Sci 14:343–351. doi: 10.1016/0305-4403(87)90024-0

Dobney K, Larson G (2006) Genetics and animal domestication: new windows on an elusive process. J Zool 269:261–271. doi: 10.1111/j.1469-7998.2006.00042.x

Doebley JF, Gaut BS, Smith BD (2006) The Molecular Genetics of Crop Domestication. Cell 127:1309–1321. doi: 10.1016/j.cell.2006.12.006

Doust AN, Lukens L, Olsen KM, et al (2014) Beyond the single gene: How epistasis and gene-by-environment effects influence crop domestication. Proc Natl Acad Sci U S A 111:6178–6183. doi: 10.1073/pnas.1308940110

Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci U S A 106:9971–9978. doi: 10.1073/pnas.0901586106

Driscoll CA, Menotti-Raymond M, Roca AL, et al (2007) The Near Eastern Origin of Cat Domestication. Science 317:519–523. doi: 10.1126/science.1139518

Duarte CM, Marbá N, Holmer M (2007) Rapid Domestication of Marine Species. Science 316:382–383. doi: 10.1126/science.1138042

Ducos P (1989) Defining domestication: a clarification. In: Clutton-Brock J (ed) The Walking larder: patterns of domestication, pastoralism, and predation. Unwin Hyman, London, pp 28–30

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for Ancient Admixture between Closely Related Populations. Mol Biol Evol 28:2239–2252. doi: 10.1093/molbev/msr048

Eda M, Lu P, Kikuchi H, et al (2016) Reevaluation of early Holocene chicken domestication in northern China. J Archaeol Sci 67:25–31. doi: 10.1016/j.jas.2016.01.012

Edwards CJ, Bollongino R, Scheu A, et al (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proceedings of the Royal Society of London B: Biological Sciences 274:1377–1385. doi: 10.1098/rspb.2007.0020

Eisenmann V, Mashkour M (2005) Chevaux de Botaï, chevaux récents et autres souches de la domestication. In: Gardeisen A (ed) Les équidés dans le monde méditerranéen antique. Edition de l'Association pour le développement de l'archéologie en Languedoc-Roussillon, Lattes, pp 41–49

Eriksson J, Larson G, Gunnarsson U, et al (2008) Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet 4:e1000010. doi: 10.1371/journal.pgen.1000010

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 27:47–73

Evin A, Cucchi T, Cardini A, et al (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

Evin A, Flink LG, Krause-Kyora B, et al (2014) Exploring the complexity of domestication: a response to Rowley-Conwy and Zeder. World Archaeol 46:825–834. doi: 10.1080/00438243.2014.953711

Fang M, Larson G, Ribeiro HS, et al (2009) Contrasting mode of evolution at a coat color locus in wild and domestic pigs. PLoS Genet 5:e1000341. doi: 10.1371/journal.pgen.1000341

Farias-Hesson E, Erikson J, Atkins A, et al (2010) Semi-automated library preparation for high-throughput DNA sequencing platforms. Biomed Res Int 2010:

Feeney A, Nilsson E, Skinner MK (2014) Epigenetics and transgenerational inheritance in domesticated farm animals. J Anim Sci Biotechnol 5:48. doi: 10.1186/2049-1891-5-48

Fernández H, Hughes S, Vigne J-D, et al (2006) Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. Proc Natl Acad Sci U S A 103:15375–15379. doi: 10.1073/pnas.0602753103

Ferrer-Admetlla A, Leuenberger C, Jensen JD, Wegmann D (2016) An Approximate Markov Model for the Wright-Fisher Diffusion and Its Application to Time Series Data. Genetics

203:831–846. doi: 10.1534/genetics.115.184598

Frantz AC, Zachos FE, Kirschning J, et al (2013a) Genetic evidence for introgression between domestic pigs and wild boars (Sus scrofa) in Belgium and Luxembourg: a comparative approach with multiple marker systems : Introgression Between Pigs and Boars. Biol J Linn Soc Lond 110:104–115. doi: 10.1111/bij.12111

Frantz LAF, Madsen O, Megens H-J, et al (2014) Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian Sus species during the Plio-Pleistocene climatic fluctuations. Mol Ecol 23:5566–5574. doi: 10.1111/mec.12958

Frantz LAF, Meijaard E, Gongora J, et al (2016a) The Evolution of Suidae. Annu Rev Anim Biosci 4:61–85. doi: 10.1146/annurev-animal-021815-111155

Frantz LAF, Mullin VE, Pionnier-Capitan M, et al (2016b) Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352:1228–1231. doi: 10.1126/science.aaf3161

Frantz LAF, Schraiber JG, Madsen O, et al (2013b) Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol 14:R107. doi: 10.1186/gb-2013-14-9-r107

Frantz LAF, Schraiber JG, Madsen O, et al (2015) Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet 47:1141–1148. doi: 10.1038/ng.3394

Freedman AH, Gronau I, Schweizer RM, et al (2014) Genome sequencing highlights the dynamic early history of dogs. PLoS Genet 10:e1004016. doi: 10.1371/journal.pgen.1004016

Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, et al (2016) Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. PLoS Genet 12:e1005851. doi: 10.1371/journal.pgen.1005851

Fu Q, Posth C, Hajdinjak M, et al (2016) The genetic history of Ice Age Europe. Nature 534:200–205. doi: 10.1038/nature17993

Gallego-Llorente M, Connell S, Jones ER, et al (2016) The genetics of an early Neolithic pastoralist from the Zagros, Iran. Sci Rep 6:31326. doi: 10.1038/srep31326

Galton F (1865) The First Steps towards the Domestication of Animals. Transactions of the Ethnological Society of London 3:122. doi: 10.2307/3014161

Galton F (1883) Inquiries into human faculty and its development. Macmillan and Company

Gamba C, Jones ER, Teasdale MD, et al (2014) Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun 5:5257. doi: 10.1038/ncomms6257

Gaunitz C, Fages A, Hanghøj K, et al (2018) Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science 360:111–114. doi: 10.1126/science.aao3297

Geigl E-M, Grange T (2018) Of Cats and Men: Ancient DNA Reveals How the Cat Conquered the Ancient World. In: Lindqvist C, Rajora OP (eds) Paleogenomics. Springer, Cham, pp 1–18

Gerbault P, Allaby RG, Boivin N, et al (2014) Storytelling and story testing in domestication. Proc Natl Acad Sci U S A 111:6159–6164. doi: 10.1073/pnas.1400425111

Germonpré M, Sablin MV, Stevens RE, et al (2009) Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. J Archaeol Sci 36:473–490. doi: 10.1016/j.jas.2008.09.033

Gibbs RA, Taylor JF, Van Tassell CP, et al (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532. doi: 10.1126/science.1167936

Giguet-Covex C, Pansu J, Arnaud F, et al (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat Commun 5:3211. doi: 10.1038/ncomms4211

Girdland Flink L, Allen R, Barnett R, et al (2014) Establishing the validity of domestication genes using DNA from ancient chickens. Proc Natl Acad Sci U S A 111:6184–6189. doi: 10.1073/pnas.1308939110

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10:381–391. doi: 10.1038/nrg2575

Gokhman D, Lavi E, Prüfer K, et al (2014) Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science 344:523–527. doi: 10.1126/science.1250368

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. doi: 10.1038/nrg.2016.49

Green EJ, Speller CF (2017) Novel Substrates as Sources of Ancient DNA: Prospects and Hurdles. Genes 8:180. doi: 10.3390/genes8070180

Green RE, Krause J, Briggs AW, et al (2010) A draft sequence of the Neandertal genome. Science 328:710–722. doi: 10.1126/science.1188021

Gregory B of T (1969) History of the Franks. Norton, New York

Haak W, Lazaridis I, Patterson N, et al (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211. doi: 10.1038/nature14317

Haile J, Holdaway R, Oliver K, et al (2007) Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? Mol Biol Evol 24:982–989. doi: 10.1093/molbev/msm016

Hammer K (1984) Das Domestikationssyndrom. Kulturpflanze 32:11–34. doi: 10.1007/BF02098682

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (Apis mellifera) originate? Ecol Evol 2:1949–1957. doi: 10.1002/ece3.312

Hanghøj K, Orlando L (2018) Ancient Epigenomics. In: Lindqvist C, Rajora OP (eds) Paleogenomics. Springer, Cham, pp 1–37

Hanghøj K, Seguin-Orlando A, Schubert M, et al (2016) Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps with epiPALEOMIX. Mol Biol Evol 33:3284–3298. doi: 10.1093/molbev/msw184

Hanon EA, Lincoln GA, Fustin J-M, et al (2008) Ancestral TSH Mechanism Signals Summer in a Photoperiodic Mammal. Curr Biol 18:1147–1152. doi: 10.1016/j.cub.2008.06.076

Hanotte O, Bradley DG, Ochieng JW, et al (2002) African Pastoralism: Genetic Imprints of Origins and Migrations. Science 296:336–339. doi: 10.1126/science.1069878

Hansen HB, Damgaard PB, Margaryan A, et al (2017) Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PLoS One 12:e0170940. doi: 10.1371/journal.pone.0170940

Harpur BA, Minaei S, Kent CF, Zayed A (2012) Management increases genetic diversity of honey bees via admixture. Mol Ecol 21:4414–4421. doi: 10.1111/j.1365-294X.2012.05614.x

Hellenthal G, Busby GBJ, Band G, et al (2014) A Genetic Atlas of Human Admixture History.

Science 343:747–751. doi: 10.1126/science.1243518

Helmer D, Gourichon L, Monchot H, et al (2005) Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In: Vigne J-D, Peters J, Helmer D (eds) The first steps of animal domestication: new archaeozoological approaches. Oxbow, Oxford, pp 86–95

Higgins D, Kaidonis J, Townsend G, et al (2013) Targeted sampling of cementum for recovery of nuclear DNA from human teeth and the impact of common decontamination measures. Investig Genet 4:18. doi: 10.1186/2041-2223-4-18

Higuchi R, Bowman B, Freiberger M, et al (1984) DNA sequences from the quagga, an extinct member of the horse family. Nature 312:282–284. doi: 10.1038/312282a0

Hofreiter M, Paijmans JLA, Goodchild H, et al (2015) The future of ancient DNA: Technical advances and conceptual shifts. Bioessays 37:284–293. doi: 10.1002/bies.201400160

Holmes M (2014) Animals in Saxon and Scandinavian England: backbones of economy and society. Sidestone Press, Leiden, Netherlands

Hongo H, Meadow RH (1998) Pig exploitation at Neolithic Cayonu Tepesi (Southeastern Anatolia). MASCA research papers in science 15:77–98

Hongo H, Pearson J, Öksüz B, Ilgezdi G (2009) The Process of Ungulate Domestication at Çayönü, Southeastern Turkey: A Multidisciplinary Approach focusing on Bos sp. and Cervus elaphus. Anthropozoologica 44:63–78. doi: 10.5252/az2009n1a3

Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13. doi: 10.1093/nar/gkn923

Huang X-H, Wu Y-J, Miao Y-W, et al (2018) Was chicken domesticated in northern China? New evidence from mitochondrial genomes. Sci Bull Fac Agric Kyushu Univ 63:743–746. doi: 10.1016/j.scib.2017.12.004

Huerta-Sánchez E, Jin X, Asan, et al (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197. doi: 10.1038/nature13408

Hu Z-L, Fritz ER, Reecy JM (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Res 35:D604–D609. doi: 10.1093/nar/gkl946

Hu Z-L, Park CA, Reecy JM (2016) Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res 44:D827–D833. doi: 10.1093/nar/gkv1233

Ibeagha-Awemu EM, Zhao X (2015) Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. Front Genet 6:302. doi: 10.3389/fgene.2015.00302

Irving-Pease EK, Frantz LAF, Sykes N, et al (2018) Rabbits and the Specious Origins of Domestication. Trends Ecol Evol 33:149–152. doi: 10.1016/j.tree.2017.12.009

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochromeb gene of mammals. J Mol Evol 32:128–144. doi: 10.1007/BF02515385

Janečka JE, Davis BW, Ghosh S, et al (2018) Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. Nat Commun 9:2945. doi: 10.1038/s41467-018-05290-6

Janowitz Koch I, Clark MM, Thompson MJ, et al (2016) The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. Mol

Ecol 25:1838–1855. doi: 10.1111/mec.13480

Jensen P (2014) Behavior Genetics and the Domestication of Animals. Annual Review of Animal Biosciences 2:85–104. doi: 10.1146/annurev-animal-022513-114135

Johnsson M, Gustafson I, Rubin C-J, et al (2012) A Sexual Ornament in Chickens Is Affected by Pleiotropic Alleles at HAO1 and BMP2, Selected during Domestication. PLoS Genet 8:e1002914. doi: 10.1371/journal.pgen.1002914

Jónsson H, Schubert M, Seguin-Orlando A, et al (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc Natl Acad Sci U S A 111:18655–18660. doi: 10.1073/pnas.1412627111

Kadwell M, Fernandez M, Stanley HF, et al (2001) Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proceedings of the Royal Society of London B: Biological Sciences 268:2575–2584. doi: 10.1098/rspb.2001.1774

Karlsson A-C, Fallahshahroudi A, Johnsen H, et al (2016) A domestication related mutation in the thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and reproduction in chickens. Gen Comp Endocrinol 228:69–78. doi: 10.1016/j.ygcen.2016.02.010

Karlsson A-C, Svemer F, Eriksson J, et al (2015) The Effect of a Mutation in the Thyroid Stimulating Hormone Receptor (TSHR) on Development, Behaviour and TH Levels in Domesticated Chickens. PLoS One 10:e0129040. doi: 10.1371/journal.pone.0129040

Kijas JW, Lenstra JA, Hayes B, et al (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biol 10:e1001258. doi: 10.1371/journal.pbio.1001258

King HD, Donaldson HH (1929) Life processes and size of the body and organs of the gray Norway rat during ten generations in captivity. American Anatomical Memoirs 14:

Kistler L, Ware R, Smith O, et al (2017) A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Res 45:6310–6320. doi: 10.1093/nar/gkx361

Kılınç GM, Omrak A, Özer F, et al (2016) The Demographic Development of the First Farmers in Anatolia. Curr Biol 26:2659–2666. doi: 10.1016/j.cub.2016.07.057

Koutsogiannouli EA, Moutou KA, Sarafidou T, et al (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. Mammalian Biology - Zeitschrift für Säugetierkunde 75:69–73. doi: 10.1016/j.mambio.2008.08.001

Krause-Kyora B, Makarewicz C, Evin A, et al (2013) Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. Nat Commun 4:2348. doi: 10.1038/ncomms3348

Kukekova AV, Johnson JL, Xiang X, et al (2018) Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. Nat Ecol Evol 2:1479–1491. doi: 10.1038/s41559-018-0611-6

Lam YM, Chen X, Pearson OM (1999) Intertaxonomic Variability in Patterns of Bone Density and the Differential Representation of Bovid, Cervid, and Equid Elements in the Archaeological Record. Am Antiq 64:343–362. doi: 10.2307/2694283

Larson G, Albarella U, Dobney K, et al (2007a) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc Natl Acad Sci U S A 104:15276–15281. doi: 10.1073/pnas.0703411104

Larson G, Bradley DG (2014) How much is that in dog years? The advent of canine population genomics. PLoS Genet 10:e1004093. doi: 10.1371/journal.pgen.1004093

Larson G, Burger J (2013) A population genetics view of animal domestication. Trends Genet 29:197–205. doi: 10.1016/j.tig.2013.01.003

Larson G, Cucchi T, Fujita M, et al (2007b) Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. Proc Natl Acad Sci U S A 104:4834–4839. doi: 10.1073/pnas.0607753104

Larson G, Dobney K, Albarella U, et al (2005) Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. Science 307:1618–1621. doi: 10.1126/science.1106927

Larson G, Fuller DQ (2014) The Evolution of Animal Domestication. Annu Rev Ecol Evol Syst 45:115–136. doi: 10.1146/annurev-ecolsys-110512-135813

Larson G, Karlsson EK, Perri A, et al (2012) Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci U S A 109:8878–8883. doi: 10.1073/pnas.1203005109

Larson G, Liu R, Zhao X, et al (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. Proc Natl Acad Sci U S A 107:7686–7691. doi: 10.1073/pnas.0912264107

Larson G, Piperno DR, Allaby RG, et al (2014) Current perspectives and the future of domestication studies. Proc Natl Acad Sci U S A 111:6139–6146. doi: 10.1073/pnas.1323964111

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of Population Structure using Dense Haplotype Data. PLoS Genet 8:e1002453. doi: 10.1371/journal.pgen.1002453

Lawson DJ, van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun 9:3258. doi: 10.1038/s41467-018-05257-7

Lazaridis I, Nadel D, Rollefson G, et al (2016) Genomic insights into the origin of farming in the ancient Near East. Nature 536:419–424. doi: 10.1038/nature19310

Lenffer J, Nicholas FW, Castle K, et al (2006) OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. Nucleic Acids Res 34:D599–D601. doi: 10.1093/nar/gkj152

Leppälä K, Nielsen SV, Mailund T (2017) admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics 33:1738–1740. doi: 10.1093/bioinformatics/btx048

Levine MA (1999) Botai and the Origins of Horse Domestication. Journal of Anthropological Archaeology 18:29–78. doi: 10.1006/jaar.1998.0332

Librado P, Gamba C, Gaunitz C, et al (2017) Ancient genomic changes associated with domestication of the horse. Science 356:442–445. doi: 10.1126/science.aam5298

Librado P, Sarkissian CD, Ermini L, et al (2015) Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. Proc Natl Acad Sci U S A 112:E6889–E6897. doi: 10.1073/pnas.1513696112

Li D, Guo Y, Shao H, et al (2010) Genetic diversity, molecular phylogeny and selection evidence of the silkworm mitochondria implicated by complete resequencing of 41 genomes. BMC Evol Biol 10:81. doi: 10.1186/1471-2148-10-81

Linderholm A, Spencer D, Battista V, et al (2016) A novel MC1R allele for black coat colour reveals

the Polynesian ancestry and hybridization patterns of Hawaiian feral pigs. Royal Society Open Science 3:160304. doi: 10.1098/rsos.160304

Lindgren G, Backström N, Swinburne J, et al (2004) Limited number of patrilines in horse domestication. Nat Genet 36:335–336. doi: 10.1038/ng1326

Lipinski MJ, Froenicke L, Baysac KC, et al (2008) The ascent of cat breeds: Genetic evaluations of breeds and worldwide random-bred populations. Genomics 91:12–21. doi: 10.1016/j.ygeno.2007.10.009

Lippold S, Knapp M, Kuznetsova T, et al (2011a) Discovery of lost diversity of paternal horse lineages using ancient DNA. Nat Commun 2:450. doi: 10.1038/ncomms1447

Lippold S, Matzke NJ, Reissmann M, Hofreiter M (2011b) Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evol Biol 11:328. doi: 10.1186/1471-2148-11-328

Lipson M, Loh P-R, Levin A, et al (2013) Efficient Moment-Based Inference of Admixture Parameters and Sources of Gene Flow. Mol Biol Evol 30:1788–1802. doi: 10.1093/molbev/mst099

Lipson M, Szécsényi-Nagy A, Mallick S, et al (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature 551:368–372. doi: 10.1038/nature24476

Liu Y-P, Wu G-S, Yao Y-G, et al (2006) Multiple maternal origins of chickens: out of the Asian jungles. Mol Phylogenet Evol 38:12–19. doi: 10.1016/j.ympev.2005.09.014

Loftus RT, MacHugh DE, Bradley DG, et al (1994) Evidence for two independent domestications of cattle. Proc Natl Acad Sci U S A 91:2757–2761. doi: 10.1073/pnas.91.7.2757

Loog L, Thomas MG, Barnett R, et al (2017) Inferring Allele Frequency Trajectories from Ancient DNA Indicates That Selection on a Chicken Gene Coincided with Changes in Medieval Husbandry Practices. Mol Biol Evol 34:1981–1990. doi: 10.1093/molbev/msx142

Luikart G, Gielly L, Excoffier L, et al (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc Natl Acad Sci U S A 98:5927–5932. doi: 10.1073/pnas.091591198

Lundin S, Stranneheim H, Pettersson E, et al (2010) Increased Throughput by Parallelization of Library Preparation for Massive Sequencing. PLoS One 5:e10029. doi: 10.1371/journal.pone.0010029

Lynch VJ, Bedoya-Reina OC, Ratan A, et al (2015) Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Rep 12:217–228. doi: 10.1016/j.celrep.2015.06.027

MacHugh DE, Larson G, Orlando L (2017) Taming the Past: Ancient DNA and the Study of Animal Domestication. Annual Review of Animal Biosciences 5:null. doi: 10.1146/annurev-animal-022516-022747

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577. doi: 10.1038/nrg2612

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold spring harbor laboratory Cold Spring Harbor, NY, New York

Margulies M, Egholm M, Altman WE, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. doi: 10.1038/nature03959

Marshall FB, Dobney K, Denham T, Capriles JM (2014) Evaluating the roles of directed breeding

and gene flow in animal domestication. Proc Natl Acad Sci U S A 111:6153–6158. doi: 10.1073/pnas.1312984110

Mathieson I, Lazaridis I, Rohland N, et al (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503. doi: 10.1038/nature16152

McCarthy MI, Abecasis GR, Cardon LR, et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–369. doi: 10.1038/nrg2344

McCue ME, Bannasch DL, Petersen JL, et al (2012) A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. PLoS Genet 8:e1002451. doi: 10.1371/journal.pgen.1002451

Meadow RH (1983) Animal Domestication in the Middle east: a View from the Eastern Margin. In: Clutton-Brock J, Grigson C (eds) Animals and archaeology. BAR, Oxford, pp 309–337

Meadows JRS, Hiendleder S, Kijas JW (2011) Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity 106:700–706. doi: 10.1038/hdy.2010.122

Meadows JRS, Kijas JW (2009) Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. Anim Genet 40:119–123. doi: 10.1111/j.1365-2052.2008.01799.x

Meyer A (1994) Shortcomings of the cytochrome b gene as a molecular marker. Trends Ecol Evol 9:278–280. doi: 10.1016/0169-5347(94)90028-0

Miao B, Wang Z, Li Y (2017) Genomic Analysis Reveals Hypoxia Adaptation in the Tibetan Mastiff by Introgression of the Gray Wolf from the Tibetan Plateau. Mol Biol Evol 34:734–743. doi: 10.1093/molbev/msw274

Miao Y-W, Peng M-S, Wu G-S, et al (2013) Chicken domestication: an updated perspective based on mitochondrial genomes. Heredity 110:277–282. doi: 10.1038/hdy.2012.83

Miller W, Drautz DI, Ratan A, et al (2008) Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456:387–390. doi: 10.1038/nature07446

Moyers BT, Morrell PL, McKay JK (2018) Genetic costs of domestication and improvement. J Hered 109:103–116. doi: 10.1093/jhered/esx069

Muir P, Li S, Lou S, et al (2016) The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol 17:53. doi: 10.1186/s13059-016-0917-0

Muir WM, Wong GK-S, Zhang Y, et al (2008) 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 105:17312–17317. doi: 10.1073/pnas.0806569105

Mullis KB, Faloona FA (1987) [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In: Methods in Enzymology. Academic Press, pp 335–350

Mwai O, Hanotte O, Kwon Y-J, Cho S (2015) African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. Asian-australas J Anim Sci 28:911–921. doi: 10.5713/ajas.15.0002R

Nachtsheim H (1936) Vom Wildtier zum Haustier. Alfred Metzner, Berlin

Naderi S, Rezaei H-R, Pompanon F, et al (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc Natl Acad Sci U S A 105:17659–17664. doi: 10.1073/pnas.0804782105

Naderi S, Rezaei H-R, Taberlet P, et al (2007) Large-Scale Mitochondrial DNA Analysis of the Domestic Goat Reveals Six Haplogroups with High Diversity. PLoS One 2:e1012. doi: 10.1371/journal.pone.0001012

Nakao N, Ono H, Yamamura T, et al (2008) Thyrotrophin in the pars tuberalis triggers photoperiodic response. Nature 452:317–322. doi: 10.1038/nature06738

Nicholas FW (2003) Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. Nucleic Acids Res 31:275–277. doi: 10.1093/nar/gkg074

Ní Leathlobhair M, Perri AR, Irving-Pease EK, et al (2018) The evolutionary history of dogs in the Americas. Science 361:81–85. doi: 10.1126/science.aao4776

Ollivier M, Tresset A, Bastian F, et al (2016) Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. Royal Society Open Science 3:160449. doi: 10.1098/rsos.160449

Olsen KM, Wendel JF (2013) A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. Annu Rev Plant Biol 64:47–70. doi: 10.1146/annurev-arplant-050312-120048

Olsen SL (2006a) Early Horse Domestication on the Eurasian Steppe. In: Zeder MA, Bradley DG, Smith BD, Emshwiller E (eds) Documenting Domestication: New Genetic and Archaeological Paradigms. University of California Press, pp 245–269

Olsen SL (2006b) Early horse domestication: Weighing the evidence. In: Olsen SL, Grant S, Choyke AM, Bartosiewicz L (eds) Horses and humans: the evolution of human-equine relationships. Archaeopress Oxford, UK, pp 81–113

Olson MV (1999) When Less Is More: Gene Loss as an Engine of Evolutionary Change. Am J Hum Genet 64:18–23. doi: 10.1086/302219

Orlando L (2018) An Ancient DNA Perspective on Horse Evolution. In: Lindqvist C, Rajora OP (eds) Paleogenomics. Springer, Cham, pp 1–27

Orlando L, Ginolhac A, Zhang G, et al (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499:74–78. doi: 10.1038/nature12323

Ottoni C, Flink LG, Evin A, et al (2013) Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. Mol Biol Evol 30:824–832. doi: 10.1093/molbev/mss261

Ottoni C, Van Neer W, Cupere BD, et al (2017) The palaeogenetics of cat dispersal in the ancient world. Nature Ecology & Evolution 1:0139. doi: 10.1038/s41559-017-0139

Outram AK, Stear NA, Bendrey R, et al (2009) The Earliest Horse Harnessing and Milking. Science 323:1332–1335. doi: 10.1126/science.1168594

Palkopoulou E, Mallick S, Skoglund P, et al (2015) Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Curr Biol 25:1395–1400. doi: 10.1016/j.cub.2015.04.007

Pang J-F, Kluetsch C, Zou X-J, 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

Park SDE, Magee DA, McGettigan PA, et al (2015) Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. Genome Biol 16:234. doi: 10.1186/s13059-015-0790-2

Patterson N, Moorjani P, Luo Y, et al (2012) Ancient Admixture in Human History. Genetics 192:1065–1093. doi: 10.1534/genetics.112.145037

Pedersen JS, Valen E, Velazquez AMV, et al (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res 24:454–466. doi: 10.1101/gr.163592.113

Pedrosa S, Uzun M, Arranz J-J, et al (2005) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. Proceedings of the Royal Society of London B: Biological Sciences 272:2211–2217. doi: 10.1098/rspb.2005.3204

Perri A (2016) A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. J Archaeol Sci 68:1–4. doi: 10.1016/j.jas.2016.02.003

Petersen JL, Mickelson JR, Cothran EG, et al (2013) Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. PLoS One 8:e54997. doi: 10.1371/journal.pone.0054997

Peters J (1997) Hahn oder Kapaun? Zur Kastration von Hähnen in der Antike. Archiv für Geflügelkunde 61:1–8

Peters J, Lebrasseur O, Deng H, Larson G (2016) Holocene cultural history of Red jungle fowl (Gallus gallus) and its domestic descendant in East Asia. Quat Sci Rev 142:102–119. doi: 10.1016/j.quascirev.2016.04.004

Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet 8:e1002967. doi: 10.1371/journal.pgen.1002967

Pinhasi R, Fernandes D, Sirak K, et al (2015) Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PLoS One 10:e0129102. doi: 10.1371/journal.pone.0129102

Pitt D, Sevane N, Nicolazzi EL, et al (2018) Domestication of cattle: Two or three events? Evol Appl 18:R157. doi: 10.1111/eva.12674

Plagnol V, Wall JD (2006) Possible ancestral structure in human populations. PLoS Genet 2:e105. doi: 10.1371/journal.pgen.0020105

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Prüfer K, Racimo F, Patterson N, et al (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49. doi: 10.1038/nature12886

Reich D, Thangaraj K, Patterson N, et al (2009) Reconstructing Indian population history. Nature 461:489–494. doi: 10.1038/nature08365

Reissmann M, Ludwig A (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. Semin Cell Dev Biol 24:576–586. doi: 10.1016/j.semcdb.2013.03.014

Reynier P, May-Panloup P, Chrétien M-F, et al (2001) Mitochondrial DNA content affects the fertilizability of human oocytes. MHR: Basic science of reproductive medicine 7:425–429. doi: 10.1093/molehr/7.5.425

Røed KH, Flagstad Ø, Nieminen M, et al (2008) Genetic analyses reveal independent domestication origins of Eurasian reindeer. Proceedings of the Royal Society of London B: Biological Sciences 275:1849–1855. doi: 10.1098/rspb.2008.0332

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 22:939–946. doi: 10.1101/gr.128124.111

Rowley-Conwy P, Zeder M (2014a) Mesolithic domestic pigs at Rosenhof – or wild boar? A critical re-appraisal of ancient DNA and geometric morphometrics. World Archaeol 46:813–824. doi: 10.1080/00438243.2014.953704

Rowley-Conwy P, Zeder M (2014b) Wild Boar or Domestic Pigs? Response to Evin et al. World Archaeol 46:835–840. doi: 10.1080/00438243.2014.953712

Rubin C-J, Megens H-J, Barrio AM, et al (2012) Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci U S A 109:19529–19536. doi: 10.1073/pnas.1217149109

Rubin C-J, Zody MC, Eriksson J, et al (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591. doi: 10.1038/nature08832

Saiki RK, Scharf S, Faloona F, et al (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354. doi: 10.1126/science.2999980

Sánchez-Villagra MR, Geiger M, Schneider RA (2016) The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. R Soc Open Sci 3:160107. doi: 10.1098/rsos.160107

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Sawyer S, Krause J, Guschanski K, et al (2012) Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLoS One 7:e34131. doi: 10.1371/journal.pone.0034131

Schaefer RJ, Schubert M, Bailey E, et al (2017) Developing a 670k genotyping array to tag \textasciitilde2M SNPs across 24 horse breeds. BMC Genomics 18:565. doi: 10.1186/s12864-017-3943-8

Scheu A, Geörg C, Schulz A, et al (2012) The arrival of domesticated animals in South-Eastern Europe as seen from ancient DNA. In: Kaiser E, Burger J, Schier W (eds) Population Dynamics in Prehistory and Early History. De Gruyter, Berlin, Boston, pp 45–54

Scheu A, Powell A, Bollongino R, et al (2015) The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet 16:54. doi: 10.1186/s12863-015-0203-2

Schraiber JG, Evans SN, Slatkin M (2016) Bayesian Inference of Natural Selection from Allele Frequency Time Series. Genetics 203:493–511. doi: 10.1534/genetics.116.187278

Schubert M, Jónsson H, Chang D, et al (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci U S A 111:E5661–9. doi: 10.1073/pnas.1416991111

Schweizer RM, Durvasula A, Smith J, et al (2018) Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. Mol Biol Evol 35:1190–1209. doi: 10.1093/molbev/msy031

Serjeantson D (2006) Birds: Food and a Mark of Status. In: Woolgar CM, Serjeantson D, Waldron T (eds) Food in Medieval England: Diet and Nutrition. Oxford University Press, Oxford, New York, pp 131–147

Shannon LM, Boyko RH, Castelhano M, et al (2015) Genetic structure in village dogs reveals a Central Asian domestication origin. Proc Natl Acad Sci U S A 112:13639–13644. doi: 10.1073/pnas.1516215112

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 25:1515–1519. doi: 10.1016/j.cub.2015.04.019

Smith BD, Zeder MA (2013) The onset of the Anthropocene. Anthropocene 4:8–13. doi: 10.1016/j.ancene.2013.05.001

Smith O, Momber G, Bates R, et al (2015a) Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. Science 347:998–1001. doi: 10.1126/science.1261278

Smith O, Momber G, Bates R, et al (2015b) Response to Comment on "Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago." Science 349:247–247. doi: 10.1126/science.aab2062

Søe MJ, Nejsum P, Seersholm FV, et al (2018) Ancient DNA from latrines in Northern Europe and the Middle East (500 BC-1700 AD) reveals past parasites and diet. PLoS One 13:e0195481. doi: 10.1371/journal.pone.0195481

Stainton J j., Charlesworth B, Haley C s., et al (2017) Use of high-density SNP data to identify patterns of diversity and signatures of selection in broiler chickens. J Anim Breed Genet 134:87–97. doi: 10.1111/jbg.12228

Staubach F, Lorenc A, Messer PW, et al (2012) Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (Mus musculus). PLoS Genet 8:e1002891. doi: 10.1371/journal.pgen.1002891

Stock F, Gifford-Gonzalez D (2013) Genetics and African Cattle Domestication. African Archaeological Review 30:51–72. doi: 10.1007/s10437-013-9131-6

Sun W, Yu H, Shen Y, et al (2012) Phylogeny and evolutionary history of the silkworm. Sci China Life Sci 55:483–496. doi: 10.1007/s11427-012-4334-7

Sun W, Zhao X-W, Zhang Z (2015) Identification and evolution of the orphan genes in the domestic silkworm, Bombyx mori. FEBS Lett 589:2731–2738. doi: 10.1016/j.febslet.2015.08.008

Sykes NJ (2007) The Norman conquest: a zoological perspective. Archaeopress, Oxford

Tapio M, Marzanov N, Ozerov M, et al (2006) Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. Mol Biol Evol 23:1776–1783. doi: 10.1093/molbev/msl043

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702. doi: 10.1038/nrg3053

Teasdale MD, van Doorn NL, Fiddyment S, et al (2015) Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. Philos Trans R Soc Lond B Biol Sci 370:20130379. doi: 10.1098/rstb.2013.0379

Thalmann O, Perri AR (2018) Paleogenomic Inferences of Dog Domestication. In: Lindqvist C, Rajora OP (eds) Paleogenomics. Springer, Cham, pp 1–34

Thalmann O, Shapiro B, Cui P, 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

Triantaphyllopoulos KA, Ikonomopoulos I, Bannister AJ (2016) Epigenetics and inheritance of phenotype variation in livestock. Epigenetics Chromatin 9:31. doi: 10.1186/s13072-016-0081-5

Troy CS, MacHugh DE, Bailey JF, et al (2001) Genetic evidence for Near-Eastern origins of

European cattle. Nature 410:1088–1091. doi: 10.1038/35074088

Trut LN, Plyusnina IZ, Oskina IN (2004) An Experiment on Fox Domestication and Debatable Issues of Evolution of the Dog. Russ J Genet 40:644–655. doi: 10.1023/B:RUGE.0000033312.92773.c1

Trut L, Oskina I, Kharlamova A (2009) Animal evolution during domestication: the domesticated fox as a model. Bioessays 31:349–360. doi: 10.1002/bies.200800070

Uauy C, Distelfeld A, Fahima T, et al (2006) A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. Science 314:1298–1301. doi: 10.1126/science.1133649

Vernot B, Akey JM (2014) Resurrecting surviving Neandertal lineages from modern human genomes. Science 343:1017–1021. doi: 10.1126/science.1245938

Vigne J-D (2011) The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. C R Biol 334:171–181. doi: 10.1016/j.crvi.2010.12.009

Vigne J-D (2008) Zooarchaeological Aspects of the Neolithic Diet Transition in the Near East and Europe, and Their Putative Relationships with the Neolithic Demographic Transition. In: The Neolithic Demographic Transition and its Consequences. Springer, Dordrecht, pp 179–205

Vigne J-D (2015) Early domestication and farming: what should we know or do for a better understanding? Anthropozoologica 50:123–150. doi: 10.5252/az2015n2a5

Vigne J-D, Briois F, Zazzo A, et al (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. Proc Natl Acad Sci U S A 109:8445–8449. doi: 10.1073/pnas.1201693109

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:S255–S271. doi: 10.1086/659306

Vigne J-D, Evin A, Cucchi T, et al (2016) Earliest "Domestic" Cats in China Identified as Leopard Cat (Prionailurus bengalensis). PLoS One 11:e0147295. doi: 10.1371/journal.pone.0147295

Vigne J-D, Guilaine J, Debue K, et al (2004) Early Taming of the Cat in Cyprus. Science 304:259–259. doi: 10.1126/science.1095335

Vilà C, Leonard JA, Götherström A, et al (2001) Widespread Origins of Domestic Horse Lineages. Science 291:474–477. doi: 10.1126/science.291.5503.474

Volf J, Kus E, Prokopova L (1991) General studbook of the Przewalski horse. Zoological Garden Prague, Prague

vonHoldt BM, Pollinger JP, Lohmueller KE, et al (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature 464:898–902. doi: 10.1038/nature08837

Wallberg A, Han F, Wellhagen G, et al (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat Genet 46:ng.3077. doi: 10.1038/ng.3077

Wang G-D, Zhai W, Yang H-C, et al (2015) Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res 26:21. doi: 10.1038/cr.2015.147

Wang X, Liu J, Zhou G, et al (2016) Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep 6.: doi: 10.1038/srep38932

Warinner C, Rodrigues JFM, Vyas R, et al (2014) Pathogens and host immunity in the ancient human oral cavity. Nat Genet 46:336–344. doi: 10.1038/ng.2906

Warinner C, Speller C, Collins MJ (2015) A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. Philos Trans R Soc Lond B Biol Sci 370:20130376. doi: 10.1098/rstb.2013.0376

Wendorf F, Schild R (2005) Are the early holocene cattle in the eastern sahara domestic or wild? Evol Anthropol 3:118–128. doi: 10.1002/evan.1360030406

West B, Zhou B-X (1988) Did chickens go North? New evidence for domestication. J Archaeol Sci 15:515–533. doi: 10.1016/0305-4403(88)90080-5

Weyrich LS, Dobney K, Cooper A (2015) Ancient DNA analysis of dental calculus. J Hum Evol 79:119–124. doi: 10.1016/j.jhevol.2014.06.018

Wheeler JC, Chikhi L, Bruford MW (2006) Genetic analysis of the origins of domestic South American camelids. Archaeology and Animal Domestication: New Genetic and Archaeological Paradigms 329–341

White S (2011) From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History. Environ Hist Durh N C 16:94–120. doi: 10.1093/envhis/emq143

Whitfield CW, Behura SK, Berlocher SH, et al (2006) Thrice Out of Africa: Ancient and Recent Expansions of the Honey Bee, Apis mellifera. Science 314:642–645. doi: 10.1126/science.1132772

Wilkins AS, Wrangham RW, Fitch WT (2014) The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. Genetics 197:795–808. doi: 10.1534/genetics.114.165423

Wood JR, Crown A, Cole TL, Wilmshurst JM (2016) Microscopic and ancient DNA profiling of Polynesian dog (kurī) coprolites from northern New Zealand. Journal of Archaeological Science: Reports 6:496–505. doi: 10.1016/j.jasrep.2016.03.020

Wright D (2015) The Genetic Architecture of Domestication in Animals. Bioinform Biol Insights 9:11–20. doi: 10.4137/BBI.S28902

Wright D, Rubin C-J, Martinez Barrio A, et al (2010) The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. Mol Ecol 19:5140–5156. doi: 10.1111/j.1365-294X.2010.04882.x

Wu D-D, Ding X-D, Wang S, et al (2018) Pervasive introgression facilitated domestication and adaptation in the Bos species complex. Nat Ecol Evol 2:1139–1145. doi: 10.1038/s41559-018-0562-y

Xiang H, Gao J, Yu B, et al (2014) Early Holocene chicken domestication in northern China. Proc Natl Acad Sci U S A 111:17564–17569. doi: 10.1073/pnas.1411882111

Xiang H, Li X, Dai F, et al (2013) Comparative methylomics between domesticated and wild silkworms implies possible epigenetic influences on silkworm domestication. BMC Genomics 14:646. doi: 10.1186/1471-2164-14-646

Xia Q, Guo Y, Zhang Z, et al (2009) Complete Resequencing of 40 Genomes Reveals Domestication Events and Genes in Silkworm (Bombyx). Science 326:433–436. doi: 10.1126/science.1176620

Yang H, Wang JR, Didion JP, et al (2011) Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet 43:648–655. doi: 10.1038/ng.847

Yang S-Y, Han M-J, Kang L-F, et al (2014) Demographic history and gene flow during silkworm

domestication. BMC Evol Biol 14:185. doi: 10.1186/s12862-014-0185-0

Zeder MA (2012a) The Domestication of Animals. J Anthropol Res 68:161–190. doi: 10.3998/jar.0521004.0068.201

Zeder MA (2016) Domestication as a model system for niche construction theory. Evol Ecol 30:325–348. doi: 10.1007/s10682-015-9801-8

Zeder MA (2012b) Pathways to animal domestication. In: Gepts PL (ed) Biodiversity in agriculture: domestication, evolution, and sustainability. Cambridge University Press, New York, pp 227–259

Zeder MA (2006) Archaeological approaches to documenting animal domestication. In: Zeder MA, Bradley DG, Smith BD, Emshwiller E (eds) Documenting Domestication: New Genetic and Archaeological Paradigms. University of California Press, pp 171–180

Zeder MA (2008) Animal domestication in the Zagros: an update and directions for future research. Publications de la Maison de l'Orient et de la Méditerranée 49:243–277

Zeder MA (2015) Core questions in domestication research. Proc Natl Acad Sci U S A 112:3191–3198. doi: 10.1073/pnas.1501711112

Zeder MA, Emshwiller E, Smith BD, Bradley DG (2006) Documenting domestication: the intersection of genetics and archaeology. Trends Genet 22:139–155. doi: 10.1016/j.tig.2006.01.007

Zeder MA, Hesse B (2000) The Initial Domestication of Goats (Capra hircus) in the Zagros Mountains 10,000 Years Ago. Science 287:2254–2257. doi: 10.1126/science.287.5461.2254

Zeuner FE (1963) A history of domesticated animals. London: Hutchinson & Co.(Publishers) Ltd.

1.12 Acknowledgments and Notes

Acknowledgments

We thank the two anonymous reviewers for their valuable comments.

Author contributions

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

1.13 Permission from co-authors

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Paleogenomics of Animal Domestication"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions:

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

Date: 24/4/19

Name(s): Hannah F Ryan

Signature(s): HRyan

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Paleogenomics of Animal Domestication"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions:

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

Date: 270919

Name(s): Alexandra Jamieson

Signature(s): AJ

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Paleogenomics of Animal Domestication"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions:

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

Date: **26-09-2019**

Name(s): **Evangelos Antonios Dimopoulos**

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Paleogenomics of Animal Domestication"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions:

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

Date: Sept 30, 2019

Name(s): Greger Larson

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Paleogenomics of Animal Domestication"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions:

E.K.I.-P. wrote chapters *1 Introduction*, *2 Sequencing Ancient DNA*, *3 Pathways to Animal Domestication*, *4.1.3 Cats*, *4.2.1 Goats*, *4.2.2 Sheep*, *4.2.3 Cattle*, *4.3.1 Horses*, *4.3.2 Rabbits*, *5 The Biological Architecture of Domestication*, *6 Future Perspectives*, and *7 Conclusion*. L.A.F.F. wrote chapters *4.1.1 Dogs* and *4.1.2 Pigs*. E.A.D. and G.L. wrote chapter *4.1.4 Chickens*. A.J. and G.L. wrote chapters *4.2.4 New World Camelids* and *4.3.3 Old World Camels*. H.R. wrote chapter *4.3.4 Insects*. E.K.I.-P. and L.A.F.F. edited the manuscript, with input from all other authors.

Date: 30/09/2019

Name(s): Laurent Frantz

Signature(s): Laurent Frantz

2 Selection Trajectories of Genetic Variants Underlying Domestic Animal Traits

2.1 Statement of Authorship

Design of Research: I designed the research, with input from of L.A.F.F and J.G.S.

Data: I curated the ancient and modern genome-wide sequencing data and performed all quality checks and filtering.

Analysis: I performed all the computational analyses, and interpreted the results with input from L.A.F.F., G.L. and J.G.S.

Manuscript: I wrote the manuscript, with input from L.A.F.F., G.L. and J.G.S.

2.2 Authors and Affiliations

Authors

Evan K. Irving-Pease[1], Laurent A. F. Frantz[1,2], Greger Larson[1], Joshua G. Schraiber[3,4]

Affiliations

[1] *The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK.*

[2] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*

[3] *Department of Biology, Temple University, Philadelphia, USA.*

[4] *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, USA.*

2.3 Abstract

The study of animal domestication is an important model system for understanding adaptive responses to changes in environmental conditions, demography, and selective pressures over time. Despite decades of genetic research into traits associated with domestication, our understanding of the underlying genetic basis of adaptation to the domestic niche is limited. Using genome-wide modern DNA, previous studies have contrasted populations of wild and domestic animals to scan for segregating signatures of selection in their respective genomes. Due to the intensive nature of modern breeding practices, it is unclear which candidate genes identified by these methods were under selection during the initial process of domestication, and which represent more recent improvement traits. Time series data, obtained from ancient DNA, can resolve these questions by directly observing changes in allele frequencies over time. Here, we model the allelic trajectory of >2,000 genome-wide association study (GWAS) variants linked to quantitative traits in taurine cattle (*Bos taurus*) and horses (*Equus ferus caballus*). Using a dataset of 369 ancient nuclear genomes, spanning >10,000 years of evolutionary history, we attempt to quantify the temporal origins and strength of selection for genetic variants associated with health, reproductive, performance, production, aesthetic and behavioural traits in cattle and horse populations. Due to technical difficulties with the Markov Chain Monte Carlo (MCMC) convergence, the majority of modelled variants (75%) did not pass quality control. Of those that met convergence metrics, only two SNPs in horses showed credible signs of selection, and none in cattle.

2.4 Introduction

The domestication of animals was one of the most important transformations in human history. It supported the transition from hunting and gathering to pastoralism and settled agricultural communities (Larson et al., 2014), and ushered in changes that ultimately transformed the entire biosphere (Vigne, 2011). Due to its central importance to evolutionary and population genetics, the domestication of animals has been widely studied (e.g. Gerbault et al., 2014; Irving-Pease et al., 2018; Jensen, 2014; Larson and

Burger, 2013). However, the underlying genetic basis of traits differentiating domestic animals them from their wild counterparts remains elusive.

Previous studies have advanced our understanding of the genetic architecture of domestication by comparing genome-wide data from populations of wild and domestic animals, to scan for segregating signatures of selection. This comparative approach has been critical in demonstrating that animal domestication did not occur through strong selection on common variants with high effect sizes. Instead, selection during animal domestication is likely to result in subtle shifts in the allele frequencies of variants at hundreds of loci, each with low effect size. For example, a recent study on European rabbit (*Oryctolagus cuniculus*) domestication compared genome-wide data from wild rabbits in France and Iberia, to six breeds of domestic rabbits (Carneiro et al., 2014). The authors scanned for signatures of selection shared across all rabbit breeds and identified more than 100 sweep regions exclusive to domestic rabbits but found very few fixed derived alleles. A similar study comparing modern sheep and goats found 90 selective sweep regions which segregated between domestic and wild populations of *Capra* and *Ovis*, but very few fixed variants (Alberto et al., 2018).

One of the principle limitations of studies of selection based on modern DNA is that they are unable to resolve the temporal component of selection, and therefore cannot distinguish between loci under selection during the early stages of domestication and those selected for in subsequent processes. For example, a genome-wide study comparing European domestic pigs and wild boar identified more than 300 selective sweep loci unique to the domestic populations (Frantz et al., 2015). However, analysis of ancient DNA from across Europe and the Near East showed that these selected haplotypes originated in the local European wild boar population, and only introgressed into domestic pigs following their arrival from the Near East, more than ~2,000 years after their initial domestication (Frantz et al., 2019). Similarly, ancient DNA studies of domestic chickens (Girdland Flink et al., 2014; Loog et al., 2017) and dogs (Arendt et al., 2016; Ollivier et al., 2016) have established that putative domestication loci, identified in modern populations, were not under selection until thousands of years after each species' initial domestication (see Section 1.8.3). More recently, the intensive nature of

modern breeding practices has made it exceedingly difficult to identify the genomic loci which were targeted by selection in the distant past.

Time series data, obtained from ancient DNA, can resolve these issues by directly observing changes in allele frequencies over time. The recent development of novel techniques for modelling serially sampled DNA (e.g. Bollback et al., 2008; Ferrer-Admetlla et al., 2016; Gory et al., 2018; Loog et al., 2017; Malaspinas et al., 2012; Mathieson and McVean, 2013; Paris et al., 2019; Sackman et al., 2019; Schraiber et al., 2016; Terhorst et al., 2015) allow inference of selection coefficients and other parameters of interest. By modelling the trajectories of alleles with known trait associations, ascertained in genome-wide association studies (GWAS), it is possible to infer the strength of selection and the age of the allele under selection for hundreds of polygenic traits. Here, we apply a modified version of the Schraiber et al. (2016) method to reconstruct the allelic trajectory of >2,000 GWAS variants linked to quantitative traits in cattle and horses. Using a dataset of 369 ancient nuclear genomes, spanning >10,000 years of evolutionary history, we attempt to quantify the temporal origins and strength of selection for genetic variants associated with health, reproductive, performance, production, aesthetic and behavioural traits in cattle and horse populations.

2.5 Methods and Materials

2.5.1 Computational Pipeline

The computational pipeline built to perform all analyses was written in the Python scripting language, using the Luigi (v. 2.8.3) (Spotify, 2019) pipeline framework and GNU parallel (Tange, 2018), with supporting code written in bash, R and SQL. All analyses are fully reproducible, with the code made available online in a GitHub repository (https://github.com/ekirving/alleletraj).

A MySQL database (v. 5.7.21) (Widenius and Axmark, 2002) was used to organise project metadata and sequencing results, as well as for hosting local builds of the AnimalQTL database (Hu and Reecy, 2007), and the Ensembl gene and variants (dbSNP) databases (Hubbard et al., 2002).

2.5.2 Reference genomes

Reference genomes for cattle (UMD3.1) (Zimin et al., 2009) and horses (EquCab2) (Wade et al., 2009) were downloaded from Ensembl (Hubbard et al., 2002), and indexed with the `samtools faidx` command from *SAMtools* (v. 1.3.1) (Li et al., 2009) and the `CreateSequenceDictionary` command from *Picard* (v. 2.5.0) (Broad Institute, 2016).

Table 2.1 Genome assemblies used for each species

Species		Assembly	GenBank Accession	
Cattle	*Bos taurus*	UMD3.1	GCA_000003055.3	(Zimin et al., 2009)
Horse	*Equus f. caballus*	EquCab2	GCA_000002305.1	(Wade et al., 2009)

Whilst newer reference genomes exist for horses (EquCab3.0) (Kalbfleisch et al., 2018), and cattle (ARS-UCD1.2) (USDA ARS, 2018), earlier builds were specifically chosen to maintain compatibility with older resources used in downstream processing.

2.5.3 Modern Samples

2.5.3.1 Published data

Modern reference panels for cattle (n=22) and horses (n=30) were compiled from publicly available whole-genome sequencing (WGS) libraries deposited with the Sequence Read Archive (SRA) (Leinonen et al., 2011). Samples were chosen to broadly represent a diversity of modern breeds. The River buffalo (*Bubalus bubalis*) and Somali wild ass (*Equus africanus somaliensis*) were used as outgroup species for cattle and horses, respectively. For the full list of *BioSample* codes for each species see Table S2.1 and Table S2.2.

Table 2.2 Summary of modern reference panels

Species	Samples	Mean DoC	Outgroup	
Cattle	22	8.78x	River buffalo	*Bubalus bubalis*

Species	Samples	Mean DoC	Outgroup	
Horse	30	15.37x	Somali wild ass	*Equus africanus somaliensis*

Individual libraries, and their corresponding SRA run accession codes, were identified by querying the SRA with the BioSample code of each respective sample, using the `esearch`, `efetch` and `xtract` workflow from the *Entrez Direct* software package (v 11.6) (Kans, 2013). SRA run accessions were downloaded and split into single- and paired-end FASTQ files, where appropriate, using the `fasterq-dump` tool from the *SRA Tools* software package (v. 2.9.6) (Leinonen et al., 2011).

2.5.3.2 Trimming

Single- and paired-end libraries were trimmed of adapters, and low-quality bases (Phred scaled base quality score < 20) were removed from both the 5' and 3’ ends using *TrimGalore!* (v. 0.4.3) (Krueger, 2015), which functions as a wrapper around *Cutadapt* (v. 1.17) (Martin, 2011). Any trimmed reads shorter than 25 bases in length were discarded from further processing. Summary statistics and quality control checks were produced for all trimmed libraries using *FastQC* (v. 0.11.7) (Andrews, 2010), with default settings, as implemented in *TrimGalore!*

2.5.3.3 Alignment

Trimmed reads were aligned to their respective reference genomes using the *Burrows-Wheeler Aligner (BWA)* (v. 0.7.5) (Li and Durbin, 2009), with the `bwa mem` command (Li, 2013), after indexing the reference genome with `bwa index -a bwtsw`.

For paired-end libraries, `bwa mem` can produce invalid FLAG information on Sequence Alignment Map (SAM) records (Li et al., 2009), so mate pair flags were reset using the `samtools fixmate` command. SAM output was filtered to remove unaligned reads using the `samtools view` command; and sorted into coordinate order and converted

into Binary Alignment Map (BAM) format using the `samtools sort` command. Sorted BAM files were indexed using the `samtools index` command.

2.5.3.4 Deduplication

Potential PCR duplicates were removed using the `MarkDuplicates` command from *Picard* (v. 2.5.0) (Broad Institute, 2016), by grouping duplicate reads using their 5' alignment, and only retaining the read with the largest sum of base quality scores.

2.5.3.5 Indel realignment

As short read sequence aligners like `bwa mem` can produce suboptimal alignments in the vicinity of insertions and deletions (indels), all BAM files had their indels realigned using the `RealignerTargetCreator` and `IndelRealigner` workflow from the *Genome Analysis ToolKit (GATK)* software package (v. 3.6-0) (McKenna et al., 2010).

2.5.3.6 Merging libraries

For samples with multiple SRA run accessions, all BAM files were merged using the `samtools merge` command and deduplicated a second time using `MarkDuplicates`. This strategy was adopted because some *BioSample* codes were linked to SRA run accessions with extensive duplication across libraries, likely caused by duplicate uploads to the SRA. To avoid systematic overestimation of sequencing depth, a small increase in the false positive rate for PCR duplicates was preferred.

All processed BAM files were checked for strict standards compliance with the `ValidateSamFile` command from *Picard*., The depth of coverage was calculated using the `samtools depth` command with the flags `-q 30` and `-Q 30,` specifying minimum Phred-scaled base and map qualities of 30.

2.5.3.7 Variant calling

Variant calling was performed jointly, across all samples for a given species, using the `bcftools mpileup` and `bcftools call` workflow from *BCFtools* (v. 1.7) (Li, 2011) and the multiallelic calling model (`-m`) (Danecek et al., 2016). Sex based ploidy files were specified for each reference genome, along with the sex of each sample, to constrain the ploidy of calls made in sex chromosomes. The command `bcftools norm` was used to left align indels and collapse all multiallelic sites, followed by `bcftools +fill-tags` to recalculate the AC and AN tags; as these are not automatically updated during normalisation. The resulting callset was stored in gzipped Variant Call Format (VCF) (Danecek et al., 2011).

2.5.3.8 Variant filtering

Variants were filtered for a minimum Phred-scaled genotype quality of 30, and a per-site read depth falling within the 5%–95% interquantile range of each respective chromosome. For each species, the `bcftools query` command was used to extract the joint read depth across all samples at each position in the genome, and the function `numpy.quantile` from the Python package *NumPy* (v. 1.16.2) (Oliphant, 2006) was used to calculate the 5% and 95% read depth quantiles of each chromosome. The `bcftools filter` command was used to apply the genotype quality and read depth filters to each chromosomal VCF.

2.5.3.9 Variant polarisation

To support downstream analysis, variants were polarised for the ancestral allele using the aligned sequence from the outgroup sample for each species. The pysam library (v. 0.15.0) (Heger and Jacobs, 2018), a Python wrapper around htslib (Li et al., 2009), was used to step through each position in the filtered callset and output a new VCF file in which the REF allele was replaced with the outgroup allele, where necessary. All sites where the outgroup had either no call, or was heterozygous, were dropped from further processing as the ancestral allele could not be reliably determined. All sites which were

private to the outgroup sample were also dropped, including cases were the REF and outgroup allele were in concordance, as these have a high chance of being mispolarised.

2.5.3.10 Polarised biallelic SNPs

The polarised VCF was further filtered with `bcftools view` to include only single nucleotide polymorphisms (SNPs) (`--types snps`), with exactly two alleles (`--min-alleles 2 --max-alleles 2`), and a minimum minor allele count of one (`--min-ac 1:minor`). All sites containing an INDEL or polyallelic SNP were excluded from further processing. This dataset is hereafter referred to as the modern polarised biallelic SNP callset.

Table 2.3 Summary of polarised biallelic SNPs

Species	Autosomal SNPs	X chrom SNPs	Total SNPs
Cattle	13,148,188	242,824	13,391,012
Horse	16,447,686	518,197	16,965,883

2.5.3.11 Derived allele frequency (DAF)

For each species, the modern polarised biallelic SNP callset was used to estimate the derived allele frequency (DAF) for every SNP in each modern population, and the results were loaded into a local MySQL database.

2.5.4 Demographic modelling

2.5.4.1 Site frequency spectra (SFS)

Site frequency spectra (SFS) were estimated from all autosomal SNPs in the modern polarised biallelic SNP callset, using the *easySFS* Python script (Overcast, 2016), with no downwards projection of samples. The folded and unfolded SFS were estimated separately for each modern population in each species.

2.5.4.2 Diffusion Approximation for Demographic Inference (∂a∂i)

To determine the population size history of each domestic species, five sequential epoch demographic models were fitted for each modern population using the *Diffusion Approximation for Demographic Inference (∂a∂i)* software package (v. 1.7.0) (Gutenkunst et al., 2009).

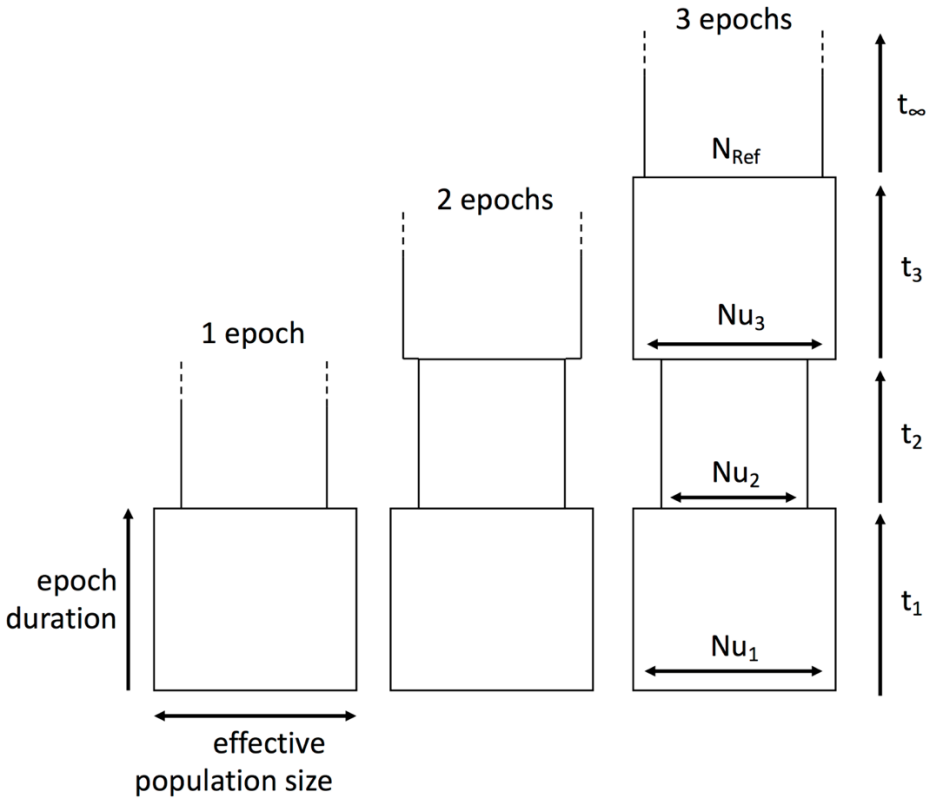


Figure 2.1 Schematic of the first three sequential epoch ∂a∂i models

Each sequential epoch in the ∂a∂i models had two parameters, an epoch duration (t_i) and a constant population size (Nu_i). Epoch transitions were modelled as instantaneous size changes in the subsequent epoch. Models with 1, 2, 3, 4 and 5 sequential epochs were fitted. To avoid models getting stuck in local maxima, each model was run 1,000 times from dispersed starting positions, drawn at random from a uniform distribution bounded by the lower and upper limits of each parameter in the model. The set of parameters with the lowest log-likelihood for each epoch model was used to choose the best overall fitting model, using the Akaike Information Criterion (AIC) (Akaike, 1973).

2.5.5 Ancient Samples

2.5.5.1 Published Data

Previously published ancient samples for horses (n=245) and cattle (n=35) were obtained from the Sequence Read Archive (SRA) (Leinonen et al., 2011). Individual libraries, and their corresponding SRA run accession codes, were identified by querying the SRA with the BioSample code of each respective sample, using the `esearch`, `efetch` and `xtract` workflow from the *Entrez Direct* software package (v 11.6) (Kans, 2013). Individual SRA run accessions were downloaded and split into single- and paired-end FASTQ files, where appropriate, using the `fasterq-dump` tool from the *SRA Tools* software package (v. 2.9.6) (Leinonen et al., 2011). The full list of *BioSample* codes for horses and cattle are provided in supplementary tables Table S2.3 and Table S2.4.

2.5.5.2 Unpublished data

Unpublished ancient cattle data (n=89) were provided by Dr Victoria E. Mullin (Natural History Museum, London) and Prof Daniel G. Bradley (Trinity College Dublin) (Mullin et al. *in prep*). Unpublished data was provided in BAM file format, and differences in pre-processing have been noted throughout.

Table 2.4 Summary of published and unpublished ancient samples

Species	Samples	Mean DoC	References
Cattle	124	1.53x	(Verdugo et al., 2019) (Mullin et al. *in prep*)
Horse	245	2.57x	(Fages et al., 2019)

2.5.5.3 Trimming

Single- and paired-end libraries were trimmed of adapters, and low-quality bases (Phred scaled base quality score < 20) were removed from both the 5' and 3’ ends using *AdapterRemoval* (v. 2.3.0) (Schubert et al., 2016). Any trimmed reads shorter than 25 bases in length were discarded from further processing. All overlapping paired-end

reads were merged into consensus reads and aligned as if they were single-end. For the pre-processed cattle data, trimming was performed by Dr Mullin using `cutadapt` (v. 1.1) (Martin, 2011).

2.5.5.4 Alignment

Trimmed reads were aligned to their respective reference genomes using *BWA* (v. 0.7.5a-r405) (Li and Durbin, 2009), with the `bwa aln` and `bwa samse` workflow, using flags `-l 1024 -n 0.01 -o 2` to improve alignment of ancient reads (Taron et al., 2018).

2.5.5.5 Deduplication

Potential PCR duplicates were removed using the `FilterUniqueSAMCons` script (Kircher, 2012), by calling a consensus sequence across reads with the same 5' alignment. For the pre-processed cattle data, deduplication was performed by Dr Mullin using the `samtools rmdup` command (version 0.1.19).

2.5.5.6 Indel realignment

For all BAM files, reads in the vicinity of indels were locally realigned using the `RealignerTargetCreator` and `IndelRealigner` workflow from *GATK* (v. 3.6-0) (McKenna et al., 2010), to correct for read misalignments from *BWA*.

2.5.5.7 Merging libraries

For samples with multiple libraries, all BAM files were merged using the `samtools merge` command and deduplicated a second time using `FilterUniqueSAMCons`. This strategy was adopted because some ancient libraries were built from the same PCR products, and therefore required deduplication after merging. BAM files prepared by Dr Mullin were not deduplicated a second time.

2.5.5.8 Base quality rescaling

All ancient BAM files were tested for characteristic patterns of ancient DNA damage using *mapDamage2* (v. 2.0.8) (Jónsson et al., 2013). To avoid misincorporation of ancient DNA damage, base quality scores were rescaled based on observed patterns of damage, using the `mapDamage --rescale` command. For the pre-processed cattle data, no base quality rescaling was performed by Dr Mullin. Instead, USER treated libraries were soft-clipped by 6 bp at read termini, and non-USER treated libraries were soft-clipped by 8 bp.

All fully processed BAM files were checked for strict standards compliance with the `ValidateSamFile` command from *Picard*. The depth of coverage was calculated for all samples using the `samtools depth` command with flags specifying all bases (`-a`), and a minimum Phred scaled base quality of 30 (`-q 30`) and map quality of 30 (`-Q 30`).

2.5.5.9 Variant calling with ANGSD

For all species, pseudo-haploid calls were made for all samples across the whole-genome with the `angsd -dohaplocall 1` command from *ANGSD* (v. 0.929) (Korneliussen et al., 2014). A base was chosen at random from each sample's pileup, provided it met the following criteria:

- Mapping quality ≥ 30 (`-minMapQ 30`)
- Base quality ≥ 30 (`-minQ 30`)
- Distance from read termini > 5 (`-trim 5`)

The haploid output from `angsd` was converted into PLINK transposed text format (TPED) using the `haploToPlink` command in *ANGSD*. Transposed text files were filtered for only those positions found in the modern polarised biallelic SNP callset, and converted into PLINK binary format (BED) using *PLINK* (v. 1.90) (Chang et al., 2015), with the flags `--missing-genotype N` and `--output-missing-genotype 0`, to correct for the non-standard representation of missing genotypes in `haploToPlink`. Any sites which were polyallelic in the resulting TPED files were dropped from further processing.

2.5.6 Combined dataset

The modern polarised biallelic SNP callset was converted from VCF format into Binary PLINK format using the `plink --vcf --make-bed` command. The modern and ancient PLINK files were merged with the `plink --merge-list` command, and any sites which were polyallelic in the merged dataset were dropped from further processing.

Table 2.5 Summary of combined modern and ancient datasets

Species	Samples	SNPs	Genotyping Rate
Cattle	147	13,258,749	0.53
Horse	276	16,824,863	0.62

2.5.7 Genome-Wide Association Studies (GWAS)

For all species, GWAS associations of quantitative traits and genomic variants were obtained from the Animal QTLdb (Hu and Reecy, 2007), release 38 (24 Apr 2019) (Hu et al., 2019). Using the public application programming interface (API) (Hu et al., 2016), records which met the following conditions were loaded into a local MySQL database:

- Study type of GWAS
- Statistically significant association
- dbSNP RefSeq (rs) ID for the association peak
- PubMed identifier for the publication

Table 2.6 Summary of unique GWAS associations, variants, and traits

Species	GWAS Associations	Unique Variants	Unique Traits
Cattle	101,952	51,054	393
Horse	1,424	1,258	46
Total	103,376	52,312	439

2.5.8 Ensembl genes and variants

For each species, a list of genes and annotated transcripts (in general transfer format [GTF]), and germline variants (in genome variation format [GVF]) were downloaded from Ensembl (Hubbard et al., 2002). In each case, the most recent Ensembl release that matched the genome assembly for that species was used.

Table 2.7 Ensembl release versions, dates, and assemblies

Species	Assembly	Ensembl	Genes	Variants	Release date
Cattle	UMD3.1	rel. 94	24,616	102,153,108	Oct 2018
Horse	EquCab2	rel. 94	26,991	21,546,500	Oct 2018

2.5.9 SNP array metadata

For each species, the metadata from the main commercial SNP arrays was downloaded from the *SNPchiMp* (v. 3) database (Nicolazzi et al., 2014), which links commercial SNPchip IDs to their respective dbSNP RefSeq IDs.

Table 2.8 Summary of commercial SNP arrays

Species	Company	Product	SNPs
Cattle	Illumina	BovineHD BeadChip	777,962
Horse	Illumina	Infinium PorcineSNP60 v2	61,565

2.5.10 Selection analysis

An analysis of the strength of selection, and the age of the allele under selection, was performed for all GWAS variants using a customised version of the *Selection* (Schraiber et al., 2016) software package. The pipeline for performing this analysis is summarised in Figure 2.2.

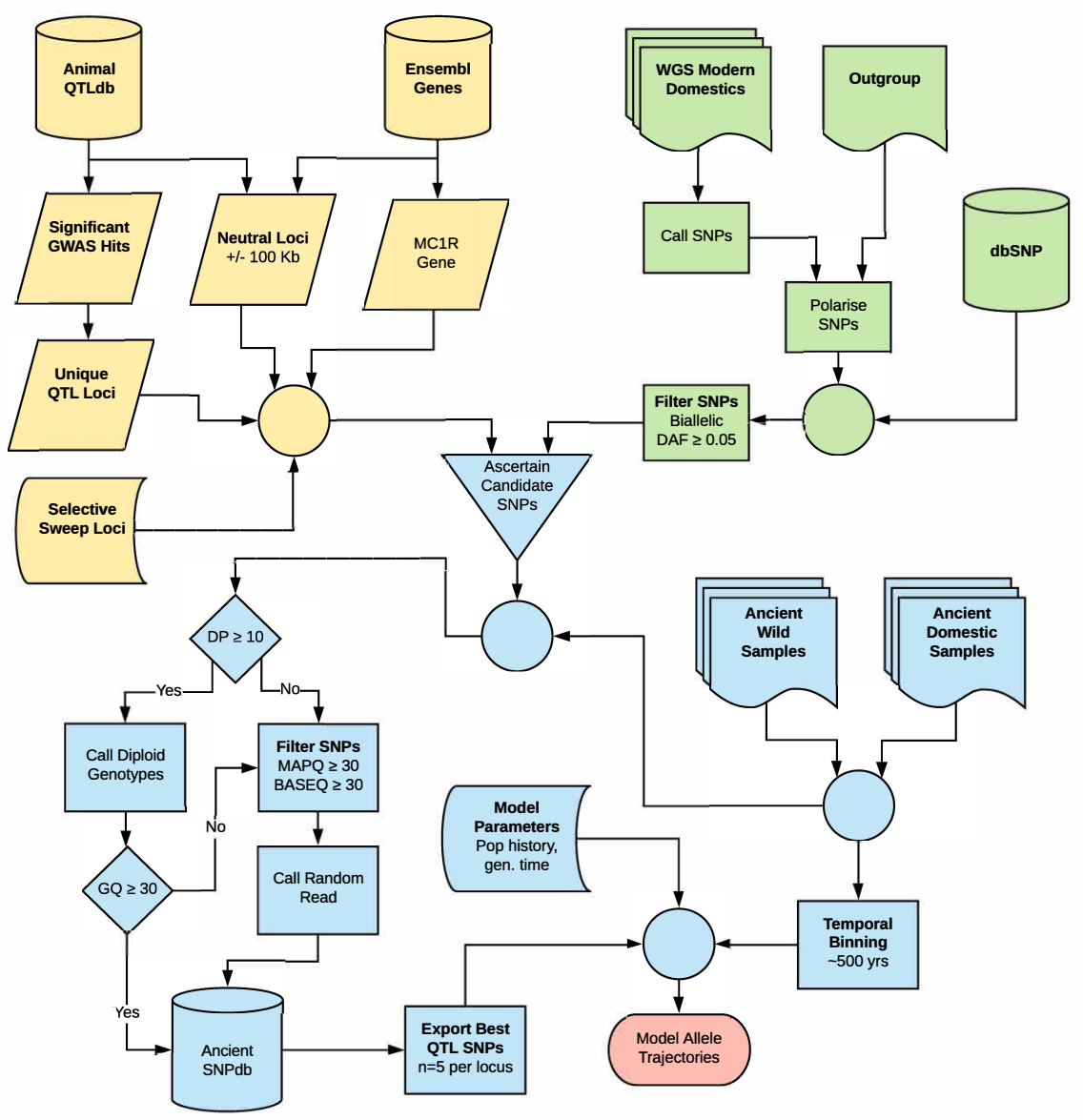


Figure 2.2 Flow chart summary of the pipeline to perform the selection analyses

2.5.10.1 Ancient SNP calling

A two-stage process was used for ancient variant calling, to extract the maximum amount of information from each sample. Sites with sufficient coverage to call unbiased diploid genotypes (i.e. ≥ 10x (Maruki and Lynch, 2017)) were genotyped with `bcftools`, and all other sites were called as pseudo-haploid genotypes—by randomly selecting a base from the pileup that passed all quality filters. This combined dataset is hereafter referred to as the ancient selection SNP callset.

2.5.10.1.1 Diploid genotyping

Diploid genotypes were called, for each sample individually, at sites with a read depth ≥ 10 using the `bcftools mpileup` and `bcftools call` workflow from BCFtools (v. 1.7) (Li, 2011), and the multiallelic calling model (`-m`) (Danecek et al., 2016). Diploid genotyping was only performed at sites corresponding to the polarised biallelic SNPs ascertained in the modern reference panels (see Section 2.5.3.10), and the calling algorithm was constrained to only call the two ascertained alleles at each site. Sites for which the diploid genotype could not be constrained were excluded from further processing. Sex based ploidy files were specified for each reference genome, along with the sex of each sample, to constrain the ploidy of calls made in sex chromosomes. Only sites with a minimum Phred-scaled genotype quality of 30 were retained for further processing. All diploid calls which passed these conditions were loaded into a local MySQL database for downstream processing.

2.5.10.1.2 Pseudo-haploid genotyping

Pseudo-haploid genotypes were called, for each sample individually, at sites with a read depth < 10, and for all sites which did not pass diploid genotype quality filters. For each target site, a random read was chosen from all reads that met the following conditions:

- Mapping quality ≥ 30
- Base quality ≥ 30
- Distance from read termini ≥ 5
- Base belongs to one of the two ascertained alleles at that site

All pseudo-haploid calls which passed these conditions were loaded into a local MySQL database for downstream processing.

2.5.10.2 Temporal binning of samples

For each ancient sample, date estimates were obtained from a mixture of direct and indirect radiocarbon dating, and cultural dating of archaeological contexts (Fages et al., 2019; Verdugo et al., 2019; Mullin et al. in prep). Each sample was grouped into

contiguous 500-year-long bins, based on the median age estimate of the sample. All modern samples were placed in a modern time bin, with no age uncertainty.

Table 2.9 Summary of cattle temporal bins

Bin #	Upper Range	Lower Range	Number of Samples	Pooled DoC
0	0 BP	0 BP	22	202.05x
1	1,000 BP	501 BP	17	19.92x
2	1,500 BP	1,001 BP	7	22.30x
3	2,000 BP	1,501 BP	8	19.98x
4	2,500 BP	2,001 BP	9	14.44x
5	3,000 BP	2,501 BP	8	21.19x
6	3,500 BP	3,001 BP	10	6.02x
7	4,000 BP	3,501 BP	4	4.87x
8	4,500 BP	4,001 BP	10	8.78x
9	5,000 BP	4,501 BP	12	27.11x
10	5,500 BP	5,001 BP	3	3.61x
11	6,000 BP	5,501 BP	2	2.07x
12	6,500 BP	6,001 BP	1	0.03x
13	7,000 BP	6,501 BP	16	11.96x
14	7,500 BP	7,001 BP	5	1.51x
15	8,000 BP	7,501 BP	8	14.81x
16	8,500 BP	8,001 BP	3	9.81x

For cattle, this resulted in 16 contiguous ancient time bins spanning the period from 500 BP to 8,500 BP, with a median of 8 samples per bin, and a median pooled coverage of 10.89x per bin.

Table 2.10 Summary of horse temporal bins

Bin #	Upper Range	Lower Range	Number of Samples	Pooled DoC
0	0 BP	0 BP	30	476.62x
1	500 BP	1 BP	17	46.78x
2	1,000 BP	501 BP	10	22.98x
3	1,500 BP	1,001 BP	43	109.40x
4	2,000 BP	1,501 BP	47	89.91x
5	2,500 BP	2,001 BP	31	87.05x
6	3,000 BP	2,501 BP	17	26.27x
7	3,500 BP	3,001 BP	16	52.30x
8	4,000 BP	3,501 BP	11	13.44x
9	4,500 BP	4,001 BP	3	1.28x
10	5,000 BP	4,501 BP	8	37.58x
11	5,500 BP	5,001 BP	28	107.43x
12	14,500 BP	14,001 BP	1	0.01x
13	16,000 BP	15,501 BP	1	6.04x
14	24,000 BP	23,501 BP	1	5.11x
15	32,000 BP	31,501 BP	1	0.01x
16	36,000 BP	35,501 BP	1	0.47x
17	43,000 BP	42,501 BP	1	21.98x

For horses, this resulted in 17 non-contiguous ancient time bins spanning the period from 1 BP to 43,000 BP, with a median of 10 samples per bin, and a median pooled coverage of 22.98x per bin.

2.5.10.3 Demographic history

The demographic history of each species was parameterised for *Selection* using the best fitting ðaði model for the autosomal modern polarised biallelic SNP callset (Section 2.5.3.10). The output from ðaði was converted into the population history file format

used by *Selection*, by cumulatively summing the ðaði epoch durations and appending a final epoch of infinite duration with an effective population size of 1.

2.5.10.4 Allele frequency estimates

For each modelled variant, input files for *Selection* were created by querying the ancient selection SNP callset. For each input file, a count of observations of the ancestral and derived alleles was made across all samples in each time bin. The age range for each bin was converted from chronological time into diffusion units, via the equations given below. Lastly, observations of the ancestral and derived alleles from the modern polarised biallelic SNP callset were appended to the input files, with an age of 0 diffusion units and no time uncertainty. SNPs with sample observations in less than three time bins were excluded from further processing.

In ðaði, time is measured in units of $\tau = t \,/\, (2Nref)$, where t is the time in generations and $Nref$ is the reference effective population size—used as a scaling unit for all other population size parameters (Gutenkunst et al., 2009). To solve for $Nref$, use $Nref = \theta \,/\, 4\mu L$, where θ is Watterson's estimator (Watterson, 1975) (estimated directly from the data using ðaði), μ is the per site mutation rate, and L is the length of sequence used to estimate the SFS, accounting for sequence dropout from mapping issues and quality filtering. To convert generations to chronological time, an estimate of the average generation time for each species was also used.

Table 2.11 Mutation rates and generation times for each species

Species	Mutation rate	Generation time	Reference
Cattle	1.260×10^{-8}	6 years	(Chen et al., 2018)
Horse	7.242×10^{-9}	8 years	(Schubert et al., 2014)

2.5.10.5 Identifying mispolarised SNPs

Sites with a high chance of being mispolarised were identified in the ancient selection SNP callset, via the following procedure. Firstly, we define $DAF_{ancient}$ to be the derived

allele frequency of the ten oldest ancient haploid observations for each SNP. Secondly, we define DAF_{recent} to be the derived allele frequency of the ten most recent ancient haploids (i.e. exclusive of all modern samples). SNPs were then flagged as potentially mispolarised which met the following two conditions:

- $DAF_{ancient} \geq 0.80$
- $DAF_{ancient} > DAF_{recent}$

2.5.10.6 Pairing GWAS variants with neutral controls

All GWAS variants were paired with a unique set of neutral controls. To be considered neutral, SNPs were required to have been ascertained in the modern polarised biallelic SNP callset, have a dbSNP RefSeq ID, and meet the following conditions:

- ≥ 100 Kb from the nearest GWAS peak
- ≥ 100 Kb from the nearest gene
- ≥ 100 Kb from the nearest selective sweep

Potential neutral SNPs were paired with GWAS variants, based on the following minimum conditions:

- Same chromosome
- Same ancestral and derived alleles
- Same modern DAF (rounded to two decimal places)
- Not flagged as mispolarised

To find SNPs with the most similar coverage across all time bins, within the set of potential neutral SNPs, we calculate the sum of the squared difference in the number of haploid observations per time bin, divided by the number of time bins, and rounded to the nearest integer.

$$f(x) = \left\lfloor \left(\sum_{i=0}^{n} (G_i - N_i)^2 \right) / n \right\rceil$$

Where n is the number of time bins, G_i is the count of ancient haploids for the GWAS SNP in the i^{th} bin, N_i is the count of ancient haploids for the neutral SNP in the i^{th} bin.

The final set of neutral SNPs were then chosen based on the following sort conditions:

- $min_{\forall x \in S}\ f(x)$, where S is the set of all SNPs considered
- Random sort order

2.5.10.7 Modifications to the Selection *software*

To account for the uncertainty of the age of the samples within a time bin, and the ascertainment scheme used to select variants, modifications were made by Dr Joshua G. Schraiber (Temple University, Philadelphia) to the published version of the *Selection* (Schraiber et al., 2016) software package. The software was extended to allow sample input files to express a date range for each set of allele frequency observations, and for these sampling dates to be jointly estimated as parameters of the Markov Chain Monte Carlo (MCMC) model. The software was further extended to account for ancient population structure, using the Balding–Nichols model (Balding and Nichols, 1995), and to explicitly model the ascertainment scheme used in this study. Specifically, that the derived allele is observed at least once in the ancient data, and all variants found within the modern reference panel have a minimum DAF of 0.05. Additionally, I made various bug fixes to the C++ code, adjusted proposal frequencies and tuning, implemented speed improvements and added gzip compression of the MCMC output. The updated code is publicly available in a forked GitHub repository (https://github.com/ekirving/selection).

2.5.10.8 Modelling selection

For each modelled variant, *Selection* was run six times, with different starting seeds, a burn-in of 50%, and the following flags:

- Infer allele age (`-a`)
- Model ascertainment of DAF ≥ 0.05 (`-A 0.05`)
- Run for 1×10^7 MCMC cycles (`-n 10000000`)
- Sample every 100$^{\text{th}}$ value from the posterior (`-s 100`)

- Assume an additive model (`-h 0.5`)
- Log every 1000th iteration (`-f 1000`)

The majority of the models were run on the University of Oxford's Advanced Research Computing (ARC) cluster (Richards, 2015), which provided free access to their computing resources.

2.5.10.9 MCMC Diagnostics

For each MCMC run, parameter mixing was assessed by calculating the effective sample size (ESS) for each parameter, and plotting the ESS vs. burn in, autocorrelation and traces, using the CODA R package (Best et al., 1995). Convergence was assessed between pairs of independent replicates, by calculating the potential scale reduction factors (PSRF) for each parameter (Gelman and Rubin, 1992) and the multivariate PSRF across all parameters (Brooks and Gelman, 1998), using CODA.

2.6 Results

2.6.1 Demographic modelling

2.6.1.1 Site frequency spectra (SFS)

For all species, running ðaði on the unfolded SFS produced unstable results across the 1,000 independent replicates. Comparisons between the folded and unfolded SFS showed a high degree of discordance between their respective maximum likelihood parameters. Under perfect conditions, with no mispolarisation error, it is expected that ðaði results from the unfolded SFS would be broadly comparable to those obtained from the folded SFS. As this was not the case, these results suggest an excess of mispolarised SNPs in the unfolded SFS. Consequently, only demographic histories estimated from the folded SFS were used for downstream analyses.

2.6.1.2 Cattle demography

For cattle, the four-epoch model provided the best fit for the folded SFS, with an AIC of 3035.46, and a relative likelihood $> 5 \times 10^{15}$ greater than the next best model (see Table 2.12), using the equation $reL = exp\big((AIC_{min} - AIC_i) / 2\big)$ (Wagenmakers and Farrell, 2004).

Table 2.12 ðaði results for cattle using the folded SFS

Epochs	Log-likelihood (lnL)	AIC	Relative likelihood (relL)
1x	-3017.6582	6039.31	0.0
2x	-1569.4508	3146.90	6.34×10^{-25}
3x	-1562.7918	3137.58	6.69×10^{-23}
4x	-1509.7337	3035.46	1
5x	-1540.6473	3101.29	5.08×10^{-15}

The best fitting model has a pre-domestication effective population size (Ne) of 233,735, which was reduced in an instantaneous bottleneck (by a factor of 21) to an Ne of 11,123 around ~7,000 years ago.

Table 2.13 Cattle demography with 4 epochs

Epoch	Effective population size (Ne)	Epoch end date
1	11,123	7,282 BP
2	233,735	16,402 BP
3	23,353	110,821 BP
4	47,099	593,596 BP
∞	19,806	-Inf

2.6.1.3 Horse demography

For horses, the five-epoch model provided the best fit for the folded SFS, with an AIC of 411.38, and a relative likelihood $> 1 \times 10^8$ greater than the next best model (see Table 2.14).

Table 2.14 ðaði results for horses using the folded SFS

Epochs	Log-likelihood (lnL)	AIC	Relative likelihood (relL)
1x	-34748.314	69500.62	0.0
2x	-501.17867	1010.35	8.59×10^{-131}
3x	-218.71305	449.42	5.48×10^{-9}
4x	-215.77722	447.55	1.40×10^{-8}
5x	-195.69056	411.38	1

The best fitting model has a pre-domestication Ne of 139,114, which was reduced in an instantaneous bottleneck (by a factor of 14) to an Ne of 9,981 around ~4,000 years ago.

Table 2.15 Horse demography with 5 epochs

Epoch	Effective population size (Ne)	Epoch end date
1	9,981	4,303 BP
2	139,114	78,553 BP
3	63,558	427,796 BP
4	1,562,401	1,163,389 BP
5	46,154	1,742,305 BP
∞	17,150	-Inf

2.6.2 Selection analysis

2.6.2.1 Identifying mispolarised SNPs

For all species, potentially mispolarised SNPs were identified by flagging SNPs with both a high ancient DAF and a negative gradient relative to the recent past (see Section 2.5.10.5). Using this conservative measure, many SNPs with a high modern DAF were flagged as mispolarised (see Table 2.16 and Table 2.17). This potentially high rate of mispolarisation is consistent with the unstable ðaði results observed when using the unfolded SFS. Consequently, all SNPs flagged as mispolarised were excluded from downstream processing due to the high chance of error.

Table 2.16 Frequency of putatively mispolarised SNPs for cattle

DAF	# SNPs	# Mispolarised SNPs	Frequency
≤ 0.05	2,828,744	238	0.0001
≤ 0.10	1,568,155	1,452	0.0009
≤ 0.15	1,036,715	2,572	0.0025
≤ 0.20	783,847	4,144	0.0053
≤ 0.25	882,893	9,340	0.0106
≤ 0.30	520,380	9,405	0.0181
≤ 0.35	446,244	12,802	0.0287
≤ 0.40	404,853	16,444	0.0406
≤ 0.45	394,208	21,713	0.0551
≤ 0.50	510,460	40,291	0.0789
≤ 0.55	292,632	30,957	0.1058
≤ 0.60	315,035	41,926	0.1331
≤ 0.65	312,356	49,760	0.1593
≤ 0.70	301,809	57,765	0.1914
≤ 0.75	421,271	98,757	0.2344
≤ 0.80	311,843	84,523	0.2710

DAF	# SNPs	# Mispolarised SNPs	Frequency
≤ 0.85	316,754	99,681	0.3147
≤ 0.90	357,687	125,113	0.3498
≤ 0.95	433,337	169,797	0.3918
≤ 1.00	951,789	456,319	0.4794

For cattle, between ~31%–48% of SNPs with a modern DAF ≥ 0.80 were flagged as putatively mispolarised. Interestingly, 1,690 SNPs with a modern DAF ≤ 0.10 were also flagged as putatively mispolarised, suggesting that there might be some mispolarised SNPs which underwent very strong selection during domestication.

Table 2.17 Frequency of putatively mispolarised SNPs for horses

DAF	# SNPs	# Mispolarised SNPs	Frequency
≤ 0.05	6,344,052	157	2.5e-5
≤ 0.10	2,360,330	187	0.0001
≤ 0.15	2,029,113	622	0.0003
≤ 0.20	878,073	796	0.0009
≤ 0.25	905,293	1,598	0.0018
≤ 0.30	438,337	1,765	0.0040
≤ 0.35	368,089	2,372	0.0064
≤ 0.40	427,650	4,849	0.0113
≤ 0.45	250,999	4,704	0.0187
≤ 0.50	324,049	8,091	0.0250
≤ 0.55	199,589	7,069	0.0354
≤ 0.60	193,382	8,888	0.0460
≤ 0.65	267,983	15,735	0.0587
≤ 0.70	185,438	13,617	0.0734
≤ 0.75	267,531	23,582	0.0881

DAF	# SNPs	# Mispolarised SNPs	Frequency
≤ 0.80	191,063	19,629	0.1027
≤ 0.85	203,194	21,736	0.1070
≤ 0.90	332,011	40,890	0.1232
≤ 0.95	264,587	34,713	0.1312
≤ 1.00	535,120	78,428	0.1466

For horses, between ~10%–15% of SNPs with a modern DAF ≥ 0.80 were flagged as mispolarised. Similar to the pattern seen in cattle, 344 SNPs with a modern DAF ≤ 0.10 were also flagged as putatively mispolarised.

2.6.2.2 Effective Sample Size (ESS)

For each MCMC run, parameter mixing was assessed by calculating the effective sample size (ESS) for each parameter. A minimum ESS threshold of 100 for each parameter was used as a cut-off for determining acceptable parameter mixing (see Table 2.18).

Table 2.18 Summary of ESS metrics

Species	MCMC chains	$ESS_{min} \geq 100$	Frequency
Cattle	9,185	2,712	0.2953
Horse	13,188	1,682	0.1275

Despite adjustments to chain length, burn in, thinning and the proposal frequency for each parameter, it was not possible to consistently raise the ESS for all models. Only 4,394 MCMC chains (19.64%) passed this threshold (including both GWAS and neutral SNPs).

2.6.2.3 Multivariate potential scale reduction factor (MPSRF)

To assess if MCMC runs which passed the minimum ESS threshold had also reached the standing distribution of the posterior, six independent replicates were run with different

starting seeds, and the Gelman–Rubin convergence diagnostic was computed for each pair of replicate chains. A minimum multivariate potential scale reduction factor (MPSRF) of ≤ 1.2 was used as a cut-off for assessing model convergence (Brooks and Gelman, 1998). Only 1,084 models (21%) had a pair of chains (out of 6 replicates) with passed both the ESS and MPSRF thresholds (including both GWAS and neutral SNPs) (see Table 2.19).

Table 2.19 Summary of MPSRF metrics

Species	MCMC Models	$\widehat{R^p} \leq 1.2$	Frequency
Cattle	2,750	698	0.2538
Horse	2,365	386	0.1632

For a complete list of the successful MCMC models for GWAS associated variants in cattle (n=408) and horses (n=183), including their dbSNP RefSeq IDs, chromosomes and positions in the relevant genome assembly, ancestral and derived alleles, and the minimum ESS and MPSRF for each model, see Table S2.5 and Table S2.6.

2.6.2.4 Maximum a posteriori

The results of each MCMC analysis are posterior probability distributions for all parameters, which can be summarised by computing the maximum *a posteriori* estimate of each parameter.

For cattle, negligible selection was detected at any successfully modelled SNP (maximum s1 = 1.36 x 10^{-4}). Of the ten SNPs with the largest maximum *a posteriori* selection coefficients, 3 were linked to milk traits (Iso-Touru et al., 2016; Olsen et al., 2016), 2 to reproduction traits (Fortes et al., 2013), 2 to production traits, 2 to exterior traits and 1 to meat and carcass traits (Berkowicz et al., 2012) (see Table 2.20). The average age of the allele across all modelled SNPs was ~125,000 BP.

Table 2.20 Ten largest maximum *a posteriori* selected traits for cattle

dbSNP ID	Class	Trait	s1	s2	Age BP	DAF
rs378261830	Milk	Milk fat yield	0.000132	0.000265	134,943	0.88
rs137567750	Milk	Milk fat percentage	0.000124	0.000247	147,549	0.88
		Milk protein percentage				
rs134839836	Reproduction	Age at puberty	0.000120	0.000239	137,737	0.91
		Scrotal circumference				
rs41623544	Production	Body depth	0.000115	0.000229	135,826	0.73
	Exterior	Stature				
	Production	Rump width				
	Exterior	Angularity				
	Meat and Carcass	Carcass weight				

For horses, moderate selection was detected for two SNPs (rs68689852 and rs68628642), with the remainder of the successfully modelled SNP showing negligible signs of selection. Of the ten SNPs with the largest maximum *a posteriori* selection coefficients, 5 were associated with the polygenic growth trait "Withers height" (Skujina et al., 2018), 2 were associated with the exterior trait "White markings" (Haase et al., 2013; Kim et al., 2017), and 1 each were associated with reproduction (Gottschalk et al., 2016), health (Metzger et al., 2012) and performance traits (Velie et al., 2018) (see Table 2.21). The average age of the allele across all modelled SNPs was ~695,000 BP.

Table 2.21 Ten largest maximum *a posteriori* selected traits for horses

dbSNP ID	Class	Trait	s1	s2	Age BP	DAF
rs68689852	Growth	Withers height	0.005025	0.010049	12,957	0.52
rs68628642	Reproduction	Number of progressively motile sperm	0.004965	0.009930	12,268	0.40
rs68619452	Growth	Withers height	0.000163	0.000326	13639	0.17
rs68967275	Growth	Withers height	0.000082	0.000164	646,778	0.79
rs68591872	Exterior	White markings	0.000082	0.000163	626,131	0.81
	Health	Guttural pouch tympany				
rs69602496	Growth	Withers height	0.000078	0.000157	692,900	0.72
rs68714332	Growth	Withers height	0.000077	0.000154	640,891	0.81
rs68713389	Exterior	White markings	0.000076	0.000153	700,918	0.67
rs397128183	Performance	Racing performance	0.000076	0.000153	751,509	0.73

2.6.2.5 Selection trajectories

To summarise the allele trajectories of the modelled variants, the maximum *a posteriori* estimate of the path was plotted against the observed allele frequencies in each temporal bin (e.g. Figure 2.3). The black line shows the maximum *a posteriori* path, and the red and green lines show the 65% and 95% credible intervals. The blue line shows the posterior for the age of the allele, which for some models is outside the 25,000-year range shown in the plots. Grey dots represent estimated allele frequencies in the past, and the dot size is proportional to the $\log_{10}$ of the sample size in that temporal bin.

2.6.2.5.1 Cattle

The SNP with the largest maximum *a posteriori* selection coefficient in cattle (s1 = 1.36 x 10^{-4}) was rs378261830 (see Figure 2.3), which has been associated with milk fat yield in Nordic Red cattle (Iso-Touru et al., 2016). This GWAS study genotyped 4,280 Nordic Red cattle bulls on the Illumina BovineSNP50 BeadChip, then imputed them to whole genomes using a two-step process. Firstly, they imputed the BovineSNP50 array to the larger 777K BovineHD array, and then imputed the BovineHD array into whole genomes (Brøndum et al., 2014; van Binsbergen et al., 2014) using data from the 1000 Bull Genomes Project (Daetwyler et al., 2014; Hayes et al., 2014). SNP rs378261830 was one of 3,594 variants significantly associated with milk fat yield—using a Bonferroni corrected threshold of $-log_{10}(p) \geq 8.50$ (Iso-Touru et al., 2016). The MCMC results for this SNP suggest it slowly rose to high frequency (DAF = 0.88) over ~135,000 years.

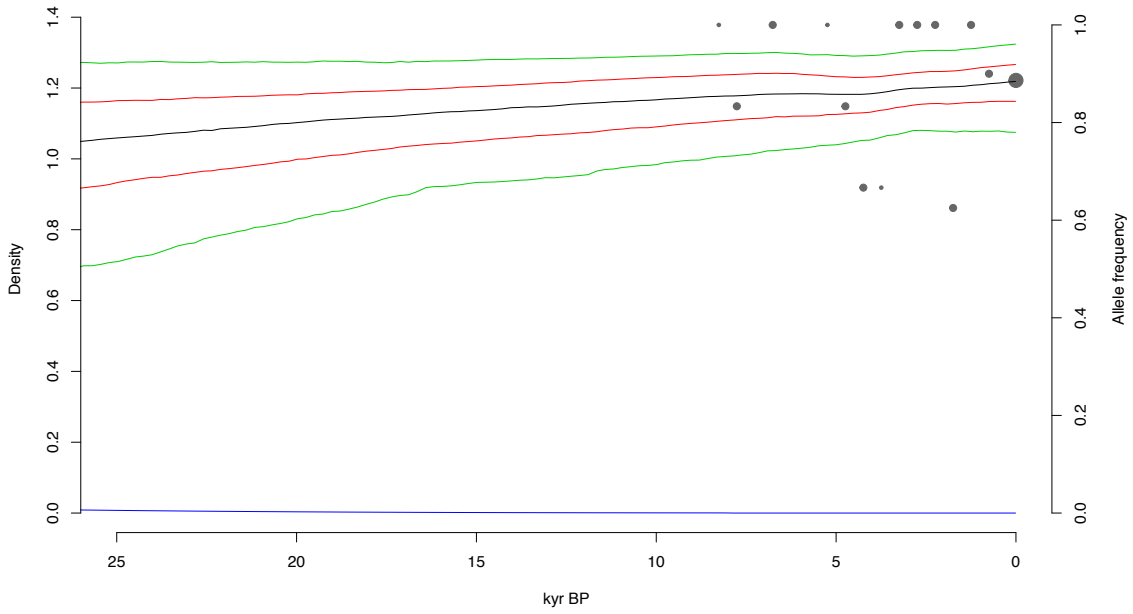


Figure 2.3 Selection trajectory for 'Milk fat yield' (rs378261830) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the log_{10} of the sample size.

The SNP with the second largest maximum *a posteriori* selection coefficient in cattle (s1 = 1.24 x 10^{-4}) was rs137567750 (see Figure 2.4), which has recently been associated with both milk fat percentage and milk protein percentage (Olsen et al., 2016). This fine mapping study investigated a QTL in chromosome 6, previously associated with clinical mastitis and milk production (Nilsen et al., 2009). Olsen et al. (2016) genotyped 3,096 Norwegian Red cattle, on a combination of the Affymetrix 25K SNP array, and the Illumina BovineSNP50 and BovineHD arrays, and imputed them to full sequence within the QTL locus. SNP rs137567750 was one of 220 variants significantly associated with milk fat percentage, and one of 381 variants significantly associated with milk protein percentage (after Bonferroni correction). The MCMC results for this SNP suggest it slowly rose to high frequency (DAF = 0.88) over ~148,000 years.

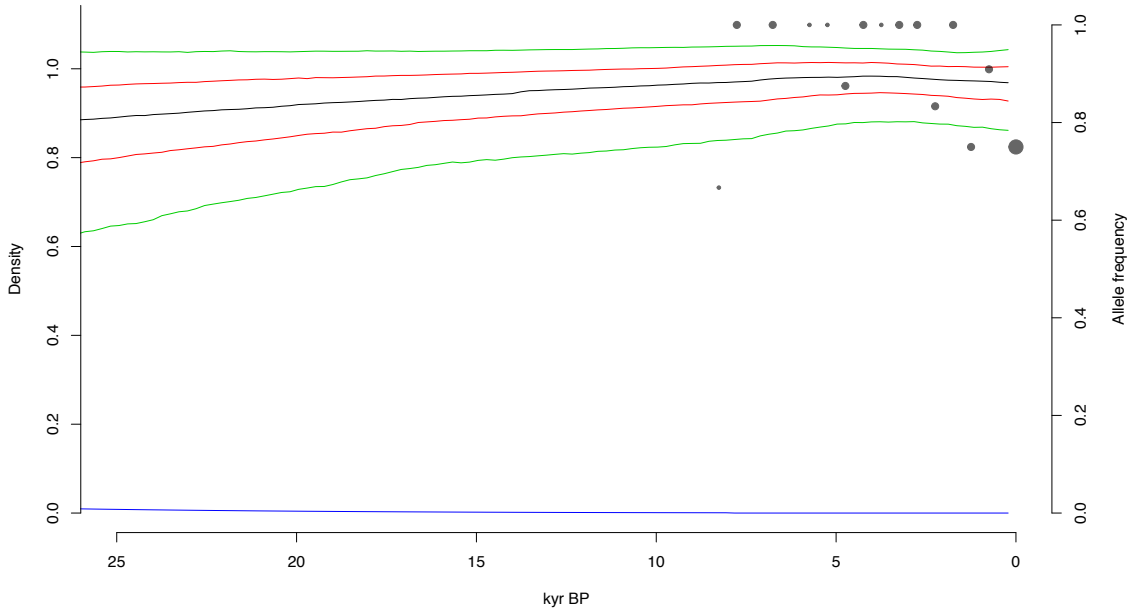


Figure 2.4 Selection trajectory for 'Milk fat percentage' and 'Milk protein percentage' (rs137567750) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.

The SNP with the third largest maximum *a posteriori* selection coefficient in cattle (s1 = 1.20 x 10^{-4}) was rs134839836 (see Figure 2.5), which has been associated with age at puberty and scrotal circumference (Fortes et al., 2013). This GWAS study genotyped 1,085 Tropical Composite bulls, on the BovineSNP50 array, and imputed them to the BovineHD array. SNP rs134839836 was one of 8,490 variants significantly associated with age at puberty, and one of 8,101 variants significantly associated with scrotal circumference (after Bonferroni correction). The MCMC results for this SNP suggest it slowly rose to high frequency (DAF = 0.91) over ~138,000 years.

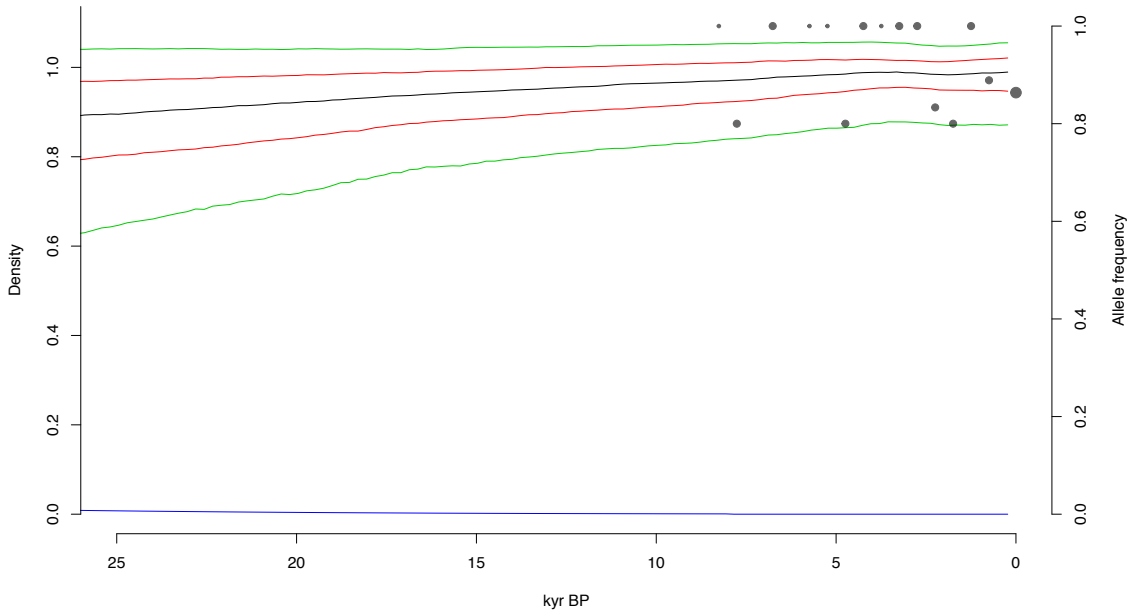


Figure 2.5 Selection trajectory for 'Age at puberty' and 'Scrotal circumference' (rs134839836) in cattle. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.

2.6.2.5.2 Horses

The SNP with the largest maximum *a posteriori* selection coefficient in horses (s1 = 5.025 x 10^{-3}) was rs68689852 (see Figure 2.6), which has been associated with withers height (Skujina et al., 2018). This GWAS study genotyped 105 horses from 4 breeds of British Isles ponies on the Illumina EquineSNP50 array. SNP rs68689852 was one of 402 variants significantly associated with withers height (after Bonferroni correction). The MCMC results for this SNP suggest it rose, under moderate selection, to mid frequency (DAF = 0.52) over ~13,000 years.

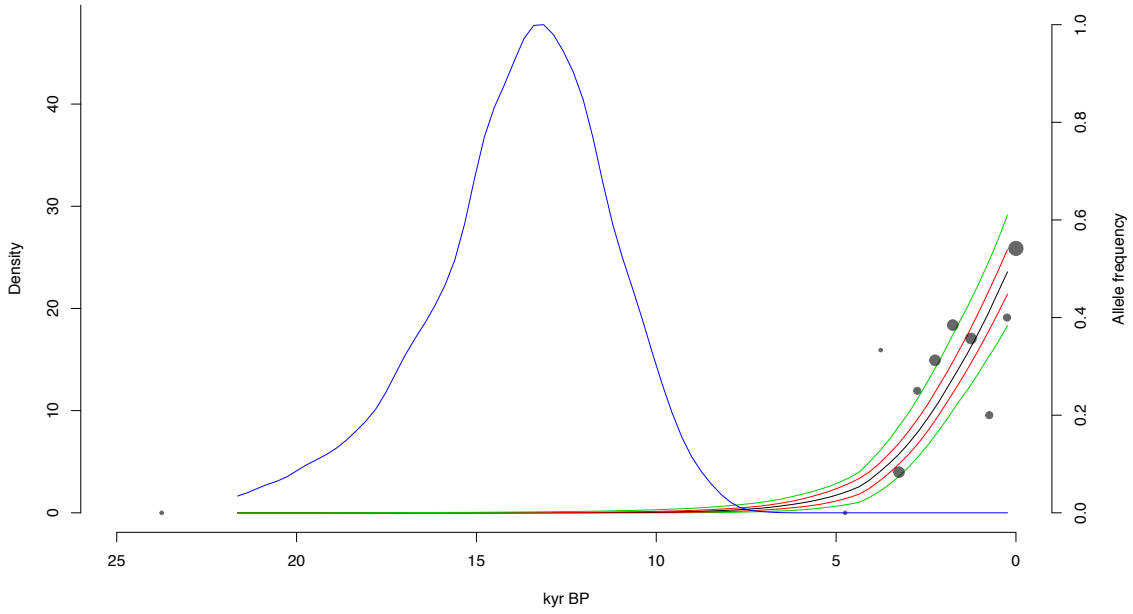


Figure 2.6 Selection trajectory for 'Withers height' (rs68689852) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.

The SNP with the second largest maximum *a posteriori* selection coefficient in horses (s1 = 4.965 x 10^{-3}) was rs68628642 (see Figure 2.7), which has been associated with the number of progressively motile sperm (Gottschalk et al., 2016). This GWAS study genotyped 109 German Warmblood stallions on the Illumina EquineSNP50 array. SNP rs68628642 was one of 29 variants significantly associated with the number of progressively motile sperm (after applying the max (T) permutation procedure (Rempala and Yang, 2013)). The MCMC results for this SNP suggest it rose, under moderate selection, to mid frequency (DAF = 0.40) over ~12,000 years.

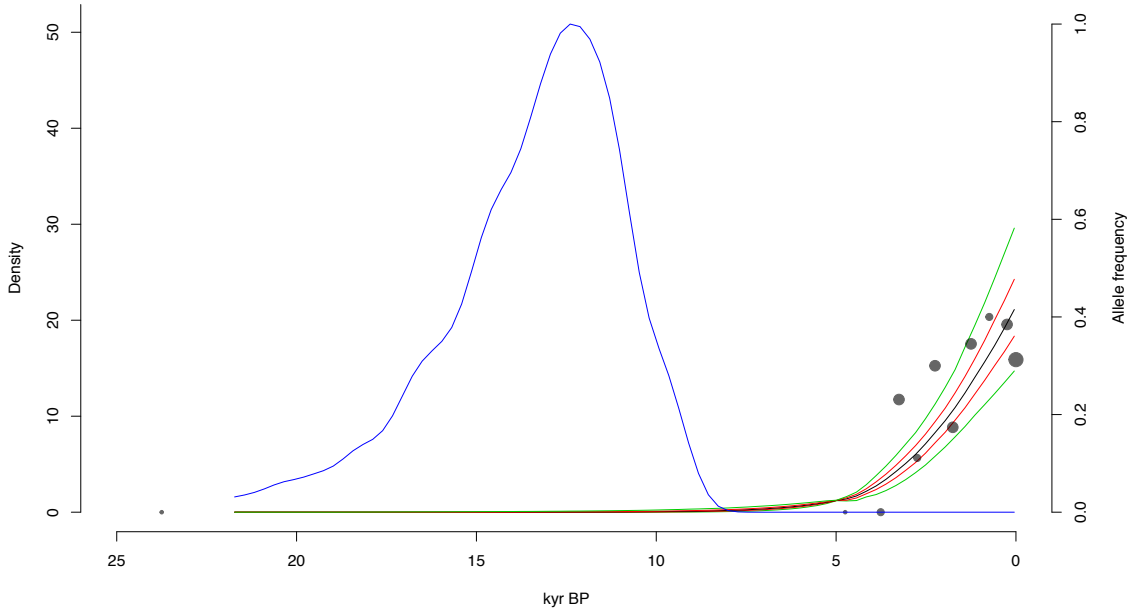


Figure 2.7 Selection trajectory for 'Number of progressively motile sperm' (rs68628642) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.

The SNP with the third largest maximum *a posteriori* selection coefficient in horses (s1 = 1.63 x 10^{-4}) was rs68619452 (see Figure 2.8), which has been associated with withers height in British Isles ponies (Skujina et al., 2018). The MCMC results for this SNP suggest it gradually rose to medium-low frequency (DAF = 0.17) over ~14,000 years.

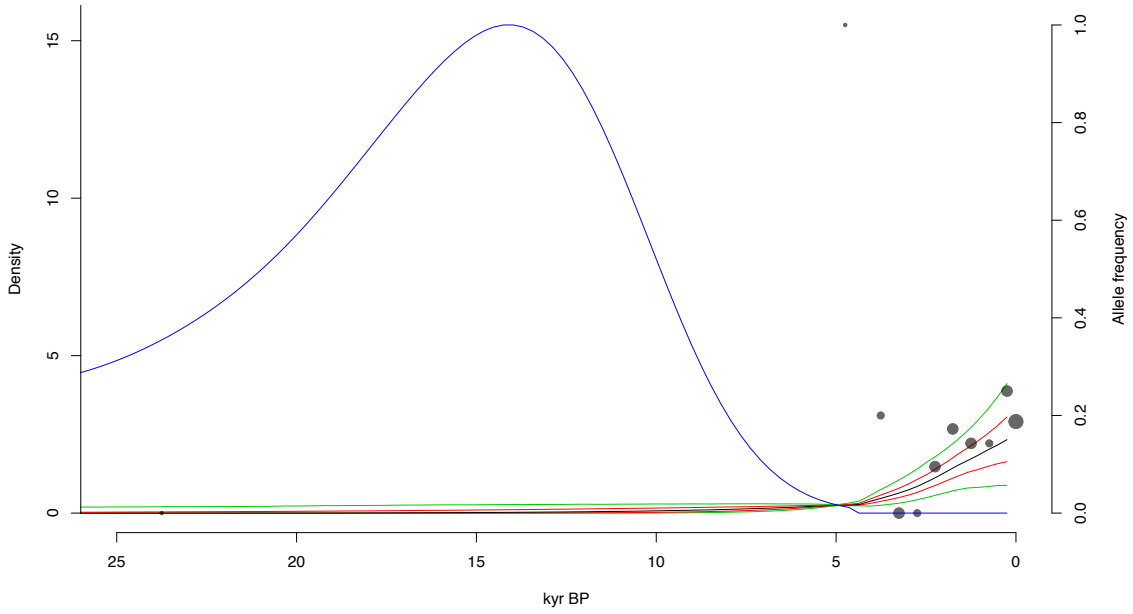


Figure 2.8 Selection trajectory for 'Withers height' (rs68619452) in horses. The black line is the maximum *a posteriori* path, the red and green lines are the 65% and 95% credible intervals, and the blue line is the posterior for the age of the allele. Grey dots show estimated DAF in each time bin, and sizes are proportional to the $\log_{10}$ of the sample size.

2.6.3 Joint posterior distributions

To summarise the strength of polygenic selection for all successfully modelled traits, we plotted the joint posterior distribution of the selection coefficient s_1 (heterozygous fitness) for all variants, grouped by their trait association (see Figure 2.9, Figure 2.10 and Figure 2.11). Neither cattle nor horses show any consistent signal of polygenic selection for the traits modelled in this study.

2.6.3.1 Cattle

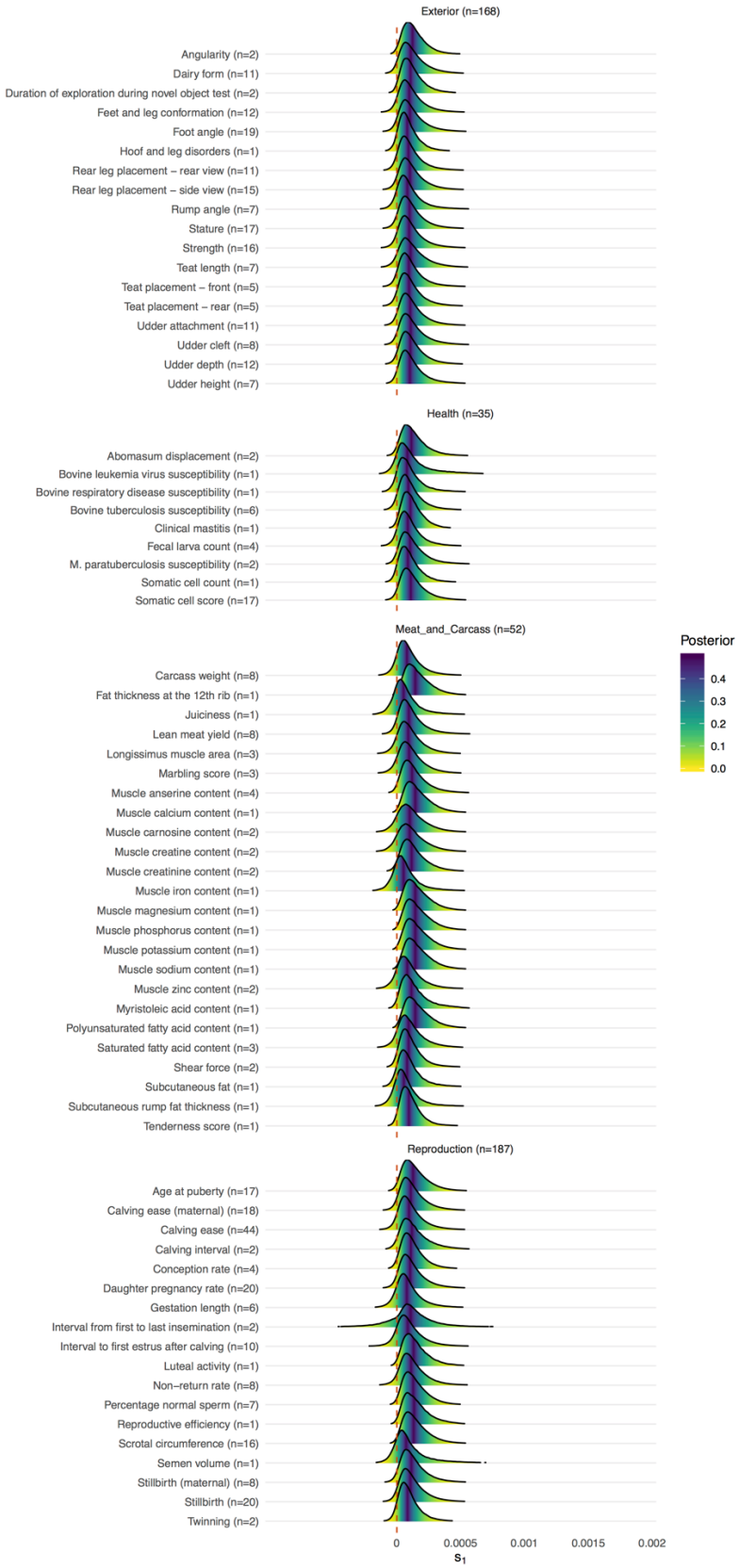


Figure 2.9 Joint posterior distribution of the selection coefficient s_1 for Exterior, Health, Meat and Carcass, and Reproduction traits in cattle

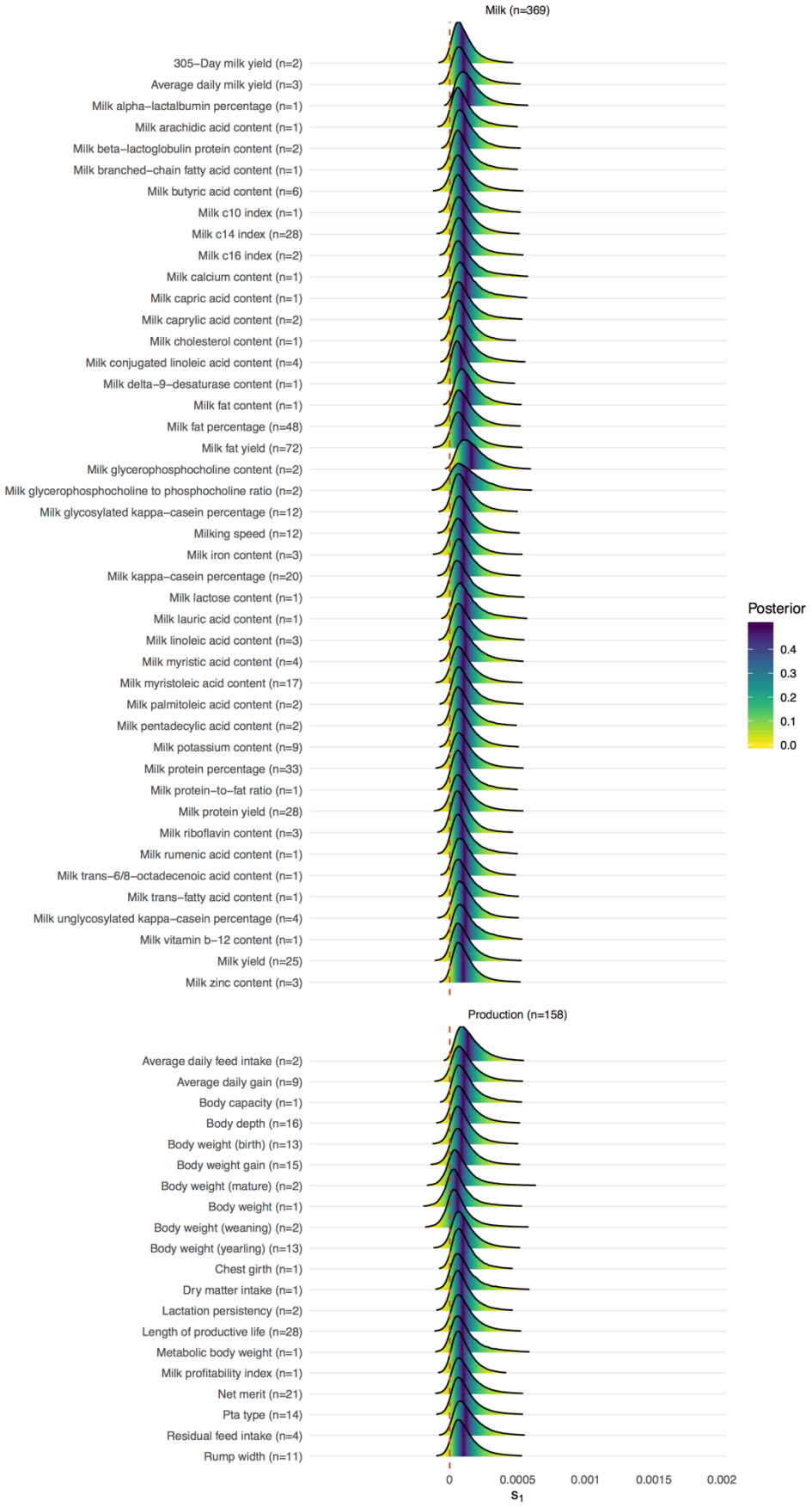


Figure 2.10 Joint posterior distribution of the selection coefficient s_1 for Milk and Production traits in cattle

2.6.3.2 Horses

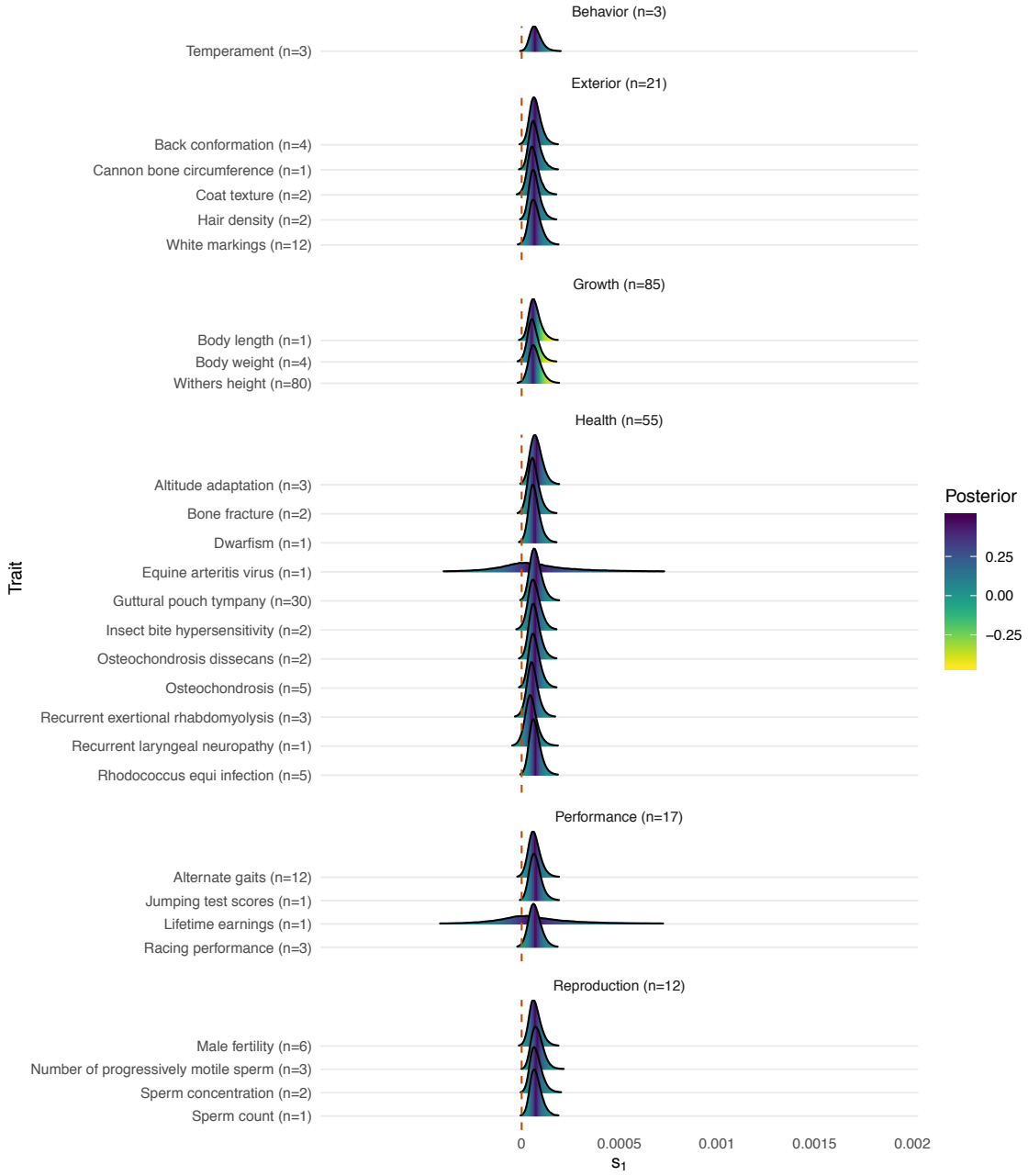


Figure 2.11 Joint posterior distribution of the selection coefficient s_1 for Behaviour, Exterior, Growth, Health, Performance and Reproduction traits in horses

2.7 Discussion

2.7.1 SNP polarisation

Due to the high rate of putatively mispolarised SNPs, many traits and associations of interest could not be reliably modelled because the ancestral allele could not be

determined. The maximum parsimony approach used here, of a single outgroup sample to polarise each allele, has several limitations. Firstly, it can mispolarise SNPs which are homozygous in the outgroup individual but segregating at the population or species level. This is especially problematic if the outgroup individual has low genetic diversity, as would be expected among marginalised wild populations. Secondly, it cannot polarise variants where the outgroup is heterozygous, leading to dropout of these variants from the polarised callset. Thirdly, it cannot polarise variants in regions where the outgroup reads cannot be aligned to the ingroup reference. Similarly, mapping issues can cause allelic dropout, making the outgroup appear falsely homozygous and leading to further mispolarisation.

Even with the conservative measure used here to exclude putatively mispolarised SNPs (see Section 2.5.10.5), it is likely that many remain in the dataset. For example, rs137567750 (see Figure 2.4) looks suspiciously like a mispolarised SNP which rose under mild selection from 0.0 to 0.22 frequency, over the last ~10 Kya. It is likely that undetected mispolarisation is a contributing factor to the overall low rate of model convergence.

An alternative, and less principled approach to polarisation, used by Ye et al. (2017) in their Schraiber et al. (2016) modelling, is to assume that the ancestral allele always increases in frequency between the first and last sampled time points, and to use that assumption to polarise the alleles. However, such an approach would systematically bias inference toward selection, by flipping neutral alleles which are drifting towards lower frequency.

Recently, a more sophisticated method for polarising variants was published, which incorporates up to three outgroup species, and fits an explicit model of DNA evolution (Keightley and Jackson, 2018). Future work might consider using such an approach to reduce the high rate of mispolarisation, which would (i) permit demographic inference using the unfolded SFS; (ii) increase the number of variants eligible for modelling in the ancient selection SNP callset; and (iii) improve model convergence by fixing presently undetected mispolarisation.

2.7.2 Sample binning

To allow us to estimate allele frequencies at discrete time points, samples were grouped into contiguous 500-year-long bins, based on the median age estimate of the sample (see Section 2.5.10.2). The results of this binning scheme are frequency observations that can vary substantially in sample size and produce noisy estimates of the ancestral allele frequency. An alternative approach, used in Fages et al. (2019), is to smooth the allele frequency estimates using a sliding window approach. They calculated allele frequencies over 1,000 year sliding windows, with a step size of 250 years. This approach counts each sample multiple times, as every sample falls into four overlapping bins. Future work might consider applying a sliding window approach, especially since the modified version of the Schraiber et al. (2016) method now explicitly models the age uncertainty of the samples. This would allow the overlapping nature of the bins to be explicitly modelled by the MCMC.

2.7.3 MCMC modelling

Overall, the MCMC models mixed very poorly, which greatly limited the number of interpretable results. Most of the MCMC chains (n=19,777; 73.62%) had below threshold ESS values for at least one parameter, as well as exhibiting generally high variance in ESS between replicate runs. This poor mixing was further evidenced by high MPSRF scores, resulting in most of the models (n=3,381; 75.72%) failing to achieve the target convergence metric, excluding them from further analysis.

Extensive testing of the MCMC hyperparameters, including (i) chain length; (ii) parameter update proposal frequency; (iii) path granularity; (iv) path update fraction; and (v) the sigma values for proposal tuning; did not lead to consistent improvements across the subset of models used for hyperparameters testing. This suggests that the systematic issues with model convergence are not related to misconfiguration of the MCMC hyperparameters.

Manual inspection of hundreds of parameter traces showed that even when chain length is increased to 100 million iterations (i.e. 100x longer than used by Schraiber et al. (2016) and Ye et al. (2017)), most models failed to escape the local maxima near to their randomly drawn starting parameters. Parameter mixing was improved when models were run with a constant population size demographic model; suggesting that the complicated demographic history of these species is a contributing factor to the poor mixing. Schraiber et al. (2016) report that under a misspecified demographic model the posterior distributions of selection coefficients can be biased. However, these biases have been shown to be relatively mild in effect (Jewett et al., 2016). In practice, other publications using the Schraiber et al. (2016) method have used a constant population demographic model (e.g. Gelabert et al., 2017; Ye et al., 2017). Future work might consider firstly modelling all variants using a constant population size and reserving the use of the fully specified demography for only those variants were significant selection is detected.

To further address the poor parameter mixing and convergence, future work might consider an implementation of the Metropolis-Coupled MCMC (MC^3) algorithm, which uses heated chains to better traverse widely dispersed local maxima (Geyer, 1991). Efficient parallelised algorithms for MC^3 have been used in phylogenetics for some time (Aberer et al., 2014; Altekar et al., 2004), and their adaptation to this problem should improve the current problem where models become quickly stuck in local optima, and cannot traverse the flat likelihood surface in the surrounding regions.

Once mixing has been improved, future work might also consider replacing the arbitrary minimum ESS threshold with a more principled multivariate ESS metric—which accounts for the cross-correlations between parameters (Vats et al., 2015). Here, we use a minimum ESS threshold of 100, whereas Schraiber et al (2016) use both 50 and 150, and Ye et al. (2017) use 50. Instead, Vats et al. (2015) provide formulae for calculating an *a priori* minimum multivariate ESS threshold ($E\hat{S}S$), based on the number of parameters in the model and the acceptable level of error in the estimator. This metric has recently been implemented in the *mcmcse* (v. 1.3-2) R package (Flegal et al., 2012). The $E\hat{S}S$

metric could, in principle, be implemented as a runtime stopping condition in the MCMC, such that instead of specifying the number of iterations in the chain, the model runs until it reaches either (i) the specified multivariate ESS threshold, or (ii) the acceptable level of error.

A separate cause of poor model convergence is the long default length of the trajectory path updates. Currently, the portion of the trajectory that is updated in each MCMC proposal is configured as a fraction of the total path length (default `-F 20`, or 5% of the path). When the path is long (i.e. the allele is old), or when there is dense sampling of the allele, this fractional length of the trajectory can easily span the entirety of the ancient observations. For horses, the mean maximum *a posteriori* inferred age of the allele is ~695,000 years old, which corresponds to a default path update length of ~35,000 years. Trajectory updates to such a long segment of the path poorly fit the finer temporal resolution of domestication and reduce the speed at which the trajectory converges on the observed frequency changes seen for many SNPs. This is especially problematic when modelling sudden changes between the most recent ancient time bin and the modern era. Due to the intensive nature of modern selective breeding, DAF_{modern} can often be substantially higher (or lower) than DAF_{recent}. As these movements occur across a single time bin, with a 500-year resolution, a proposal update length of ~35,000 years is disproportionate. However, when the path update size is decreased (e.g. `-F 1000`, or 0.1% of the path), the model requires a much longer chain length for the path to converge, as each update moves a much smaller portion of the trajectory. Additionally, the model adds by default (`-M 10`) an extra 0.01 diffusion units onto the fractional path update length—equating to 2,377 years for cattle and 2,744 years for horses. Consequently, even for very large values of `-F`, the minimum update size prevents updates shorter than ~2.5 thousand years.

A potential solution to this problem would be to tune the length of the trajectory updates, in a manner similar to the tuning of sigma values in the prior distributions of other parameters. For example, where the acceptance rate for trajectory updates falls below the target of 0.234 (Roberts et al., 1997), then the length of the update would be

shortened to attempt to increase the acceptance rate. As with other parameters, the step size of tuning changes would decay with chain length, such that tuning ceases in time to ensure proper MCMC convergence. This would address a commonly seen problem where the modelled trajectory fails to fit the observed changes in allele frequency because they occur too rapidly for long trajectory updates to properly accommodate.

Another potential improvement would be to increase the sample size of the modern reference panels. The extent to which any single estimate of the allele frequency constrains the posterior distribution of the trajectory is proportional to the number of haploid observations. Consequently, an increase in the number of modern samples will force the trajectory to converge on the high-confidence estimate of the modern frequency. This may help address the current problem where the inferred modern allele frequency can vary considerably from the empirical estimate. However, as DAF_{recent} and DAF_{modern} can differ substantially, future work might also consider replicating each model without DAF_{modern} to see how this changes the estimates of the parameters of interest.

2.7.4 GWAS associations

Modelling selection on variants ascertained via GWAS has the advantage that the trait association of each variant is already known—making it relatively straightforward to interpret selection coefficients without having to rely on Gene Ontology (GO) term enrichment analysis, or other *post hoc* methods of annotating putatively selected loci. However, there are also distinct limitations to this approach.

There is a systematic bias in the animal quantitative genetics literature towards traits with contemporary commercial relevance. Whilst it is plausible that many traits—especially those related to health, development, or reproduction—may also have been of interest to early farmers and pastoralists, many other traits have such modern specificity that their relevance outside of contemporary animal breeding is limited (e.g. lifetime earnings (Velie et al., 2018), or drip loss (Ponsuksili et al., 2014)). As such, many

GWAS traits are unlikely to have undergone selection outside of modern commercial breeding.

Within much of the animal quantitative genetics literature, GWAS associations have been performed in medium or high-density SNP array panels. The cost effectiveness of such an approach has facilitated large sample sizes, which in turn have increased statistical power to identify variants with small effect sizes. However, the use of SNP array data is reliant on the causal variant being in linkage disequilibrium (LD) with the observed genotype (Visscher et al., 2017). This dependency on LD can be problematic if haplotype structure differed substantially in the past (e.g. Frantz et al., 2019), in which case the trajectory of the modelled variant might differ greatly from that of the causal variant.

Similarly, many of the GWAS associations used here were ascertained in specific domestic breeds or populations, and it is unclear how robustly GWAS associations hold outside of the tested populations (e.g. Kim et al., 2018; Kuchenbaecker et al., 2019; Palmer and Pe'er, 2017). This is especially problematic in ancient individuals with potentially divergent genetic background. Epistasis is likely to be a major confounding factor in applying trait associations outside of their test populations. Similarly, gene–environment interactions (GxE) are likely to play a confounding role on trait associations as domestic animals spread outside of their natural ranges during the Neolithic expansion.

Another limitation of this approach is that GWAS can only be performed on traits which are variable in the test populations. This is a major issue for behavioural traits which are already fixed in modern domestic populations (e.g. loss of aggression). This makes them unsuitable for GWAS discovery unless wild or primitive populations are included in the studies. Consequently, behavioural traits represent a mere 0.02% of the total GWAS associations for livestock specues in the AnimalQTL database (Hu and Reecy, 2007).

2.7.5 Modelling assumptions

The Schraiber et al. (2016) method has two inference modes: (i) where the age of the allele is inferred, by modelling the complete trajectory, from the mutation of the allele to the present (`-a`); and (ii) where the trajectory of the allele is modelled only within the range of the sample observations, and the starting frequency of the allele is drawn from a uniform prior distribution. When inferring the age of the allele (`-a`), this method assumes that the selection coefficients apply uniformly across the history of the allele. As we are specifically sampling from within the timeframe of domestication, the assumption of constant selection through time will systematically bias the inferred age of the allele towards a more recent origin.

This method also assumes that our ancient samples are drawn from a single panmictic population with genetic continuity through time. Whilst care has been taken in selecting samples which best fit these assumptions—e.g. by only using horses from the lineage leading to modern domestics (Fages et al., 2019) and excluding cattle with Zebu admixture (Verdugo et al., 2019)—we cannot rule out the possibility of geographic or temporal structure creating spurious signals. Similarly, this method cannot model the effect of admixture from a divergent population (e.g. Frantz et al., 2019). If an introgressing population has a significantly different allele frequency, then changes to the ingroup frequency are sensitive to the degree of admixture, independently of selection.

Lastly, the Schraiber et al. (2016) method uses a diffusion approximation to the Wright–Fisher model (Fisher, 1930; Wright, 1931). It has been shown that in scenarios with skewed offspring distributions, inference based on the Wright–Fisher model can be strongly biased (Eldon and Wakeley, 2006; Matuszewski et al., 2018). The extent to which this is a problem among ancient domestic animal populations remains unclear, however, modern breeding practices suggest this might be a cause for concern—e.g. the use of sires in cattle and horse breeding. Recently, a time-series method has been developed to specifically model demography and selection under skewed offspring

distributions (Sackman et al., 2019), and future work might consider testing how such a model effects inference of selection in these datasets.

2.8 Conclusions

In this study we modelled the allelic trajectory of >2,000 GWAS variants linked to quantitative traits in cattle and horses. Using a dataset of 369 ancient nuclear genomes, spanning >10,000 years of evolutionary history, we attempted to quantify the temporal origins and strength of selection for genetic variants associated with health, reproductive, performance, production, aesthetic and behavioural traits in cattle and horse populations. By modelling the strength and timing of selection for GWAS variants in these species, we aimed to discover which traits were important during the initial phase of domestication and which traits were selected for more recently.

Due to technical difficulties with MCMC convergence, the majority of modelled variants (75%) did not pass quality control. Of those that met convergence metrics, only two SNPs in horses showed any signs of selection and none in cattle. To address the technical issues with this modelling, we have made detailed suggestions for how best to advance this work in the future.

Specifically, we have identified substantial issues with mispolarisation, for which we intend to use the method of Keightley and Jackson (2018) that incorporates three outgroup species and fits an explicit model of DNA evolution. To address variable sampling density, we intend to use a sliding window approach to sample binning, similar to Fages et al. (2019). To improve MCMC parameter mixing, we intend to implement the Metropolis-Coupled MCMC (MC^3) algorithm (Geyer, 1991), and to use a minimum multivariate ESS (Vats et al., 2015) as the stopping condition for the chain lengths. We also intend to implement a system for tuning the length of trajectory updates to achieve the target acceptance rate.

Given the low rate of model convergence in our results, no broad conclusions about selection for GWAS traits in ancient domestic animal populations can be made. Whilst

we have identified very little selection in this dataset, it remains possible that model convergence issues have systematically biased against identifying selection. Similarly, the large number of putatively mispolarised variants may be obscuring many selected traits.

2.9 References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556. https://doi.org/10.1093/molbev/msu236

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Petrov, B.N., Caski, F. (Eds.), Proceedings of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Alberto, F.J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H.R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M.W., Taberlet, P., Pompanon, F., 2018. Convergent genomic signatures of domestication in sheep and goats. Nat. Commun. 9, 813. https://doi.org/10.1038/s41467-018-03206-y

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415. https://doi.org/10.1093/bioinformatics/btg427

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Arendt, M., Cairns, K.M., Ballard, J.W.O., Savolainen, P., Axelsson, E., 2016. Diet adaptation in dog reflects spread of prehistoric agriculture. Heredity 117, 301–306. https://doi.org/10.1038/hdy.2016.48

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å., Lindblad-Toh, K., 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495, 360–364. https://doi.org/10.1038/nature11837

Balding, D.J., Nichols, R.A., 1995. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 96, 3–12. https://doi.org/10.1007/bf01441146

Berkowicz, E.W., Magee, D.A., Berry, D.P., Sikora, K.M., Howard, D.J., Mullen, M.P., Evans, R.D., Spillane, C., MacHugh, D.E., 2012. Single nucleotide polymorphisms in the imprinted bovine insulin-like growth factor 2 receptor gene (IGF2R) are associated with body size traits in Irish Holstein-Friesian cattle. Anim. Genet. 43, 81–87. https://doi.org/10.1111/j.1365-2052.2011.02211.x

Best, N., Cowles, M.K., Vines, K., 1995. CODA* convergence diagnosis and output analysis software for Gibbs sampling output Version 0.30. MRC Biostatistics Unit, Cambridge 52.

Bollback, J.P., York, T.L., Nielsen, R., 2008. Estimation of 2Nes from temporal allele frequency data. Genetics 179, 497–502. https://doi.org/10.1534/genetics.107.085019

Botigué, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., Taravella, A.M., Seregély, T., Zeeb-Lanz, A., Arbogast, R.-M., Bobo, D., Daly, K., Unterländer, M., Burger, J., Kidd, J.M., Veeramah, K.R., 2017. Ancient European dog genomes reveal continuity since the Early Neolithic. Nat. Commun. 8, 16082. https://doi.org/10.1038/ncomms16082

Broad Institute, 2016. Picard Tools [WWW Document]. URL http://broadinstitute.github.io/picard/

Brøndum, R.F., Guldbrandtsen, B., Sahana, G., Lund, M.S., Su, G., 2014. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. BMC Genomics 15, 728. https://doi.org/10.1186/1471-2164-15-728

Brooks, S.P., Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. J. Comput. Graph. Stat. 7, 434–455. https://doi.org/10.1080/10618600.1998.10474787

Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F.W., Alföldi, J., Martinez Barrio, A., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D., Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L., Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C., Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M., Lander, E.S., Ferrand, N., Lindblad-Toh, K., Andersson, L., 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345, 1074–1079. https://doi.org/10.1126/science.1253714

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., Zheng, Z., Song, W., Ma, Z., Ma, Y., Dang, R., Zhang, Z., Xu, L., Jia, Y., Liu, S., Yue, X., Deng, W., Zhang, X., Sun, Z., Lan, X., Han, J., Chen, H., Bradley, D.G., Jiang, Y., Lei, C., 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat. Commun. 9, 2337. https://doi.org/10.1038/s41467-018-04737-0

Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E., Orlando, L., 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. Mol. Ecol. Resour. 17, 508–522. https://doi.org/10.1111/1755-0998.12595

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.J., Anderson, C., VanTassell, C.P., Hulsegge, I., Goddard, M.E., Guldbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R., Hayes, B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46, 858–865. https://doi.org/10.1038/ng.3034

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Schiffels, S., Durbin, R., 2016. Multiallelic calling model in bcftools (-m). samtools.github.io.

Eldon, B., Wakeley, J., 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172, 2621–2633. https://doi.org/10.1534/genetics.105.052175

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., Alfarhan, A.H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J.H., Bayarsaikhan, J., Benecke, N.,

Bernáldez-Sánchez, E., Berrocal-Rangel, L., Biglari, F., Boessenkool, S., Boldgiv, B., Brem, G., Brown, D., Burger, J., Crubézy, E., Daugnora, L., Davoudi, H., de Barros Damgaard, P., de Los Ángeles de Chorro Y de Villa-Ceballos, M., Deschler-Erb, S., Detry, C., Dill, N., do Mar Oom, M., Dohr, A., Ellingvåg, S., Erdenebaatar, D., Fathi, H., Felkel, S., Fernández-Rodríguez, C., García-Viñas, E., Germonpré, M., Granado, J.D., Hallsson, J.H., Hemmer, H., Hofreiter, M., Kasparov, A., Khasanov, M., Khazaeli, R., Kosintsev, P., Kristiansen, K., Kubatbek, T., Kuderna, L., Kuznetsov, P., Laleh, H., Leonard, J.A., Lhuillier, J., Liesau von Lettow-Vorbeck, C., Logvin, A., Lõugas, L., Ludwig, A., Luis, C., Arruda, A.M., Marques-Bonet, T., Matoso Silva, R., Merz, V., Mijiddorj, E., Miller, B.K., Monchalov, O., Mohaseb, F.A., Morales, A., Nieto-Espinet, A., Nistelberger, H., Onar, V., Pálsdóttir, A.H., Pitulko, V., Pitskhelauri, K., Pruvost, M., Rajic Sikanjic, P., Rapan Papeša, A., Roslyakova, N., Sardari, A., Sauer, E., Schafberg, R., Scheu, A., Schibler, J., Schlumbaum, A., Serrand, N., Serres-Armero, A., Shapiro, B., Sheikhi Seno, S., Shevnina, I., Shidrang, S., Southon, J., Star, B., Sykes, N., Taheri, K., Taylor, W., Teegen, W.-R., Trbojević Vukičević, T., Trixl, S., Tumen, D., Undrakhbold, S., Usmanova, E., Vahdati, A., Valenzuela-Lamas, S., Viegas, C., Wallner, B., Weinstock, J., Zaibert, V., Clavel, B., Lepetz, S., Mashkour, M., Helgason, A., Stefánsson, K., Barrey, E., Willerslev, E., Outram, A.K., Librado, P., Orlando, L., 2019. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell 177, 1419–1435.e31. https://doi.org/10.1016/j.cell.2019.03.049

Fang, M., Larson, G., Ribeiro, H.S., Li, N., Andersson, L., 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. PLoS Genet. 5, e1000341. https://doi.org/10.1371/journal.pgen.1000341

Ferrer-Admetlla, A., Leuenberger, C., Jensen, J.D., Wegmann, D., 2016. An Approximate Markov Model for the Wright-Fisher Diffusion and Its Application to Time Series Data. Genetics 203, 831–846. https://doi.org/10.1534/genetics.115.184598

Fisher, R.A., 1930. The genetical theory of natural selection. Clarendon Press.

Flegal, J.M., Hughes, J., Vats, D., Dai, N., 2012. mcmcse: Monte Carlo standard errors for MCMC. Riverside, CA and Minneapolis, MN. R package version 1–0.

Fortes, M.R.S., Reverter, A., Kelly, M., McCulloch, R., Lehnert, S.A., 2013. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. Andrology 1, 644–650. https://doi.org/10.1111/j.2047-2927.2013.00101.x

Frantz, L.A.F., Haile, J., Lin, A.T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V.E., Daly, K.G., Battista, V.M., Price, M., Gron, K.J., Alexandri, P., Arbogast, R.-M., Arbuckle, B., Bălăşescu, A., Barnett, R., Bartosiewicz, L., Baryshnikov, G., Bonsall, C., Borić, D., Boroneanţ, A., Bulatović, J., Çakirlar, C., Carretero, J.-M., Chapman, J., Church, M., Crooijmans, R., De Cupere, B., Detry, C., Dimitrijevic, V., Dumitraşcu, V., du Plessis, L., Edwards, C.J., Erek, C.M., Erim-Özdoğan, A., Ervynck, A., Fulgione, D., Gligor, M., Götherström, A., Gourichon, L., Groenen, M.A.M., Helmer, D., Hongo, H., Horwitz, L.K., Irving-Pease, E.K., Lebrasseur, O., Lesur, J., Malone, C., Manaseryan, N., Marciniak, A., Martlew, H., Mashkour, M., Matthews, R., Matuzeviciute, G.M., Maziar, S., Meijaard, E., McGovern, T., Megens, H.-J., Miller, R., Mohaseb, A.F., Orschiedt, J., Orton, D., Papathanasiou, A., Pearson, M.P., Pinhasi, R., Radmanović, D., Ricaut, F.-X., Richards, M., Sabin, R., Sarti, L., Schier, W., Sheikhi, S., Stephan, E., Stewart, J.R., Stoddart, S., Tagliacozzo, A., Tasić, N., Trantalidou, K., Tresset, A., Valdiosera, C., van den Hurk, Y., Van Poucke, S., Vigne, J.-D., Yanevich, A., Zeeb-Lanz, A., Triantafyllidis, A., Gilbert, M.T.P., Schibler, J., Rowley-Conwy, P., Zeder, M., Peters, J., Cucchi, T., Bradley, D.G., Dobney, K., Burger, J., Evin, A., Girdland-Flink, L., Larson, G., 2019. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. Proc. Natl. Acad. Sci. U. S. A.

https://doi.org/10.1073/pnas.1901169116

Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É.A., Sablin, M.V., Bréhard, S., Mashkour, M., Bălăşescu, A., Gillet, B., Hughes, S., Chassaing, O., Hitte, C., Vigne, J.-D., Dobney, K., Hänni, C., Bradley, D.G., Larson, G., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352, 1228–1231. https://doi.org/10.1126/science.aaf3161

Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.-J., Bosse, M., Paudel, Y., Semiadi, G., Meijaard, E., Li, N., Crooijmans, R.P.M.A., Archibald, A.L., Slatkin, M., Schook, L.B., Larson, G., Groenen, M.A.M., 2013. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 14, R107. https://doi.org/10.1186/gb-2013-14-9-r107

Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R.P.M.A., Larson, G., Groenen, M.A.M., 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat. Genet. 47, 1141–1148. https://doi.org/10.1038/ng.3394

Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A.R., Parker, H.G., Lee, C., Tadigotla, V., Wilton, A., Siepel, A., Bustamante, C.D., Harkins, T.T., Nelson, S.F., Ostrander, E.A., Marques-Bonet, T., Wayne, R.K., Novembre, J., 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10, e1004016. https://doi.org/10.1371/journal.pgen.1004016

Gelabert, P., Olalde, I., de-Dios, T., Civit, S., Lalueza-Fox, C., 2017. Malaria was a weak selective force in ancient Europeans. Sci. Rep. 7, 1377. https://doi.org/10.1038/s41598-017-01534-5

Gelman, A., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. 7, 457–472. https://doi.org/10.1214/ss/1177011136

Gerbault, P., Allaby, R.G., Boivin, N., Rudzinski, A., Grimaldi, I.M., Pires, J.C., Climer Vigueira, C., Dobney, K., Gremillion, K.J., Barton, L., Arroyo-Kalin, M., Purugganan, M.D., Rubio de Casas, R., Bollongino, R., Burger, J., Fuller, D.Q., Bradley, D.G., Balding, D.J., Richerson, P.J., Gilbert, M.T.P., Larson, G., Thomas, M.G., 2014. Storytelling and story testing in domestication. Proc. Natl. Acad. Sci. U. S. A. 111, 6159–6164. https://doi.org/10.1073/pnas.1400425111

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood, in: Proceedings of the 23rd Symposium on the Interface Computing Science and Statistics. Interface Foundation of North America, pp. 156–163.

Girdland Flink, L., Allen, R., Barnett, R., Malmström, H., Peters, J., Eriksson, J., Andersson, L., Dobney, K., Larson, G., 2014. Establishing the validity of domestication genes using DNA from ancient chickens. Proc. Natl. Acad. Sci. U. S. A. 111, 6184–6189. https://doi.org/10.1073/pnas.1308939110

Gory, J.J., Herbei, R., Kubatko, L.S., 2018. Bayesian inference of selection in the Wright-Fisher diffusion model. Stat. Appl. Genet. Mol. Biol. 17. https://doi.org/10.1515/sagmb-2017-0046

Gottschalk, M., Metzger, J., Martinsson, G., Sieme, H., Distl, O., 2016. Genome-wide association study for semen quality traits in German Warmblood stallions. Anim. Reprod. Sci. 171, 81–

86. https://doi.org/10.1016/j.anireprosci.2016.06.002

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Nguyen, D.T., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J.V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., Schook, L.B., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491, 393–398. https://doi.org/10.1038/nature11622

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Haase, B., Signer-Hasler, H., Binns, M.M., Obexer-Ruff, G., Hauswirth, R., Bellone, R.R., Burger, D., Rieder, S., Wade, C.M., Leeb, T., 2013. Accumulating mutations in series of haplotypes at the KIT and MITF loci are major determinants of white markings in Franches-Montagnes horses. PLoS One 8, e75071. https://doi.org/10.1371/journal.pone.0075071

Hayes, B.J., MacLeod, I.M., Daetwyler, H.D., Bowman, P.J., Chamberlian, A.J., Vander Jagt, C.J., Capitan, A., Pausch, H., Stothard, P., Liao, X., Schrooten, C., Mullaart, E., Fries, R., Guldbrandtsen, B., Lund, M.S., Boichard, D.A., Veerkamp, R.F., VanTassell, C.P., Gredler, B., Druet, T., Bagnato, A., Vilkki, J., deKoning, D.J., Santus, E., Goddard, M.E., 2014. Genomic prediction from whole genome sequence in livestock: the 1000 bull genomes project, in: 10th World Congress on Genetics Applied to Livestock Production (WCGALP).

Heger, A., Jacobs, K., 2018. pysam. Github. https://github.com/pysam-developers/pysam

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., Clamp, M., 2002. The Ensembl genome database project. Nucleic Acids Res. 30, 38–41. https://doi.org/10.1093/nar/30.1.38

Hu, Z.-L., Park, C.A., Reecy, J.M., 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res. 47, D701–D710. https://doi.org/10.1093/nar/gky1084

Hu, Z.-L., Park, C.A., Reecy, J.M., 2016. Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res. 44, D827–33. https://doi.org/10.1093/nar/gkv1233

Hu, Z.-L., Reecy, J.M., 2007. Animal QTLdb: beyond a repository. Mamm. Genome 18, 1–4.

https://doi.org/10.1007/s00335-006-0105-8

Irving-Pease, E.K., Ryan, H., Jamieson, A., Dimopoulos, E.A., Larson, G., Frantz, L.A.F., 2018. Paleogenomics of animal domestication, in: Lindqvist, C., Rajora, O.P. (Eds.), Paleogenomics: Genome-Scale Analysis of Ancient DNA, Population Genomics. Springer, pp. 225–272.

Iso-Touru, T., Sahana, G., Guldbrandtsen, B., Lund, M.S., Vilkki, J., 2016. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. BMC Genet. 17, 55. https://doi.org/10.1186/s12863-016-0363-8

Jensen, P., 2014. Behavior genetics and the domestication of animals. Annu Rev Anim Biosci 2, 85–104. https://doi.org/10.1146/annurev-animal-022513-114135

Jewett, E.M., Steinrücken, M., Song, Y.S., 2016. The Effects of Population Size Histories on Estimates of Selection Coefficients from Time-Series Genetic Data. Mol. Biol. Evol. 33, 3002–3027. https://doi.org/10.1093/molbev/msw173

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kalbfleisch, T.S., Rice, E.S., DePriest, M.S., Walenz, B.P., Hestand, M.S., Vermeesch, J.R., O'Connell, B.L., Fiddes, I.T., Vershinina, A.O., Petersen, J.L., Finno, C.J., Bellone, R.R., McCue, M.E., Brooks, S.A., Bailey, E., Orlando, L., Green, R.E., Miller, D.C., Antczak, D.F., MacLeod, J.N., 2018. EquCab3, an Updated Reference Genome for the Domestic Horse. bioRxiv. https://doi.org/10.1101/306928

Kans, J., 2013. Entrez Direct: E-utilities on the UNIX Command Line. National Center for Biotechnology Information (US).

Keightley, P.D., Jackson, B.C., 2018. Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. Genetics 209, 897–906. https://doi.org/10.1534/genetics.118.301120

Keinan, A., Mullikin, J.C., Patterson, N., Reich, D., 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat. Genet. 39, 1251–1255. https://doi.org/10.1038/ng2116

Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J., Lachance, J., 2018. Genetic disease risks can be misestimated across global populations. Genome Biol. 19, 179. https://doi.org/10.1186/s13059-018-1561-7

Kim, N.Y., Bhuiyan, M.S.A., Chae, H.S., Baek, K.S., Son, J.K., Shin, S.M., Woo, J.H., Park, S.H., Lee, S.H., 2017. Genome-wide association study for tobiano spotting coat color in Korean Jeju × Thoroughbred horse population. Anim. Genet. 48, 728–729. https://doi.org/10.1111/age.12596

Kircher, M., 2012. Analysis of High-Throughput Ancient DNA Sequencing Data, in: Shapiro, B., Hofreiter, M. (Eds.), Ancient DNA: Methods and Protocols. Humana Press, Totowa, NJ, pp. 197–228. https://doi.org/10.1007/978-1-61779-516-9_23

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Krueger, F., 2015. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Kuchenbaecker, K., Telkar, N., Reiker, T., Walters, R.G., Lin, K., Eriksson, A., Gurdasani, D., Gilly,

A., Southam, L., Tsafantakis, E., Karaleftheri, M., Seeley, J., Kamali, A., Asiki, G., Millwood, I.Y., Holmes, M., Du, H., Guo, Y., Kumari, M., Dedoussis, G., Li, L., Chen, Z., Sandhu, M.S., Zeggini, E., Benzeval, M., Burton, J., Buck, N., Jäckle, A., Laurie, H., Lynn, P., Pudney, S., Rabe, B., Wolke, D., Understanding Society Scientific Group, 2019. The transferability of lipid loci across African, Asian and European cohorts. Nat. Commun. 10, 4330. https://doi.org/10.1038/s41467-019-12026-7

Larson, G., Burger, J., 2013. A population genetics view of animal domestication. Trends Genet. 29, 197–205. https://doi.org/10.1016/j.tig.2013.01.003

Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M., Barton, L., Vigueira, C.C., Denham, T., Dobney, K., Doust, A.N., Gepts, P., Gilbert, M.T.P., Gremillion, K.J., Lucas, L., Lukens, L., Marshall, F.B., Olsen, K.M., Pires, J.C., Richerson, P.J., Casas, R.R. de, Sanjur, O.I., Thomas, M.G., Fuller, D.Q., 2014. Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. U. S. A. 111, 6139–6146. https://doi.org/10.1073/pnas.1323964111

Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration, 2011. The sequence read archive. Nucleic Acids Res. 39, D19–21. https://doi.org/10.1093/nar/gkq1019

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Loog, L., Thomas, M.G., Barnett, R., Allen, R., Sykes, N., Paxinos, P.D., Lebrasseur, O., Dobney, K., Peters, J., Manica, A., Larson, G., Eriksson, A., 2017. Inferring Allele Frequency Trajectories from Ancient DNA Indicates That Selection on a Chicken Gene Coincided with Changes in Medieval Husbandry Practices. Mol. Biol. Evol. 34, 1981–1990. https://doi.org/10.1093/molbev/msx142

Malaspinas, A.-S., Malaspinas, O., Evans, S.N., Slatkin, M., 2012. Estimating allele age and selection coefficient from time-serial data. Genetics 192, 599–607. https://doi.org/10.1534/genetics.112.140939

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Maruki, T., Lynch, M., 2017. Genotype Calling from Population-Genomic Sequencing Data. G3 7, 1393–1404. https://doi.org/10.1534/g3.117.039008

Mathieson, I., McVean, G., 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics 193, 973–984. https://doi.org/10.1534/genetics.112.147611

Matuszewski, S., Hildebrandt, M.E., Achaz, G., Jensen, J.D., 2018. Coalescent Processes with Skewed Offspring Distributions and Nonequilibrium Demography. Genetics 208, 323–338. https://doi.org/10.1534/genetics.117.300499

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Metzger, J., Ohnesorge, B., Distl, O., 2012. Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. PLoS One 7, e41640. https://doi.org/10.1371/journal.pone.0041640

Nicolazzi, E.L., Picciolini, M., Strozzi, F., Schnabel, R.D., Lawley, C., Pirani, A., Brew, F., Stella, A., 2014. SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics 15, 123. https://doi.org/10.1186/1471-2164-15-123

Nilsen, H., Olsen, H.G., Hayes, B., Nome, T., Sehested, E., Svendsen, M., Meuwissen, T.H.E., Lien, S., 2009. Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6. Anim. Genet. 40, 701–712. https://doi.org/10.1111/j.1365-2052.2009.01908.x

Oliphant, T.E., 2006. A guide to NumPy. Trelgol Publishing USA.

Ollivier, M., Tresset, A., Bastian, F., Lagoutte, L., Axelsson, E., Arendt, M.-L., Bălăşescu, A., Marshour, M., Sablin, M.V., Salanova, L., Vigne, J.-D., Hitte, C., Hänni, C., 2016. Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. Royal Society Open Science 3, 160449. https://doi.org/10.1098/rsos.160449

Olsen, H.G., Knutsen, T.M., Lewandowska-Sabat, A.M., Grove, H., Nome, T., Svendsen, M., Arnyasi, M., Sodeland, M., Sundsaasen, K.K., Dahl, S.R., Heringstad, B., Hansen, H.H., Olsaker, I., Kent, M.P., Lien, S., 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. Genet. Sel. Evol. 48, 79. https://doi.org/10.1186/s12711-016-0257-2

Overcast, I., 2016. easySFS. Github. https://github.com/isaacovercast/easySFS

Palmer, C., Pe'er, I., 2017. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. PLoS Genet. 13, e1006916. https://doi.org/10.1371/journal.pgen.1006916

Paris, C., Servin, B., Boitard, S., 2019. Inference of selection from genetic time series using various parametric approximations to the Wright-Fisher model. BioRxiv.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M., Wimmers, K., 2014. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. Int. J. Biol. Sci. 10, 327–337. https://doi.org/10.7150/ijbs.8134

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rempala, G.A., Yang, Y., 2013. On Permutation Procedures for Strong Control in Multiple Testing with Gene Expression Data. Stat. Interface 6. https://doi.org/10.4310/SII.2013.v6.n1.a8

Richards, A., 2015. University of Oxford Advanced Research Computing. https://doi.org/10.5281/zenodo.22558

Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Probab. 7, 110–120. https://doi.org/10.1214/aoap/1034625254

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591. https://doi.org/10.1038/nature08832

Sackman, A.M., Harris, R.B., Jensen, J.D., 2019. Inferring Demography and Selection in Organisms Characterized by Skewed Offspring Distributions. Genetics 211, 1019–1028. https://doi.org/10.1534/genetics.118.301684

Schraiber, J.G., Evans, S.N., Slatkin, M., 2016. Bayesian Inference of Natural Selection from Allele Frequency Time Series. Genetics 203, 493–511. https://doi.org/10.1534/genetics.116.187278

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T.S., Velazquez, A.M.V., Stenderup, J., Hoover, C.A., Rubin, C.-J., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., MacHugh, D.E., Kalbfleisch, T., MacLeod, J.N., Rubin, E.M., Sicheritz-Ponten, T., Andersson, L., Hofreiter, M., Marques-Bonet, T., Gilbert, M.T.P., Nielsen, R., Excoffier, L., Willerslev, E., Shapiro, B., Orlando, L., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc. Natl. Acad. Sci. U. S. A. 111, E5661–9. https://doi.org/10.1073/pnas.1416991111

Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88. https://doi.org/10.1186/s13104-016-1900-2

Skujina, I., Winton, C.L., Hegarty, M.J., McMahon, R., Nash, D.M., Davies Morel, M.C.G., McEwan, N.R., 2018. Detecting genetic regions associated with height in the native ponies of the British Isles by using high density SNP genotyping. Genome 61, 767–770. https://doi.org/10.1139/gen-2018-0006

Spotify, 2019. Luigi. https://github.com/spotify/luigi.

Tange, O., 2018. GNU Parallel 2018. https://doi.org/10.5281/zenodo.1146014

Taron, U.H., Lell, M., Barlow, A., Paijmans, J.L.A., 2018. Testing of Alignment Parameters for Ancient Samples: Evaluating and Optimizing Mapping Parameters for Ancient Samples Using the TAPAS Tool. Genes 9. https://doi.org/10.3390/genes9030157

Terhorst, J., Schlötterer, C., Song, Y.S., 2015. Multi-locus analysis of genomic time series data from experimental evolution. PLoS Genet. 11, e1005069. https://doi.org/10.1371/journal.pgen.1005069

USDA ARS, 2018. ARS-UCD1.2 - Genome - Assembly - NCBI [WWW Document]. URL https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2

van Binsbergen, R., Bink, M.C., Calus, M.P., van Eeuwijk, F.A., Hayes, B.J., Hulsegge, I., Veerkamp, R.F., 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genet. Sel. Evol. 46, 41. https://doi.org/10.1186/1297-9686-46-41

Vats, D., Flegal, J.M., Jones, G.L., 2015. Multivariate Output Analysis for Markov chain Monte Carlo. arXiv [math.ST].

Velie, B.D., Fegraeus, K.J., Solé, M., Rosengren, M.K., Røed, K.H., Ihler, C.-F., Strand, E., Lindgren, G., 2018. A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. BMC Genet. 19, 80. https://doi.org/10.1186/s12863-018-0670-3

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., Hesse, P., Fulton, D., Sauer, E.W., Mohaseb, F.A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Vahidi, S.M.F., MacHugh, D.E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B.S., Mashkour, M., Orton, D., Horwitz, L.K., Teasdale, M.D., Bradley, D.G., 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 365, 173–176. https://doi.org/10.1126/science.aav1002

Vigne, J.-D., 2011. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. C. R. Biol., On the trail of domestications, migrations and invasions in agriculture 334, 171–181. https://doi.org/10.1016/j.crvi.2010.12.009

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., Blöcker, H., Distl, O., Edgar, R.C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J.N., Penedo, M.C.T., Raison, J.M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D.F., Biagi, T., Binns, M.M., Chowdhary, B.P., Coleman, S.J., Della Valle, G., Fryc, S., Guérin, G., Hasegawa, T., Hill, E.W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J.R., Murray, J., Nergadze, S.G., Onofrio, R., Pedroni, S., Piras, M.F., Raudsepp, T., Rocchi, M., Røed, K.H., Ryder, O.A., Searle, S., Skow, L., Swinburne, J.E., Syvänen, A.C., Tozaki, T., Valberg, S.J., Vaudin, M., White, J.R., Zody, M.C., Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Lander, E.S., Lindblad-Toh, K., 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326, 865–867. https://doi.org/10.1126/science.1178158

Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. Psychon. Bull. Rev. 11, 192–196.

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D.M., Billis, K., Chow, W., Eory, L., Finlayson, H.A., Flicek, P., Girón, C.G., Griffin, D.K., Hall, R., Hannum, G., Hourlier, T., Howe, K., Hume, D.A., Izuogu, O., Kim, K., Koren, S., Liu, H., Manchanda, N., Martin, F.J., Nonneman, D.J., O'Connor, R.E., Phillippy, A.M., Rohrer, G.A., Rosen, B.D., Rund, L.A., Sargent, C.A., Schook, L.B., Schroeder, S.G., Schwartz, A.S., Skinner, B.M., Talbot, R., Tseng, E., Tuggle, C.K., Watson, M., Smith, T.P.L., Archibald, A.L., 2019. An improved pig reference genome sequence to enable pig genetics and genomics research. bioRxiv. https://doi.org/10.1101/668921

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Widenius, M., Axmark, D., 2002. Mysql Reference Manual, 1st ed. O'Reilly & Associates, Inc., Sebastopol, CA, USA.

Wright, S., 1931. Evolution in Mendelian Populations. Genetics 16, 97–159.

Ye, K., Gao, F., Wang, D., Bar-Yosef, O., Keinan, A., 2017. Dietary adaptation of FADS genes in Europe varied across time and geography. Nat Ecol Evol 1, 167. https://doi.org/10.1038/s41559-017-0167

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A., Salzberg, S.L., 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10, R42. https://doi.org/10.1186/gb-2009-10-4-r42

2.10 Supplementary Materials

2.10.1 Sample accessions

2.10.1.1 Modern cattle samples

Table S2.1 Modern cattle samples

Sample	BioProject	BioSample	Sex	Reference
Angus_19	PRJNA238491	SAMN02671514	M	(Daetwyler et al., 2014)
Angus_28	PRJNA238491	SAMN02671523	M	(Daetwyler et al., 2014)
Angus_33	PRJNA238491	SAMN02671528	M	(Daetwyler et al., 2014)
Angus_39	PRJNA238491	SAMN02671534	M	(Daetwyler et al., 2014)
Angus_40	PRJNA238491	SAMN02671535	M	(Daetwyler et al., 2014)
Fleckvieh_39	PRJNA238491	SAMN02671663	M	(Daetwyler et al., 2014)
Fleckvieh_40	PRJNA238491	SAMN02671664	M	(Daetwyler et al., 2014)
Fleckvieh_42	PRJNA238491	SAMN02671666	M	(Daetwyler et al., 2014)
Fleckvieh_5	PRJNA238491	SAMN02671629	M	(Daetwyler et al., 2014)
Holstein_116	PRJNA238491	SAMN02671669	M	(Daetwyler et al., 2014)
Holstein_71	PRJNA238491	SAMN02671465	M	(Daetwyler et al., 2014)
Holstein_79	PRJNA238491	SAMN02671473	M	(Daetwyler et al., 2014)
Holstein_88	PRJNA238491	SAMN02671482	M	(Daetwyler et al., 2014)
Holstein_90	PRJNA238491	SAMN02671484	M	(Daetwyler et al., 2014)
Jersey_12	PRJNA238491	SAMN02671608	M	(Daetwyler et al., 2014)
Jersey_2	PRJNA238491	SAMN02671598	M	(Daetwyler et al., 2014)
Jersey_4	PRJNA238491	SAMN02671600	M	(Daetwyler et al., 2014)
Jersey_5	PRJNA238491	SAMN02671601	M	(Daetwyler et al., 2014)
Jersey_7	PRJNA238491	SAMN02671603	M	(Daetwyler et al., 2014)
Aln_12	PRJEB31621	SAMEA5577153	M	(Verdugo et al., 2019)
Hl_8059	PRJEB31621	SAMEA5577150	M	(Verdugo et al., 2019)
Wag_1a	PRJEB31621	SAMEA5577012	M	(Verdugo et al., 2019)
Bbub	PRJNA350833	SAMN05949030	F	(Whitacre et al., 2017)

2.10.1.2 Modern horse samples

Table S2.2 Modern horse samples

Sample	Bioproject	Biosample	Sex	Reference
Prze_0150A	PRJEB10098	SAMEA3498579	F	(Der Sarkissian et al., 2015)
Prze_0151A	PRJEB10098	SAMEA3498580	M	(Der Sarkissian et al., 2015)
Prze_0157A	PRJEB10098	SAMEA3498586	M	(Der Sarkissian et al., 2015)
Prze_0158A	PRJEB10098	SAMEA3498587	F	(Der Sarkissian et al., 2015)
Prze_0159A	PRJEB10098	SAMEA3498588	M	(Der Sarkissian et al., 2015)
Prze_0160A	PRJEB10098	SAMEA3498589	M	(Der Sarkissian et al., 2015)
Arab_0237A	PRJNA230019	SAMN02439777	M	(Metzger et al., 2014)
Conn_0004A	PRJNA273402	SAMN03291442	M	(Finno et al., 2015)
Duel_0238A	PRJNA230019	SAMN02422919	F	(Metzger et al., 2014)
Frie_0296A	PRJEB13863	SAMEA3951218	M	(Leegwater et al., 2016)
FrMo_0065A	PRJEB10098	SAMEA3498888	M	(Der Sarkissian et al., 2015)
Hano_0235A	PRJNA230019	SAMN02439779	M	(Metzger et al., 2014)
Heav_0269A	PRJNA291776	SAMN03955412	M	(Metzger et al., 2014)
Icel_0144A	N/A	SAMN00857894	M	(Andersson et al., 2012)
Icel_0247A	PRJEB8911	SAMEA3311207	M	(Haase et al., 2015)
Jeju_0275A	PRJNA169102	SAMN01057172	M	(Do et al., 2014)
Marw_0239A	PRJNA246445	SAMN02767683	M	(Jun et al., 2014)
Mong_0153A	PRJEB10098	SAMEA3498582	M	(Der Sarkissian et al., 2015)
Mong_0215A	PRJNA277815	SAMN04002338	M	(Do et al., 2014)
Morg_0096A	PRJEB10098	SAMEA3499838	F	(Der Sarkissian et al., 2015)
Quar_0073A	PRJEB10098	SAMEA3499837	M	(Der Sarkissian et al., 2015)
Shet_0249A	PRJEB9269	SAMEA3367610	F	(Frischknecht et al., 2015)
Shet_0250A	PRJEB9267	SAMEA3367609	F	(Frischknecht et al., 2015)
Sorr_0236A	PRJNA230019	SAMN02439778	M	(Metzger et al., 2014)
Stan_0081A	PRJEB10098	SAMEA3499832	M	(Warmuth et al., 2011)
Thor_0145A	PRJNA277815	SAMN04002340	F	(Orlando et al., 2013)

Sample	Bioproject	Biosample	Sex	Reference
Thor_0290A	PRJNA168142	SAMN01047706	M	(Kim et al., 2013)
Yaku_0163A	PRJEB10854	SAMEA3542237	M	(Librado et al., 2015)
Yaku_0170A	PRJEB10854	SAMEA3542244	M	(Librado et al., 2015)
Yaku_0171A	PRJEB10854	SAMEA3542245	M	(Librado et al., 2015)
Esom_0226A	PRJEB7446	SAMEA2802531	F	(Jónsson et al., 2014)

2.10.1.3 Ancient cattle samples

Table S2.3 Ancient cattle samples

Sample	BioProject	BioSample	Sex	Reference
Ace1	PRJEB31621	SAMEA5577346	F	(Verdugo et al., 2019)
Bal1	N/A	N/A	F	(Mullin et al. in prep)
Bal2	N/A	N/A	F	(Mullin et al. in prep)
Bel1	PRJEB31621	SAMEA5577349	F	(Verdugo et al., 2019)
Bel2	PRJEB31621	SAMEA5577350	M	(Verdugo et al., 2019)
Bis1	N/A	N/A	M	(Mullin et al. in prep)
Bis2	N/A	N/A	M	(Mullin et al. in prep)
Bla1	PRJEB31621	SAMEA5577353	M	(Verdugo et al., 2019)
Bla2	PRJEB31621	SAMEA5577354	M	(Verdugo et al., 2019)
Bor1	N/A	N/A	M	(Mullin et al. in prep)
Bri1	N/A	N/A	M	(Mullin et al. in prep)
Bub1	PRJEB31621	SAMEA5577355	F	(Verdugo et al., 2019)
Bun1	N/A	N/A	F	(Mullin et al. in prep)
Bun2	N/A	N/A	M	(Mullin et al. in prep)
Ch22	PRJEB31621	SAMEA5577358	M	(Verdugo et al., 2019)
Cla1	N/A	N/A	F	(Mullin et al. in prep)
Cla2	N/A	N/A	M	(Mullin et al. in prep)
Cla3	N/A	N/A	F	(Mullin et al. in prep)
Cla4	N/A	N/A	F	(Mullin et al. in prep)

Sample	BioProject	BioSample	Sex	Reference
Cla5	N/A	N/A	F	(Mullin et al. in prep)
Cla6	N/A	N/A	F	(Mullin et al. in prep)
Cla7	N/A	N/A	F	(Mullin et al. in prep)
Cla8	N/A	N/A	F	(Mullin et al. in prep)
Cla9	N/A	N/A	F	(Mullin et al. in prep)
Da1	N/A	N/A	F	(Mullin et al. in prep)
Da2	N/A	N/A	M	(Mullin et al. in prep)
Da3	N/A	N/A	F	(Mullin et al. in prep)
Da4	N/A	N/A	F	(Mullin et al. in prep)
Da5	N/A	N/A	F	(Mullin et al. in prep)
Da6	N/A	N/A	F	(Mullin et al. in prep)
Dro1	N/A	N/A	M	(Mullin et al. in prep)
Dub1	N/A	N/A	F	(Mullin et al. in prep)
Dub2	N/A	N/A	M	(Mullin et al. in prep)
Dur1	N/A	N/A	F	(Mullin et al. in prep)
Dur2	N/A	N/A	F	(Mullin et al. in prep)
Dur3	N/A	N/A	F	(Mullin et al. in prep)
Dyr1	N/A	N/A	F	(Mullin et al. in prep)
Dzh1	N/A	N/A	M	(Mullin et al. in prep)
Dzh2	N/A	N/A	F	(Mullin et al. in prep)
Els1	N/A	N/A	M	(Mullin et al. in prep)
Far1	PRJEB31621	SAMEA5577360	F	(Verdugo et al., 2019)
Fir1	N/A	N/A	F	(Mullin et al. in prep)
Fis1	N/A	N/A	F	(Mullin et al. in prep)
Fis2	N/A	N/A	F	(Mullin et al. in prep)
Gen1	N/A	N/A	F	(Mullin et al. in prep)
Gil1	PRJEB31621	SAMEA5577361	F	(Verdugo et al., 2019)
Gyu2	PRJEB31621	SAMEA5577362	M	(Verdugo et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
Has3	PRJEB31621	SAMEA5577364	M	(Verdugo et al., 2019)
HF1	N/A	N/A	F	(Mullin et al. in prep)
Hou1	N/A	N/A	M	(Mullin et al. in prep)
Hou2	N/A	N/A	F	(Mullin et al. in prep)
Hxh1	N/A	N/A	F	(Mullin et al. in prep)
Kie1	N/A	N/A	M	(Mullin et al. in prep)
Kie2	N/A	N/A	M	(Mullin et al. in prep)
Kil1	N/A	N/A	F	(Mullin et al. in prep)
Lud1	N/A	N/A	F	(Mullin et al. in prep)
Lud2	N/A	N/A	F	(Mullin et al. in prep)
Lud3	N/A	N/A	M	(Mullin et al. in prep)
Lud4	N/A	N/A	F	(Mullin et al. in prep)
Mar1	N/A	N/A	M	(Mullin et al. in prep)
Men1	PRJEB31621	SAMEA5577376	M	(Verdugo et al., 2019)
Men2	PRJEB31621	SAMEA5577377	F	(Verdugo et al., 2019)
Mon1	PRJEB31621	SAMEA5577378	F	(Verdugo et al., 2019)
Nah1	PRJEB31621	SAMEA5577379	M	(Verdugo et al., 2019)
Ness2	N/A	N/A	F	(Mullin et al. in prep)
Ness3	N/A	N/A	F	(Mullin et al. in prep)
Ness4	N/A	N/A	F	(Mullin et al. in prep)
Ness5	N/A	N/A	M	(Mullin et al. in prep)
New1	N/A	N/A	M	(Mullin et al. in prep)
New2	N/A	N/A	M	(Mullin et al. in prep)
New3	N/A	N/A	M	(Mullin et al. in prep)
New4	N/A	N/A	F	(Mullin et al. in prep)
New5	N/A	N/A	M	(Mullin et al. in prep)
Otb1	N/A	N/A	M	(Mullin et al. in prep)
Par1	N/A	N/A	F	(Mullin et al. in prep)

Sample	BioProject	BioSample	Sex	Reference
Par2	N/A	N/A	M	(Mullin et al. in prep)
Plo1	PRJEB31621	SAMEA5577380	F	(Verdugo et al., 2019)
Plo2	PRJEB31621	SAMEA5577381	F	(Verdugo et al., 2019)
Plo3	PRJEB31621	SAMEA5577382	M	(Verdugo et al., 2019)
Plo4	PRJEB31621	SAMEA5577383	M	(Verdugo et al., 2019)
Plo5	PRJEB31621	SAMEA5577384	M	(Verdugo et al., 2019)
Plo6	PRJEB31621	SAMEA5577385	M	(Verdugo et al., 2019)
Plo7	PRJEB31621	SAMEA5577386	F	(Verdugo et al., 2019)
Plo8	PRJEB31621	SAMEA5577387	F	(Verdugo et al., 2019)
Pot1	N/A	N/A	F	(Mullin et al. in prep)
Pot2	N/A	N/A	F	(Mullin et al. in prep)
Pot3	N/A	N/A	M	(Mullin et al. in prep)
Pot4	N/A	N/A	F	(Mullin et al. in prep)
Pot5	N/A	N/A	F	(Mullin et al. in prep)
Pot6	N/A	N/A	M	(Mullin et al. in prep)
Pro1	PRJEB31621	SAMEA5577388		(Verdugo et al., 2019)
Qaz1	PRJEB31621	SAMEA5577389	F	(Verdugo et al., 2019)
Rou1	N/A	N/A	F	(Mullin et al. in prep)
Rou2	N/A	N/A	F	(Mullin et al. in prep)
Rou3	N/A	N/A	F	(Mullin et al. in prep)
Rou4	N/A	N/A	F	(Mullin et al. in prep)
Sac3	PRJEB31621	SAMEA5577390	F	(Verdugo et al., 2019)
Sar38	PRJEB31621	SAMEA5577391		(Verdugo et al., 2019)
Sch1	N/A	N/A	F	(Mullin et al. in prep)
Snu1	N/A	N/A	F	(Mullin et al. in prep)
Ste1	N/A	N/A	F	(Mullin et al. in prep)
Stu1	PRJEB31621	SAMEA5577392	M	(Verdugo et al., 2019)
Sub1	PRJEB31621	SAMEA5577393	M	(Verdugo et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
Tda1	PRJEB31621	SAMEA5577394	M	(Verdugo et al., 2019)
Thr1	PRJEB31621	SAMEA5577396	M	(Verdugo et al., 2019)
Tqa1	PRJEB31621	SAMEA5577401	M	(Verdugo et al., 2019)
Tqa2	PRJEB31621	SAMEA5577402	F	(Verdugo et al., 2019)
Tqa3	PRJEB31621	SAMEA5577403	F	(Verdugo et al., 2019)
Tsa3	PRJEB31621	SAMEA5577409	F	(Verdugo et al., 2019)
Viz1	N/A	N/A	M	(Mullin et al. in prep)
Viz2	N/A	N/A	M	(Mullin et al. in prep)
Win1	N/A	N/A	F	(Mullin et al. in prep)
Yor1	N/A	N/A	F	(Mullin et al. in prep)
Yor10	N/A	N/A	F	(Mullin et al. in prep)
Yor11	N/A	N/A	F	(Mullin et al. in prep)
Yor12	N/A	N/A	M	(Mullin et al. in prep)
Yor2	N/A	N/A	M	(Mullin et al. in prep)
Yor3	N/A	N/A	M	(Mullin et al. in prep)
Yor4	N/A	N/A	F	(Mullin et al. in prep)
Yor5	N/A	N/A	M	(Mullin et al. in prep)
Yor6	N/A	N/A	F	(Mullin et al. in prep)
Yor7	N/A	N/A	M	(Mullin et al. in prep)
Yor8	N/A	N/A	M	(Mullin et al. in prep)
Yor9	N/A	N/A	F	(Mullin et al. in prep)

2.10.1.4 Ancient horse samples

Table S2.4 Ancient horse samples

Sample	BioProject	BioSample	Sex	Reference
Borly4_PAVH11	PRJEB22390	SAMEA104233651	F	(Gaunitz et al., 2018)
Borly4_PAVH4	PRJEB22390	SAMEA104233652	F	(Gaunitz et al., 2018)
Borly4_PAVH6	PRJEB22390	SAMEA104233653	F	(Gaunitz et al., 2018)

Sample	BioProject	BioSample	Sex	Reference
Borly4_PAVH8	PRJEB22390	SAMEA104233654	M	(Gaunitz et al., 2018)
Borly4_PAVH9	PRJEB22390	SAMEA104233655	M	(Gaunitz et al., 2018)
Botai_1	PRJEB22390	SAMEA104233656	M	(Gaunitz et al., 2018)
Botai_2	PRJEB22390	SAMEA104233657	M	(Gaunitz et al., 2018)
Botai_3	PRJEB22390	SAMEA104233658	M	(Gaunitz et al., 2018)
Botai_4	PRJEB22390	SAMEA104233659	M	(Gaunitz et al., 2018)
Botai_5	PRJEB22390	SAMEA104233660	M	(Gaunitz et al., 2018)
Botai_6	PRJEB22390	SAMEA104233661	M	(Gaunitz et al., 2018)
Botai_8	PRJEB22390	SAMEA104233662	F	(Gaunitz et al., 2018)
Botai_A	PRJEB31613	SAMEA5408166	F	(Fages et al., 2019)
Botai_B	PRJEB31613	SAMEA5408167	F	(Fages et al., 2019)
Botai_C	PRJEB22390	SAMEA104233663	M	(Fages et al., 2019)
Botai_D1	PRJEB22390	SAMEA104233664	M	(Gaunitz et al., 2018)
Botai_D2	PRJEB31613	SAMEA5408168	F	(Fages et al., 2019)
Botai_D4	PRJEB22390	SAMEA104233665	F	(Gaunitz et al., 2018)
Botai_D5	PRJEB22390	SAMEA104233666	M	(Gaunitz et al., 2018)
Botai_D6	PRJEB22390	SAMEA104233667	F	(Gaunitz et al., 2018)
Botai_E	PRJEB31613	SAMEA5408169	F	(Fages et al., 2019)
Botai_F	PRJEB22390	SAMEA104233668	M	(Gaunitz et al., 2018)
Botai_G	PRJEB22390	SAMEA104233669	M	(Fages et al., 2019)
Botai_I	PRJEB22390	SAMEA104233670	M	(Fages et al., 2019)
Botai_K	PRJEB22390	SAMEA104233671	M	(Fages et al., 2019)
Botai_L	PRJEB22390	SAMEA104233672	F	(Gaunitz et al., 2018)
Botai_N	PRJEB31613	SAMEA5408170	F	(Fages et al., 2019)
Botai_NB18	PRJEB31613	SAMEA5408171	F	(Fages et al., 2019)
Botai_O	PRJEB31613	SAMEA5408172	M	(Fages et al., 2019)
Botai_P	PRJEB22390	SAMEA104233673	M	(Gaunitz et al., 2018)
Botai_Petrous	PRJEB22390	SAMEA104233674	F	(Gaunitz et al., 2018)

Sample	BioProject	BioSample	Sex	Reference
Botai_R	PRJEB22390	SAMEA104233675	F	(Gaunitz et al., 2018)
Botai_T	PRJEB31613	SAMEA5408173	F	(Fages et al., 2019)
Actiparc_GVA124	PRJEB31613	SAMEA5408142	M	(Fages et al., 2019)
Actiparc_GVA307	PRJEB31613	SAMEA5408143	M	(Fages et al., 2019)
Actiparc_GVA308	PRJEB31613	SAMEA5408144	F	(Fages et al., 2019)
Actiparc_GVA309	PRJEB31613	SAMEA5408145	M	(Fages et al., 2019)
Actiparc_GVA310	PRJEB22390	SAMEA104233649	F	(Fages et al., 2019)
Actiparc_GVA311	PRJEB31613	SAMEA5408146	M	(Fages et al., 2019)
Altata_NB31	PRJEB31613	SAMEA5408147	M	(Fages et al., 2019)
ArzhanI_Arz3	PRJEB31613	SAMEA5408149	M	(Fages et al., 2019)
ArzhanI_I-K2_Arz1	PRJEB31613	SAMEA5408150	M	(Librado et al., 2017)
ArzhanI_I-K3_Arz2	PRJEB31613	SAMEA5408151	M	(Librado et al., 2017)
ArzhanII_Arz15	PRJEB31613	SAMEA5408152	M	(Fages et al., 2019)
ArzhanII_Arz17	PRJEB31613	SAMEA5408148	M	(Fages et al., 2019)
ArzhanII_Rus11	PRJEB19970 PRJEB20000	SAMEA103910522	M	(Fages et al., 2019)
ArzhanII_Rus9	PRJEB19970 PRJEB20000	SAMEA103910523	M	(Fages et al., 2019)
AugustaRaurica_JG160	PRJEB31613	SAMEA5408154	M	(Fages et al., 2019)
AugustaRauricaSchmidmatt_NBxK9279	PRJEB31613	SAMEA5408155	M	(Fages et al., 2019)
AugustaRauricaSchmidmatt_NBxP9261	PRJEB31613	SAMEA5408153	M	(Fages et al., 2019)
Balagansk_Rus19	PRJEB31613	SAMEA5408156	M	(Fages et al., 2019)
BapskaGradac_BAPSKA	PRJEB31613	SAMEA5408157	M	(Fages et al., 2019)
Bateni_Rus14	PRJEB22390	SAMEA104233650	M	(Gaunitz et al., 2018)
Bateni_Rus16	PRJEB31613	SAMEA5408158	M	(Fages et al., 2019)
Beauvais_GVA122	PRJEB31613	SAMEA5408159	M	(Fages et al., 2019)
Beauvais_GVA375	PRJEB31613	SAMEA5408160	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
Belgheis_TrBWBX116	PRJEB31613	SAMEA5408161	M	(Fages et al., 2019)
Belkaragay_NB13	PRJEB31613	SAMEA5408162	M	(Fages et al., 2019)
Belkaragay_NB15	PRJEB31613	SAMEA5408163	F	(Fages et al., 2019)
Berel_BER01_A	PRJEB19970	SAMEA103910511	M	(Librado et al., 2017)
Berel_BER02_B	PRJEB19970	SAMEA103910512	M	(Librado et al., 2017)
Berel_BER04_D	PRJEB19970 PRJEB20000	SAMEA103910513	M	(Librado et al., 2017)
Berel_BER05_E	PRJEB19970 PRJEB20000	SAMEA103910514	M	(Librado et al., 2017)
Berel_BER06_F	PRJEB19970 PRJEB20000	SAMEA103910515	M	(Librado et al., 2017)
Berel_BER07_G	PRJEB20000 PRJEB19970	SAMEA103910516	M	(Librado et al., 2017)
Berel_BER08_H	PRJEB19970 PRJEB20000	SAMEA103910517	M	(Librado et al., 2017)
Berel_BER09_I	PRJEB19970 PRJEB20000	SAMEA103910518	M	(Librado et al., 2017)
Berel_BER10_K	PRJEB20000 PRJEB19970	SAMEA103910519	M	(Librado et al., 2017)
Berel_BER11_L	PRJEB19970 PRJEB20000	SAMEA103910520	M	(Librado et al., 2017)
Berel_BER12_M	PRJEB19970 PRJEB20000	SAMEA103910521	M	(Librado et al., 2017)
Berufjordur_VHR102	PRJEB31613	SAMEA5408164	M	(Fages et al., 2019)
Boves_GVA191	PRJEB31613	SAMEA5408174	F	(Fages et al., 2019)
BozAdyr_KYRH10	PRJEB31613	SAMEA5408175	M	(Fages et al., 2019)
BozAdyr_KYRH8	PRJEB31613	SAMEA5408176	M	(Fages et al., 2019)
BroughOfDeerness_VHR010	PRJEB31613	SAMEA5408177	F	(Fages et al., 2019)
BroughOfDeerness_VHR011	PRJEB31613	SAMEA5408178	M	(Fages et al., 2019)
BroughOfDeerness_VHR037	PRJEB31613	SAMEA5408179	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
BroughOfDeerness_VHR062	PRJEB31613	SAMEA5408180	F	(Fages et al., 2019)
Bruszcewo_Bru4	PRJEB31613	SAMEA5408181	M	(Fages et al., 2019)
CaminoDeLasYeseras_CdY2	PRJEB31613	SAMEA5408182	M	(Fages et al., 2019)
Cantorella_UE2275x2	PRJEB31613	SAMEA5408183	F	(Fages et al., 2019)
Capesterre_LIS2	PRJEB31613	SAMEA5408184	F	(Fages et al., 2019)
Capote_Cap102	PRJEB31613	SAMEA5408185	F	(Fages et al., 2019)
Charregass_NBxRa849	PRJEB31613	SAMEA5408186	M	(Fages et al., 2019)
Chartres_GVA1	PRJEB31613	SAMEA5408187	M	(Fages et al., 2019)
Chartres_GVA111	PRJEB31613	SAMEA5408188	F	(Fages et al., 2019)
Chartres_GVA112	PRJEB31613	SAMEA5408189	M	(Fages et al., 2019)
Chartres_GVA115	PRJEB31613	SAMEA5408190	M	(Fages et al., 2019)
Chartres_GVA26	PRJEB31613	SAMEA5408192	M	(Fages et al., 2019)
Chartres_GVA28	PRJEB31613	SAMEA5408193	M	(Fages et al., 2019)
Chartres_GVA36	PRJEB31613	SAMEA5408194	M	(Fages et al., 2019)
Chartres_GVA4	PRJEB31613	SAMEA5408196	M	(Fages et al., 2019)
Chartres_GVA43	PRJEB31613	SAMEA5408197	F	(Fages et al., 2019)
Chartres_GVA47	PRJEB31613	SAMEA5408198	M	(Fages et al., 2019)
Chartres_GVA48	PRJEB31613	SAMEA5408199	M	(Fages et al., 2019)
Chartres_GVA53	PRJEB31613	SAMEA5408200	M	(Fages et al., 2019)
Chartres_GVA56	PRJEB31613	SAMEA5408201	M	(Fages et al., 2019)
Chartres_GVA60	PRJEB31613	SAMEA5408202	M	(Fages et al., 2019)
Chartres_GVA75	PRJEB31613	SAMEA5408206	M	(Fages et al., 2019)
Chartres_GVA81	PRJEB31613	SAMEA5408208	F	(Fages et al., 2019)
Chartres_GVA9	PRJEB31613	SAMEA5408209	M	(Fages et al., 2019)
Dariali_Georgia2	PRJEB31613	SAMEA5408211	F	(Fages et al., 2019)
Derkul_NB2	PRJEB31613	SAMEA5408212	M	(Fages et al., 2019)
Derkul_NB4	PRJEB31613	SAMEA5408213	F	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
Dunaujvaros_Duk2	PRJEB22390	SAMEA104233676	M	(Gaunitz et al., 2018)
ElAcequion_Spain38	PRJEB31613	SAMEA5408214	F	(Fages et al., 2019)
ElAcequion_Spain39	PRJEB31613	SAMEA5408215	M	(Fages et al., 2019)
ElsVilars_UE4618	PRJEB31613	SAMEA5408216	F	(Fages et al., 2019)
Evreux_GVA133	PRJEB31613	SAMEA5408219	M	(Fages et al., 2019)
Evreux_GVA135	PRJEB31613	SAMEA5408220	M	(Fages et al., 2019)
Evreux_GVA140	PRJEB31613	SAMEA5408221	M	(Fages et al., 2019)
Fengtai_Fen4	PRJEB31613	SAMEA5408222	M	(Fages et al., 2019)
Fmontauban_GVA126	PRJEB22390	SAMEA104233677	M	(Gaunitz et al., 2018)
FrankfurtHeddenheim_Fr1	PRJEB31613	SAMEA5408223	M	(Fages et al., 2019)
Garbovat_Gar3	PRJEB22390	SAMEA104233678	M	(Gaunitz et al., 2018)
GolModII_Mon23	PRJEB31613	SAMEA5408224	F	(Fages et al., 2019)
GolModII_Mon24	PRJEB31613	SAMEA5408225	F	(Fages et al., 2019)
GolModII_Mon25	PRJEB31613	SAMEA5408226	F	(Fages et al., 2019)
GolModII_Mon26	PRJEB31613	SAMEA5408227	F	(Fages et al., 2019)
GolModII_Mon27	PRJEB31613	SAMEA5408228	F	(Fages et al., 2019)
GolModII_Mon28	PRJEB22390	SAMEA104233679	M	(Gaunitz et al., 2018)
Goyet_Vert293	PRJEB31613	SAMEA5408229	F	(Fages et al., 2019)
Goyet_Vert300	PRJEB31613	SAMEA5408230	F	(Fages et al., 2019)
Goyet_Vert304	PRJEB31613	SAMEA5408231	F	(Fages et al., 2019)
Granastadir_VHR031	PRJEB31613	SAMEA5408233	F	(Fages et al., 2019)
Gregorevka4_PAVH2	PRJEB22390	SAMEA104233680	M	(Gaunitz et al., 2018)
Halvai_KSH4	PRJEB31613	SAMEA5408234	F	(Fages et al., 2019)
Halvai_KSH5	PRJEB31613	SAMEA5408235	M	(Fages et al., 2019)
Haunstetten	PRJEB22390	SAMEA104233681	M	(Gaunitz et al., 2018)
Khatuu_Kha2_t1	PRJEB31613	SAMEA5408236	M	(Fages et al., 2019)
Khotont_UCIE2012x85	PRJEB31613	SAMEA5408237	M	(Fages et al., 2019)
Kokorevo_Rus3	PRJEB31613	SAMEA5408238	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
KoulianCave_MV178	PRJEB31613	SAMEA5408242	F	(Fages et al., 2019)
KrasnayaGorka_Rus48	PRJEB31613	SAMEA5408239	M	(Fages et al., 2019)
Krasnokamenka_NB10	PRJEB31613	SAMEA5408240	M	(Fages et al., 2019)
Krasnokamenka_NB9	PRJEB31613	SAMEA5408241	F	(Fages et al., 2019)
LebyazhinkaIV_NB35	PRJEB31613	SAMEA5408243	M	(Fages et al., 2019)
LongueilAnnel_GVA129	PRJEB31613	SAMEA5408244	M	(Fages et al., 2019)
Macon_GVA201	PRJEB31613	SAMEA5408245	M	(Fages et al., 2019)
Mainz_Mzr1	PRJEB31613	SAMEA5408246	M	(Fages et al., 2019)
Marvele_1	PRJEB31613	SAMEA5408247	M	(Fages et al., 2019)
Marvele_16	PRJEB31613	SAMEA5408250	M	(Fages et al., 2019)
Marvele_18	PRJEB31613	SAMEA5408251	M	(Fages et al., 2019)
Marvele_2	PRJEB31613	SAMEA5408248	M	(Fages et al., 2019)
Marvele_21	PRJEB31613	SAMEA5408252	M	(Fages et al., 2019)
Marvele_22	PRJEB31613	SAMEA5408253	M	(Fages et al., 2019)
Marvele_27	PRJEB31613	SAMEA5408254	M	(Fages et al., 2019)
Marvele_32	PRJEB31613	SAMEA5408255	M	(Fages et al., 2019)
Marvele_5	PRJEB31613	SAMEA5408249	M	(Fages et al., 2019)
MerzlyYar_Rus45	PRJEB31613	SAMEA5408256	M	(Fages et al., 2019)
Metz_GVA321	PRJEB31613	SAMEA5408257	M	(Fages et al., 2019)
Miciurin_Mic2	PRJEB31613	SAMEA5408258	M	(Fages et al., 2019)
Museum_Earb5	PRJEB31613	SAMEA5408259	M	(Fages et al., 2019)
Museum_Earb6	PRJEB31613	SAMEA5408260	M	(Fages et al., 2019)
Noyon_GVA123	PRJEB31613	SAMEA5408261	F	(Fages et al., 2019)
Nustar_4	PRJEB31613	SAMEA5408262	M	(Fages et al., 2019)
Nustar_5	PRJEB31613	SAMEA5408263	M	(Fages et al., 2019)
Oktyabrsky_Rus37	PRJEB31613	SAMEA5408264	M	(Fages et al., 2019)
Oktyabrsky_Rus38	PRJEB31613	SAMEA5408265	M	(Fages et al., 2019)
OlonKurinGol_OKG1	PRJEB31613	SAMEA5408266	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
OlonKurinGol_OKG2	PRJEB31613	SAMEA5408267	M	(Fages et al., 2019)
Otepaa_Ote2	PRJEB31613	SAMEA5408268	M	(Fages et al., 2019)
Otok_OTOK16	PRJEB31613	SAMEA5408269	F	(Fages et al., 2019)
PotapovkaI_1	PRJEB31613	SAMEA5408270	F	(Fages et al., 2019)
Quoygrew_VHR017	PRJEB31613	SAMEA5408271	M	(Fages et al., 2019)
Ridala_Rid1	PRJEB22390	SAMEA104233682	M	(Gaunitz et al., 2018)
Ridala_Rid2	PRJEB31613	SAMEA5408272	M	(Fages et al., 2019)
Saadjarve_Saa1	PRJEB31613	SAMEA5408273	M	(Fages et al., 2019)
Sagzabad_SAGS27	PRJEB31613	SAMEA5408274	M	(Fages et al., 2019)
SaintJust_GVA242	PRJEB31613	SAMEA5408279	M	(Fages et al., 2019)
SaintQuentin_GVA237	PRJEB31613	SAMEA5408280	M	(Fages et al., 2019)
Santarem_254	PRJEB31613	SAMEA5408282	F	(Fages et al., 2019)
Sayangorsk_Rus41	PRJEB31613	SAMEA5408283	F	(Fages et al., 2019)
Schloßvippach_Svi6	PRJEB31613	SAMEA5408284	M	(Fages et al., 2019)
Sebastovce_131	PRJEB31613	SAMEA5408285	M	(Fages et al., 2019)
SharIQumis_AM115	PRJEB31613	SAMEA5408286	M	(Fages et al., 2019)
SharIQumis_AM181	PRJEB22390	SAMEA104233683	M	(Gaunitz et al., 2018)
Sintashta_NB44	PRJEB31613	SAMEA5408287	M	(Fages et al., 2019)
Sintashta_NB45	PRJEB31613	SAMEA5408288	M	(Fages et al., 2019)
Sintashta_NB46	PRJEB19970 PRJEB10532	SAMEA3514287	F	(Librado et al., 2017)
SolothurnVigier_NB175	PRJEB31613	SAMEA5408289	M	(Fages et al., 2019)
SolothurnVigier_NB63	PRJEB31613	SAMEA5408290	F	(Fages et al., 2019)
Syrgal_Syr1t1c3	PRJEB22390	SAMEA104233684	M	(Gaunitz et al., 2018)
Syrgal_Syr1t1c4	PRJEB31613	SAMEA5408292	M	(Fages et al., 2019)
TachtiPerda_TP4	PRJEB22390	SAMEA104233685	M	(Gaunitz et al., 2018)
TavanTolgoi_GEP13	PRJEB31613	SAMEA5408293	M	(Fages et al., 2019)
TavanTolgoi_GEP14	PRJEB31613	SAMEA5408294	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
TavanTolgoi_GEP21	PRJEB31613	SAMEA5408295	M	(Fages et al., 2019)
TepeHasanlu_1140	PRJEB31613	SAMEA5408296	F	(Fages et al., 2019)
TepeHasanlu_2327	PRJEB31613	SAMEA5408297	M	(Fages et al., 2019)
TepeHasanlu_2405	PRJEB22390	SAMEA104233686	M	(Fages et al., 2019)
TepeHasanlu_2529	PRJEB31613	SAMEA5408298	F	(Fages et al., 2019)
TepeHasanlu_2689	PRJEB31613	SAMEA5408299	F	(Fages et al., 2019)
TepeHasanlu_3394	PRJEB31613	SAMEA5408300	M	(Fages et al., 2019)
TepeHasanlu_3398	PRJEB31613	SAMEA5408301	F	(Fages et al., 2019)
TepeHasanlu_3461	PRJEB31613	SAMEA5408303	F	(Fages et al., 2019)
TepeHasanlu_368	PRJEB31613	SAMEA5408304	M	(Fages et al., 2019)
Tumeski_CGG101397	PRJNA225855	SAMN02383958	M	(Der Sarkissian et al., 2014)
Uppsala_Upps02	PRJEB31613	SAMEA5408306	M	(Fages et al., 2019)
UushgiinUvur_Mon37	PRJEB31613	SAMEA5408307	M	(Fages et al., 2019)
UushgiinUvur_Mon39	PRJEB31613	SAMEA5408308	F	(Fages et al., 2019)
UushgiinUvur_Mon40	PRJEB31613	SAMEA5408309	M	(Fages et al., 2019)
UushgiinUvur_Mon41	PRJEB31613	SAMEA5408310	M	(Fages et al., 2019)
UushgiinUvur_Mon42	PRJEB31613	SAMEA5408311	F	(Fages et al., 2019)
UushgiinUvur_Mon43	PRJEB31613	SAMEA5408312	F	(Fages et al., 2019)
UushgiinUvur_Mon44	PRJEB31613	SAMEA5408313	M	(Fages et al., 2019)
UushgiinUvur_Mon45	PRJEB31613	SAMEA5408314	F	(Fages et al., 2019)
UushgiinUvur_Mon79	PRJEB31613	SAMEA5408315	M	(Fages et al., 2019)
UushgiinUvur_Mon84	PRJEB22390	SAMEA104233687	M	(Gaunitz et al., 2018)
UushgiinUvur_Mon86	PRJEB22390	SAMEA104233688	M	(Gaunitz et al., 2018)
UushgiinUvur_Mon87	PRJEB31613	SAMEA5408316	F	(Fages et al., 2019)
UushgiinUvur_Mon89	PRJEB31613	SAMEA5408317	M	(Fages et al., 2019)
Vermand_GVA199	PRJEB31613	SAMEA5408318	F	(Fages et al., 2019)
Vicerrectorado_VIR175	PRJEB31613	SAMEA5408319	M	(Fages et al., 2019)

Sample	BioProject	BioSample	Sex	Reference
WhitehallRomanVilla_UK08	PRJEB31613	SAMEA5408320	M	(Fages et al., 2019)
WitterPlace_UK15	PRJEB31613	SAMEA5408321	F	(Fages et al., 2019)
WitterPlace_UK16	PRJEB31613	SAMEA5408322	F	(Fages et al., 2019)
WitterPlace_UK17	PRJEB31613	SAMEA5408323	F	(Fages et al., 2019)
WitterPlace_UK18	PRJEB31613	SAMEA5408324	F	(Fages et al., 2019)
WitterPlace_UK19	PRJEB31613	SAMEA5408325	M	(Fages et al., 2019)
WitterPlace_UK20	PRJEB31613	SAMEA5408326	F	(Fages et al., 2019)
Yenikapi_Tur140	PRJEB31613	SAMEA5408328	M	(Fages et al., 2019)
Yenikapi_Tur141	PRJEB31613	SAMEA5408329	M	(Fages et al., 2019)
Yenikapi_Tur142	PRJEB31613	SAMEA5408330	M	(Fages et al., 2019)
Yenikapi_Tur145	PRJEB31613	SAMEA5408332	M	(Fages et al., 2019)
Yenikapi_Tur146	PRJEB31613	SAMEA5408333	M	(Fages et al., 2019)
Yenikapi_Tur150	PRJEB31613	SAMEA5408336	M	(Fages et al., 2019)
Yenikapi_Tur170	PRJEB31613	SAMEA5408339	M	(Fages et al., 2019)
Yenikapi_Tur171	PRJEB31613	SAMEA5408340	M	(Fages et al., 2019)
Yenikapi_Tur172	PRJEB22390	SAMEA104233689	M	(Gaunitz et al., 2018)
Yenikapi_Tur173	PRJEB31613	SAMEA5408341	M	(Fages et al., 2019)
Yenikapi_Tur175	PRJEB31613	SAMEA5408342	M	(Fages et al., 2019)
Yenikapi_Tur176	PRJEB31613	SAMEA5408343	M	(Fages et al., 2019)
Yenikapi_Tur181	PRJEB31613	SAMEA5408344	M	(Fages et al., 2019)
Yenikapi_Tur193	PRJEB31613	SAMEA5408347	M	(Fages et al., 2019)
Yenikapi_Tur194	PRJEB31613	SAMEA5408348	M	(Fages et al., 2019)
Yenikapi_Tur229	PRJEB31613	SAMEA5408350	M	(Fages et al., 2019)
Yenikapi_Tur243	PRJEB31613	SAMEA5408351	M	(Fages et al., 2019)
Yerqorqan_YER28	PRJEB31613	SAMEA5408358	M	(Fages et al., 2019)
Zhanaturmus_Issyk1	PRJEB22390	SAMEA104233690	M	(Gaunitz et al., 2018)
Przewalski_Paratype	PRJEB10098	SAMEA3498593	F	(Der Sarkissian et al., 2015)

Sample	BioProject	BioSample	Sex	Reference
Batagai	PRJNA222593	SAMN02371510	M	(Librado et al., 2015)
Goyet_Vert311	PRJEB31613	SAMEA5408232	F	(Fages et al., 2019)
Taymyr_CGG10022	PRJEB7537	SAMEA2821680	F	(Schubert et al., 2014)
Taymyr_CGG10023	PRJEB7537	SAMEA2821681	M	(Schubert et al., 2014)

2.10.2 MCMC convergence metrics

2.10.2.1 Cattle GWAS SNPs

Table S2.5 MCMC convergence metrics for cattle GWAS SNPs

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs41637122	1	9852198	A	G	1.04	335
rs29024890	1	68175028	C	T	1.02	237
rs110607430	1	78564007	A	G	1.01	393
rs42355462	1	109284636	A	G	1.01	322
rs134242409	1	111056711	A	G	1.07	296
rs41751194	1	112551212	G	A	1.02	426
rs42218645	1	113745976	G	A	1.01	294
rs110308751	1	119749415	T	C	1.05	245
rs133868000	2	80276795	C	T	1.02	243
rs133782951	2	94268065	A	G	1.01	283
rs110959205	2	96410491	C	A	1.02	401
rs110348122	2	105139011	C	A	1.00	356
rs42201608	2	128589016	G	A	1.03	232
rs136647965	2	128672069	T	C	1.02	214
rs132678958	3	3427943	A	G	1.04	232
rs110423687	3	8945826	G	A	1.02	295
rs41565903	3	10640386	G	A	1.01	331
rs109543144	3	13689610	T	C	1.04	291
rs109428825	3	21128533	C	T	1.03	295
rs110855192	3	34098046	C	T	1.02	305
rs109502268	3	34310875	A	C	1.02	436
rs29019303	3	65924617	G	A	1.02	312
rs43357991	3	89311070	C	T	1.01	225
rs43350961	3	92015484	C	T	1.06	348
rs110127380	4	47737653	G	A	1.01	382

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs42756348	4	48028619	G	A	1.03	242
rs110713885	4	56846638	G	A	1.03	234
rs29004485	4	93261877	G	C	1.01	283
rs29019448	4	102109006	G	A	1.01	360
rs41652792	4	113584710	T	G	1.09	291
rs109290308	4	115946071	C	T	1.06	307
rs41621927	5	9837147	A	G	1.01	328
rs42448768	5	10856853	T	C	1.01	251
rs43431798	5	11314799	G	A	1.01	287
rs110895486	5	12438670	T	C	1.01	246
rs29003499	5	30374517	T	C	1.02	287
rs111008601	5	37901579	T	G	1.04	273
rs111032218	5	86641601	A	G	1.04	318
rs41653536	5	90255244	T	C	1.02	428
rs380754745	5	92383941	G	T	1.02	483
rs378261830	5	92414844	T	A	1.02	267
rs109538481	5	92417032	T	C	1.03	426
rs109061046	5	92763518	G	T	1.04	272
rs110818412	5	92764049	G	C	1.03	291
rs135938702	5	93768316	A	G	1.02	264
rs110187948	5	93773787	C	T	1.01	344
rs209397140	5	93945058	T	A	1.01	408
rs110292191	5	113468743	C	T	1.08	322
rs43007076	6	426582	C	T	1.03	324
rs41257336	6	3574803	T	C	1.01	365
rs109137267	6	14597357	G	A	1.01	296
rs385849404	6	16458976	A	G	1.01	245
rs42189414	6	20578947	T	G	1.01	303

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs42189387	6	20582550	T	C	1.00	334
rs41649876	6	27433375	T	C	1.02	274
rs41620119	6	28213461	T	G	1.01	459
rs41595105	6	32904847	C	T	1.03	366
rs109566008	6	34457042	C	T	1.05	311
rs111012626	6	34889220	C	T	1.01	252
rs109243604	6	37870319	G	A	1.01	274
rs135260513	6	38818283	T	C	1.01	401
rs43458270	6	39837065	A	G	1.08	307
rs137567750	6	42924558	G	A	1.01	234
rs136984149	6	44294488	C	T	1.03	287
rs41665298	6	44459263	A	G	1.01	414
rs43461031	6	44692207	T	C	1.01	313
rs41595994	6	44975057	G	A	1.03	244
rs43458005	6	45017700	A	C	1.07	254
rs110253547	6	46774053	T	C	1.01	321
rs109495237	6	46777671	G	A	1.01	379
rs109109118	6	47054628	G	T	1.01	321
rs42675214	6	51588791	G	A	1.02	347
rs110746806	6	59620476	T	C	1.13	330
rs41653413	6	61364700	C	T	1.00	297
rs135578329	6	63648858	A	G	1.02	341
rs109594486	6	65093783	C	T	1.01	382
rs135668414	6	67204004	T	C	1.04	262
rs133180598	6	67555985	T	C	1.04	462
rs110875592	6	68059441	G	T	1.09	367
rs41605943	6	73601110	C	T	1.01	314
rs110220717	6	74849823	G	A	1.03	223

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs134254312	6	76822977	T	C	1.05	248
rs110985084	6	78298832	C	T	1.03	236
rs109101513	6	82792569	C	T	1.01	313
rs41597420	7	8860921	C	T	1.01	416
rs109003743	7	20500709	G	T	1.02	383
rs136935442	7	25923497	G	T	1.03	438
rs42361379	7	43029354	T	C	1.02	244
rs43521213	7	48089488	G	A	1.02	272
rs110569150	7	55345542	G	A	1.01	270
rs42284419	7	58255324	G	T	1.00	394
rs41656938	7	90655911	T	G	1.01	304
rs110059753	7	93218452	A	T	1.01	359
rs137018675	7	94181023	A	G	1.02	267
rs109324223	7	97292917	T	C	1.03	445
rs383592398	7	97544465	C	A	1.02	273
rs110803813	7	103074895	C	T	1.02	426
rs41597368	7	109970008	C	G	1.02	425
rs136844821	8	1782458	A	G	1.08	327
rs137257775	8	2200430	A	G	1.02	315
rs134487663	8	5800547	C	T	1.02	334
rs41654633	8	37604683	T	C	1.02	325
rs41858642	8	42551771	C	A	1.01	295
rs42718723	8	68340517	G	A	1.05	393
rs43565085	8	73680263	A	G	1.02	257
rs41658330	8	83048502	T	C	1.01	230
rs43584717	9	2048367	T	C	1.06	298
rs137136823	9	9276802	C	A	1.03	263
rs109910252	9	36191579	T	C	1.02	309

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs110184218	9	36960364	C	A	1.03	377
rs136046475	9	37724159	G	A	1.03	502
rs41593345	9	45303296	T	G	1.02	288
rs109113936	9	86556291	G	A	1.04	304
rs41623544	9	97703867	G	A	1.04	223
rs110798626	X	20758679	T	C	1.02	216
rs135614039	X	32201695	G	A	1.00	246
rs137194221	X	38433216	G	A	1.02	273
rs132710921	X	53616449	G	A	1.02	291
rs132650062	X	54033526	C	T	1.03	435
rs137168395	X	54040600	C	A	1.02	337
rs109469328	X	54617594	A	G	1.02	238
rs137243134	X	58910783	A	C	1.04	371
rs135982667	X	58982113	G	A	1.01	252
rs110393463	X	60454458	C	T	1.03	242
rs134346798	X	61378939	C	T	1.01	294
rs41609600	X	62311454	C	T	1.04	239
rs134879327	X	63103584	C	T	1.02	331
rs109296672	X	64798892	T	C	1.03	257
rs137022787	X	66211397	C	T	1.01	263
rs133787631	X	80744475	A	C	1.03	310
rs134839836	X	93368909	C	T	1.03	289
rs109360280	X	94399529	C	T	1.03	257
rs110285027	10	3343140	C	T	1.03	357
rs136980328	10	12669210	C	T	1.02	246
rs134854665	10	34784433	T	C	1.01	334
rs43710033	10	43669993	A	G	1.03	349
rs136986107	10	76832220	G	A	1.02	332

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs134979955	10	77104114	G	A	1.02	376
rs109887835	10	86036359	G	A	1.02	336
rs110386638	10	86089652	C	A	1.01	245
rs43657898	11	3589846	A	T	1.02	241
rs43672740	11	16412384	G	A	1.05	382
rs110744803	11	47629102	G	T	1.02	368
rs41629959	11	56112585	T	C	1.03	364
rs109636797	11	67059602	G	A	1.00	405
rs133187703	11	71401938	G	A	1.04	213
rs109018444	11	75352649	C	T	1.03	249
rs43708493	11	77235847	C	T	1.00	385
rs134776118	11	96445983	C	T	1.09	274
rs108956157	11	100883976	T	G	1.05	216
rs41622482	12	23340411	G	A	1.02	230
rs110146441	12	88202582	G	A	1.01	229
rs41659130	13	6434606	G	A	1.02	306
rs41682296	13	16562560	G	A	1.02	235
rs41679667	13	19697650	G	T	1.03	319
rs109275885	13	20788173	T	G	1.02	292
rs133954577	13	29027517	G	A	1.01	381
rs41701255	13	67309183	G	A	1.10	536
rs209058440	14	1752281	T	C	1.03	307
rs465770218	14	1797137	C	T	1.02	334
rs110749653	14	2138926	C	T	1.03	332
rs110716848	14	2311270	T	A	1.01	239
rs110058092	14	2546280	C	T	1.03	324
rs108992746	14	2951045	C	T	1.02	308
rs137068630	14	3640627	C	G	1.04	420

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs109264011	14	4734579	G	A	1.02	231
rs109230014	14	4943963	G	A	1.03	386
rs110006921	14	4984674	A	G	1.02	386
rs109567244	14	5769804	T	C	1.01	495
rs108938638	14	6964980	A	G	1.02	333
rs136062085	14	7084443	G	A	1.03	263
rs43138052	14	7203134	T	C	1.01	411
rs110737431	14	7478147	G	T	1.02	303
rs109176086	14	7566056	C	T	1.02	348
rs110547220	14	9508873	G	C	1.01	362
rs109383111	14	9606819	A	G	1.02	277
rs137730758	14	21069733	T	C	1.01	374
rs29021868	14	31014368	C	T	1.00	373
rs41730924	14	46921191	A	G	1.02	284
rs110813155	14	48158499	A	G	1.03	267
rs41727727	14	64689471	C	T	1.02	251
rs133892263	15	28013280	A	G	1.00	260
rs208575577	15	34802991	C	T	1.02	264
rs109092953	16	1690484	T	C	1.01	314
rs41609061	16	20099741	G	A	1.01	309
rs41802694	16	36001082	G	A	1.03	312
rs41640569	16	57529654	A	G	1.03	246
rs41625414	16	67703949	T	G	1.02	284
rs41634183	16	71208967	G	A	1.02	430
rs41634367	17	6295259	T	C	1.01	231
rs41838716	17	12288221	A	G	1.03	236
rs41575617	17	14357124	A	C	1.02	381
rs110897513	17	20336068	A	G	1.01	302

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs110937773	17	35247491	T	C	1.00	292
rs136889580	17	40959536	C	T	1.02	466
rs43328777	17	45255651	T	G	1.02	225
rs43326617	17	45284754	C	T	1.02	314
rs109640232	17	46405498	C	T	1.02	364
rs136937699	17	49004586	T	C	1.03	232
rs110623921	17	53718022	C	T	1.02	212
rs41845088	17	59554302	G	A	1.01	209
rs110293277	17	66725739	T	C	1.05	427
rs41850492	17	66736462	C	T	1.01	253
rs42277010	17	68788808	G	A	1.03	262
rs109646517	17	71120078	C	T	1.01	374
rs42684745	17	71853112	C	T	1.02	455
rs134679535	17	72718021	A	G	1.06	377
rs110199636	17	72739135	G	A	1.01	235
rs110402225	17	72747746	C	T	1.10	235
rs29009603	18	19037363	A	G	1.04	338
rs41636773	18	53970861	G	A	1.03	309
rs132769295	18	58123525	C	T	1.03	334
rs135009925	18	58669053	T	C	1.02	256
rs135456047	19	1942110	A	G	1.02	293
rs136325239	19	8833056	C	T	1.02	234
rs207768378	19	10371440	C	T	1.02	241
rs109165512	19	29811964	A	G	1.14	398
rs108976591	19	33941547	C	T	1.02	341
rs41636123	19	46075773	C	T	1.05	295
rs133786569	19	58597255	T	C	1.00	362
rs29013174	20	11139158	G	A	1.01	299

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs42246125	20	12879068	G	C	1.10	336
rs109397111	20	24985074	G	A	1.06	258
rs110576856	20	29995228	C	T	1.01	518
rs41639260	20	31904078	C	A	1.04	254
rs43021487	20	36152608	T	C	1.03	438
rs41581068	20	36336225	C	T	1.01	354
rs210962013	21	900664	G	A	1.13	397
rs109223213	21	1766094	G	A	1.10	476
rs42964821	21	1794556	G	A	1.01	237
rs134650763	21	2790290	G	T	1.03	288
rs136134264	21	2807802	C	T	1.02	285
rs41643809	21	2831075	G	A	1.03	269
rs134748436	21	2927369	A	G	1.04	289
rs42769474	22	2086012	C	T	1.01	291
rs110331907	22	25076967	C	T	1.02	267
rs41587560	23	2965430	T	C	1.03	322
rs108941415	23	3972340	C	T	1.01	314
rs110248306	23	7398527	C	A	1.00	212
rs109408667	23	8714121	G	A	1.02	268
rs41583098	23	10167082	G	A	1.01	342
rs109121210	23	13623324	A	G	1.02	287
rs110495370	23	14394214	C	A	1.03	305
rs41640104	23	15768513	T	C	1.01	297
rs109916764	23	28054444	G	A	1.01	357
rs110450042	23	30096862	A	G	1.02	223
rs29011699	23	32975635	G	A	1.01	315
rs110684599	23	35114464	G	T	1.01	334
rs41641410	23	40703070	T	C	1.05	292

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs109752755	24	1401886	A	G	1.03	349
rs42037422	24	3471386	T	C	1.03	267
rs41612383	24	5486959	A	G	1.02	477
rs110828071	24	26974278	T	C	1.01	298
rs29027126	24	43500292	T	C	1.06	396
rs109205642	25	1104544	A	G	1.01	317
rs110066931	25	1105508	G	A	1.01	430
rs109985869	25	1111293	T	C	1.01	393
rs110480449	25	1132326	C	T	1.02	345
rs109087616	25	1135306	G	A	1.01	357
rs110446187	25	1136024	A	G	1.05	377
rs110028146	25	1147232	A	G	1.04	253
rs136525746	25	1185329	G	A	1.05	323
rs110963327	25	1194125	G	A	1.04	386
rs109981439	25	1212489	C	T	1.02	380
rs111016512	25	1331969	G	A	1.06	297
rs111008794	25	1431881	T	C	1.03	299
rs109720251	25	1461390	T	G	1.01	330
rs136531545	25	1474546	G	A	1.00	644
rs110465674	25	1549991	C	T	1.01	350
rs110112679	25	1611399	C	T	1.01	339
rs109513627	25	1632935	G	A	1.04	227
rs110914965	25	1665327	A	C	1.02	294
rs209546511	25	1824775	C	T	1.02	420
rs210431907	25	1825003	T	C	1.01	252
rs384900477	25	1891613	A	G	1.01	320
rs110749311	25	3498960	G	A	1.01	431
rs109595393	25	3555431	C	T	1.00	291

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs109393691	25	3559329	C	T	1.01	280
rs110250832	25	4589105	C	T	1.02	340
rs385294664	25	8864280	A	G	1.01	305
rs444166770	25	9348092	T	C	1.02	307
rs110303715	25	9350410	C	T	1.04	315
rs385201321	25	9563772	G	A	1.01	406
rs137614716	25	9710233	C	T	1.02	234
rs133199082	25	9838940	C	T	1.02	318
rs137356061	25	9890595	C	T	1.03	228
rs137050873	25	9898361	A	G	1.03	284
rs135397399	25	9898664	G	A	1.01	257
rs135478256	25	9901100	C	T	1.01	348
rs134383773	25	9901890	C	T	1.03	334
rs136164611	25	9903378	C	A	1.04	316
rs380088064	25	9903605	C	T	1.02	360
rs136457906	25	9940813	T	C	1.01	273
rs381368598	25	9960972	C	G	1.01	332
rs380047300	25	9965071	G	A	1.03	336
rs211395087	25	9969596	A	C	1.02	422
rs379376340	25	9979371	C	T	1.01	387
rs135211343	25	9983745	T	C	1.02	301
rs384553361	25	9984599	G	A	1.01	275
rs137105505	25	10077046	C	T	1.04	287
rs109934816	25	13510306	C	T	1.03	251
rs110969500	25	15539265	C	T	1.03	314
rs110933959	25	24331537	G	A	1.00	306
rs42071221	25	26316640	A	G	1.00	334
rs41587739	25	33345774	T	C	1.03	412

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs111009159	25	33852697	C	T	1.01	332
rs41567529	25	35342491	C	T	1.05	302
rs135783257	26	6259435	A	G	1.01	293
rs134797277	26	6970122	A	G	1.03	359
rs29015731	26	8769852	G	A	1.01	232
rs42366847	26	9078964	C	T	1.02	330
rs42375643	26	12067863	G	A	1.02	317
rs133664558	26	13243687	T	C	1.00	313
rs137399384	26	13637163	A	G	1.01	219
rs110287393	26	16662006	T	C	1.01	326
rs42087774	26	17042328	T	C	1.02	283
rs43047902	26	17125312	G	A	1.04	322
rs42088194	26	17638682	T	C	1.02	319
rs42088152	26	17646892	G	A	1.00	271
rs42084524	26	18259297	G	A	1.02	298
rs137226762	26	18488870	T	C	1.01	297
rs42088561	26	18691383	G	T	1.08	414
rs110982357	26	19065046	A	G	1.04	297
rs132764360	26	19090353	A	G	1.02	426
rs110406912	26	20391606	G	A	1.03	377
rs42092324	26	20474308	A	G	1.05	262
rs41255691	26	21144884	C	T	1.01	262
rs41255688	26	21148111	T	C	1.03	448
rs42092174	26	21363670	G	A	1.05	250
rs42092165	26	21365778	A	C	1.00	240
rs42094442	26	21800887	G	A	1.08	253
rs134960396	26	23922332	G	A	1.02	279
rs134408889	26	23989648	A	G	1.04	359

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs109195079	26	24378302	C	T	1.00	248
rs137744436	26	24380331	C	T	1.02	280
rs42091614	26	24387381	G	C	1.01	339
rs41646524	26	24391410	G	A	1.02	321
rs42091567	26	24394671	T	G	1.01	311
rs42091548	26	24399559	A	G	1.01	303
rs42091534	26	24408492	G	A	1.05	315
rs42090833	26	24537058	G	A	1.02	238
rs42095892	26	24544308	G	A	1.02	358
rs209052351	26	24647414	G	A	1.01	348
rs42095368	26	24747566	A	G	1.02	342
rs207663344	26	25126927	G	C	1.01	339
rs42094411	26	25143992	C	A	1.03	230
rs109257502	26	25253444	G	A	1.02	205
rs110570159	26	25321229	G	A	1.02	378
rs109650179	26	25322990	C	T	1.03	236
rs110275411	26	29972396	T	G	1.06	283
rs110212716	26	30329201	T	G	1.02	351
rs42699577	26	30339282	T	C	1.06	306
rs42699555	26	30352858	C	T	1.01	398
rs42099466	26	34270114	A	G	1.01	271
rs134696790	26	38192805	G	A	1.05	313
rs109930382	26	47852501	C	T	1.02	244
rs42106668	26	47957153	G	A	1.02	418
rs109490826	26	49394472	G	A	1.01	334
rs110781068	27	3081686	G	A	1.01	229
rs42116249	27	14650855	T	C	1.04	288
rs109615803	27	20757482	G	A	1.01	309

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs42123527	27	24516466	G	A	1.06	329
rs42123979	27	25850995	C	T	1.07	420
rs41593538	27	28298862	T	C	1.02	390
rs457527504	27	35930734	G	A	1.02	335
rs137486809	27	36131079	C	T	1.05	328
rs209855549	27	36211708	C	T	1.02	333
rs109248310	27	36258043	G	A	1.01	226
rs137583429	27	36276567	G	A	1.02	400
rs109929384	27	37019464	C	T	1.01	239
rs110072942	27	37030084	G	A	1.02	391
rs210032345	27	37073667	A	G	1.01	265
rs383134052	27	37480137	A	G	1.03	311
rs209674186	27	38034553	C	T	1.01	246
rs134451152	27	38188191	G	A	1.01	301
rs382585984	27	40100234	C	T	1.07	464
rs41586304	27	41872925	G	A	1.01	278
rs43680506	28	5506637	A	G	1.01	357
rs110674951	28	6556095	T	C	1.02	413
rs42138037	28	7823597	A	C	1.03	332
rs109873278	28	8346709	G	T	1.00	295
rs41612930	28	18173916	G	A	1.02	373
rs41587042	28	20197058	G	A	1.04	271
rs41606880	28	24423785	G	A	1.03	391
rs42315935	28	36289049	C	T	1.06	426
rs209251861	28	39080987	G	A	1.01	241
rs109830972	28	43350234	C	T	1.06	236
rs41621937	29	5265850	A	G	1.03	313
rs134861001	29	9063897	G	A	1.01	347

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs109969230	29	13242840	T	C	1.02	288
rs110849642	29	13277661	G	A	1.05	328
rs134834406	29	13335947	C	T	1.03	294
rs109611361	29	13456917	C	T	1.02	380
rs43046837	29	13935802	T	C	1.02	257
rs109352104	29	15110952	A	G	1.02	309
rs41613164	29	15487637	G	A	1.00	380
rs108965864	29	19234709	C	T	1.02	280
rs42169108	29	19332759	A	G	1.00	279
rs42169423	29	20587268	C	A	1.01	331
rs42181608	29	29254185	G	A	1.01	469
rs110950145	29	37515087	A	G	1.01	221
rs29016922	29	40430062	A	G	1.01	368
rs42189852	29	41778946	G	A	1.01	280
rs383678427	29	41895289	G	A	1.00	237
rs41615211	29	42763775	G	C	1.02	281
rs42194242	29	43130312	G	A	1.01	315
rs109943243	29	44279347	G	A	1.05	348
rs42190003	29	44612074	A	G	1.01	342

2.10.2.2 Horse GWAS SNPs

Table S2.6 MCMC convergence metrics for horse GWAS SNPs

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs68478481	1	1644004	C	T	1.04	303
rs68481232	1	2176819	C	T	1.00	367
rs68527356	1	12056924	A	C	1.03	331
rs68609471	1	33484987	C	A	1.01	280
rs68619452	1	37686973	G	A	1.03	291

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs68628642	1	55909321	T	C	1.12	231
rs68512519	1	77958896	T	G	1.03	314
rs68444275	1	94215030	G	A	1.00	271
rs68516240	1	104634430	G	A	1.05	336
rs68445321	1	149273182	C	T	1.01	319
rs68487449	1	160479717	C	T	1.01	273
rs68497754	1	161466540	G	A	1.03	260
rs68534139	1	178879422	C	T	1.02	285
rs68552412	2	17767350	A	G	1.02	233
rs68629874	2	51218408	T	C	1.03	260
rs68704438	2	53350245	A	C	1.02	356
rs68618144	3	16221997	G	A	1.01	305
rs68458115	3	33216203	T	C	1.01	218
rs68458424	3	33889026	G	A	1.01	255
rs68651622	3	37540091	C	T	1.04	242
rs68651630	3	37548550	C	T	1.02	265
rs68651630	3	37548550	C	T	1.02	265
rs68651700	3	37789700	G	A	1.01	301
rs68676984	3	38050768	A	G	1.11	303
rs68678346	3	38502764	A	G	1.01	242
rs68678346	3	38502764	A	G	1.01	242
rs68667770	3	39861536	T	C	1.02	258
rs68595457	3	40523770	T	G	1.00	265
rs68597974	3	41631052	A	G	1.02	337
rs68624770	3	43022497	G	A	1.06	262
rs68626066	3	43265818	T	C	1.02	242
rs68625745	3	44944877	C	T	1.02	256
rs68626831	3	45228417	A	C	1.02	240
rs68562819	3	47219906	A	G	1.02	250

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs68562866	3	47452823	A	G	1.01	321
rs68550262	3	47766129	T	C	1.01	231
rs68559755	3	49923098	T	G	1.00	234
rs68498097	3	51729578	T	C	1.02	360
rs68499582	3	51782286	G	A	1.02	334
rs68588906	3	52272374	T	C	1.00	279
rs68591872	3	52638506	C	T	1.01	345
rs68557684	3	53396596	A	G	1.01	348
rs68528525	3	55940167	T	C	1.02	325
rs68528525	3	55940167	T	C	1.02	325
rs68539950	3	58026750	T	C	1.02	321
rs68546930	3	58379826	G	A	1.01	266
rs68599693	3	59874495	A	G	1.00	386
rs68525931	3	70453544	T	C	1.02	306
rs68469741	3	72464545	T	C	1.02	345
rs68540522	3	75416339	G	A	1.01	237
rs68631483	3	76206156	G	A	1.02	343
rs68713389	3	78402249	C	T	1.02	282
rs68512502	3	88493417	T	C	1.03	308
rs68596759	3	104403770	T	G	1.03	263
rs68517783	3	110933709	T	C	1.07	304
rs69595642	4	37862469	T	C	1.03	331
rs69595664	4	38074016	C	T	1.04	225
rs69592767	4	43678861	G	A	1.01	297
rs69567454	4	55937477	C	T	1.01	348
rs69602444	4	72076243	C	T	1.01	240
rs69602496	4	72263223	A	G	1.03	266
rs69578919	4	75637735	C	T	1.01	328
rs69578919	4	75637735	C	T	1.01	328

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs69581644	4	76807199	G	A	1.02	295
rs69587062	4	77084810	C	A	1.04	323
rs69588583	4	77454283	C	T	1.05	285
rs69577877	4	78412704	G	A	1.04	333
rs69480556	5	2449200	A	G	1.03	224
rs69497377	5	6851934	C	T	1.03	261
rs69481428	5	25902912	A	G	1.10	233
rs69570288	5	90129827	G	A	1.01	324
rs68679421	6	4645464	G	A	1.02	298
rs1147380069	6	20020914	T	C	1.05	698
rs68706495	6	75454449	G	A	1.01	286
rs68689852	7	19856060	C	T	1.07	228
rs68812621	7	76899332	G	A	1.01	332
rs68696531	7	81884331	G	A	1.02	251
rs68750361	8	11014804	A	G	1.03	251
rs68681009	8	19118089	T	C	1.01	353
rs68862716	8	36632827	C	A	1.02	341
rs68781333	8	48391504	C	T	1.01	318
rs68714332	8	50351947	T	C	1.02	238
rs68770018	8	51551764	A	G	1.03	222
rs68766799	8	54131804	T	C	1.04	249
rs68767947	8	56115677	A	G	1.08	287
rs68728843	8	68601294	G	A	1.02	272
rs68767154	8	69525181	T	C	1.01	344
rs68760204	8	70132806	G	A	1.01	318
rs68719723	9	2793185	C	T	1.02	225
rs68801114	9	19578673	T	C	1.01	289
rs68851316	9	34689915	T	C	1.01	297
rs68956271	9	34966747	G	A	1.00	260

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs68748134	9	74798143	G	A	1.03	306
rs69559518	X	23410847	A	G	1.04	256
rs68976249	10	30107493	C	A	1.08	302
rs396750816	11	21579177	T	C	1.01	312
rs1141844779	11	21751182	T	C	1.02	261
rs1150445524	11	22006773	A	G	1.02	352
rs68947757	11	32753122	T	C	1.03	284
rs1148094638	11	35414844	T	G	1.07	282
rs68964369	11	42975749	C	T	1.08	315
rs68967275	11	43560763	C	T	1.01	367
rs68960562	11	44420199	T	C	1.02	251
rs68910867	11	46064651	C	T	1.02	267
rs68888914	11	49589952	G	A	1.08	696
rs69030304	12	15778263	G	A	1.06	293
rs68903614	14	8134649	G	A	1.01	265
rs68948168	14	12656702	C	T	1.01	287
rs68986591	14	17534553	C	T	1.02	348
rs69029523	14	57685603	C	T	1.01	282
rs69029571	14	57995649	T	C	1.03	290
rs69029571	14	57995649	T	C	1.03	290
rs68974670	14	87254688	C	A	1.01	349
rs68975893	14	87408235	A	C	1.02	256
rs68995618	15	6057803	G	A	1.01	266
rs69013380	15	10927348	C	T	1.01	245
rs69031610	15	27542702	A	G	1.02	212
rs69050876	15	30758520	C	A	1.02	268
rs69017477	15	54419913	A	C	1.01	265
rs69074138	15	64551633	C	T	1.02	276
rs69000510	15	65519182	T	C	1.05	282

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs69007375	15	66547035	A	C	1.04	312
rs68969310	15	73009720	G	A	1.02	291
rs69008495	15	76004165	G	A	1.04	268
rs69023779	15	80654797	T	C	1.01	278
rs69031080	15	81143358	C	A	1.02	339
rs69052679	15	88693017	C	T	1.01	302
rs69038642	16	4488665	A	G	1.03	267
rs69057893	16	19887946	A	G	1.14	311
rs69027064	16	21282004	G	A	1.01	303
rs69022802	16	27201329	G	A	1.07	296
rs69089470	16	31837350	C	T	1.03	364
rs69089471	16	31837487	T	G	1.02	243
rs69090736	16	32011126	T	A	1.02	265
rs69012389	16	32561589	C	A	1.01	244
rs69015221	16	32879992	T	C	1.01	254
rs69109811	16	34541548	A	G	1.02	283
rs69227354	16	38564627	T	C	1.02	324
rs69068932	16	73393693	A	C	1.03	248
rs397128183	17	20813164	A	G	1.03	250
rs69108105	17	42932974	C	T	1.05	260
rs69149880	17	54213676	G	T	1.03	402
rs69178248	18	21601340	T	C	1.03	266
rs69171012	18	49758616	A	G	1.05	298
rs69125208	18	54328550	C	T	1.03	222
rs69186944	18	55476961	C	T	1.03	287
rs69192318	18	62054146	A	G	1.04	166
rs69196531	18	64252426	T	C	1.02	277
rs69223815	19	33380654	A	C	1.02	265
rs69250631	19	37674757	T	C	1.01	322

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs69135967	20	42679616	G	A	1.01	341
rs69127806	20	43195488	C	A	1.01	264
rs69127835	20	43319368	C	T	1.05	370
rs69129203	20	43520722	T	C	1.08	241
rs69167803	20	47313526	A	G	1.03	256
rs69234285	20	54561140	G	T	1.02	239
rs69201713	20	55757789	C	T	1.03	283
rs69155494	20	60629010	C	T	1.04	270
rs69201241	21	38275055	G	T	1.04	259
rs69257978	21	49202410	T	C	1.01	313
rs69216416	22	8781707	T	G	1.01	246
rs69307505	22	39170079	A	G	1.03	230
rs69450070	23	16796559	G	T	1.01	261
rs69272233	23	21277387	C	T	1.03	231
rs69272273	23	21304138	G	T	1.05	392
rs69274511	23	21512716	T	C	1.00	296
rs69275936	23	21689609	C	T	1.01	225
rs1142335494	23	22461979	T	C	1.01	316
rs1142335494	23	22461979	T	C	1.01	316
rs69236907	23	23202421	A	G	1.10	313
rs69238354	23	23389693	C	T	1.01	325
rs69328176	23	27480326	C	T	1.02	265
rs69324033	23	28542968	T	C	1.01	298
rs69267408	23	30185805	T	G	1.03	334
rs69271986	23	32891960	C	A	1.01	321
rs69364573	23	37851841	T	G	1.07	255
rs69364573	23	37851841	T	G	1.07	255
rs69338284	23	45845318	T	C	1.01	373
rs69261914	24	37501259	C	T	1.02	383

dbSNP RefSeq ID	Chr	Position	Ancestral	Derived	MPSRF	Min ESS
rs69291607	25	12939295	A	C	1.03	264
rs69291607	25	12939295	A	C	1.03	264
rs69319908	25	34218757	A	C	1.01	277
rs69312200	25	38546385	C	T	1.02	298
rs69342642	26	14009076	A	G	1.01	302
rs69269074	26	35703955	C	T	1.01	292
rs69359229	27	10676851	C	T	1.09	258
rs69384495	27	14840127	A	G	1.02	218
rs69332109	28	32996239	C	T	1.01	222
rs69374801	28	40749346	T	C	1.02	285
rs69429256	29	16812990	G	A	1.04	246
rs69450917	31	17012751	T	C	1.01	290

2.10.3 Supplementary References

Andersson, L.S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.-J., Patra, K., Arnason, T., Wellbring, L., Hjälm, G., Imsland, F., Petersen, J.L., McCue, M.E., Mickelson, J.R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., Lindgren, G., Andersson, L., Kullander, K., 2012. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature 488, 642–646.

Bosse, M., Megens, H.-J., Frantz, L.A.F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R.P.M.A., Groenen, M.A.M., 2014. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat. Commun. 5, 4392.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.J., Anderson, C., VanTassell, C.P., Hulsegge, I., Goddard, M.E., Guldbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R., Hayes, B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46, 858–865.

Der Sarkissian, C., Ermini, L., Jónsson, H., Alekseev, A.N., Crubezy, E., Shapiro, B., Orlando, L., 2014. Shotgun microbial profiling of fossil remains. Mol. Ecol. 23, 1780–1798.

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M.A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G.K., Albrechtsen, A., Vieira, F.G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., Neuditschko, M., Jagannathan, V., Feh, C., Greenblatt, C.L., Ludwig, A., Abramson, N.I., Zimmermann, W., Schafberg, R., Tikhonov, A., Sicheritz-Ponten, T., Willerslev, E., Marques-Bonet, T., Ryder, O.A., McCue, M., Rieder, S., Leeb, T., Slatkin, M., Orlando, L., 2015. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Curr. Biol. 25, 2577–2583.

Do, K.-T., Lee, J.-H., Lee, H.-K., Kim, J., Park, K.-D., 2014. Estimation of effective population size using single-nucleotide polymorphism (SNP) data in Jeju horse. Hanguk Tongmul Chawon Kwahakhoe Chi 56, 28.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., Alfarhan, A.H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J.H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., Berrocal-Rangel, L., Biglari, F., Boessenkool, S., Boldgiv, B., Brem, G., Brown, D., Burger, J., Crubézy, E., Daugnora, L., Davoudi, H., de Barros Damgaard, P., de Los Ángeles de Chorro Y de Villa-Ceballos, M., Deschler-Erb, S., Detry, C., Dill, N., do Mar Oom, M., Dohr, A., Ellingvåg, S., Erdenebaatar, D., Fathi, H., Felkel, S., Fernández-Rodríguez, C., García-Viñas, E., Germonpré, M., Granado, J.D., Hallsson, J.H., Hemmer, H., Hofreiter, M., Kasparov, A., Khasanov, M., Khazaeli, R., Kosintsev, P., Kristiansen, K., Kubatbek, T., Kuderna, L., Kuznetsov, P., Laleh, H., Leonard, J.A., Lhuillier, J., Liesau von Lettow-Vorbeck, C., Logvin, A., Lõugas, L., Ludwig, A., Luis, C., Arruda, A.M., Marques-Bonet, T., Matoso Silva, R., Merz, V., Mijiddorj, E., Miller, B.K., Monchalov, O., Mohaseb, F.A., Morales, A., Nieto-Espinet, A., Nistelberger, H., Onar, V., Pálsdóttir, A.H., Pitulko, V., Pitskhelauri, K., Pruvost, M., Rajic Sikanjic, P., Rapan Papeša, A., Roslyakova, N., Sardari, A., Sauer, E., Schafberg, R., Scheu, A., Schibler, J., Schlumbaum, A., Serrand, N., Serres-Armero, A., Shapiro, B., Sheikhi Seno, S., Shevnina, I., Shidrang, S., Southon, J., Star, B., Sykes, N., Taheri, K., Taylor, W., Teegen, W.-R., Trbojević Vukičević, T., Trixl, S., Tumen, D., Undrakhbold, S., Usmanova, E., Vahdati, A., Valenzuela-Lamas, S., Viegas, C., Wallner, B., Weinstock, J., Zaibert, V., Clavel, B., Lepetz, S., Mashkour, M., Helgason, A., Stefánsson, K., Barrey, E., Willerslev, E., Outram, A.K., Librado, P., Orlando, L., 2019. Tracking Five Millennia of Horse Management with

Extensive Ancient Genome Time Series. Cell 177, 1419–1435.e31.

Finno, C.J., Stevens, C., Young, A., Affolter, V., Joshi, N.A., Ramsay, S., Bannasch, D.L., 2015. SERPINB11 frameshift variant associated with novel hoof specific phenotype in Connemara ponies. PLoS Genet. 11, e1005122.

Frantz, L.A.F., Haile, J., Lin, A.T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V.E., Daly, K.G., Battista, V.M., Price, M., Gron, K.J., Alexandri, P., Arbogast, R.-M., Arbuckle, B., Bălăşescu, A., Barnett, R., Bartosiewicz, L., Baryshnikov, G., Bonsall, C., Borić, D., Boroneanţ, A., Bulatović, J., Çakirlar, C., Carretero, J.-M., Chapman, J., Church, M., Crooijmans, R., De Cupere, B., Detry, C., Dimitrijevic, V., Dumitraşcu, V., du Plessis, L., Edwards, C.J., Erek, C.M., Erim-Özdoğan, A., Ervynck, A., Fulgione, D., Gligor, M., Götherström, A., Gourichon, L., Groenen, M.A.M., Helmer, D., Hongo, H., Horwitz, L.K., Irving-Pease, E.K., Lebrasseur, O., Lesur, J., Malone, C., Manaseryan, N., Marciniak, A., Martlew, H., Mashkour, M., Matthews, R., Matuzeviciute, G.M., Maziar, S., Meijaard, E., McGovern, T., Megens, H.-J., Miller, R., Mohaseb, A.F., Orschiedt, J., Orton, D., Papathanasiou, A., Pearson, M.P., Pinhasi, R., Radmanović, D., Ricaut, F.-X., Richards, M., Sabin, R., Sarti, L., Schier, W., Sheikhi, S., Stephan, E., Stewart, J.R., Stoddart, S., Tagliacozzo, A., Tasić, N., Trantalidou, K., Tresset, A., Valdiosera, C., van den Hurk, Y., Van Poucke, S., Vigne, J.-D., Yanevich, A., Zeeb-Lanz, A., Triantafyllidis, A., Gilbert, M.T.P., Schibler, J., Rowley-Conwy, P., Zeder, M., Peters, J., Cucchi, T., Bradley, D.G., Dobney, K., Burger, J., Evin, A., Girdland-Flink, L., Larson, G., 2019. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. Proc. Natl. Acad. Sci. U. S. A. 116, 17231–17238.

Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R.P.M.A., Larson, G., Groenen, M.A.M., 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat. Genet. 47, 1141–1148.

Frischknecht, M., Jagannathan, V., Plattet, P., Neuditschko, M., Signer-Hasler, H., Bachmann, I., Pacholewska, A., Drögemüller, C., Dietschi, E., Flury, C., Rieder, S., Leeb, T., 2015. A Non-Synonymous HMGA2 Variant Decreases Height in Shetland Ponies and Other Small Horses. PLoS One 10, e0140749.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I.J., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A.F., Davoudi, H., Alquraishi, S., Alfarhan, A.H., Al-Rasheid, K.A.S., Crubézy, E., Benecke, N., Olsen, S., Brown, D., Anthony, D., Massy, K., Pitulko, V., Kasparov, A., Brem, G., Hofreiter, M., Mukhtarova, G., Baimukhanov, N., Lõugas, L., Onar, V., Stockhammer, P.W., Krause, J., Boldgiv, B., Undrakhbold, S., Erdenebaatar, D., Lepetz, S., Mashkour, M., Ludwig, A., Wallner, B., Merz, V., Merz, I., Zaibert, V., Willerslev, E., Librado, P., Outram, A.K., Orlando, L., 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science 360, 111–114.

Haase, B., Jagannathan, V., Rieder, S., Leeb, T., 2015. A novel KIT variant in an Icelandic horse with white-spotted coat colour. Anim. Genet. 46, 466.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T.S., Vilstrup, J.T., Lear, T., Myka, J.L., Lundquist, J., Miller, D.C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Stagegaard, J., Strauss, G., Bertelsen, M.F., Sicheritz-Ponten, T., Antczak, D.F., Bailey, E., Nielsen, R., Willerslev, E., Orlando, L., 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc. Natl. Acad. Sci. U. S. A. 111, 18655–18660.

Jun, J., Cho, Y.S., Hu, H., Kim, H.-M., Jho, S., Gadhvi, P., Park, K.M., Lim, J., Paek, W.K., Han, K.,

Manica, A., Edwards, J.S., Bhak, J., 2014. Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. BMC Genomics 15 Suppl 9, S4.

Kim, H., Lee, T., Park, W., Lee, J.W., Kim, J., Lee, B.-Y., Ahn, H., Moon, S., Cho, S., Do, K.-T., Kim, H.-S., Lee, H.-K., Lee, C.-K., Kong, H.-S., Yang, Y.-M., Park, J., Kim, H.-M., Kim, B.C., Hwang, S., Bhak, J., Burt, D., Park, K.-D., Cho, B.-W., Kim, H., 2013. Peeling back the evolutionary layers of molecular mechanisms responsive to exercise-stress in the skeletal muscle of the racing horse. DNA Res. 20, 287–298.

Leegwater, P.A., Vos-Loohuis, M., Ducro, B.J., Boegheim, I.J., van Steenbeek, F.G., Nijman, I.J., Monroe, G.R., Bastiaansen, J.W.M., Dibbits, B.W., van de Goor, L.H., Hellinga, I., Back, W., Schurink, A., 2016. Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. BMC Genomics 17, 839.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M.A., Gamba, C., Seguin-Orlando, A., Mortensen, C.D., Petersen, B., Hoover, C.A., Lorente-Galdos, B., Nedoluzhko, A., Boulygina, E., Tsygankova, S., Neuditschko, M., Jagannathan, V., Thèves, C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Sicheritz-Ponten, T., Popov, R., Grigoriev, S., Alekseev, A.N., Rubin, E.M., McCue, M., Rieder, S., Leeb, T., Tikhonov, A., Crubézy, E., Slatkin, M., Marques-Bonet, T., Nielsen, R., Willerslev, E., Kantanen, J., Prokhortchouk, E., Orlando, L., 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. Proc. Natl. Acad. Sci. U. S. A. 112, E6889–97.

Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., Fages, A., Khan, N., Schubert, M., Jagannathan, V., Serres-Armero, A., Kuderna, L.F.K., Povolotskaya, I.S., Seguin-Orlando, A., Lepetz, S., Neuditschko, M., Thèves, C., Alquraishi, S., Alfarhan, A.H., Al-Rasheid, K., Rieder, S., Samashev, Z., Francfort, H.-P., Benecke, N., Hofreiter, M., Ludwig, A., Keyser, C., Marques-Bonet, T., Ludes, B., Crubézy, E., Leeb, T., Willerslev, E., Orlando, L., 2017. Ancient genomic changes associated with domestication of the horse. Science 356, 442–445.

Metzger, J., Tonda, R., Beltran, S., Agueda, L., Gut, M., Distl, O., 2014. Next generation sequencing gives an insight into the characteristics of highly selected breeds versus non-breed horses in the course of domestication. BMC Genomics 15, 562.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P.L.F., Fumagalli, M., Vilstrup, J.T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C.D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A.M.V., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G.D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J.F., Weinstock, J., Gregersen, K., Røed, K.H., Eisenmann, V., Rubin, C.J., Miller, D.C., Antczak, D.F., Bertelsen, M.F., Brunak, S., Al-Rasheid, K.A.S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M.T.P., Kjær, K., Sicheritz-Ponten, T., Jensen, L.J., Olsen, J.V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., Willerslev, E., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499, 74–78.

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T.S., Velazquez, A.M.V., Stenderup, J., Hoover, C.A., Rubin, C.-J., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., MacHugh, D.E., Kalbfleisch, T., MacLeod, J.N., Rubin, E.M., Sicheritz-Ponten, T., Andersson, L., Hofreiter, M., Marques-Bonet, T., Gilbert, M.T.P., Nielsen, R., Excoffier, L., Willerslev, E., Shapiro, B., Orlando, L., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc. Natl. Acad. Sci. U. S. A. 111, E5661–9.

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., Hesse, P., Fulton, D., Sauer, E.W., Mohaseb, F.A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Vahidi, S.M.F., MacHugh, D.E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B.S., Mashkour, M., Orton, D., Horwitz, L.K., Teasdale, M.D., Bradley, D.G., 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 365, 173–176.

Warmuth, V., Eriksson, A., Bower, M.A., Cañon, J., Cothran, G., Distl, O., Glowatzki-Mullis, M.-L., Hunt, H., Luís, C., do Mar Oom, M., Yupanqui, I.T., Ząbek, T., Manica, A., 2011. European domestic horses originated in two holocene refugia. PLoS One 6, e18194.

Whitacre, L.K., Hoff, J.L., Schnabel, R.D., Albarella, S., Ciotola, F., Peretti, V., Strozzi, F., Ferrandi, C., Ramunno, L., Sonstegard, T.S., Williams, J.L., Taylor, J.F., Decker, J.E., 2017. Elucidating the genetic basis of an oligogenic birth defect using whole genome sequence data in a non-model organism, Bubalus bubalis. Sci. Rep. 7, 39719.

2.11 Acknowledgments and Notes

Acknowledgments: We thank J. Haile, V. E. Mullin, K. G. Daly, D. G. Bradley, A. Fages and L. Orlando for providing ancient DNA sequencing data, and J. Enk for assistance with the in-solution capture assay. We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility for providing computing time.

Author contributions: L.A.F.F., G.L. and J.G.S. conceived of the project; E.K.I.P, L.A.F.F. and J.G.S. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F., G.L. and J.G.S.; and E.K.I.P. wrote the paper with input from all other authors.

2.12 Permission from Co-authors

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Selection Trajectories of Genetic Variants Underlying Domestic Animal Traits"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L. and J.G.S. conceived of the project; E.K.I.P, L.A.F.F. and J.G.S. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F., G.L. and J.G.S.; and E.K.I.P. wrote the paper with input from all other authors.

Date: Sept 30 2019

Name(s): Greger Larson

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Selection Trajectories of Genetic Variants Underlying Domestic Animal Traits"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L. and J.G.S. conceived of the project; E.K.I.P, L.A.F.F. and J.G.S. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F., G.L. and J.G.S.; and E.K.I.P. wrote the paper with input from all other authors.

Date: October 7, 2019

Name(s): Joshua Schraiber

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Selection Trajectories of Genetic Variants Underlying Domestic Animal Traits"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L. and J.G.S. conceived of the project; E.K.I.P, L.A.F.F. and J.G.S. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F., G.L. and J.G.S.; and E.K.I.P. wrote the paper with input from all other authors.

Date: 30/09/2019

Name(s): Laurent Frantz

Signature(s): Laurent Frantz

3 The Evolutionary History of Dogs in the Americas

This chapter was published on 06 Jul 2018 in *Science* Vol. 361, Issue 6397, pp. 81-85 https://doi.org/10.1126/science.aao4776.

3.1 Statement of Authorship

This work was the product of a large international collaboration, with multiple co-first authors leading separate parts of the project. A full list of all author contributions is given in Section 0.

Design of Research

I participated in the design of the computational analyses, under the supervision of L.A.F.F.

Data

I curated the published and novel genome-wide and SNP array data, and performed quality checks and filtering of the data, under the supervision of L.A.F.F.

Analysis

I led the ancestry analyses of the ancient nuclear DNA, under the supervision of L.A.F.F.; I wrote the bioinformatic pipeline to perform the computational analyses of the ancient data. I performed the principle component analyses (SmartPCA), the clustering analyses (ADMIXTURE), the neighbour joining tree analyses (NJ tree), the admixture graph analyses (TreeMix and qpGraph), and the formal statistical admixture tests (qpDstat and qpF4ratio).

I developed a novel software tool for the automated fitting of admixture graphs, using f4 statistics, to iteratively build the most parsimonious admixture graph for a given set of populations. I further developed an unsupervised hierarchal clustering algorithm, to group similar admixture graph topologies based on their graph edit distance.

I interpreted the ancestry results with M.N.L., L.A.F.F. and G.L.

Manuscript

I contributed to the writing of the main text with L.A.F.F., G.L., E.P.M., M.N.L. and A.P.; I wrote the 'Nuclear ancestry analyses' section in the Supplementary Materials with L.A.F.F. and M.N.L.; I generated Figs. 3.1c and 3.1d in the main text, and Figs. S3.8, S3.9, S3.10, S3.11, S3.12, S3.21, S3.22, S3.23, S3.24, S3.25, S3.26 in the Supplementary Materials.

3.2 Authors and Affiliations

Authors: Máire Ní Leathlobhair[1#], Angela R. Perri[2,3#], Evan K. Irving-Pease[4#], Kelsey E. Witt[5#], Anna Linderholm[4,6#], James Haile[4,7], Ophelie Lebrasseur[4], Carly Ameen[8], Jeffrey Blick[9,†], Adam R. Boyko[10], Selina Brace[11], Yahaira Nunes Cortes[12], Susan J. Crockford[13], Alison Devault[14], Evangelos A. Dimopoulos[4], Morley Eldridge[15], Jacob Enk[14], Shyam Gopalakrishnan[7], Kevin Gori[1], Vaughan Grimes[16], Eric Guiry[17] , Anders J. Hansen[7,18], Ardern Hulme-Beaman[4,8], John Johnson[19], Andrew Kitchen[20], Aleksei K. Kasparov[21], Young-Mi Kwon[1], Pavel A. Nikolskiy[21,22], Carlos Peraza Lope[23], Aurélie Manin[24,25], Terrance Martin[26], Michael Meyer[27], Kelsey Noack Myers[28], Mark Omura[29], Jean-Marie Rouillard[14,30], Elena Y. Pavlova[21,31], Paul Sciulli[32], Mikkel-Holger S. Sinding[7,18,33], Andrea Strakova[1], Varvara V. Ivanova[34] , Christopher Widga[35], Eske Willerslev[7], Vladimir V. Pitulko[21], Ian Barnes[11], M. Thomas P. Gilbert[7,36], Keith M. Dobney[8], Ripan S. Malhi[37,38], Elizabeth P. Murchison[1,a,§], Greger Larson[4,a,§] and Laurent A. F. Frantz[4,39,a,§]

Affiliations:

[1] *Transmissible Cancer Group, Department of Veterinary Medicine, University of Cambridge, Cambridge, U.K.*

[2] *Department of Archaeology, Durham University, Durham, U.K.*

[3] *Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

[4] *The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK.*

[5] *School of Integrative Biology, University of Illinois, Urbana-Champaign, USA*

[6] *Department of Anthropology, Texas A&M University, College Station, USA*

[7] *Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark*

[8] *Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK*

[9] *Department of Government and Sociology, Georgia College and State University, USA*

[10] *Department of Biomedical Sciences, Cornell University, Ithaca, USA*

[11] *Department of Earth Sciences, Natural History Museum, London, UK*

[12] *Department of Anthropology, University at Albany-SUNY, Albany, New York, USA*

[13] *Pacific Identifications Inc., Victoria, Canada*

[14] *Arbor Biosciences, Ann Arbor, USA.*

[15] *Millennia Research Limited, Victoria, Canada*

[16] *Department of Archaeology, Memorial University, Queen's College, St. John's, Canada*

[17] *Department of Anthropology, University of British Columbia, Vancouver, Canada*

[18] *The Qimmeq Project, University of Greenland, Nuussuaq, Greenland*

[19] *Department of Anthropology, Santa Barbara Museum of Natural History, USA*

[20] *Department of Anthropology, University of Iowa, Iowa City, USA*

[21] *Institute for the Material Culture History, Russian Academy of Sciences, St Petersburg, Russia*

[22] *Geological Institute, Russian Academy of Sciences, Moscow, Russia*

[23] *Centro INAH Yucatán, Mérida, Yucatán, México*

[24] *Department of Archaeology, BioArCh, University of York, York, UK*

[25] *UMR 7209, Archéozoologie, Archéobotanique. Muséum national d'Histoire naturelle, Paris, France*

[26] *Research and Collections Center, Illinois State Museum, Springfield, USA*

[27] *Touray & Meyer Vet Clinic, Serrekunda, Gambia*

[28] *Glenn A. Black Laboratory of Anthropology, Indiana University Bloomington, USA*

[29] *Department of Mammalogy, Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA*

[30] *Chemical Engineering Department, University of Michigan, Ann Arbor, USA.*

[31] *Arctic & Antarctic Research Institute, St Petersburg, Russia*

[32] *Department of Anthropology, Ohio State University, Columbus, USA*

[33] *Natural History Museum, University of Oslo, Oslo, Norway*

[34] *VNIIOkeangeologia Research Institute, 1 Angliisky Ave., St Petersburg, 190021, Russia*

[35] *Center of Excellence in Paleontology, East Tennessee State University, Gray, USA*

[36] *Norwegian University of Science and Technology, University Museum, Trondheim, Norway*

[37] *University of Illinois at Urbana-Champaign Department of Anthropology, USA*

[38] *University of Illinois at Urbana-Champaign Carl R. Woese Institute for Genomic Biology, USA*

[39] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*

[#] These authors contributed equally to this work

[a] These authors co-supervised this work

[†] deceased

[§] Corresponding authors: Laurent A. F. Frantz – laurent.frantz@arch.ox.ac.uk; Greger Larson – greger.larson@arch.ox.ac.uk; Elizabeth P. Murchison - epm27@cam.ac.uk

3.3 Abstract

Dogs were present in the Americas prior to the arrival of European colonists, but the origin and fate of these pre-contact dogs are largely unknown. We sequenced 71 mitochondrial and seven nuclear genomes from ancient North American and Siberian dogs spanning ~9,000 years. Our analysis indicates that American dogs were not domesticated from North American wolves. Instead, American dogs form a monophyletic lineage that likely originated in Siberia and dispersed into the Americas alongside people. After the arrival of Europeans, native American dogs almost completely disappeared, leaving a minimal genetic legacy in modern dog populations. Remarkably, the closest detectable extant lineage to pre-contact American dogs is the canine transmissible venereal tumour, a contagious cancer clone derived from an individual dog that lived up to 8,000 years ago.

3.4 Main Text

The history of the global dispersal of dogs remains contentious (Larson et al., 2012). In North America, the earliest confirmed dog remains have been radiocarbon dated to ~9,900 calibrated years before present (cal. BP) (Koster, Illinois; (Driggers, 2015; Perri, n.d.)), approximately 6,000 years after the earliest unambiguous evidence of humans arriving in North America (Goebel et al., 2008). While these early dogs were most likely not domesticated *in situ* (Leonard et al., 2002), the timing of their arrival and their geographic origins are unknown. Studies of the control region of mitochondrial DNA have suggested that the pre-contact American dog population was largely replaced following the introduction of European dogs after the arrival of Europeans, and Eurasian Arctic dogs (e.g., Siberian huskies) during the Alaskan gold rush (Brown et al., 2015; Leonard et al., 2002; Witt et al., 2015). It remains possible, however, that some modern American dogs retain a degree of ancestry from the pre-contact population (Shannon et al., 2015; van Asch et al., 2013).

We sequenced complete mitochondrial genomes (mitogenomes) from 71 archaeological dog remains collected in North America and Siberia (Figure 3.1a; Table S3.1) and analysed these with 145 mitogenomes derived from a global dataset of

modern and ancient canids (Driggers, 2015). A phylogenetic tree constructed from the mitogenomes indicated that all sampled pre-contact dogs (spanning ~9,000 years) formed a monophyletic group within dog haplogroup A (Figure 3.1b; Figure S3.3; Figure S3.6), which we refer to as pre-contact dogs (PCD). This analysis indicated that the most closely related mitochondrial lineage to the PCD clade are ~9,000 year-old dogs from Zhokhov Island in Eastern Siberia (Driggers, 2015) (Figure 3.1b; Figure S3.3; Figure S3.6). In addition, molecular clock analyses suggest that all PCD dogs shared a common ancestor ~14,600 years ago (95% high posterior density [HPD]: 16,484-12,965; Figure 3.1b; Figure S3.6), which diverged from a shared ancestor with the Zhokhov Island dogs ~1,000 years earlier (95% HPD:17,646-13,739; Figure 3.1b; Figure S3.6). Interestingly, these time frames are broadly coincident with early migrations into the Americas (Graf and Buvit, 2017; Moreno-Mayar et al., 2018; Raghavan et al., 2015).

To further investigate the evolutionary history of PCD, we generated low coverage nuclear genome sequences (~0.005-2.0x) from seven pre-contact dogs sampled in six locations in North America spanning ~9,000 years (Table S3.1). We analysed these nuclear data alongside publicly available datasets including 45 modern canid whole genomes sampled from Eurasia and the Americas (Table S3.2) (Fan et al., 2016; Frantz et al., 2016; Freedman et al., 2014; Wang et al., 2016). A neighbour-joining tree constructed using single nucleotide polymorphism (SNP) revealed that, like the mitogenome phylogeny, PCD individuals clustered in a distinct monophyletic lineage that is more closely related to dogs than to either Eurasian or North American wolves (Figure 3.1c). Furthermore, our nuclear genome analysis indicated that the closest-related sister clade to PCD consists of modern Arctic dogs from the Americas (including Alaskan malamutes, Greenland dogs and Alaskan huskies) and Eurasia (Siberian huskies; Figure 3.1c). TreeMix (Driggers, 2015) (Figure 3.1d), outgroup f3-statistics (Figure S3.13) and D-statistics (Figure S3.14; Figure S3.15) also supported this phylogenetic structure. Combined, our mitochondrial and nuclear results indicate that PCD were not domesticated *in situ* from North American wolves, but were instead introduced by people into the Americas via Beringia from a population that was related to modern Arctic dogs.

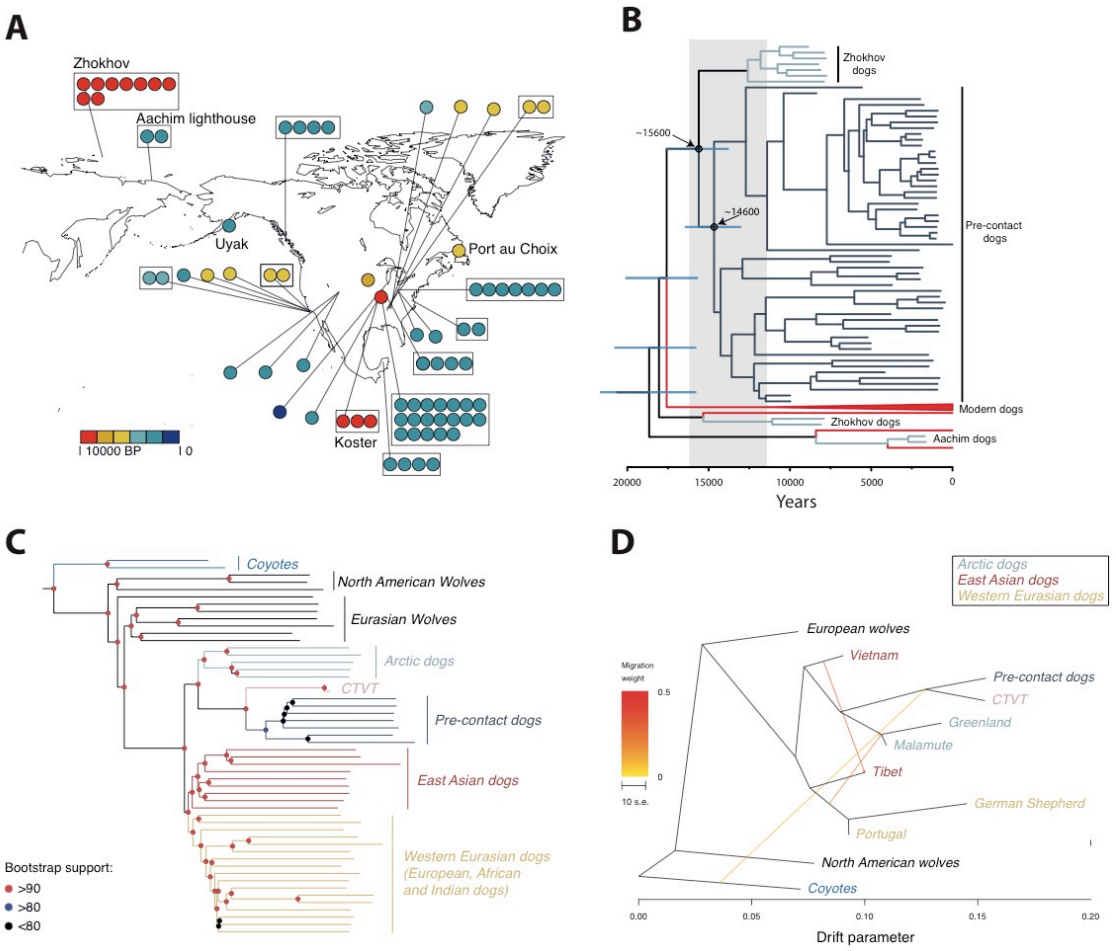


Figure 3.1 Sample location and ancestry of pre-contact dogs. a. A map depicting the location and age of the archaeological remains analysed in this study. Each dot represents a single sample, and multiple samples per archaeological site are grouped in boxes. Sites mentioned in the text are labelled. b. A tip calibrated Bayesian mitochondrial phylogenetic tree of dogs, within haplogroup A. This analysis was conducted with 71 novel ancient mitogenomes together with 145 publicly available mitogenomes from both modern and ancient canids (Driggers, 2015) (Figure S3.6). Red branches represent modern dogs. Blue horizontal bars on nodes represent 95% High Density Posterior age. The grey shaded area represents the time frame during which people entered the Americas (Graf and Buvit, 2017; Moreno-Mayar et al., 2018; Raghavan et al., 2015) c. A neighbour-joining tree built with whole genomes (Driggers, 2015). d An admixture graph constructed with TreeMix (based on transversions; Supplementary Material) depicting the relationship between PCD (including the Port Du Choix [AL3194] and Weyanoke Old Town [AL3223] samples) and other dog and wolf populations. We only used Greenland dogs and Malamute (American Arctic dogs) for this analysis as these are the least admixed with Western Eurasian dogs (Driggers, 2015).

Studies of nuclear data have identified two modern clades of global dogs: an East Asian clade (including dingoes) and a Western Eurasian clade (including European, Indian, and African dogs) (Frantz et al., 2016; Shannon et al., 2015; Wang et al., 2016). These analyses placed modern Arctic dogs with either Western Eurasian (Frantz et al., 2016; Vonholdt et al., 2010) or East Asian dogs (Shannon et al., 2015; Wang et al., 2016). Our analyses of nuclear data revealed a close relationship between Arctic dogs and PCD which together form a clade (PCD/Arctic) that is basal to both Western Eurasian and East Asian dogs and suggests the existence of a third monophyletic clade of dogs (Figure 3.1c). Though all three clades are well-supported, the relationships between them are ambiguous. For example, our outgroup f3-statistics analysis (Figure S3.13) indicated that the PCD/Arctic clade is basal to the two other Eurasian dog clades. However, when excluding specific East Asian dogs that possess evidence of gene flow from European dogs (Table S3.7; (Wang et al., 2016)), East Asian dogs became the most basal clade in a neighbour joining tree, and the PCD/Arctic clade became the sister clade to Western Eurasian dogs (Figure S3.11). Conversely, admixture graphs ((Driggers, 2015); Figure S3.25) and *TreeMix* (Pickrell and Pritchard, 2012) (Figure 3.1d) suggested that the PCD/Arctic clade is closest to East Asian dogs and West Eurasian dogs are the most basal. Conflicting phylogenies based on nuclear data have been reported on numerous occasions (Frantz et al., 2016; Larson et al., 2012; Wang et al., 2016), and these inconsistent topologies could result either from substantial post-divergence gene flow among Eurasian dogs (Figure 3.1c; Figure S3.25; (Driggers, 2015; Wang et al., 2016)), or from a near simultaneous divergence of all three lineages.

Our nuclear data indicates that modern Arctic dogs sampled from both Siberia and North America cluster in a distinct phylogenetic group that forms a sister taxon to PCD (Figure 3.1c). This close phylogenetic relationship between modern American Arctic dogs (Alaskan malamutes, Alaskan huskies, and Greenland dogs) and modern Eurasian Arctic dogs (Siberian huskies; Figure 3.1c; Figure S3.11; Figure S3.13) suggests that PCD are not the direct ancestor of modern American Arctic dogs. It is possible that modern American Arctic dogs are the descendants of dogs brought by the Paleo-Eskimo (~6,000 years ago) or by the Thule (~1,000 years ago) (Raghavan et al., 2014). However, both mitogenomic and low coverage nuclear data from a late Paleo-Eskimo dog from Kodiak Island, Alaska

(Uyak: AL3198; Figure 3.1a; Table S3.1) indicate that this dog is more closely related to PCD than to modern American Arctic dogs (Figure S3.10; Figure S3.4). This suggests that modern American Arctic dogs are not the descendants of Paleo-Eskimo dogs and that Paleo-Eskimos likely acquired local dogs in North America or brought Siberian dogs that were genetically indistinguishable from PCD. Our sampling did not include dogs from sites associated with the Thule culture, so it is plausible that the modern American Arctic dogs included in our analysis, such as Alaskan malamutes and Greenland dogs, are the descendants of dogs introduced by the Thule. Alternatively, the modern American Arctic dogs that we sampled may be the descendants of recently introduced Eurasian Arctic dogs, many of which were introduced during the 19th-century Alaskan gold rush and as sled dog racing stock (Brown et al., 2015).

Interestingly, genomic analyses of canine transmissible venereal tumour (CTVT) genomes indicated a close affinity with modern Arctic dogs (Murchison et al., 2014). CTVT is a contagious cancer clone that manifests as genital tumours and spreads between dogs by the transfer of living cancer cells during mating. This clone first originated from the cells of an individual dog, the "CTVT founder dog," which lived several thousand years ago, and still carries the genome of this individual (Murchison et al., 2014). To investigate the relationship between the CTVT founder dog and PCD, we analysed two CTVT genomes alongside a panel of modern and ancient canid genomes.

In order to accommodate for the fact that CTVT is a cancer, and to limit the impact of somatic mutations, we confined our genotyping analysis to SNPs which mapped to genomic regions that have retained both parental chromosomal copies in CTVT (Murchison et al., 2014), and excluded singleton SNPs exclusively called in CTVT genomes. Remarkably, CTVT clustered with PCD on neighbour-joining trees (Figure 3.1c; Figure S3.10; Figure S3.11), a Bayesian tree (Figure S3.12), TreeMix (Figure 3.1d) and admixture graphs (Figure S3.25). This result is further supported by both outgroup f3 (Figure S3.13) and D-statistics (Figure S3.14; Figure S3.15). These findings indicate that the CTVT founder dog is more closely related to PCD than to modern Arctic dogs. Multiple horizontal transfers of mitochondrial genomes from dog hosts to CTVT tumours has led to the replacement of the founder dog's mitogenome (Rebbeck et al., 2011;

Strakova et al., 2016), thus we could not determine the mitochondrial haplogroup of the CTVT founder dog and we limited our analyses to the nuclear genome.

To assess whether the CTVT founder dog lived prior to, or after dogs entered North America, we re-estimated its temporal origin by sequencing the nuclear genomes of two CTVT tumours, 608T and 609T. 608T is a CTVT tumour from the skin of a ten-month-old puppy which was likely engrafted from its mother's vaginal tumour (609T) during birth. We identified mutations with a clock-like mutational process which were present in 608T, but not detectable in 609T, and used these to derive a lower bound for a somatic mutation rate for CTVT (Driggers, 2015). Applying this rate to the total burden of clock-like somatic mutation in the CTVT lineage (Driggers, 2015), we estimated that the CTVT founder dog lived up to 8,225 years ago (Driggers, 2015). This time frame postdates the initial arrival of dogs into the Americas, raising the possibility that CTVT may have originated in a dog living in North America.

To further assess this scenario, we quantified the degree of introgression between North American endemic canids (coyotes and North American wolves), PCD dogs, modern Arctic dogs, and the CTVT founder dog. Our analyses indicated that, unlike Arctic dogs, PCD dogs share a number of derived alleles with coyotes and North American wolves, indicative of admixture (Figure S3.16; Figure S3.17). The CTVT founder dog also showed some weak evidence of coyote ancestry, but did not appear to possess admixture with North American wolves (Figure S3.16; Figure S3.17). Because coyotes are restricted to North America, this suggests that CTVT may have originated there. Since we did not ascertain the degree of coyote ancestry in ancient PCD-related dogs in Northern Siberia (such as the Zhokhov Island dogs, Figure 3.1), however, this analysis does not establish the location in which CTVT originated. Furthermore, studies that used somatic mutations to reconstruct the phylogeography of the CTVT clone indicated a deep divergence in Asia and a recent introduction to the Americas (Strakova et al., 2016). Altogether, these results suggest a scenario in which CTVT originated in Asia from a dog that was closely related to PCD, although we cannot exclude the possibility that the clone arose in America, then dispersed early into Asia before being reintroduced to America.

The legacy of PCD in modern American dog populations is uncertain. It has been suggested that some North American wolves obtained a mutation leading to black coat colour possibly via admixture with early American dogs (Anderson et al., 2009). This allele was not present, however, in either of the two higher coverage ancient PCD dogs in this study (Driggers, 2015) or in CTVT (Murchison et al., 2014). Additional ancient genomes are necessary to determine if this allele was present in the PCD population.

In addition, previous studies have argued that some modern American dog populations possess a genetic signature from indigenous American dogs (Parker et al., 2017; Shannon et al., 2015; van Asch et al., 2013). To test this hypothesis, we analysed nuclear data obtained from more than 5,000 modern dogs (including American village dogs) genotyped on a 180K SNP array (Shannon et al., 2015). We found 7-20% PCD ancestry in modern American Arctic dogs using f4 ratios (Alaskan husky, Alaskan malamute, and Greenland dogs; Table S3.10 & S3.11; Supplementary Material). This result, however, could reflect ancient population substructure in Arctic dogs rather than genuine admixture (Supplementary Material). Our *f4* ratio analysis did not detect a significant admixture signal from PCD into any modern American dogs of European ancestry (Table S3.10).

Our ADMIXTURE analysis detected varying degrees of PCD/Arctic ancestry in three individual Carolina dogs (0-33%; Figure S3.20). This analysis, however, could not distinguish between PCD and Arctic ancestry, and we cannot rule out that this was result of admixture from modern Arctic dogs and not from PCD (Driggers, 2015).The majority of modern American dog populations, including 138 village dogs from South America and multiple “native” breeds (e.g., hairless dogs and Catahoulas), possess no detectable traces of PCD ancestry (Figure S3.20; Table S3.10; Figure 3.2a), though this analysis may suffer from ascertainment bias.

To further assess the contribution of PCD to modern American dog populations, we also analysed 590 additional modern dog mitogenomes, including 169 village and breed dogs that were sampled in North and South America (Strakova et al., 2016). We identified

two modern American dogs (a chihuahua and a mixed breed dog from Nicaragua) that carried PCD mitochondrial haplotypes (Figure S3.5); consistent with a limited degree of PCD ancestry (<2%) in modern American dogs. We also identified three East Asian dogs that carried a PCD haplotype, possibly as a result of ancient population substructure or recent dog dispersal (Figure S3.5; (Driggers, 2015)). Although greater degrees of PCD ancestry may remain in American dogs which have not yet been sampled, our results suggest that European dogs almost completely replaced native American dog lineages. This near disappearance of PCD likely resulted from the arrival of Europeans, which led to shifts in cultural preferences and the persecution of indigenous dogs (Derr, 2005). Introduced European dogs may also have brought infectious diseases to which PCD were susceptible.

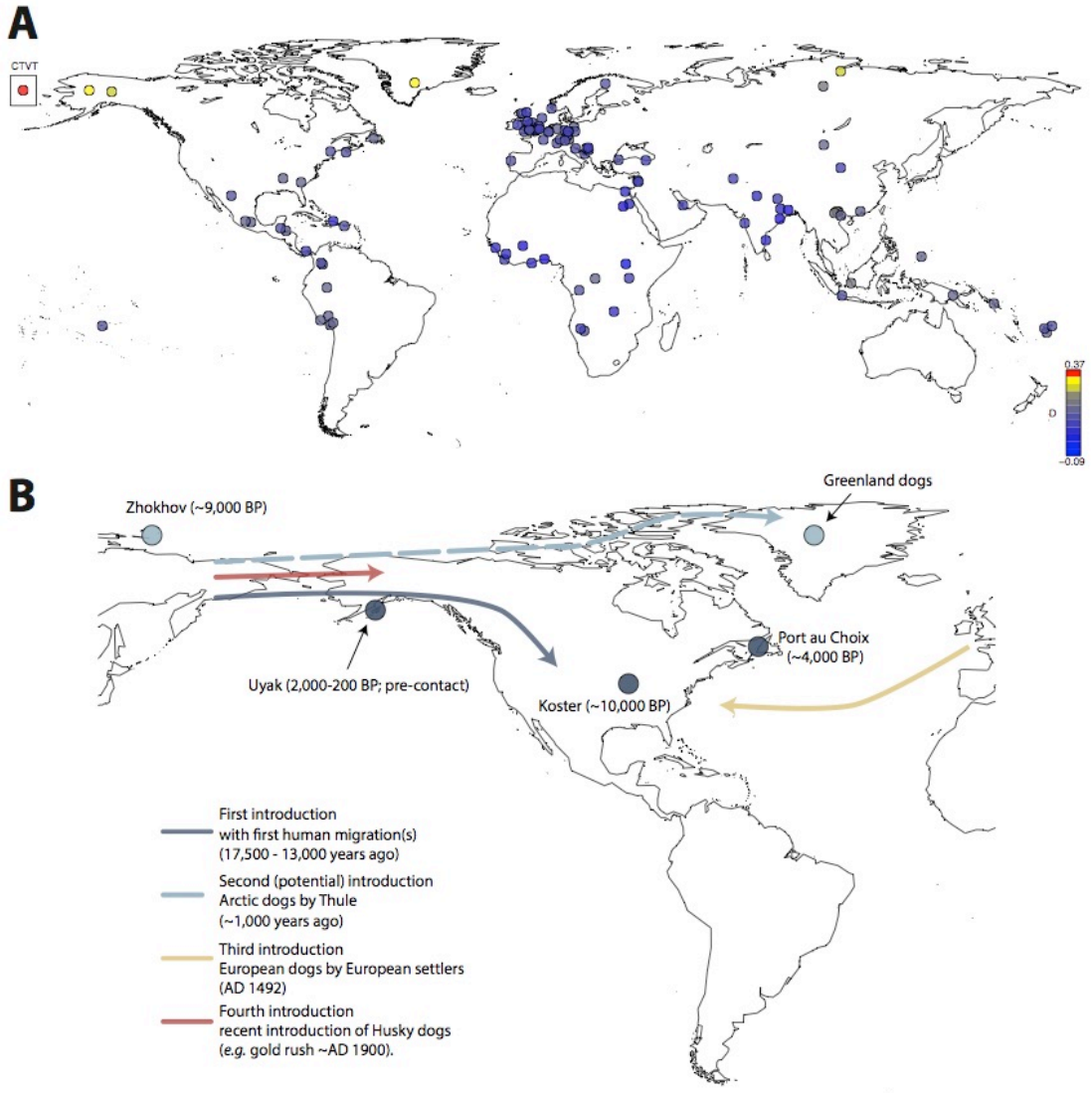


Figure 3.2 Legacy of pre-contact dogs in modern American dogs. a. A map showing the locations of dog populations obtained from (Shannon et al., 2015) and their degree of

relatedness (D-statistics) with the ~4ky old Port au Choix dog (AL3194; see (Driggers, 2015) and Figure S3.14). Higher values (in red) represent closer relatedness. b. A map depicting the multiple introductions of dogs into the Americas.

The first appearance of dogs in the North American archaeological record occurs ~6,000 years after the earliest evidence of human activity (Goebel et al., 2008; Graf and Buvit, 2017). In addition, our molecular clock analysis indicates that the PCD lineage appeared ~6,500 years after North American human lineages (Figure 3.1b) (Raghavan et al., 2015). These discrepancies suggest that dogs may not have arrived into the Americas alongside the first human migration. A recent human genetic study suggests that Northern Native American populations admixed with an East Siberian population ~11,500 years ago(Moreno-Mayar et al., 2018). This timing is compatible with both the archaeological record and our PCD divergence time estimate and suggests a scenario in which dogs were brought to the Americas several thousand years after the first people arrived.

This initial dog population entered North America then dispersed throughout the Americas where it remained isolated for at least 9,000 years. Within the past 1,000 years, however, there have been at least three independent re-introductions of dogs. The first may have consisted of Arctic dogs that arrived with the Thule culture ~1,000 years ago (Brown et al., 2015). Then, beginning in the 15th century, Europeans brought a second wave of dogs that appear to have almost completely replaced indigenous dogs. Lastly, Siberian huskies were introduced to the American Arctic during the Alaskan gold rush (Derr, 2005). As a result of these more recent introductions, the modern American dog population is largely derived from Eurasian breeds, and the closest known extant vestige of the first American dogs now exists as a worldwide transmissible cancer.

3.5 References

Agogino, G.A., Frankforter, W.D., 1960. A Paleo-Indian Bison-Kill in Northwestern Iowa. Am. Antiq. 25, 414–415.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., Stratton, M.R., 2015. Clock-like mutational processes in human somatic cells. Nat. Genet. 47, 1402–1407.

Allen, G.M., 1920. Dogs of the American aborigines. Museum [of Comparative Zoology].

Allen, G.M., 1939. Dog Skulls from Uyak Bay, Kodiak Island. J. Mammal. 20, 336–340.

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P.D., Dąbrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Population genomics of Bronze Age Eurasia. Nature 522, 167–172.

Anderson, T.M., vonHoldt, B.M., Candille, S.I., Musiani, M., Greco, C., Stahler, D.R., Smith, D.W., Padhukasahasram, B., Randi, E., Leonard, J.A., Bustamante, C.D., Ostrander, E.A., Tang, H., Wayne, R.K., Barsh, G.S., 2009. Molecular and evolutionary history of melanism in North American gray wolves. Science 323, 1339–1343.

Bai, B., Zhao, W.-M., Tang, B.-X., Wang, Y.-Q., Wang, L., Zhang, Z., Yang, H.-C., Liu, Y.-H., Zhu, J.-W., Irwin, D.M., Wang, G.-D., Zhang, Y.-P., 2014. DoGSD: the dog and wolf genome SNP database. Nucleic Acids Res. 43, D777–D783.

Barkalow, F.S., Jr., 1972. Vertebrate Remains from Archaeological Sites in the Tennessee Valley of Alabama. South. Indian Stud. 24, 3–41.

Beverley, R., 1705. The History and Present State of Virginia. The University of North Carolina Press, Chapel Hill.

Björnerfeldt, S., Webster, M.T., Vilà, C., 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication 990–994.

Blaine, D.P., 1810. A Domestic Treatise on the Diseases of Horses and Dogs. T. Boosey.

Blick, J.P., 1988. A preliminary report on the osteometric analysis of some aboriginal dogs (Canis familiaris) from Weyanoke Old Town, 44 PG 51, Prince George County, Virginia. Bull. Am. Assoc. Hist. Nurs. 43.

Blick, J.P., 2000. The Archaeology and Ethnohistory of the Dog in Virginia Algonquian Culture as Seen from Weyanoke Old Town. Algonquian Papers-Archive 31.

Bluhm, E.A., Liss, A., 1961. Anker Site, in: Chicago Area Archaeology. Illinois Archaeological Survey, Urbana.

Boon, A.L., 2013. A Faunal Analysis of the Eleventh Horizon of the Koster Site (11GE4). Indiana University of Pennsylvania.

Borgic, Q.L., Galloy, J.M., 2004. Domesticated Dog Remains from the Janey B. Goode Site.

Brereton, J., 1602. True Relation of the Discoveries of the North Part of Virginia.

Brown, J.A., Vierra, R.K., 1983. What happened in the Middle Archaic?: Introduction to an ecological approach to Koster Site Archaeology. Academic Press Orlando.

Brown, S.K., Darwent, C.M., Wictum, E.J., Sacks, B.N., 2015. Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. Heredity 115, 488–495.

Buikstra, J.E., 1981. Mortuary practices, paleodemography and paleopathology: a case study from the Koster site (Illinois).

Butzer, K.W., 1978. Changing Holocene Environments at the Koster Site: A Geo-Archaeological Perspective. Am. Antiq. 43, 408–413.

Church, F., Nass, J.P., Jr, Hart, J.P., Reith, C.B., 2002. Central Ohio Valley during the Late Prehistoric Period: Subsistence-Settlement Systems' Responses to Risk. Northeast Subsistence-Settlement Change: AD 700--1300 11.

Clark, D.W., 1974. Contributions to the Later Prehistory of Kodiak Island, Alaska. National Museum of Man Mercury Series, No. 20. National Museums of Canada: Ottawa.

Concannon, P., Whaley, S., Lein, D., Wissler, R., 1983. Canine gestation length: variation related to time of mating and fertile life of sperm. Am. J. Vet. Res. 44, 1819–1821.

Cook, R.A., 2012. DOGS OF WAR: POTENTIAL SOCIAL INSTITUTIONS OF CONFLICT, HEALING, AND DEATH IN A FORT ANCIENT VILLAGE. Am. Antiq. 77, 498–523.

Crockford, S.J., 2014. Analysis of the vertebrate fauna from neighbouring Prince Rupert Harbour sites GbTo-54 and GbTo-13: prehistoric mountain goat capital of North America. Pacific IDentification for Millenia Research Limited.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U. S. A. 110, 15758–15763.

Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., Allentoft, M.E., Sarkissian, C.D., Prüfer, K., Krause, J., Meyer, M., Reich, D., Raghavan, M., Raghavan, M., Rasmussen, M., Rasmussen, M., Orlando, L., Fu, Q., Seguin-Orlando, A., Gamba, C., Keller, A., Lazaridis, I., Olalde, I., Skoglund, P., Carpenter, M.L., Rizzi, E., Lari, M., Gigli, E., Bellis, G.D., Caramelli, D., Poinar, H.N., Sarkissian, C., Campos, P.F., Schwarz, C., Jans, M.M.E., Collins, M.J., Ginolhac, A., Orlando, L., Adler, C.J., Haak, W., Donlon, D., Cooper, A., Higgins, D., Austin, J.J., Higgins, D., Kaidonis, J., Townsend, G., Hughes, T., Austin, J.J., Trivedi, R., Chattopadhyay, P., Kashyap, V.K., Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M., Barnes, I., Willerslev, E., Cooper, A., Briggs, A.W., Meyer, M., Kircher, M., Lindgreen, S., Li, H., Durbin, R., Schubert, M., Li, H., Deagle, B.E., Eveson, J.P., Jarman, S.N., Allentoft, M.E., Jónsson, H., Allentoft, M.E., Fu, Q., Maricic, T., Whitten, M., Pääbo, S., Gansauge, M.-T., Meyer, M., Gansauge, M.-T., Meyer, M., 2015. Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. 5, 11184.

Decker, B., Davis, B.W., Rimbault, M., Long, A.H., Karlins, E., Jagannathan, V., Reiman, R., Parker, H.G., Drögemüller, C., Corneveaux, J.J., Chapman, E.S., Trent, J.M., Leeb, T., Huentelman, M.J., Wayne, R.K., Karyadi, D.M., Ostrander, E.A., 2015. Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumour. Genome Res. 25, 1646–1655.

Derr, M., 2005. A Dog's History of America: How Our Best Friend Explored, Conquered, and Settled a Continent. Farrar, Straus and Giroux.

Driggers, R., 2015. Publishing supplementary material: editorial. Appl. Opt. 54, ED12.

Drögemüller, C., Karlsson, E.K., Hytönen, M.K., Perloski, M., Dolf, G., Sainio, K., Lohi, H., Lindblad-Toh, K., Leeb, T., 2008. A mutation in hairless dogs implicates FOXI3 in ectodermal development. Science 321, 1462.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Eldridge, M., Parker, A., Mueller, C., Crockford, S., 2014. Archaeological investigations at Ya asqalu'i/Kaien Siding, Prince Rupert Harbour. Millennia Research Limited for CN.

Equihua, J.C., 2008. Proyecto Salvamento Arqueológico Tizayuca 2006. Informe. Secuencias estratigráficas y conjuntos arquitectónicos, tomo I. Technical report submitted to the INAH.

Eriksson, A., Manica, A., 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proc. Natl. Acad. Sci. U. S. A. 109, 13956–13960.

Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., Ramirez, O., Pollinger, J., Galaverni, M., Ortega Del-Vecchyo, D., Du, L., Zhang, W., Zhang, Z., Xing, J., Vilá, C., Marques-Bonet, T., Godinho, R., Yue, B., Wayne, R.K., 2015. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. gr.197517.115–.

Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., Ramirez, O., Pollinger, J., Galaverni, M., Ortega Del-Vecchyo, D., Du, L., Zhang, W., Zhang, Z., Xing, J., Vilà, C., Marques-Bonet, T., Godinho, R., Yue, B., Wayne, R.K., 2016. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. 26, 163–173.

Feest, C.F., 1967. The Virginia Indian in Pictures, 1612-1624. Smithsonian Institution.

Feest, C.F., 1978. Virginia Algonquians. Smithsonian Institution.

Fowler, M.L., 1959. Modoc Rock Shelter: An Early Archaic Site in Southern Illinois. Am. Antiq. 24, 257–270.

Fraley, C., Raftery, A.E., Murphy, T.B., Scrucca, L., n.d. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. 2012. University of Washington: Seattle.

Frankforter, W.D., Agogino, G.A., 1960. The Simonsen Site: Report for the Summer of 1959. Plains Anthropol. 10, 65–70.

Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É.A., Sablin, M.V., Bréhard, S., Mashkour, M., Bălăşescu, A., Gillet, B., Hughes, S., Chassaing, O., Hitte, C., Vigne, J.-D., Dobney, K., Hänni, C., Bradley, D.G., Larson, G., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352, 1228–1231.

Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É.A., Sablin, M.V., Bréhard, S., Mashkour, M., Bălăşescu, A., Gillet, B., Hughes, S., Chassaing, O., Hitte, C., Vigne, J.-D.,

Dobney, K., Hänni, C., Bradley, D.G., Larson, G., 2016a. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352, 1228–1231.

Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É.A., Sablin, M.V., Bréhard, S., Mashkour, M., Bălăşescu, A., Gillet, B., Hughes, S., Chassaing, O., Hitte, C., Vigne, J.-D., Dobney, K., Hänni, C., Bradley, D.G., Larson, G., 2016b. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352, 1228–1231.

Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A.R., Parker, H.G., Lee, C., Tadigotla, V., Siepel, A., Bustamante, C.D., Harkins, T.T., Nelson, S.F., Ostrander, E.A., Marques-Bonet, T., Wayne, R.K., Novembre, J., 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10, e1004016.

Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A.R., Parker, H.G., Lee, C., Tadigotla, V., Wilton, A., Siepel, A., Bustamante, C.D., Harkins, T.T., Nelson, S.F., Ostrander, E.A., Marques-Bonet, T., Wayne, R.K., Novembre, J., 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10, e1004016.

Futato, E., 2002. Middle and Late Archaic Settlement at the Perry Site, 1LU25, Lauderdale County, Alabama. Journal of Alabama Archaeology 48, 80–92.

Galloy, J.M., 2010. PART V : WOODLAND PERIOD The Janey B . Goode Site (11S1232): Highlights of Investigations at a Massive Late Prehistoric Site in the American Bottom. Illinois Archaeology 22, 529–552.

Gavrilovic, B.B., Andersson, K., Linde Forsberg, C., 2008. Reproductive patterns in the domestic dog--a retrospective study of the Drever breed. Theriogenology 70, 783–794.

Goebel, T., Waters, M.R., O'Rourke, D.H., 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319, 1497–1502.

Graf, K.E., Buvit, I., 2017. Human Dispersal from Siberia to Beringia: Assessing a Beringian Standstill in Light of the Archaeological Evidence. Curr. Anthropol. 58, S583–S603.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. a., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A draft sequence of the Neandertal genome. Science 328, 710–722.

Gregory, E.S., 1986. Weyanoke Old Town. Archeological Society of Virginia Quarterly Bulletin 41, 49–71.

Gregory, L.B., 1980. The Hatch site: A preliminary report. Virginia Archeological Society, Quarterly Bulletin 34, 239–248.

Griffin, J.B., 1978. Late prehistory of the Ohio Valley. Handbook of North American Indians 15, 547–559.

Griffin, J.B., 1992. Fort Ancient Has No Class: The Absence of an Elite Group in Mississippian Societies in the Central Ohio Valley. Archeological Papers of the American Anthropological Association 3, 53–59.

Groves, C.P., 1999. The advantages and disadvantages of being domesticated. Perspectives in Human Biology 4, 1–12.

Guiry, E.J., Grimes, V., 2013. Domestic dog (Canis familiaris) diets among coastal Late Archaic groups of northeastern North America: A case study for the canine surrogacy approach. Journal of Anthropological Archaeology 32, 732–745.

Ha, G., Roth, A., Lai, D., Bashashati, A., Ding, J., 2012. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast …. Genome.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M.A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211.

Hajic, E.R., 1990. Koster site archaeology I: stratigraphy and landscape evolution. Center for Amer Archeology Pr.

Hajic, E.R., 2017. Koster Site, Illinois, in: Gilbert, A.S. (Ed.), Encyclopedia of Geoarchaeology, Encyclopedia of Earth Sciences Series. Springer Netherlands, pp. 457–458.

Hao, Z., Zhang, Q., Qu, B., 2016. The complete mitochondrial genome of the Chinese indigenous dog. Mitochondrial DNA A DNA Mapp Seq Anal 27, 88–89.

Harriot, T., 1588. A Brief and True Report of the New Found Land of Virginia. digitalcommons.unl.edu, London.

Heizer, R.F., Hewes, G.W., Hrdlicka, A., 1956. nd Archaeology of the Uyak Site, Kodiak Island, Alaska.

Hill, F.C., 1972. A Middle Archaic Dog Burial in Illinois. Foundation for Illinois Archaeology.

Hood, B.C., 1993. The Maritime Archaic Indians of Labrador: Investigating Prehistoric Social Organization. Newfoundland and Labrador Studies 9.

Houart, G.L., 1971. Koster: a stratified Archaic site in the Illinois valley.

Hrdlička, A., 1944. The Anthropology of Kodiak Island. Wistar Institute of Anatomy and Biology.

Hulton, P., 1984. America 1585: The Complete Drawings of John White.

Jeske, R.J., Lurie, R., 1993. THE ARCHAEOLOGICAL VISIBILITY OF BIPOLAR TECHNOLOGY: AN EXAMPLE FROM THE KOSTER SITE. MidCont. J. Archaeol. 18, 131–160.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684.

Kane, A.E., Robinson, C.K., 1988. Dolores Archaeological Program: Anasazi Communities at Dolores: McPhee Village. USDI Bureau of Reclamation, Engineering and Research Center., Denver.

Katoh, K., Asimenos, G., Toh, H., n.d. Multiple Alignment of DNA Sequences with MAFFT, in:

Posada, D. (Ed.), Bioinformatics for DNA Sequence Analysis, Methods in Molecular Biology. Humana Press, pp. 39–64.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kerber, J.E., 1997. NATIVE AMERICAN TREATMENT OF DOGS IN NORTHEASTERN NORTH AMERICA: ARCHAEOLOGICAL AND ETHNOHISTORICAL PERSPECTIVES. Archaeology of Eastern North America 25, 81–95.

Kerber, J.E., Leveillee, A.D., Greenspan, R.L., 1989. AN UNUSUAL DOG BURIAL FEATURE AT THE LAMBERT FARM SITE, WARWICK, RHODE ISLAND: PRELIMINARY OBSERVATIONS. Archaeology of Eastern North America 17, 165–174.

Kircher, M., 2012. Analysis of high-throughput ancient DNA sequencing data. Methods Mol. Biol. 840, 197–228.

Kuckelman, K.A., 2003. The Archaeology of Yellow Jacket Pueblo (Site 5MT5): Excavations at a Large Community Center in Southwestern Colorado [WWW Document]. URL http://www.crowcanyon.org/yellowjacket.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol. Biol. Evol. 34, 772–773.

Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y.W., Peters, J., Stahl, P.W., Piper, P.J., Lingaas, F., Fredholm, M., Comstock, K.E., Modiano, J.F., Schelling, C., Agoulnik, A.I., Leegwater, P.A., Dobney, K., Vigne, J.-D., Vilà, C., Andersson, L., Lindblad-Toh, K., 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc. Natl. Acad. Sci. U. S. A. 109, 8878–8883.

Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y.W., Peters, J., Stahl, P.W., Piper, P.J., Lingaas, F., Fredholm, M., Comstock, K.E., Modiano, J.F., Schelling, C., Agoulnik, A.I., Leegwater, P.A., Dobney, K., Vigne, J.-D., Vila, C., Andersson, L., Lindblad-Toh, K., 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proceedings of the National Academy of Sciences 1203005109–.

Lee, E.J., Merriwether, D.A., Kasparov, A.K., Nikolskiy, P.A., Sotnikova, M.V., Pavlova, E.Y., Pitulko, V.V., 2015. Ancient DNA analysis of the oldest canid species from the Siberian Arctic and genetic contribution to the domestic dog. PLoS One 10, e0125759.

Leonard, J.A., Wayne, R.K., Wheeler, J., Valadez, R., Guillén, S., Vilà, C., 2002. Ancient DNA evidence for Old World origin of New World dogs. Science 298, 1613–1616.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50, 913–925.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment / Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. Data Processing 1–2.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.-W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A.,

Hitte, C., Kim, L., Koepfli, K.-P., Parker, H.G., Pollinger, J.P., Searle, S.M.J., Sutter, N.B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A.C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.-P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A., Lander, E.S., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819.

Lindgreen, S., 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res. Notes 5, 337.

Lipe, W.D., Morris, J.N., Kohler, T.A., 1988. Dolores Archaeological Program: Anasazi Communities at Dolores: Grass Mesa Village. USDI Bureau of Reclamation, Engineering and Research Center, Denver.

Lischer, H.E.L., Excoffier, L., 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299.

Lope, C.P., Masson, M.A., Hare, T.S., Pedro, C., Delgado Ku, P.C., 2006. The Late Postclassic chronology of Mayapan: new radiocarbon evidence. Ancient Mesoamerica 17, 153–176.

Manin, A., 2015. Utilisation matérielle et symbolique des animaux dans le nord de La Mésoamérique entre le Classique et la Conquête (200-1521 Apr. J.-C.) (PhD). Muséum National d'Histoire Naturelle.

Masson, M.A., Lope, C.P., 2008. Animal use at the Postclassic Maya center of Mayapan. Quat. Int. 191, 170–183.

McCary, B.C., 1957. Indians in Seventeenth-Century Virginia. University Press of Virginia, Charlottesville.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010, db.prot5448.

Mills, W.C., 1906. Explorations of the Baum prehistoric village site. FJ Heer.

Monaghan, G., Peebles, C., 2010. The construction, use, and abandonment of Angel Site Mound A: tracing the history of a Middle Mississippian town through its earthworks. Am. Antiq. 75, 935–953.

Moore, C.B., 1915. Aboriginal sites on the Tennessee river. Journal of the Academy for Natural Sciences of Philadelphia 16, 169–428.

Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspinas, A.-S., Sikora, M., Reuther, J.D., Irish, J.D., Malhi, R.S., Orlando, L., Song, Y.S., Nielsen, R., Meltzer, D.J., Willerslev, E., 2018. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature 553, 203–207.

Morey, D.F., 2006/2. Burying key evidence: the social bond between dogs and people. J. Archaeol. Sci. 33, 158–175.

Morey, D.F., Wiant, M.D., 1992. Early Holocene Domestic Dog Burials From the North American Midwest. Curr. Anthropol. 33, 224–229.

Murchison, E.P., Wedge, D.C., Alexandrov, L.B., Fu, B., Martincorena, I., Ning, Z., Tubio, J.M.C., Werner, E.I., Allen, J., De Nardi, A.B., Donelan, E.M., Marino, G., Fassati, A., Campbell, P.J., Yang, F., Burt, A., Weiss, R.A., Stratton, M.R., 2014. Transmissible [corrected] dog cancer genome reveals the origin and history of an ancient cell lineage. Science 343, 437–440.

Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., Weiss, R.A., 2006. Clonal origin and evolution of a transmissible cancer. Cell 126, 477–487.

Nolan, K., Sculli, P., 2014. Rejoinder to Sciulli and Purcell: Two Late Prehistoric Dogs from the Reinhardt Site (33PI880), Pickaway County, Ohio. Pa. Archaeol. 84, 65–73.

Nolan, K.C., 2009. Archaeological Survey of the Reinhardt Tract Property through a Certified Local Government (CLG) Grant on behalf of the City of Columbus in Harrison Township, Pickaway County, Ohio, Volume I: Survey Results. Report Submitted to the Ohio Historic Preservation Office, Columbus, OH.

Nolan, K.C., 2011. Distributional Survey of the Reinhardt Site (33PI880), Pickaway County, Ohio: A Strategy for Deciphering the Community Structure of a Fort Ancient Village. MidCont. J. Archaeol. 36, 105–130.

Nolan, K.C., Cook, R.A., 2010/3. An evolutionary model of social change in the Middle Ohio Valley: Was social complexity impossible during the late woodland but mandatory during the late prehistoric? Journal of Anthropological Archaeology 29, 62–79.

Okkens, A.C., Hekerman, T.W., de Vogel, J.W., van Haaften, B., 1993. Influence of litter size and breed on variation in length of gestation in the dog. Vet. Q. 15, 160–161.

Pang, J.-F., Kluetsch, C., Zou, X.-J., Zhang, A.-B., Luo, L.-Y., Angleby, H., Ardalan, A., Ekström, C., Sköllermo, A., Lundeberg, J., Matsumura, S., Leitner, T., Zhang, Y.-P., Savolainen, P., 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.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290.

Parker, A., Eldridge, M., 2014. Archaeology in the Third and Fourth Dimensions: A Case Study of

3D Data Collection and Analysis from Prince Rupert, BC, Canada, in: Giligny, F., Djindjian, F., Costa, L., Moscati, P., Rober, S. (Eds.), Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Archaeopress, Oxford, pp. 114–122.

Parker, H.G., Dreger, D.L., Rimbault, M., Davis, B.W., Mullen, A.B., Carpintero-Ramirez, G., Ostrander, E.A., 2017. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. Cell Rep. 19, 697–708.

Parmalee, P.W., 1959. Appendix II: animals remains from the Modoc Rock Shelter site, Randolph County, Illinois. Illinois State Museum, Springfield, Illinois.

Parmalee, P.W., Paloumpis, A.A., Wilson, N., 1972. Animals Utilized by Woodland Peoples Occupying the Apple Creek Site, Illinois, Research Papers 5. Illinois State Museum, Illinois Valley Archaeological Program, Springfield, I.

Parmalee, P.W., Paloumpis, A.A., Wilson, N., 1972. Peoples occupying the Apple Creek Site, Illinois. Illinois State Museum Reports of Investigations 23.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, e190.

Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L.Z., Beaudoin, A.B., Zutter, C., Larsen, N.K., Potter, B.A., Nielsen, R., Rainville, R.A., Orlando, L., Meltzer, D.J., Kjær, K.H., Willerslev, E., 2016. Postglacial viability and colonization in North America's ice-free corridor. Nature 537, 45–49.

Perri, A., 2013. Global Hunting Adaptations to Early Holocene Temperate Forests: Intentional Dog Burials as Evidence of Hunting (PhD). Durham University.

Perri, A., 2017. A typology of dog deposition in archaeological contexts, in: Rowley-Conwy, P., Serjeantson, D., Halstead, P. (Eds.), Economic Zooarchaeology: Studies in Hunting, Herding and Early Agriculture. Oxbow.

Perri, A., n.d. NEW EVIDENCE OF THE EARLIEST DOMESTIC DOGS IN THE AMERICAS.

Peterson, S., 2010. Townscape Archaeology At Angel Mounds, Indiana: Mississippian Spatiality and Community.

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8, e1002967.

Pitulko, V.V., Kasparov, A., 1996. Ancient Arctic Hunters: Material Culture and Survival Strategy. Arctic Anthropol. 33, 1–36.

Pitulko, V.V., Kasparov, A.K., 2017/6. Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. Journal of Archaeological Science: Reports 13, 491–515.

Potter, M.A., Baby, R.S., 1964. Hopewellian Dogs. Ohio J. Sci. 64, 36–40.

Proskouriakoff, T., 1962. Civic and Religious Structures of Mayapan, in: Mayapan, Yucatan, Mexico, Carnegie Institute of Washington Publication 619. p. 87.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Putnam, F.W., 1896. The prehistoric dog of America, in: Moore, C.B. (Ed.), Additional Mounds of

Duval and Clay Counties, Florida. Journal of the Academy of Natural Sciences of Philadelphia, Philadelphia, pp. 26–27.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., Fitzhugh, W., Malmström, H., Rasmussen, S., Olsen, J., Melchior, L., Fuller, B.T., Fahrni, S.M., Stafford, T., Jr, Grimes, V., Renouf, M.A.P., Cybulski, J., Lynnerup, N., Lahr, M.M., Britton, K., Knecht, R., Arneborg, J., Metspalu, M., Cornejo, O.E., Malaspinas, A.-S., Wang, Y., Rasmussen, M., Raghavan, V., Hansen, T.V.O., Khusnutdinova, E., Pierre, T., Dneprovsky, K., Andreasen, C., Lange, H., Hayes, M.G., Coltrain, J., Spitsyn, V.A., Götherström, A., Orlando, L., Kivisild, T., Villems, R., Crawford, M.H., Nielsen, F.C., Dissing, J., Heinemeier, J., Meldgaard, M., Bustamante, C., O'Rourke, D.H., Jakobsson, M., Gilbert, M.T.P., Nielsen, R., Willerslev, E., 2014. The genetic prehistory of the New World Arctic. Science 345, 1255832.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J.R., Wall, J., Cornejo, O.E., Moreno-Mayar, J.V., Korneliussen, T.S., Pierre, T., Rasmussen, M., Campos, P.F., de Barros Damgaard, P., Allentoft, M.E., Lindo, J., Metspalu, E., Rodríguez-Varela, R., Mansilla, J., Henrickson, C., Seguin-Orlando, A., Malmström, H., Stafford, T., Jr, Shringarpure, S.S., Moreno-Estrada, A., Karmin, M., Tambets, K., Bergström, A., Xue, Y., Warmuth, V., Friend, A.D., Singarayer, J., Valdes, P., Balloux, F., Leboreiro, I., Vera, J.L., Rangel-Villalobos, H., Pettener, D., Luiselli, D., Davis, L.G., Heyer, E., Zollikofer, C.P.E., Ponce de León, M.S., Smith, C.I., Grimes, V., Pike, K.-A., Deal, M., Fuller, B.T., Arriaza, B., Standen, V., Luz, M.F., Ricaut, F., Guidon, N., Osipova, L., Voevoda, M.I., Posukh, O.L., Balanovsky, O., Lavryashina, M., Bogunov, Y., Khusnutdinova, E., Gubina, M., Balanovska, E., Fedorova, S., Litvinov, S., Malyarchuk, B., Derenko, M., Mosher, M.J., Archer, D., Cybulski, J., Petzelt, B., Mitchell, J., Worl, R., Norman, P.J., Parham, P., Kemp, B.M., Kivisild, T., Tyler-Smith, C., Sandhu, M.S., Crawford, M., Villems, R., Smith, D.G., Waters, M.R., Goebel, T., Johnson, J.R., Malhi, R.S., Jakobsson, M., Meltzer, D.J., Manica, A., Durbin, R., Bustamante, C.D., Song, Y.S., Nielsen, R., Willerslev, E., 2015. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349, aab3884.

Rebbeck, C.A., Leroi, A.M., Burt, A., 2011. Mitochondrial capture by a transmissible cancer. Science 331, 303.

Rebbeck, C.A., Thomas, R., Breen, M., Leroi, A.M., Burt, A., 2009. Origins and evolution of a transmissible cancer. Evolution 63, 2340–2349.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., 2009. Reconstructing Indian population history. Nature 461, 489–494.

Renouf, M.A.P., 1993. Palaeoeskimo Seal Hunters at Port au Choix, Northwestern Newfoundland. Newfoundland and Labrador Studies 9.

Rick, T.C., Erlandson, J.M., Vellanoweth, R.L., Braje, T.J., 2005. From pleistocene mariners to complex hunter-gatherers: The archaeology of the california channel islands. Journal of World Prehistory 19, 169–228.

Rick, T.C., Walker, P.L., Willis, L.M., Noah, a. C., Erlandson, J.M., Vellanoweth, R.L., Braje, T.J., Kennett, D.J., 2008. Dogs, humans and island ecosystems: the distribution, antiquity and ecology of domestic dogs (Canis familiaris) on California's Channel Islands, USA. Holocene 18, 1077–1087.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., WGS500 Consortium, Wilkie, A.O.M., McVean, G., Lunter, G., 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. 46, 912–918.

Ritchie, W.A., 1945. An Early Site in Cayug a County, New York. Research Records 7.

Ritchie, W.A., 1965. The Archaeology of New York State. Knopf Doubleday Publishing Group.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rountree, H.C., 1990. Pocahontas's People: The Powhatan Indians of Virginia Through Four Centuries. University of Oklahoma Press.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A.S., Willerslev, E., Krogh, A., Orlando, L., 2012. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics 13, 178.

Scott, G.R., 1992. Affinities of Prehistoric and Modern Kodiak Islanders and the Question of Kachemak-Koniag Biological Continuity. Arctic Anthropol. 29, 150–166.

Shannon, L.M., Boyko, R.H., Castelhano, M., Corey, E., Hayward, J.J., McLean, C., White, M.E., Abi Said, M., Anita, B.A., Bondjengo, N.I., Calero, J., Galov, A., Hedimbi, M., Imam, B., Khalap, R., Lally, D., Masta, A., Oliveira, K.C., Pérez, L., Randall, J., Tam, N.M., Trujillo-Cornejo, F.J., Valeriano, C., Sutter, N.B., Todhunter, R.J., Bustamante, C.D., Boyko, A.R., 2015. Genetic structure in village dogs reveals a Central Asian domestication origin. Proc. Natl. Acad. Sci. U. S. A. 112, 13639–13644.

Simoons, F.J., 1994. Eat Not this Flesh: Food Avoidances from Prehistory to the Present. Univ of Wisconsin Press.

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, 1515–1519.

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M.P., Friedlaender, F.R., Friedlaender, J.S., Kivisild, T., Koki, G., Kusuma, P., Andrew Merriwether, D., Ricaut, F.-X., Wee, J.T.S., Patterson, N., Krause, J., Pinhasi, R., Reich, D., 2016. Genomic insights into the peopling of the Southwest Pacific. Nature 538, 510–513.

Smith, A.L., 1962. Residential and Associated structures at Mayapan, in: Mayapan, Yucatan, Mexico, Carnegie Institute of Washington Publication 619. p. 619.

Smith, J., 1624. The Generall Historie of Virginia, New-England, and the Summer Isles.

Squier, E.G., Davis, E.H., 1848. Ancient Monuments of the Mississippi Valley: Comprising the Results of Extensive Original Surveys and Explorations. Smithsonian Institution.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Steffian, A.F., 1992. Fifty years after Hrdlicka: Further excavation of the Uyak site, Kodiak Island, Alaska. Contributions to the anthropology of southcentral and southwestern Alaska 141–164.

Strachey, W., 1612. The History of Travel into Virginia Britannia. Hakluyt Society, London.

Strakova, A., Ní Leathlobhair, M., Wang, G.-D., Yin, T.-T., Airikkala-Otter, I., Allen, J.L., Allum, K.M., Bansse-Issa, L., Bisson, J.L., Castillo Domracheva, A., de Castro, K.F., Corrigan, A.M.,

Cran, H.R., Crawford, J.T., Cutter, S.M., Delgadillo Keenan, L., Donelan, E.M., Faramade, I.A., Flores Reynoso, E., Fotopoulou, E., Fruean, S.N., Gallardo-Arrieta, F., Glebova, O., Häfelin Manrique, R.F., Henriques, J.J., Ignatenko, N., Koenig, D., Lanza-Perea, M., Lobetti, R., Lopez Quintana, A.M., Losfelt, T., Marino, G., Martincorena, I., Martínez Castañeda, S., Martínez-López, M.F., Meyer, M., Nakanwagi, B., De Nardi, A.B., Neunzig, W., Nixon, S.J., Onsare, M.M., Ortega-Pacheco, A., Peleteiro, M.C., Pye, R.J., Reece, J.F., Rojas Gutierrez, J., Sadia, H., Schmeling, S.K., Shamanova, O., Ssuna, R.K., Steenland-Smit, A.E., Svitich, A., Thoya Ngoka, I., Viţălaru, B.A., de Vos, A.P., de Vos, J.P., Walkinton, O., Wedge, D.C., Wehrle-Martinez, A.S., van der Wel, M.G., Widdowson, S.A., Murchison, E.P., 2016. Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. Elife 5.

Strong, J.A., 1985. Late Woodland dog ceremonialism on Long Island in comparative and temporal perspective. The Bulletin of the Journal of the New York State Archaeological Association 91, 32–38.

Struever, S., Holton, F.A., 1979. Koster: Americans in search of their prehistoric past.

Tacon, P., Pardoe, C., 2002. Dogs make us human.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpré, M.B., Sablin, M.V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H.-P., Loponte, D.M., Acosta, A. a., Giemsch, L., Schmitz, R.W., Worthington, B., Buikstra, J.E., Druzhkova, A., Graphodatsky, a. S., Ovodov, N.D., Wahlberg, N., Freedman, a. H., Schweizer, R.M., Koepfli, K.-P., Leonard, J. a., Meyer, M., Krause, J., Pääbo, S., Green, R.E., Wayne, R.K., 2013. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science 342, 871–874.

Tuck, J.A., 1970. Archaic Indian Cemetery in Newfoundland. Sci. Am. 222, 112–121.

Tuck, J.A., 1971. An Archaic Cemetery at Port Au Choix, Newfoundland. Am. Antiq. 36, 343–358.

Tuck, J.A., 1976. Ancient people of Port au Choix.

van Asch, B., Zhang, A.-B., Oskarsson, M.C.R., Klütsch, C.F.C., Amorim, A., Savolainen, P., 2013. Pre-Columbian origins of Native American dog breeds, with only limited replacement by European dogs, confirmed by mtDNA analysis. Proc. Biol. Sci. 280, 20131142.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E.H., Hansen, M.S.T., Lawley, C.T., Karlsson, E.K., LUPA Consortium, Bannasch, D., Vilà, C., Lohi, H., Galibert, F., Fredholm, M., Häggström, J., Hedhammar, A., André, C., Lindblad-Toh, K., Hitte, C., Webster, M.T., 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genet. 7, e1002316.

Vonholdt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J.C., Mosher, D.S., Spady, T.C., Elkahloun, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C., Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P.G., Qian, Z., Huang, W., Ding, Z.-L., Zhang, Y.-P., Bustamante, C.D., Ostrander, E.A., Novembre, J., Wayne, R.K., 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature 464, 898–902.

Wallis, W.D., 1955. The Micmac Indians of Eastern Canada. U of Minnesota Press.

Wang, G.-D., Zhai, W., Yang, H.-C., Wang, L., Zhong, L., Liu, Y.-H., Fan, R.-X., Yin, T.-T., Zhu, C.-L., Poyarkov, A.D., Irwin, D.M., Hytonen, M.K., Lohi, H., Wu, C.-I., Savolainen, P., Zhang, Y.-P., 2016. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res.

Warren, D.M., 2004. Skeletal biology and paleopathology of domestic dogs from prehistoric .

Waszak, S.M., Tiao, G., Zhu, B., Rausch, T., Muyas, F., Rodriguez-Martin, B., Rabionet, R., Yakneen, S., Escaramis, G., Li, Y., Saini, N., Roberts, S.A., Demidov, G.M., Pitkanen, E., Delaneau, O., Heredia-Genestar, J.M., Weischenfeldt, J., Shringarpure, S.S., Chen, J., Nakagawa, H., Alexandrov, L.B., Drechsel, O., Jonathan Dursi, L., Segre, A.V., Garrison, E., Erkek, S., Habermann, N., Urban, L., Khurana, E., Cafferkey, A., Hayashi, S., Imoto, S., Aaltonen, L.A., Alvarez, E.G., Baez-Ortega, A., Bailey, M., Bosio, M., Bruzos, A.L., Buchhalter, I., Bustamante, C.D., Calabrese, C., DiBiase, A., Gerstein, M., Holik, A.Z., Hua, X., Huang, K.-L., Letunic, I., Klimczak, L.J., Koster, R., Kumar, S., McLellan, M., Mashl, J., Mirabello, L., Newhouse, S., Prasad, A., Raetsch, G., Schlesner, M., Schwarz, R., Sharma, P., Shmaya, T., Sidiropoulos, N., Song, L., Susak, H., Tanskanen, T., Tojo, M., Wedge, D.C., Wright, M., Wu, Y., Ye, K., Yellapantula, V.D., Zamora, J., Butte, A.J., Getz, G., Simpson, J., Ding, L., Marques-Bonet, T., Navarro, A., Brazma, A., Campbell, P., Chanock, S.J., Chatterjee, N., Stegle, O., Siebert, R., Ossowski, S., Harismendy, O., Gordenin, D.A., Tubio, J.M.C., De La Vega, F.M., Easton, D.F., Estivill, X., Korbel, J., PCAWG Germline Working Group, ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Net, 2017. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. bioRxiv.

Webb, K.M., Allard, M.W., 2009. Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region. J. Forensic Sci. 54, 275–288.

Webb, W.S., 1946. Indian Knoll, Site Oh 2, Ohio County, Kentucky: The University of Kentucky Reports in Anthropology and Archaeology. Lexington: Department of Anthropology 111–365.

Webb, W.S., DeJarnette, D.L., 1948. Little Bear Creek Site, Museum Paper 26.

Webb, W.S., DeJarnette, D.L., 1948a. The Flint River Site, Museum Paper 23. Geological Survey of Alabama, Tuscaloosa, Alabama.

Webb, W.S., DeJarnette, D.L., 1948b. The Perry Site, 1Lu25. Geological Survey of Alabama, Alabama Natural History Museum Paper 25.

West, C.F., France, C.A., 2015. Human and Canid Dietary Relationships: Comparative Stable Isotope Analysis From the Kodiak Archipelago, Alaska. J. Ethnobiol. 35, 519–535.

West, C.F., Jarvis, K.N., 2015. Osteometric Variation in Domestic Dogs (Canis familiaris) from the Kodiak Archipelago, Alaska. Int. J. Osteoarchaeol. 25, 289–298.

Wiant, M.D., Farnsworth, K., Hajic, E.R., Emerson, T.E., McElrath, D.L., Fortier, A.C., 2009. The archaic period in the lower Illinois river basin. Archaic Societies: Diversity and Complexity across the Midcontinent 229–286.

Witt, K.E., Judd, K., Kitchen, A., Grier, C., Kohler, T. a., Ortman, S.G., Kemp, B.M., Malhi, R.S., 2015. DNA analysis of ancient dogs of the Americas: Identifying possible founding haplotypes and reconstructing population histories. J. Hum. Evol. 79, 105–118.

Witt, K.E., Judd, K., Kitchen, A., Grier, C., Kohler, T.A., Ortman, S.G., Kemp, B.M., Malhi, R.S., 2015. DNA analysis of ancient dogs of the Americas: identifying possible founding haplotypes and reconstructing population histories. J. Hum. Evol. 79, 105–118.

Zhao, J.-H., Liu, W., 2016. The complete mitochondrial genome of the Simao Chinese indigenous dog. Mitochondrial DNA A DNA Mapp Seq Anal 27, 545–546.

3.6 Supplementary Materials

The following is an extract from the Supplementary Materials, covering the areas of analysis which I led. The complete 97-page supplement can be viewed online. http://science.sciencemag.org/content/suppl/2018/07/03/361.6397.81.DC1

3.6.1 Nuclear ancestry analyses

3.6.1.1 PCA

Using SmartPCA (Patterson et al., 2006) we performed Principal Components Analysis (PCA) using various projections and data sets on our 2.03M SNPs:

- All canids (including wolves and coyotes) - PCD samples projected (Figure S3.7)
- Only dogs (excluding wolves and coyotes) - PCD samples projected (Figure S3.8)
- Only dogs (excluding wolves and coyotes) - PCD and CTVT samples projected (Figure S3.9)

For PCD we used all 7 samples for which we could call at least 10,000 sites (minimum number of sites suggested for ancient DNA analysis (Allentoft et al., 2015)). We used all available sites (sites covered in at least 1 ancient sample; ~1.5M SNPs) to compute the eigenvectors and then projected PCD onto that space. We also projected CTVT to ensure that their placement was not an artefact of somatic mutations.

Figure S3.7 shows that PCD are more closely related to dogs (except for one sample, AL2135 from Koster; see below) than wolves or coyotes. It also shows that dogs are less variable than wolves or coyotes. Figure S3.8 shows a distinction between Arctic, East Asian, and European dogs. Lastly, Figure S3.9 recapitulates the same results demonstrating that this result is not induced by somatic mutations in CTVT and also shows how PCD and CTVT are more closely related to each other, and fall in between Arctic dogs and all other dogs. A PCD sample (AL2135; Koster, Illinois) was projected in between dogs and wild canids (Figure S3.7). This suggests that this sample is admixed with wild canids (see D-statistics analyses below). Its mtDNA haplotype, however, clusters with other PCD dogs (Figure S3.3).

3.6.1.2 Neighbour joining tree

We used plink v1.9 (Purcell et al., 2007) to compute an Identity By State (IBS) matrix using all 2.03M SNPs. This matrix was used to build a neighbour joining tree (NJ) using the R package "ape" (Paradis et al., 2004); (Figure S3.10). The tree recapitulates the deep split between East Asian and Western Eurasian dogs (Frantz et al., 2016a) and confirms that CTVT is more closely related to the PCD dogs than to any other dog population (bootstrap=100). It also shows that CTVT/PCD form a monophyletic group with Arctic breeds that fall outside of the rest of the dogs. The tree also confirms that PCD form a monophyletic clade with high support (bootstrap = 100).

Admixture analyses (see D-statistics below) show that all East Asian dogs, excluding Vietnamese Village dogs, are significantly admixed with European dog populations. Such disproportionate admixture could affect the topology of the tree. To test this, we built a tree excluding all East Asian dogs except Vietnamese. This tree shows a different topology with Vietnamese still grouping with Dingoes, however, PCD, Arctic dogs and CTVT are now more closely related to Western dogs than to Asian Dogs (Figure S3.11).

3.6.1.3 Bayesian Tree

We built a phylogeny using nuclear genotypes with MrBayes 3.2 (Ronquist and Huelsenbeck, 2003). To do so we used PGDSpider 2.0.9.2 (Lischer and Excoffier, 2012) to build a Nexus file with discrete SNP format (0=reference, 1=heterozygous, 2=homozygous alternative). We used the Mkv model (Lewis, 2001) implemented in MrBayes (Ordered character), which provides a likelihood framework for data sets that contain only variable characters. We also imposed a minimum distance of 10Kb between SNPs to limit the influence of linkage disequilibrium (LD) and lastly included only PCD samples with higher coverage (AL3194 and AL3223; Table S3.1; ~26K SNP total).

We ran two independent runs of four MCMC chains with two million samples. Trees were summarized discarding 25% as burn-in. To limit biases from missing data we

limited this analysis to transversions that were covered in 90% of our samples. Convergence was assessed by ensuring that average standard deviation of split frequencies was below 0.01 and that the potential scale reduction factor was close to 1 for all parameters. This analysis confirms that CTVT and PCD are monophyletic with high support (Posterior probability [PP] = 1; Figure S3.12) and the basal placement of the CTVT/PCD clade. However, this analysis suggests that modern Arctic dogs are more closely related to Eurasian dogs than to PCD. This is most likely due to the complex ancestry of Arctic dogs such as admixture from European dogs (see below; Table S3.4).

3.6.1.4 f3 statistics

We computed outgroup *f3*-statistics as f3(pre-contact dogs [PCD], X; outgroup) using ADMIXTOOLS (Patterson et al., 2012) where X is any other dog population (see Table S3.2), to quantify the amount of genetic drift shared between pre-contact dogs and other dogs using only transversions (Figure S3.13). For this analysis, we used only two PCD samples (AL3194 and AL3223; Table S3.1), with ~1.9x and ~0.5x coverage, respectively. Our results support our NJ tree (Figure S3.10) demonstrating that PCD is more closely related to CTVT and Arctic dogs than any other dog population. These results also support the observation that PCD/CTVT and Arctic breeds are equally related to all other dogs, except for Basenji and one Indian dog, which could be due to admixture from wolves into these two samples (e.g. Indian wolf or golden wolf).

3.6.1.5 D-statistics

We only used two PCD samples with ~1.9x and ~0.5x coverage (AL3194 and AL3223; Table S3.1) for these analyses, except when explicitly mentioned (e.g. Koster dog AL2135; see below). We used all 2.03M SNPs.

3.6.1.5.1 PCD is more closely related to CTVT and Arctic breeds

We computed D(Outgroup, PCD, 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 any possible other sample. We plotted, as box plots, the results of

these combinations (Figure S3.14; Figure S3.15). Positive values imply that PCD shares more derived alleles with the population on the y axis, while negative values imply that pre-contact dogs are closer to the other dog populations. The results indicate that PCD do not share any more derived alleles with European dogs than they do with Asian dogs, suggesting that they are equally related to both. This result supports our f3-statistics and NJ tree finding that PCD and CTVT are equally related to European and Asian dogs (Figure S3.10; Figure S3.13). Arctic breeds also appear more closely related to PCD/CTVT than any other dog population (Figure S3.14; Figure S3.15; Figure S3.13).

3.6.1.5.2 Admixture between Coyote / North American wolves and PCD

We tested for admixture from wild North American canids into higher coverage PCD genomes (AL3194; AL3223). To do so we computed D(Outgroup, Coyote/North American Wolf, Pop3, Pop4) where Pop3/4 can be any dog genome. We found that in both cases Z values were mostly above 3 in most cases (Figure S3.16; Figure S3.17) for both AL3194 and AL3223 indicating admixture from Coyotes / North American Wolves in PCD samples. We also tested for extra admixture from wild canids into our higher coverage PCD genomes (AL3194; AL3223). To do so we computed D(Outgroup, Coyote, AL3194, AL3223) and D(Outgroup, American Wolf, AL3194, AL3223). We found no evidence of extra admixture from wild canids into these samples (Table S3.3).

We computed D(Outgroup, Coyote, CTVT, B) and D(Outgroup, American Wolf, CTVT, B) where B represented every possible pair of populations to determine whether there was any detectable admixture from Coyote and American wolf populations into the CTVT founder dog (Figure S3.16; Figure S3.17). We also found evidence that the CTVT founder dog shared more derived alleles with Coyotes than other non-PCD population suggestive of admixture. This is consistent with TreeMix and qpGraph analyses (see below). We note, however, that this pattern could be consistent with admixture in both direction (dogs to wolves / wolves to dogs; e.g. see (Anderson et al., 2009)).

3.6.1.5.3 Estimating Eurasian ancestry in Arctic dogs

Using D-statistics based on whole genome data we found evidence that all Arctic breeds are a mixture of the basal lineage (that leads to CTVT and PCD) and of the Eurasian dog lineage (Table S3.4).

3.6.1.5.4 Taimyr admixture into Arctic dogs, PCD and CTVT

We used whole genome data to assess Taimyr wolf admixture into PCD, Arctic dogs and CTVT (Skoglund et al., 2015). We found few values with $|Z| > 3$ (AL3194 and Alaskan malamute; Table S3.5). Lowering the threshold to $|Z|>2.5$, we found admixture from the Taimyr wolf into PCD (both AL3194 and AL3223) as well as in all Arctic dogs (husky, Greenland sledge dog, and Alaskan malamute) and CTVT. We find no evidence for additional admixture from the Taimyr wolf into either CTVT, PCD or Arctic breed (Table S3.6). This suggests that Taimyr admixture into Arctic dogs suggested in (Skoglund et al., 2015) may have taken place after the PCD, Arctic dog and CTVT lineage diverged from Eurasian dogs but before the divergence of the Arctic dog and PCD/CTVT lineages.

3.6.1.5.5 Admixture from European dogs into East Asian dogs

We tested for admixture from European dogs into east Asian dogs. Following (Frantz et al., 2016a; Shannon et al., 2015) we used Vietnamese village dogs as the reference East Asian population to test for by computing D(Outgroup, Portugal, Vietnam, X). We found evidence of admixture in all East Asian populations test in this study (Table S3.7).

3.6.1.5.6 Potential evidence for Coyote admixture in Koster dog (AL2135)

Our PCA analysis suggests that AL2135 is admixed with wild canids. To test this hypothesis, we computed D(Outgroup, North American canid / Taimyr wolf, AL2135, AL3194). We restricted this analysis to AL3194 as it is the highest coverage PCD dog available in this study. We found borderline significant results ($|Z|>2$; Table S3.8) suggestive of admixture from Coyote into AL2135. This sample, however, is very low coverage (only ~17K SNPs were called). Its placement on the PCA and this positive admixture signal might therefore be due to this low coverage.

3.6.1.6 Estimating pre-contact ancestry in modern North American and Arctic dogs

We used the SNP array data obtained from (Shannon et al., 2015) to assess the degree to which modern dog populations found in North America retained ancestry from pre-contact dogs. This SNP panel contained 28 genotyped populations from North America, such as Peruvian village dogs, Alaskan village dogs or Carolina dogs (see Table S3.9 for the full list). We computed f4 ratios using ADMIXTOOLS (Patterson et al., 2012; Reich et al., 2009) to estimate admixture proportion (α) from pre-contact dogs into these populations by computing:

$$\alpha = f4(A, O; X, C) \div f4(A, O; B, C)$$

Where A is CTVT, O is the Andean fox (outgroup), B is PCD (AL3194 and AL3223), C is any European or East Asian population (see Table S3.9) and X is any American dog (see Table S3.9 and Figure S3.18). We computed α for all combinations of European/Asian and modern North American populations (jack-knifing was performed with a block sizes of 1 cM).

Besides the Alaskan Village dogs, we found no significant signal of pre-contact ancestry in modern North American populations (α always <4% and Z always < 3; Table S3.10). Alaskan Village dogs, on the other hand, have ~17% (11-20% and Z always > 4.5; Table S3.10) ancestry derived from pre-contact dogs. We also used outgroup f3 statistics to assess the degree of shared drift between various populations available on the SNP array and PCD (Figure 2b).

To further assess this result we used ADMIXTURE (Alexander et al., 2009) on a subset of the SNP array samples including all modern North American populations as well as Arctic dogs, "basal" breeds (Larson et al., 2012) and selected European and Asian populations (e.g. Boxer and Chow-Chow). K=4 was selected as the best K value based on 10 fold cross validation (Figure S3.19). This analysis support previous f4 ratio analysis showing that most modern North American dog populations have little pre-contact ancestry (<4%;

Figure S3.20). ADMIXTURE, however, detects some evidence of limited PCD/Arctic ancestry in Carolina dogs ranging from 0-33% (Figure S3.20; population CD). Such signal might not have been detected by our F4 analysis as a result of the variable amount of ancestry in this population. This analysis also reveals an affinity between Chinook and PCD/Arctic breeds (12-15%; Figure S3.20). This is not surprising given that Chinook dogs are considered as Sledge dogs. With K=4, however, we cannot distinguish between PCD and Arctic dogs' ancestry. This PCD/Arctic component in Carolina dogs and Chinook might therefore be the result of admixture with Arctic dogs rather than PCD. To test this, we tried to separate PCD/Arctic ancestry with higher K values. Both K=10 and K=15, however, failed to differentiate PCD and Arctic dog ancestry (Figure S3.20) but instead differentiated New World Arctic dogs (Alaskan malamute and Greenland sledge dogs) from Old World Arctic dogs/PCD (Figure S3.20).

Alaskan Village dogs were the population of north American village dog with the most PCD admixture. This is not surprising as these are closely related to Arctic breeds (Brown et al., 2015; Shannon et al., 2015). The f4 ratio conducted above is thus not appropriate for this population (as it assumes close relatedness to Eurasian dogs; see Figure S3.18). Here we wanted to test whether these dogs have any pre-contact ancestry (interbred with pre-contact dogs). To do so we computed every possible combination of the same f4 ratio as above but using only Arctic breeds. We found that both Alaskan malamute and Greenland sledge dog have a significant amount of ancestry from PCD (~4-14%; Table S3.11). This however, might be the result of substructure among Arctic dogs (see below).

We assessed whether these results could be affected by the ascertainment of the SNP array by repeating the analysis above (PCD admixture fraction into Alaskan malamute and Greenland sledge dogs) using whole genome data. We used only transversions for this analysis (~600K SNPs). We found very little difference in admixture fraction (~7-14%; Table S3.11) indicating that the ascertainment of the array did not introduce much bias.

We also used D-statistics on genome-wide data to test for admixture from PCD into Arctic breeds since their MRCA. As for the f4 ratio we found that both Alaskan malamute

and Greenland sledge dogs have a significant amount of ancestry from PCD (Table S3.12). We tested whether this signal could be due to admixture from Eurasian dogs into Siberian husky dogs (making derived alleles in Alaskan malamute and Greenland sledge dogs (GSD) match PCD more often; Table S3.13). We found evidence that the Siberian husky and Alaskan malamute genomes that we analysed here received gene-flow from European dogs (Table S3.13). However, we found no evidence that GSD received gene-flow since the MRCA of Arctic breeds. This suggests that Eurasian admixture did not affect our result. As stated above, this could also be the result of ancient substructure within Arctic dogs.

We found, however, no signal that either Alaskan malamute or Greenland sledge dogs shared an excess of derived alleles with PCD compared with each other (D(Outgroup, AL3194, Alaskan malamute, Greenland sledge dog)=-0.0001, sd=-0.010). This shows that these dog lineages did not receive additional gene-flow from PCD since their divergence from each other. This result suggest that the signal detected above (excess shared derived alleles between American Arctic dogs and PCD) is due to ancient substructure within Arctic dogs (Eriksson and Manica, 2012). More precisely, we hypothesise that the Eurasian Arctic dogs that were recently brought into the Americas, all the way to Greenland, originated from a population that was more closely related to PCD dogs than other Arctic dogs. The high degree of mtDNA divergence within ancient Eurasian Arctic dogs from Zhokhov (~9,000 BP; Figure 1b) suggests that ancient substructure with Arctic dogs is a plausible scenario.

3.6.1.7 TreeMix

In order to test the topology suggested by our phylogenetics and f3 statistics analyses we used TreeMix (Pickrell and Pritchard, 2012) to build a tree with admixture edges. We only used 3 representatives from each major dog group:

- *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS)
- *East Asian dogs* - Vietnamese village dogs (DVN) because they lack admixture from European dogs (see above) and Tibetan village dogs (DTI)

- *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223; Table S3.1)
- *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL) because they seem least admixed with Western dogs (see above) and
- *CTVT* - (79T and 24T)
- *Eurasian wolves* (WEU) from Spain and Portugal
- *North American wolves* (WAM) from Yellowstone
- *Coyotes* (COY) as an outgroup

We only used transversions in order to limit the effect of DNA damage on the analysis and only used sites that were covered in all samples (~60,000 SNPs).

The results of these analyses are presented in Figure S3.21, Figure S3.22, Figure S3.23 and Figure S3.24. The placement of the East Asian dog (DVN) population was affected by adding admixture edges (Figure S3.22, Figure S3.23). With two admixture edges DVN outgroup PCD/CTVT and Arctic dogs but has strong admixture into DTI. This support results from our D-statistics analysis (Table S3.7) and NJ analysis (Figure S3.10; Figure S3.11) that suggests that DTI is mixed with European ancestry. We also found evidence for European ancestry in Arctic dogs, supporting our D-statistics analyses (Table S3.4). Lastly, we also found admixture from COY into PCD/CTVT, consistent with D-statistics (Figure S3.16).

3.6.1.8 qpGraph

We used qpGraph (Patterson et al., 2012) to fit admixture graphs to nine populations representing PCD, CTVT, and each of the three major dog groups, plus wolves and coyotes.

Western Eurasian dogs - Portuguese village dogs (DEU)
East Asian dogs - Vietnamese village dogs (DVN)
Pre-contact dogs (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223; Table S3.1)

Canine transmissible venereal tumour (CTVT), including C_24T, C_79T and C_399T

Arctic dogs - Alaskan malamute (DMA)

Eurasian wolves (WEU) from Spain and Portugal

North American wolves (WAM) from Yellowstone

Coyotes (COY) from California

Andean Fox (OUT) as the outgroup

We only used transversions in order to limit the effect of DNA damage on the analysis. This resulted in 600,991 high quality SNPs.

To explore the space of all possible admixture graphs we implemented a heuristic search algorithm. Given an outgroup with which to root the graph, a stepwise addition order algorithm was used for adding leaf nodes to the graph. At each step, insertion of a new node was tested at all branches of the graph, except the outgroup branch. Where a node could not be inserted without producing f4 outliers (i.e. |Z| >=3) then all possible admixture combinations were also attempted. If a node could not be inserted via either approach, that sub-graph was discarded. If the node was successfully inserted, the remaining nodes were recursively inserted into that graph. All possible starting node orders were attempted to ensure full coverage of the graph space.

As the number of possible graphs grows super-exponentially with each additional leaf node, we initially excluded CTVT from the search space and looked for models with fit the remaining eight populations. We fitted 480,166 unique admixture graphs for these 8 populations and recorded the 892 graphs that left no f4 outliers (i.e. |Z| < 3). We then fitted a further 309,525 unique models, testing all possible insertions of CTVT into the 892 eight-population graphs, and recorded the 1,655 graphs that left no f4 outliers.

TreeMix analysis was also performed using the same nine populations, with six admixture edges (the maximum number seen in the qpGraph analyses). We chose the most plausible qpGraph model (Figure S3.25) by comparing all fitted models to the TreeMix tree with the same sampling (Figure S3.26), Neighbour joining tree (Figure S3.10), Bayesian tree (Figure S3.12) and D-statistics analyses (see above).

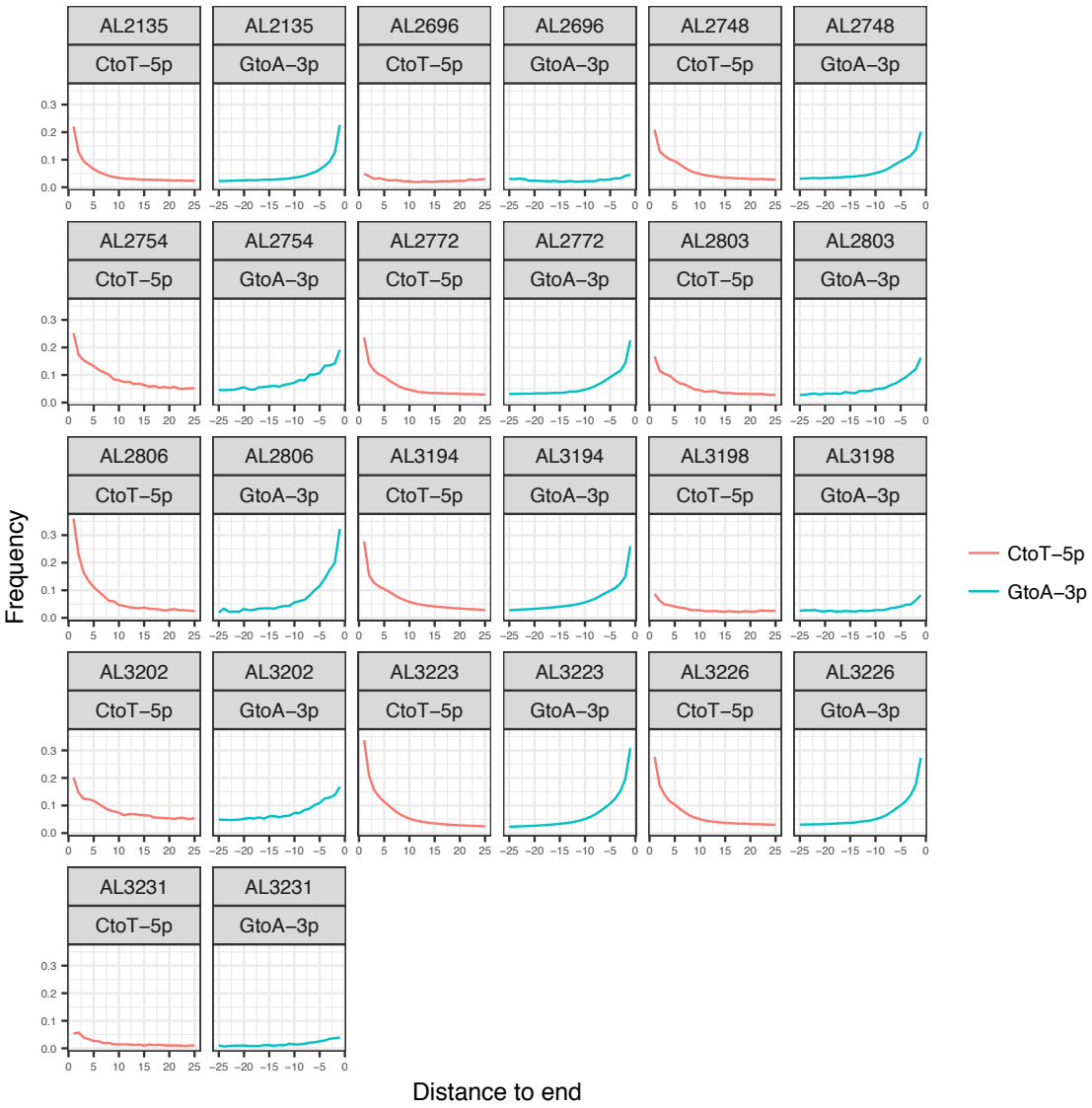


Figure S3.1 Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in nuclear genome analyses.

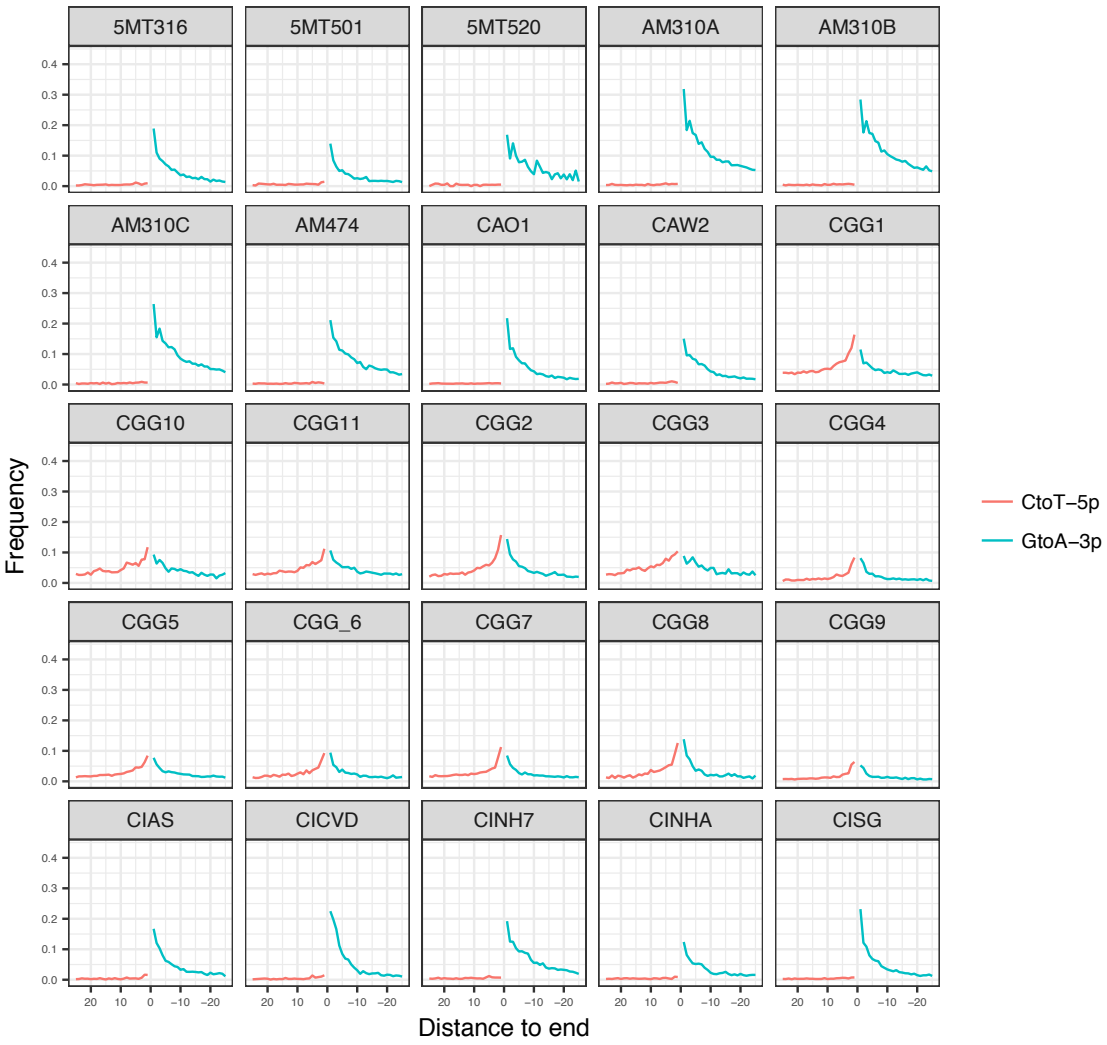


Figure S3.2 Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in mtDNA analyses. Lack of 5' damage in some libraries is due to library preparation protocol (see Ancient DNA Illinois section).

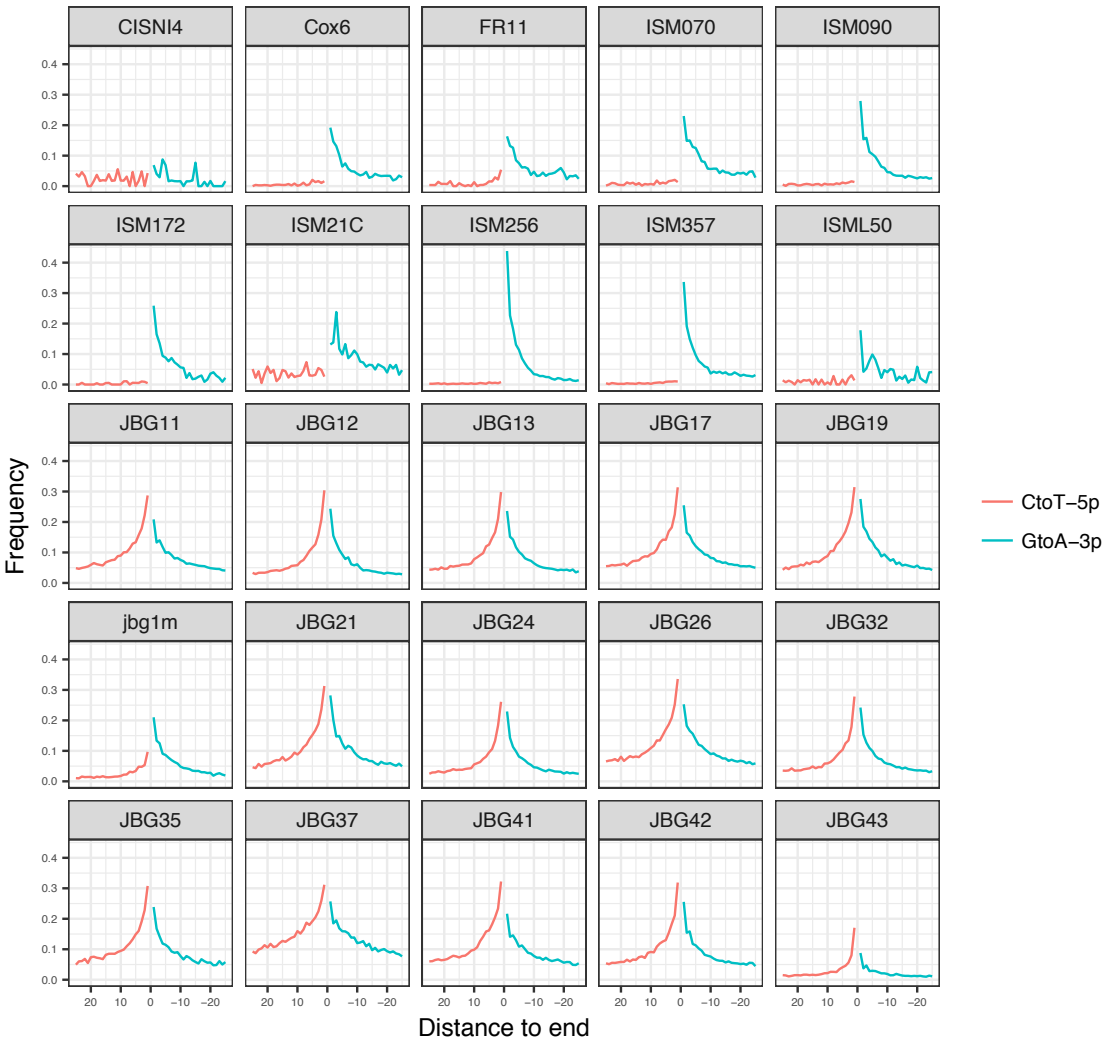


Figure S3.2 (continued) Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in mtDNA analyses. Lack of 5' damage in some libraries is due to library preparation protocol (see Ancient DNA Illinois section).

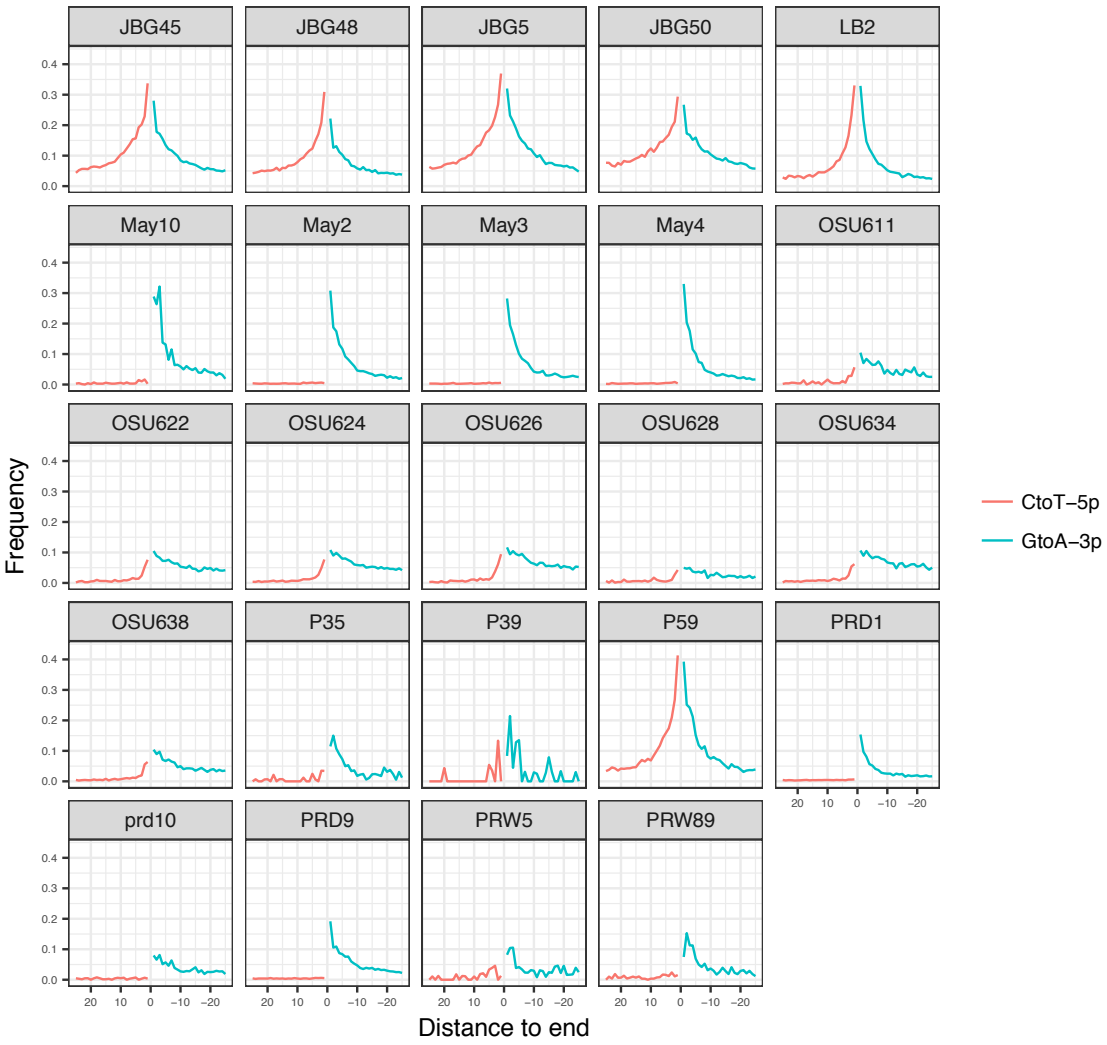


Figure S3.2 (continued) Per library C to T (red) and G to A (blue) frequency of mis-incorporation at 3' and 5' end of read for samples used in mtDNA analyses. Lack of 5' damage in some libraries is due to library preparation protocol (see Ancient DNA Illinois section).

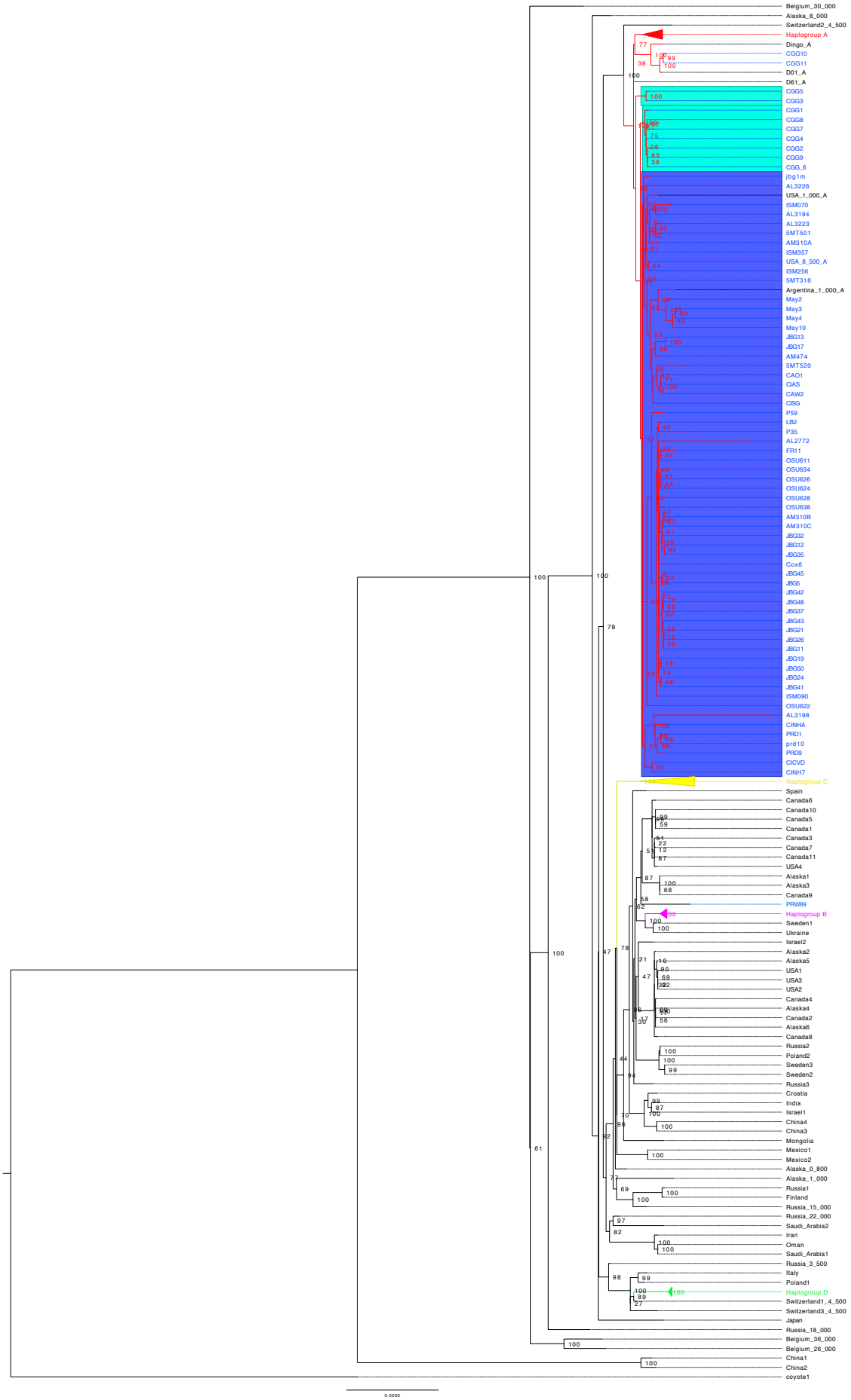

Belgium_30_000
Alaska_8_000
Switzerland2_4_500
Haplogroup A
Dingo_A
CGG10
CGG11
D01_A
D61_A
CGG5
CGG3
CGG1
CGG8
CGG7
CGG4
CGG2
CGG9
CGG_6
jbg1m
AL3226
USA_1_000_A
ISM070
AL3194
AL3223
5MT501
AM310A
ISM357
USA_8_500_A
ISM256
5MT316
Argentina_1_000_A
May2
May3
May4
May10
JBG13
JBG17
AM474
5MT520
CAO1
CIAS
CAW2
CISG
P59
LB2
P35
AL2772
FR11
OSU611
OSU634
OSU626
OSU624
OSU628
OSU638
AM310B
AM310C
JBG32
JBG12
JBG35
Cox6
JBG45
JBG5
JBG42
JBG48
JBG37
JBG43
JBG21
JBG26
JBG11
JBG19
JBG50
JBG24
JBG41
ISM090
OSU622
AL3198
CINHA
PRD1
prd10
PRD9
CICVD
CINH7
Haplogroup C
Spain
Canada6
Canada10
Canada5
Canada1
Canada3
Canada7
Canada11
USA4
Alaska1
Alaska3
Canada9
PRW89
Haplogroup B
Sweden1
Ukraine
Israel2
Alaska2
Alaska5
USA1
USA3
USA2
Canada4
Alaska4
Canada2
Alaska6
Canada8
Russia2
Poland2
Sweden3
Sweden2
Russia3
Croatia
India
Israel1
China4
China3
Mongolia
Mexico1
Mexico2
Alaska_0_800
Alaska_1_000
Russia1
Finland
Russia_15_000
Russia_22_000
Saudi_Arabia2
Iran
Oman
Saudi_Arabia1
Russia_3_500
Italy
Poland1
Haplogroup D
Switzerland1_4_500
Switzerland3_4_500
Japan
Russia_18_000
Belgium_36_000
Belgium_26_000
China1
China2
coyote1
0.0050

Figure S3.3 Maximum likelihood tree based on mtDNA data. The four major dog haplogroups are indicated: A (red; includes all but one pre-contact dogs), B (purple), C (yellow), D (green). Blue tip label represents newly sequenced samples (this study). Dark blue highlighted clade represents American dogs (monophyletic, bootstrap support value=87). Light blue highlighted clades (CGG1-10) represent Zhokhov Island samples (~9Kya sled dogs from Eastern Siberia; see Table S3.1). CGG10-11 (outside of the Zhokhov / pre-contact clade) are more recent sled dogs from Siberia (~1.5kya; Table S3.1). Node labels indicate bootstrap replicates.

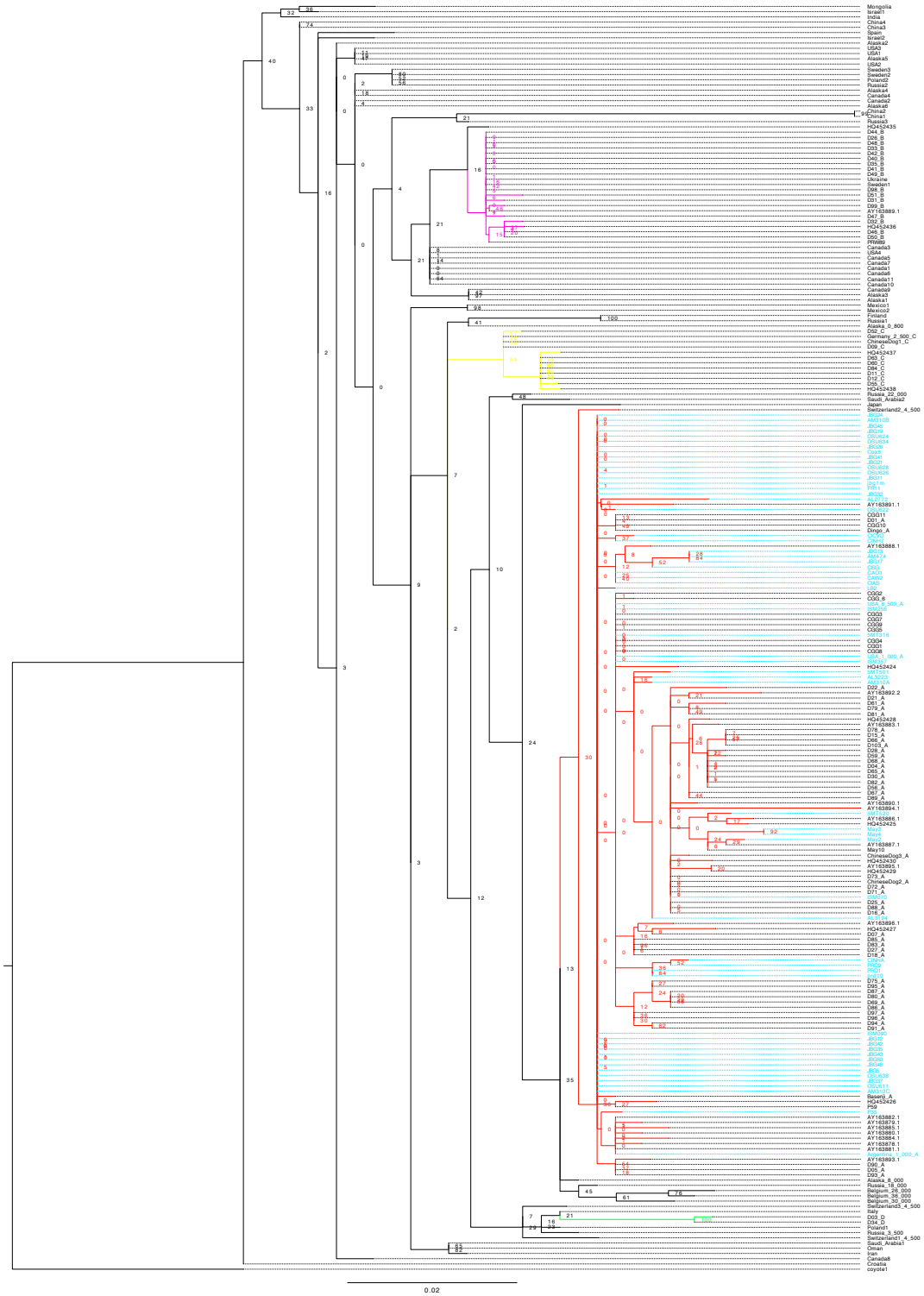


Figure S3.4 Maximum likelihood tree based on 605bp of the control region. All samples starting with prefix HQ were obtained from (*7*) while all samples starting with prefix AY were obtained from (Leonard et al., 2002). The four major dog haplogroups are indicated with different branch colours: A (red; includes all but one pre-contact dogs), B (purple), C (yellow), D (green). All pre-Columbian dogs from this study are highlighted in light blue. Node labels indicate bootstrap replicates.

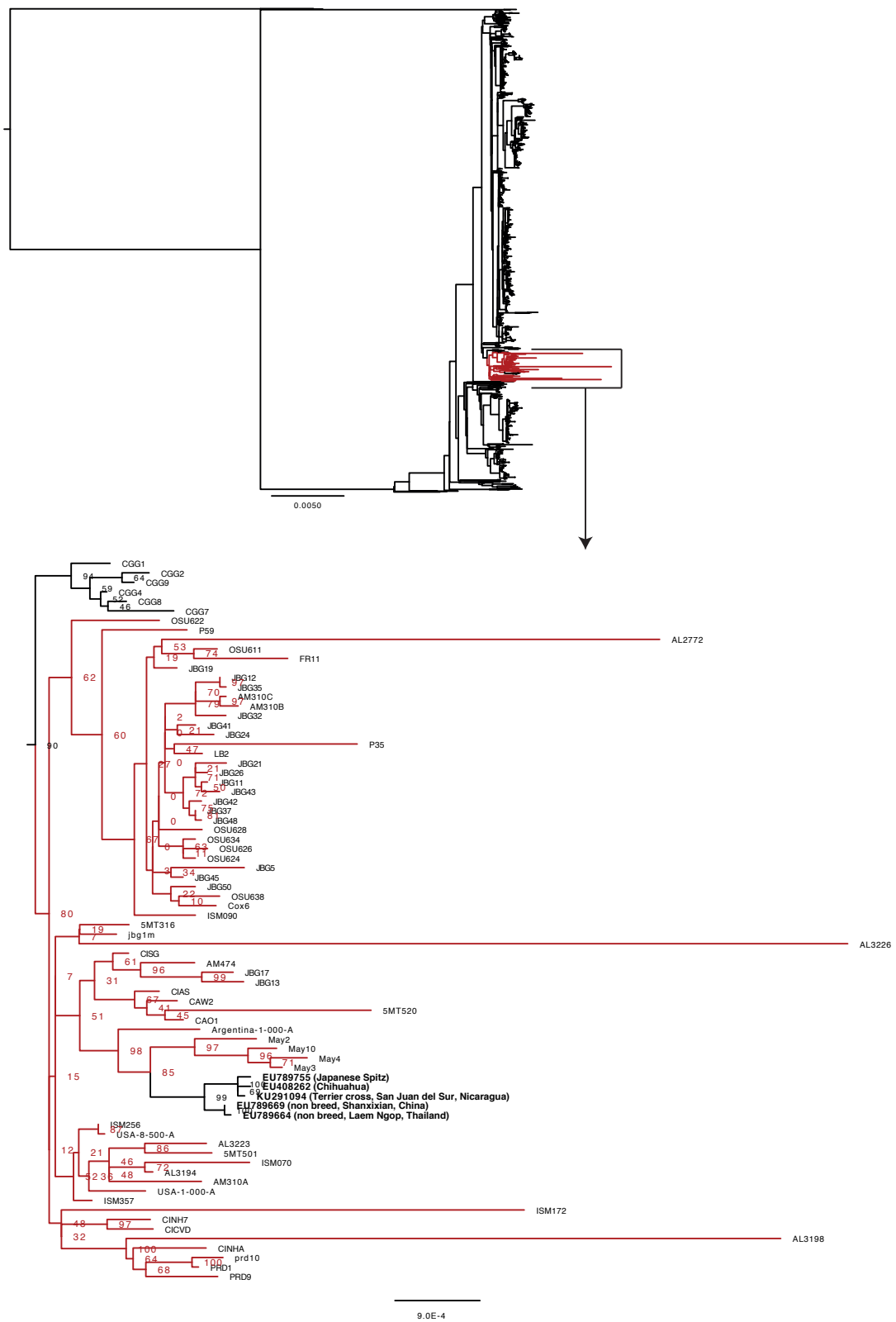


Figure S3.5 Maximum likelihood tree based on mtDNA data including data from (Strakova et al., 2016). Red branches represent ancient pre-contact dogs. Node labels indicate bootstrap replicates.

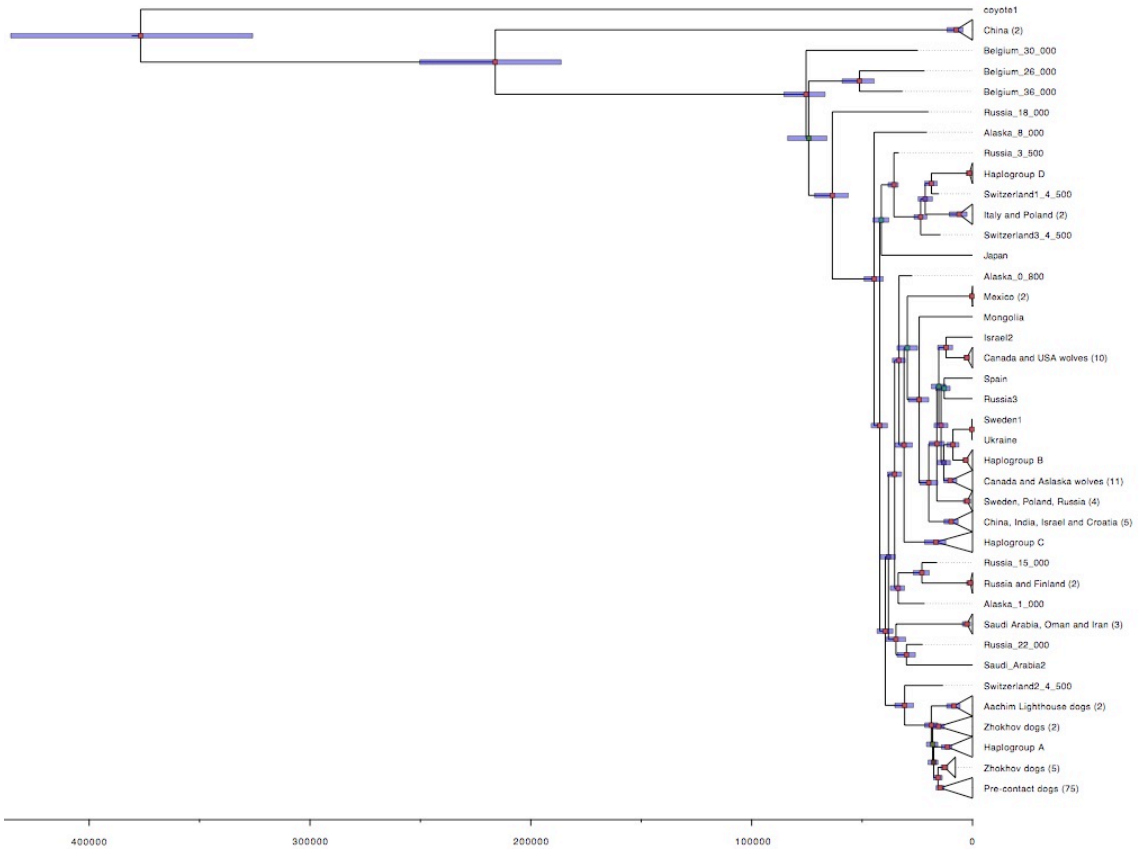


Figure S3.6 Bayesian tree (BEAST) of mtDNA data. Red, purple, and green circle represent nodes with >0.9, >0.7 and >0.5 posterior probability respectively. Blue bar represents confidence interval of divergence time (scaled in year before present).

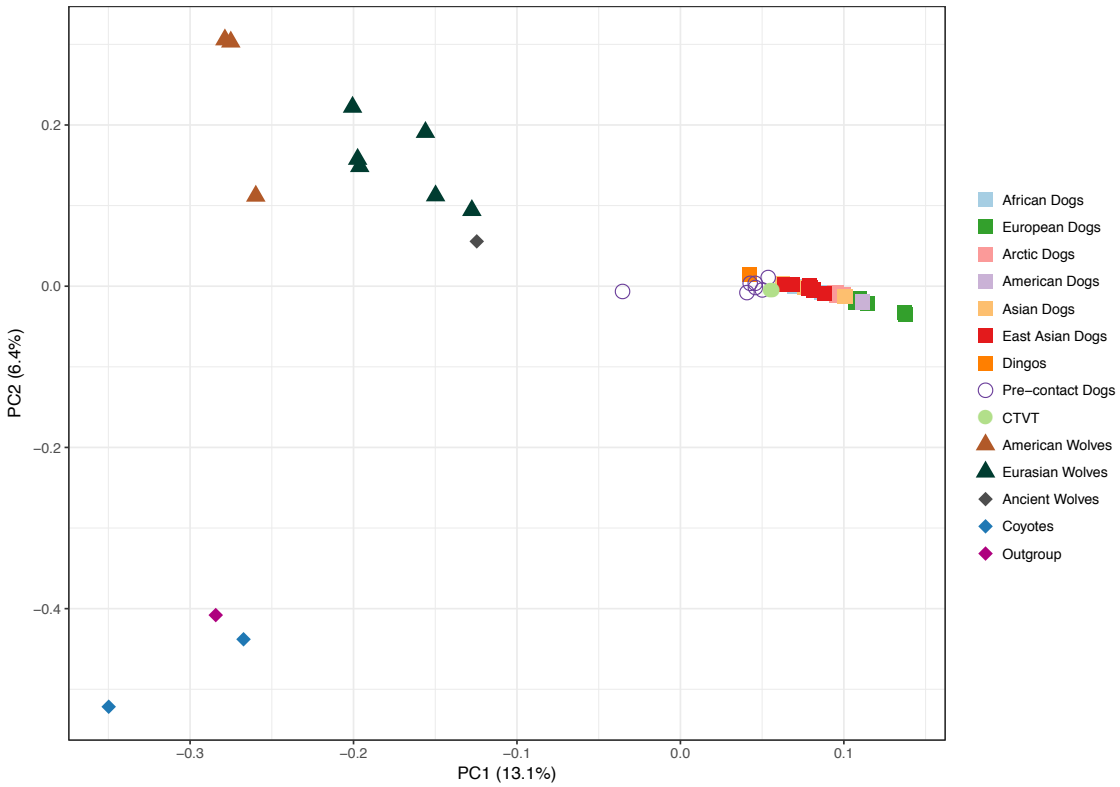


Figure S3.7 Principal Components Analysis (PC1 versus PC2) of 57 canid samples (including wolves and coyotes) based on 2,063,129 SNPs ascertained using the genome-wide data-set. All pre-contact dog samples were projected.

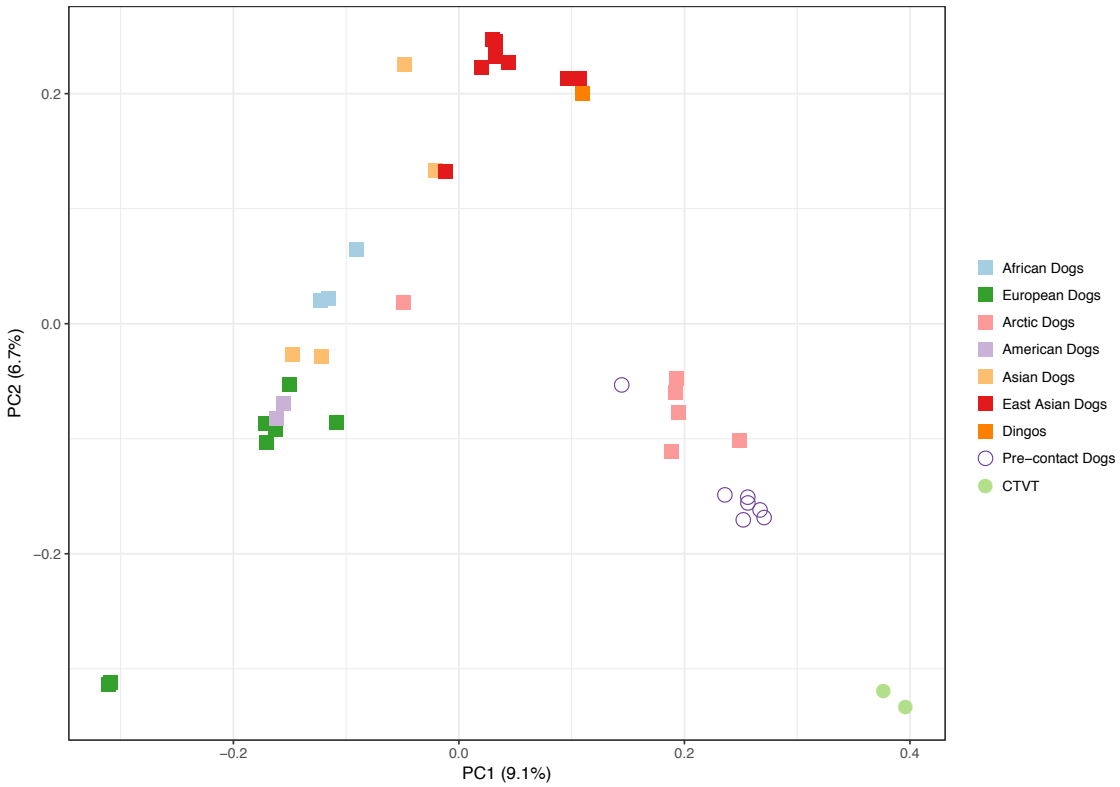


Figure S3.8 Principal Components Analysis (PC1 versus PC2) of 44 dog samples (excluding wolves and coyotes) based on 2,063,129 SNPs ascertained using the genome-wide data-set. All pre-contact dog samples were projected.

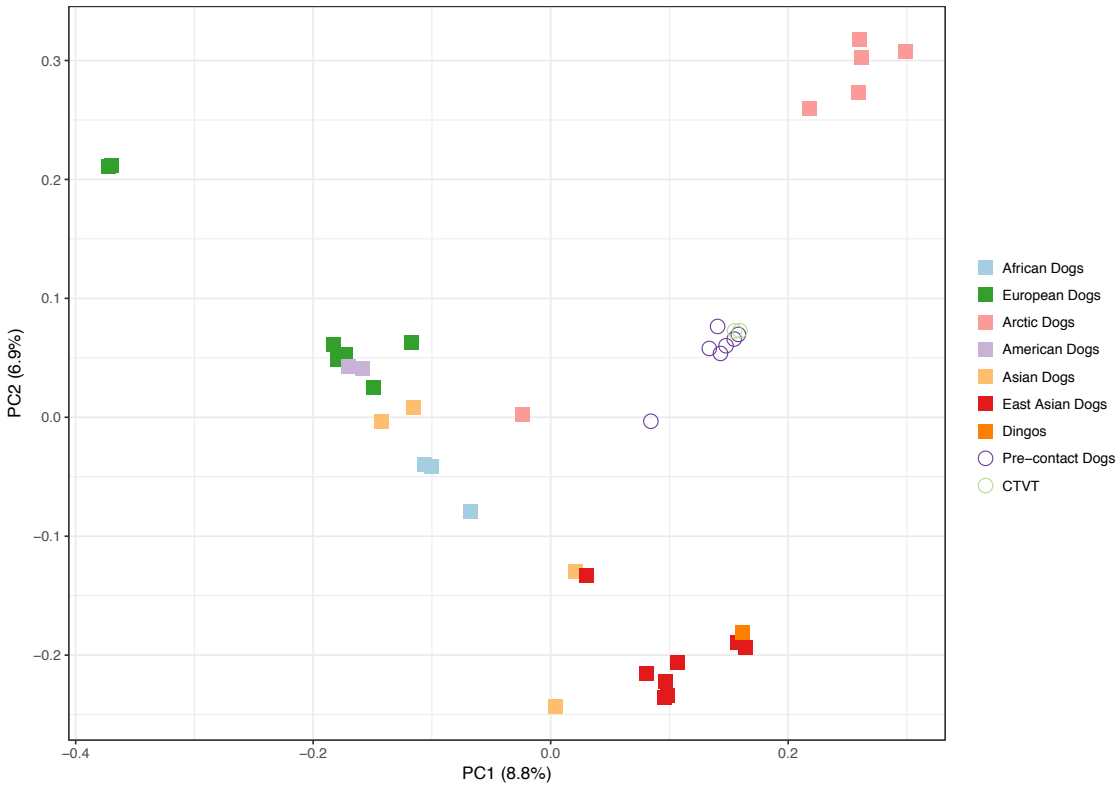


Figure S3.9 Principal Components Analysis (PC1 versus PC2) of 44 dog samples (excluding wolves and coyotes) based on 2,063,129 SNPs ascertained using the genome-wide data-set. All pre-contact dog and CTVT samples were projected.

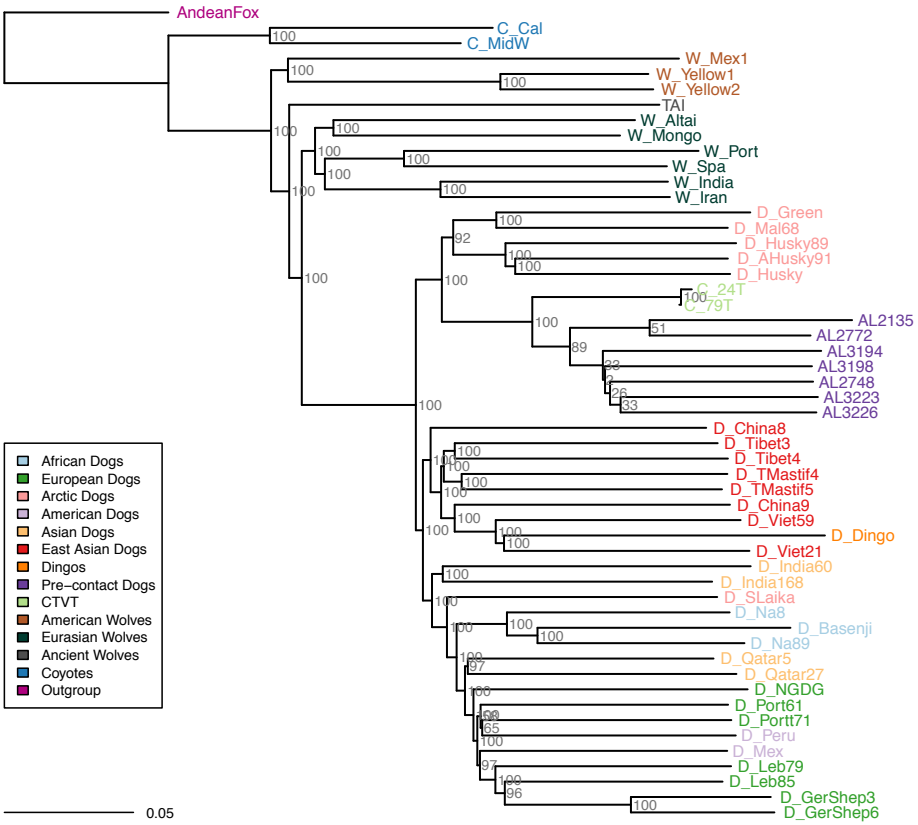


Figure S3.10 Neighbour Joining (NJ) tree based on Identity By State (IBS). This figure is the same as in Figure 1c. Confirms that CTVT is more closely related to pre-contact dogs than any other dog population. Confirms that PCD form a monophyletic clade.

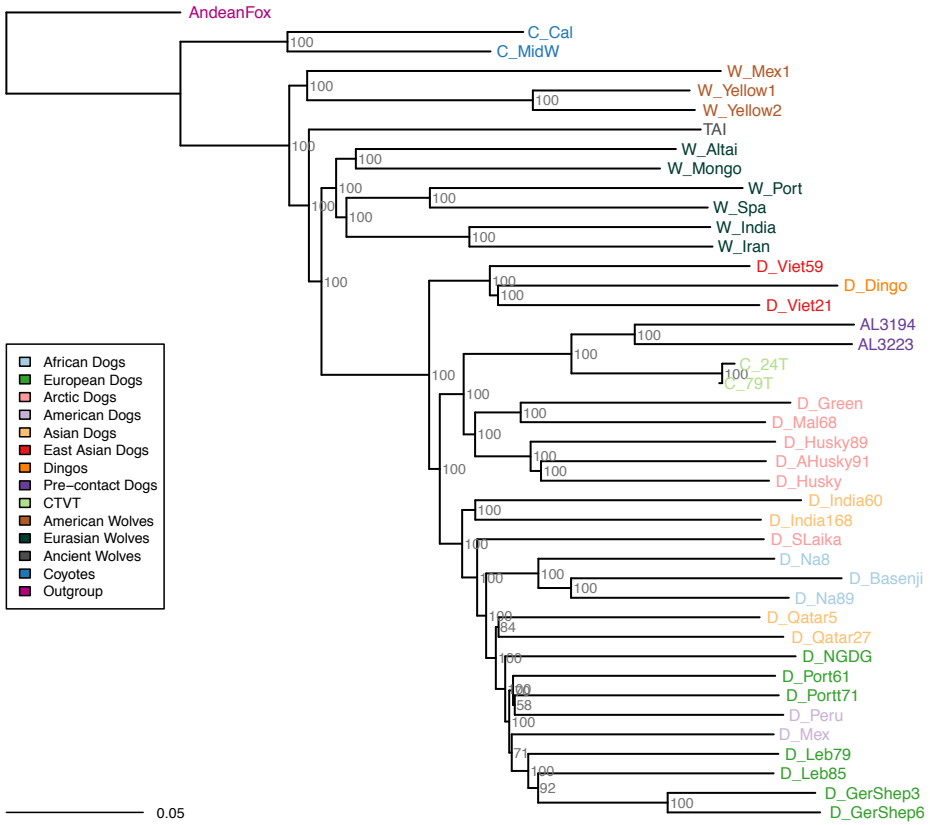


Figure S3.11 Neighbour Joining (NJ) tree based on Identity By State (IBS). Same as Figure S3.3 but without East Asian dogs that are admixed with European dogs i.e. excluding all East Asian dogs except Vietnamese. PCD, Arctic dogs and CTVT founder now appear more closely related to Western dogs.

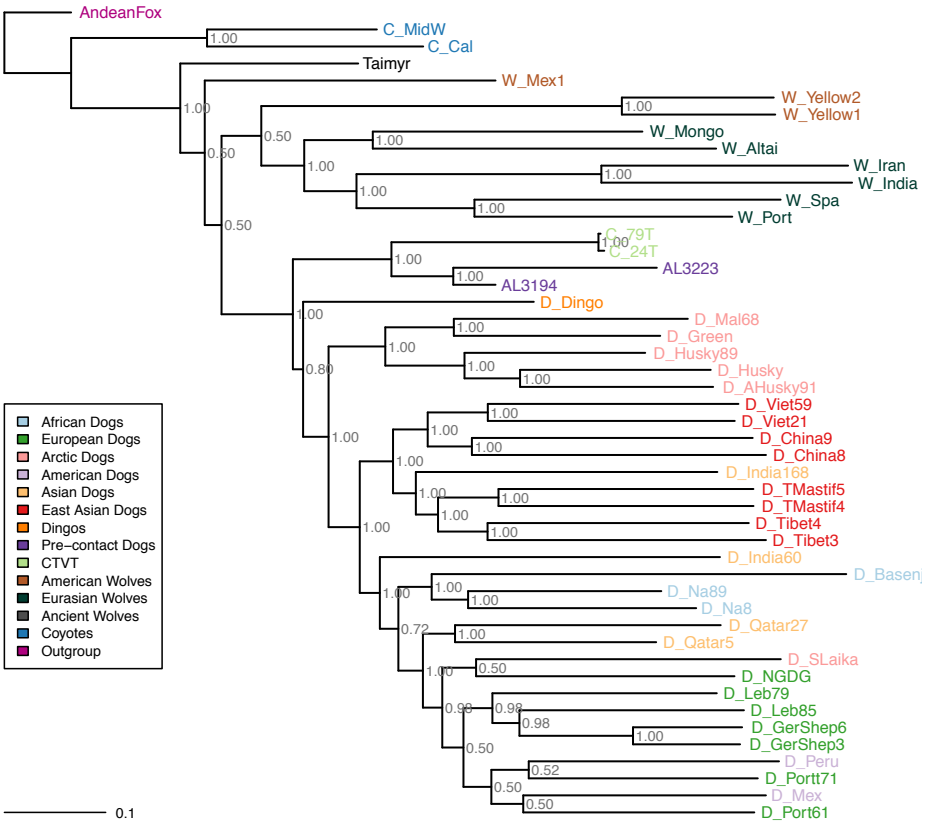


Figure S3.12 Bayesian tree based on ~26K transversions. Confirms that CTVT and PCD are monophyletic with high support and supports the basal placement of the CTVT/PCD clade. Support values represent posterior probability.

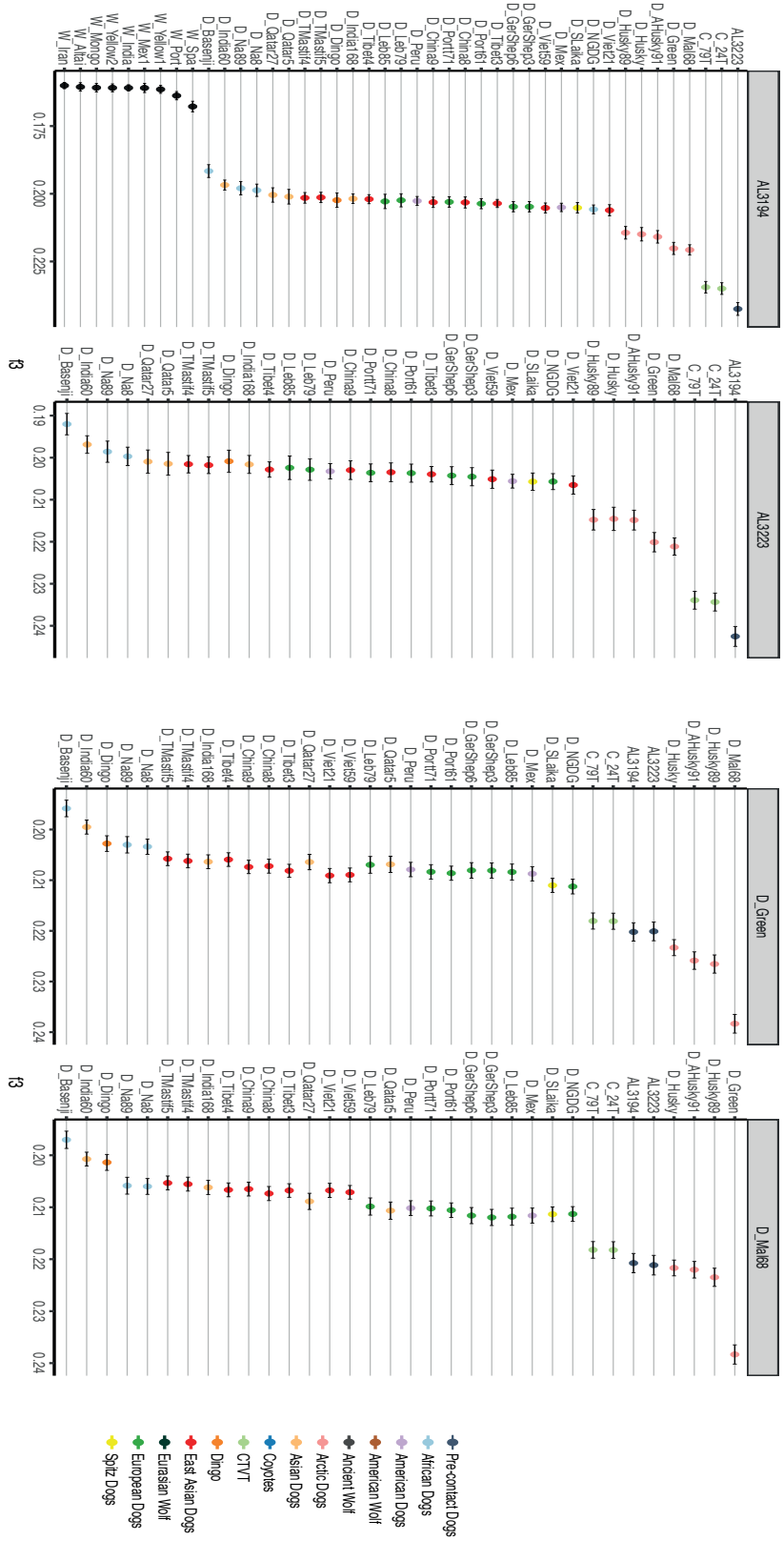


Figure S3.13 Shared genetic drift measured by f_3(Outgroup; Y, X) where Y is either Port au Choix dog (AL3194), Weyanoke Old town dog (AL3223), Alaskan Malamute (D_Mal68), Greenland sledge dog (D_Green) and X represents modern dog populations. Error bars represent 1 SE.

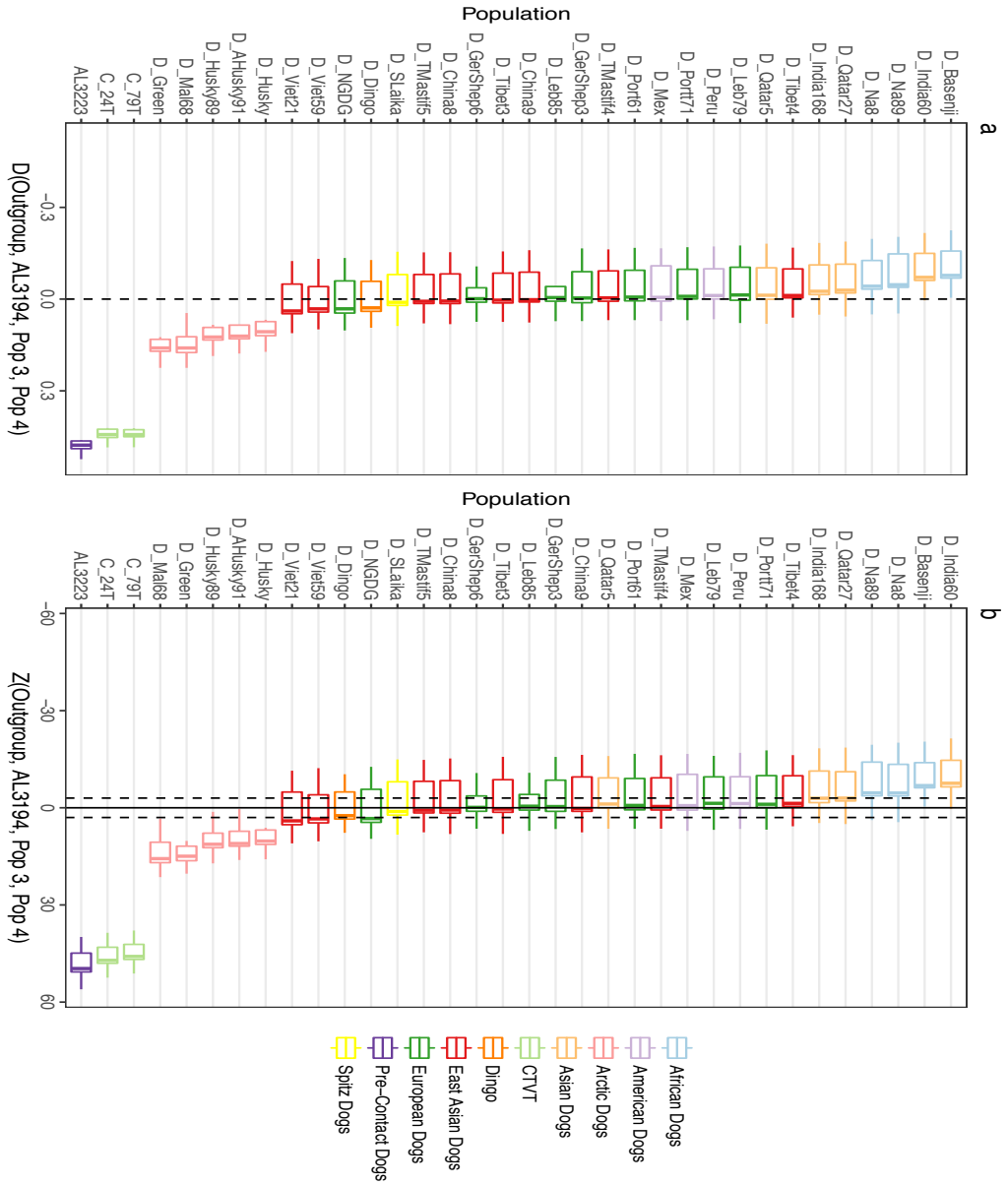


Figure S3.14 Box plot representing a. D-statistics and b. significance of D-statistics (Z) for every combination of D(Outgroup, AL3194[Port au Choix], Pop3, Pop4), where Pop3 is fixed and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and PCD while negative values imply PCD are closer to other dog populations. If Pop3 is not admixed with PCD, we expect -4<Z<4 (x-axis).

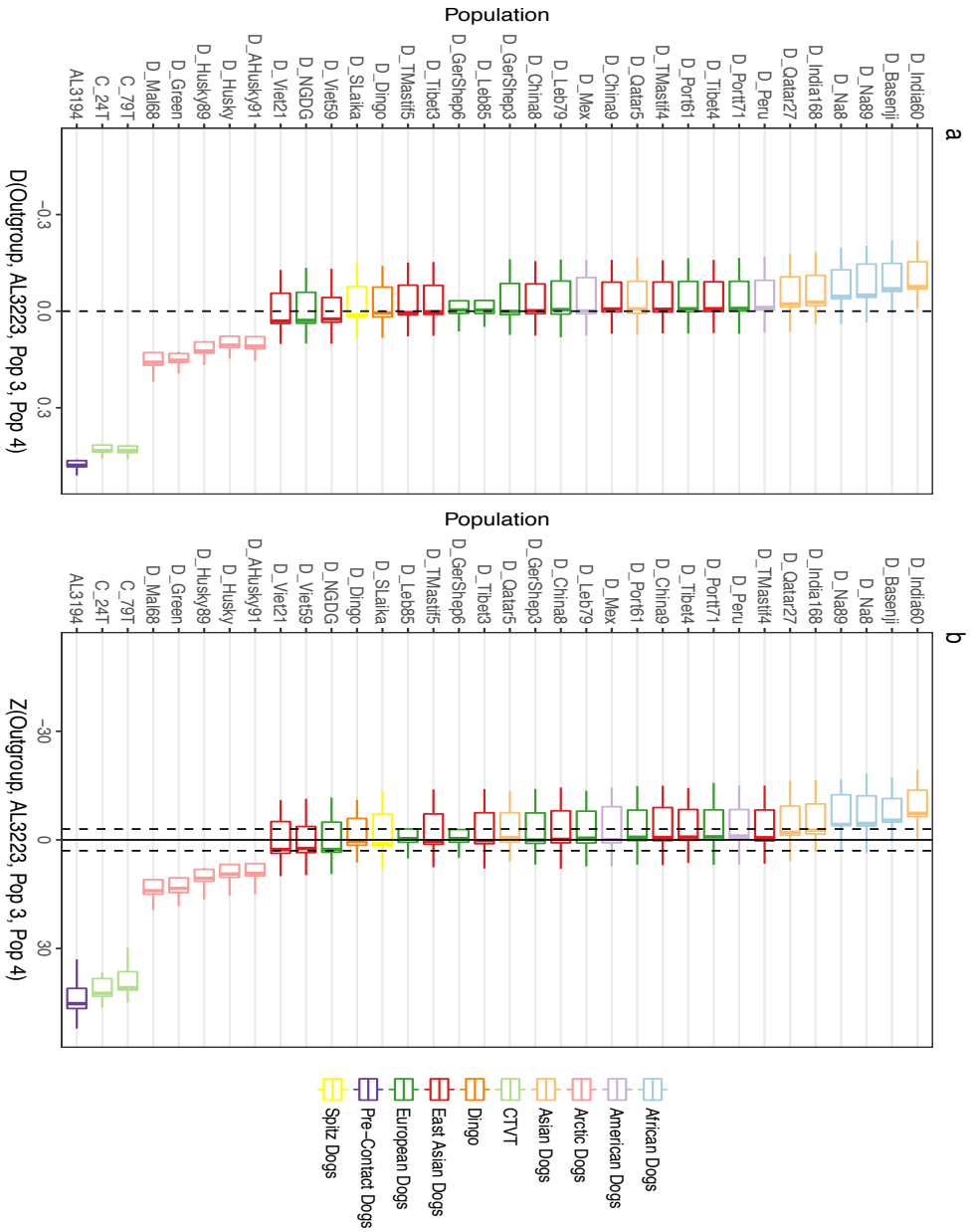


Figure S3.15 Box plot representing a. D-statistics and b. significance of D-statistics (Z) for every combination of D(Outgroup, AL3223[Weyanoke Old town], Pop3, Pop4), where Pop3 is fixed and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and PCD while negative values imply PCD are closer to other dog populations. If Pop3 is not admixed with PCD, we expect -4<Z<4 (x-axis).

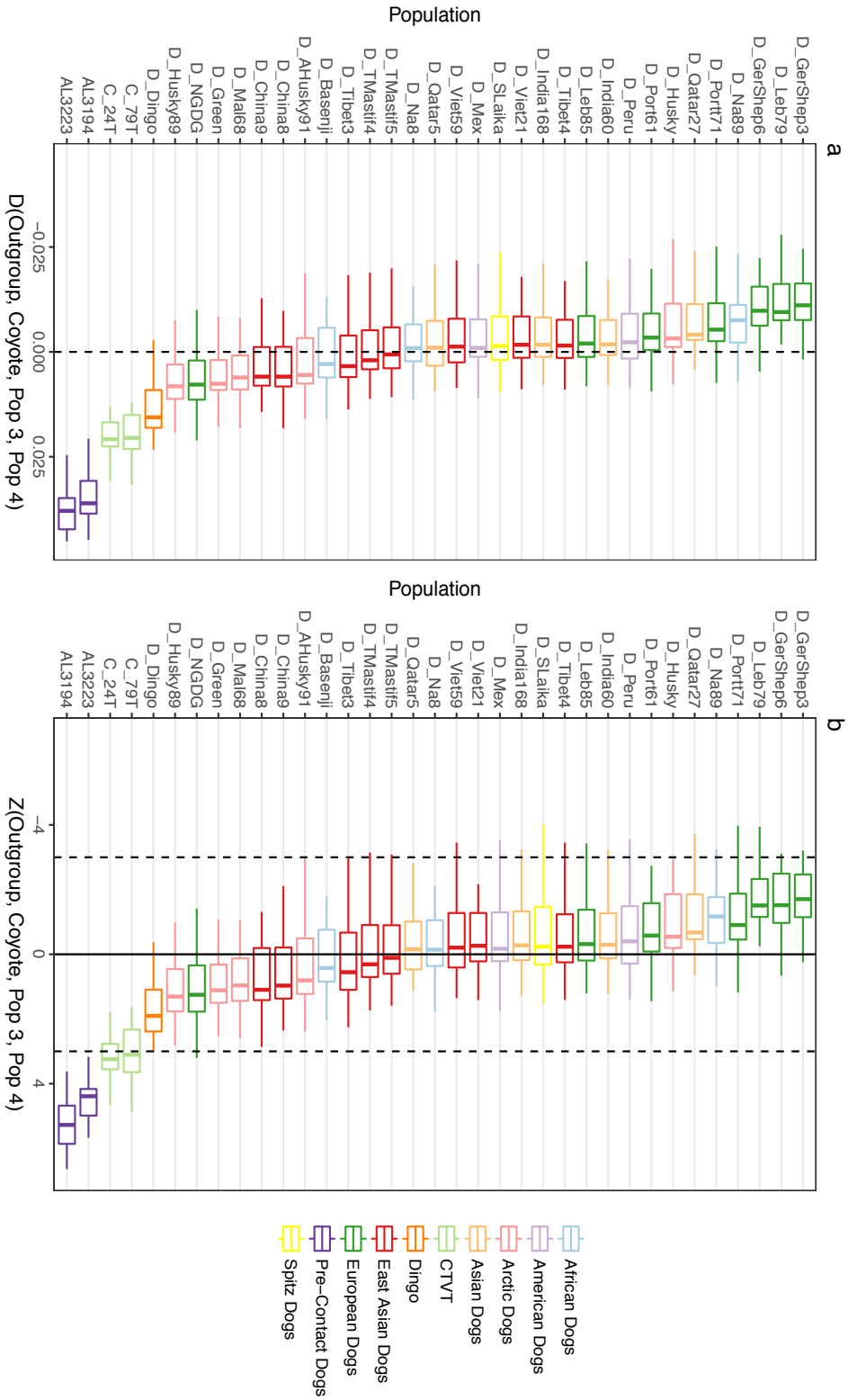


Figure S3.16 Box plot representing a. D-statistics and b. significance of D-statistics (Z) for every combination of D(Outgroup, Coyote, Pop3, Pop4), where Pop3 is fixed (x-axis) and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and coyotes while negative values imply coyotes are closer to other dog populations. If Pop3 is not admixed with coyotes, we expect -4<Z<4 (x-axis).

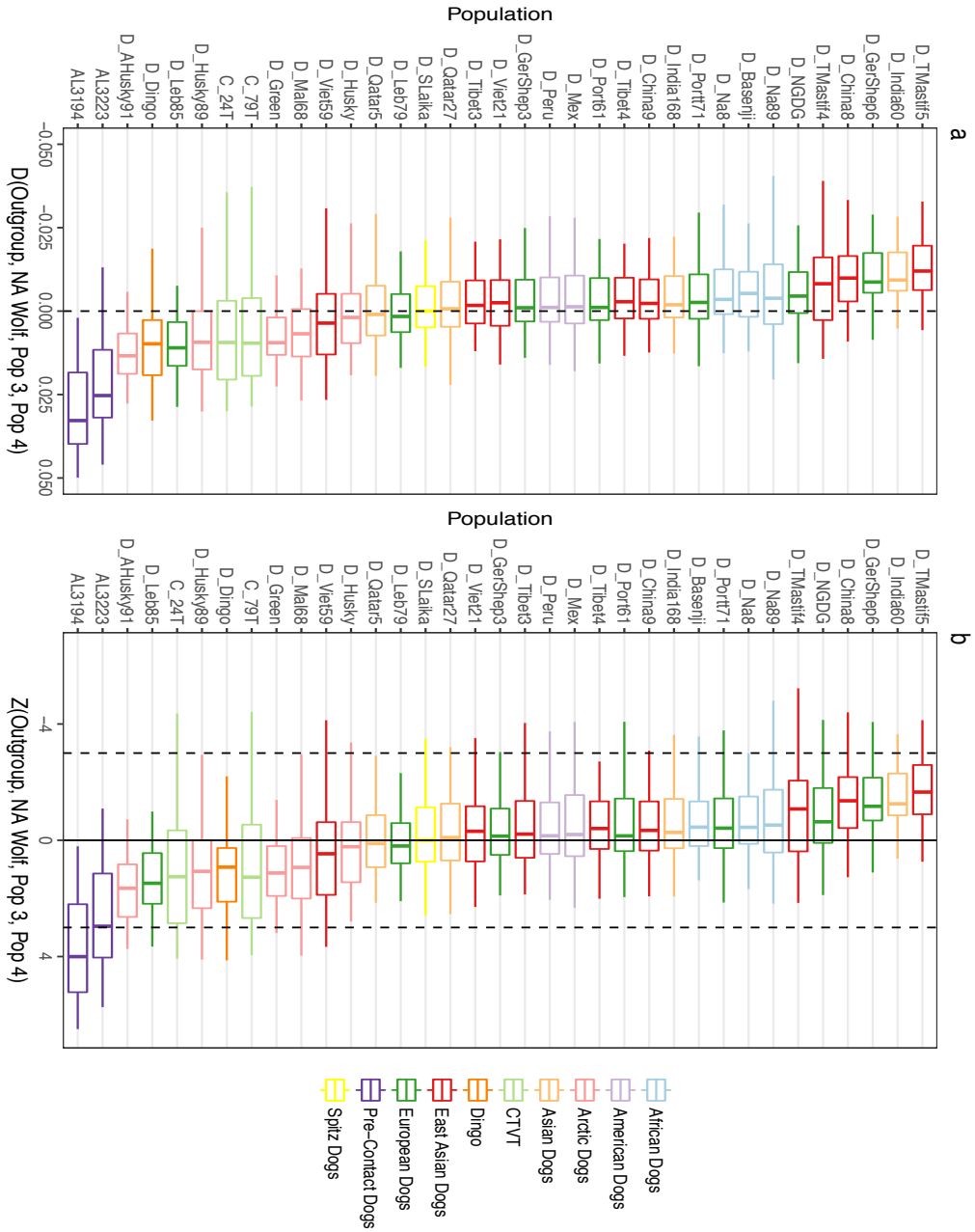


Figure S3.17 Box plot representing a. D-statistics b. significance of D-statistics (Z) for every combination of D(Outgroup, north American wolf, Pop3, Pop4), where Pop3 is fixed and Pop4 represents any other genome. Positive values support a close relationship between Pop3 and north American wolves while negative values imply NA wolves are closer to other dog populations. If Pop3 is not admixed with north American wolves, we expect -4<Z<4 (x-axis).

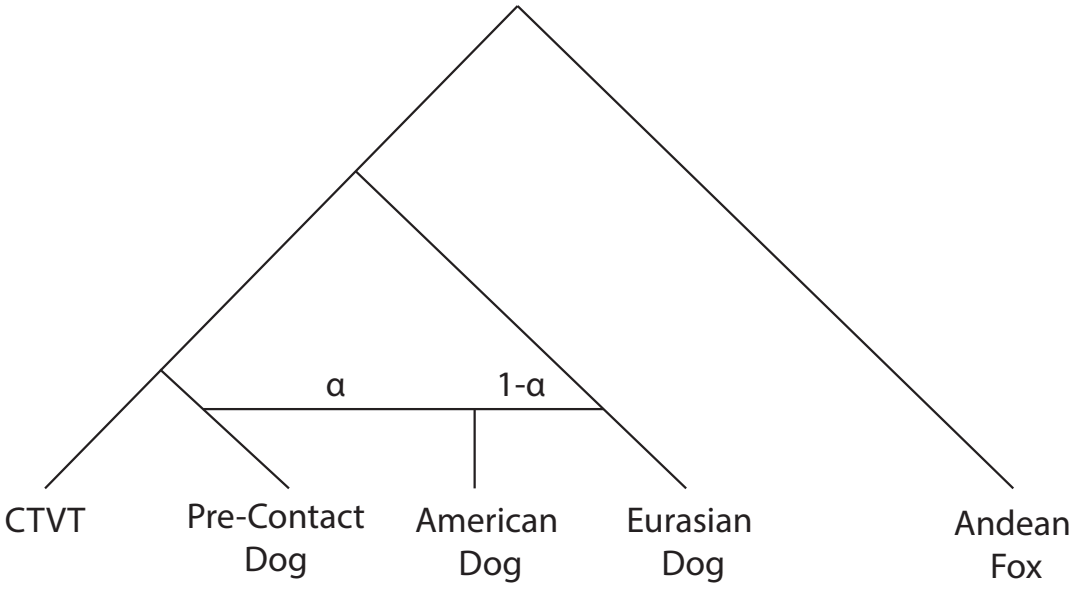


Figure S3.18 Schematic representation of the assumed phylogeny for the f4 ratio test used to estimate pre-contact ancestry into modern North American dogs. Alpha represents the degree of ancestry from pre-contact dogs.

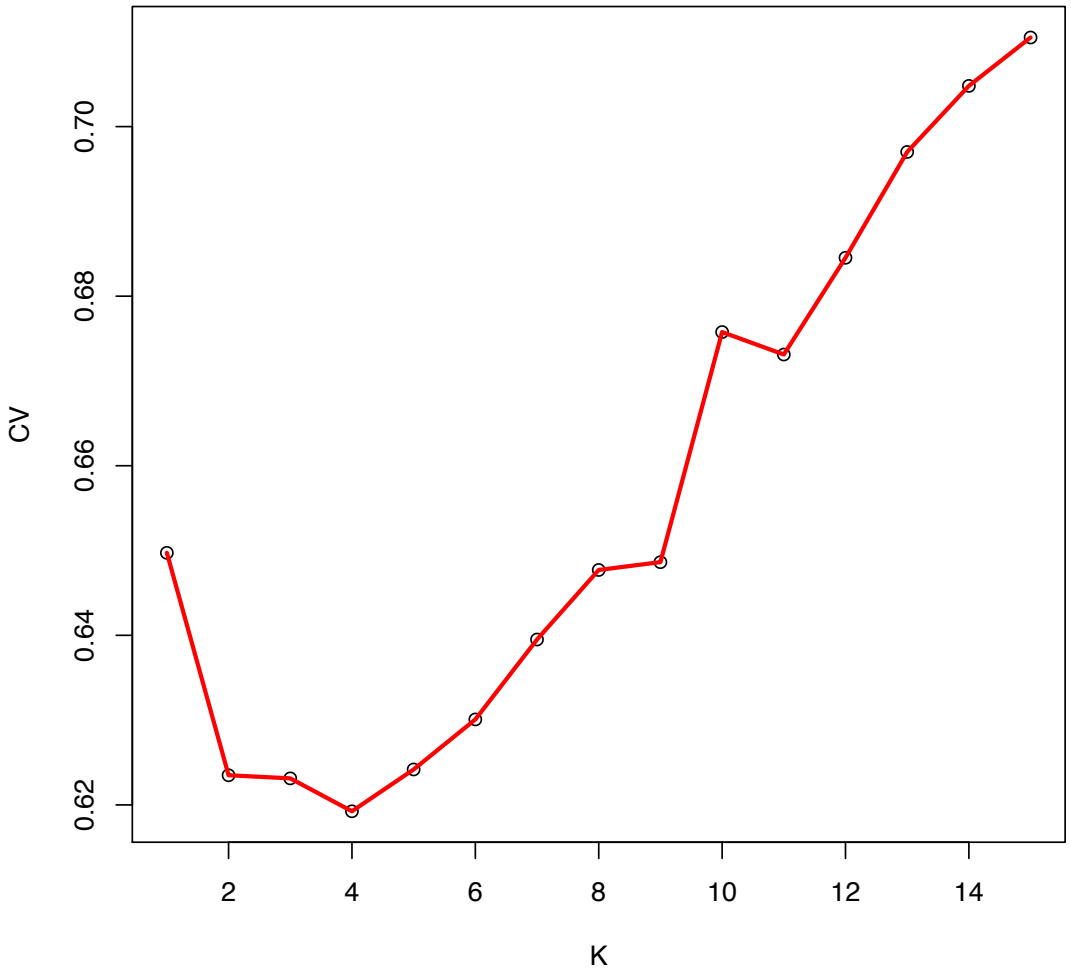


Figure S3.19 Cross validation (CV) values for ADMIXTURE analysis of SNP array data.

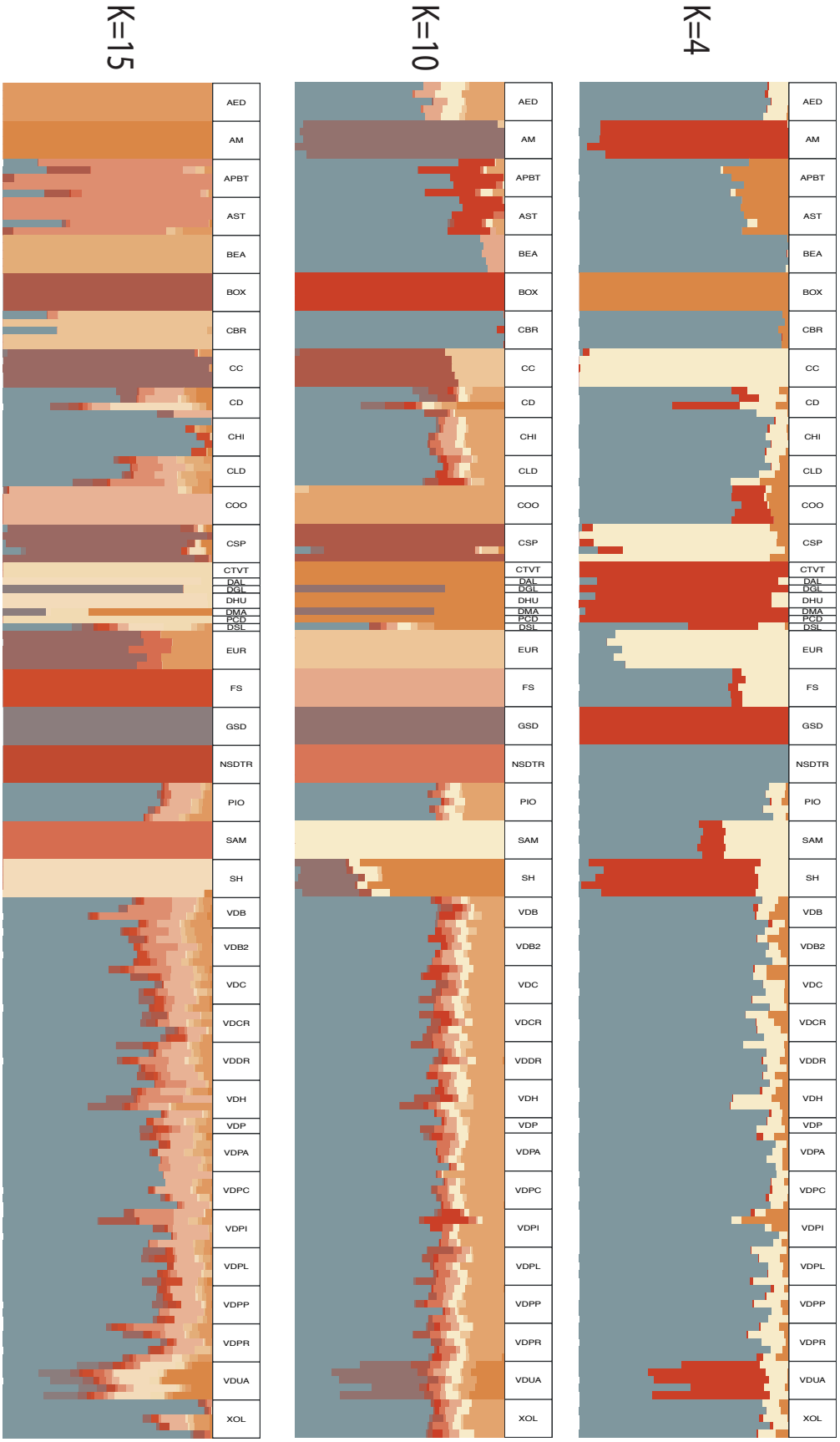


Figure S3.20 ADMIXTURE results based on SNP array data for K=4, 10 and 15 (see Table S3.9 for population codes).

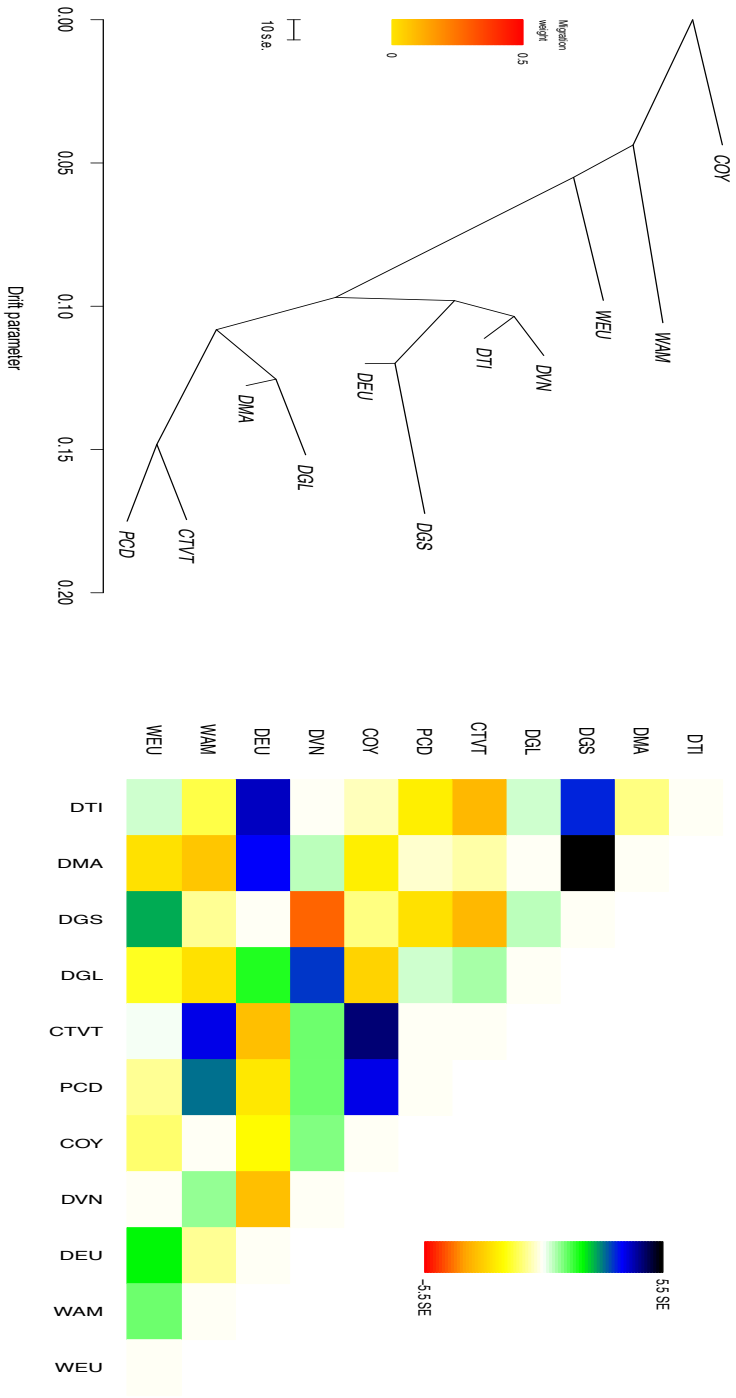


Figure S3.21 Admixture graph without migration edge and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

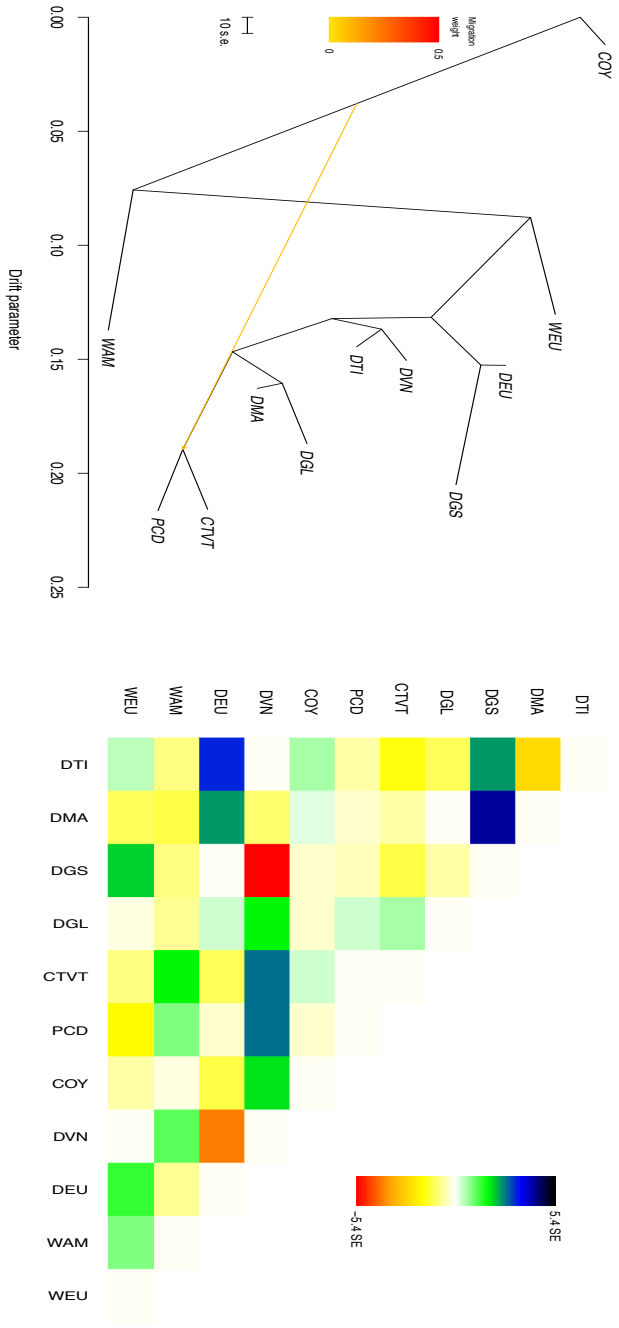


Figure S3.22 Admixture graph with a single migration edge and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). We see evidence for admixture from coyotes (COY) into the pre-contact dog lineage (PCD/CTVT), consistent with Figure S3.16. *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

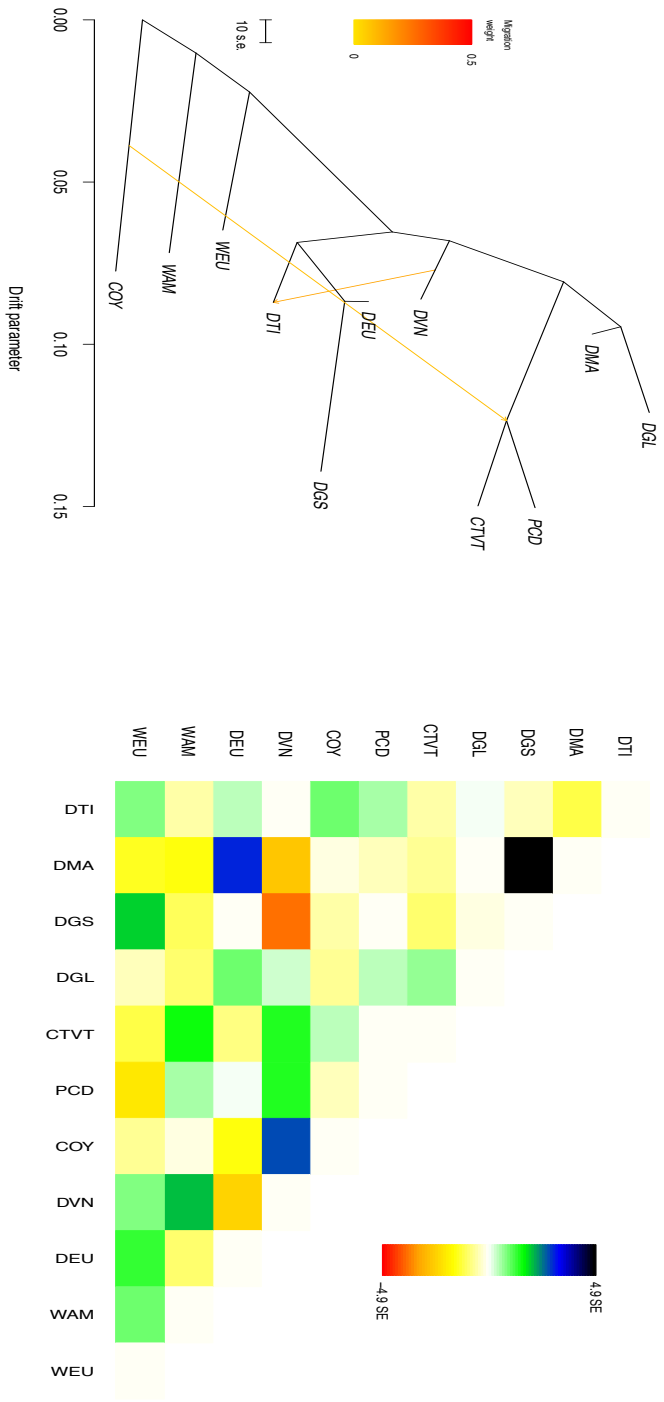


Figure S3.23 Admixture graph with two migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

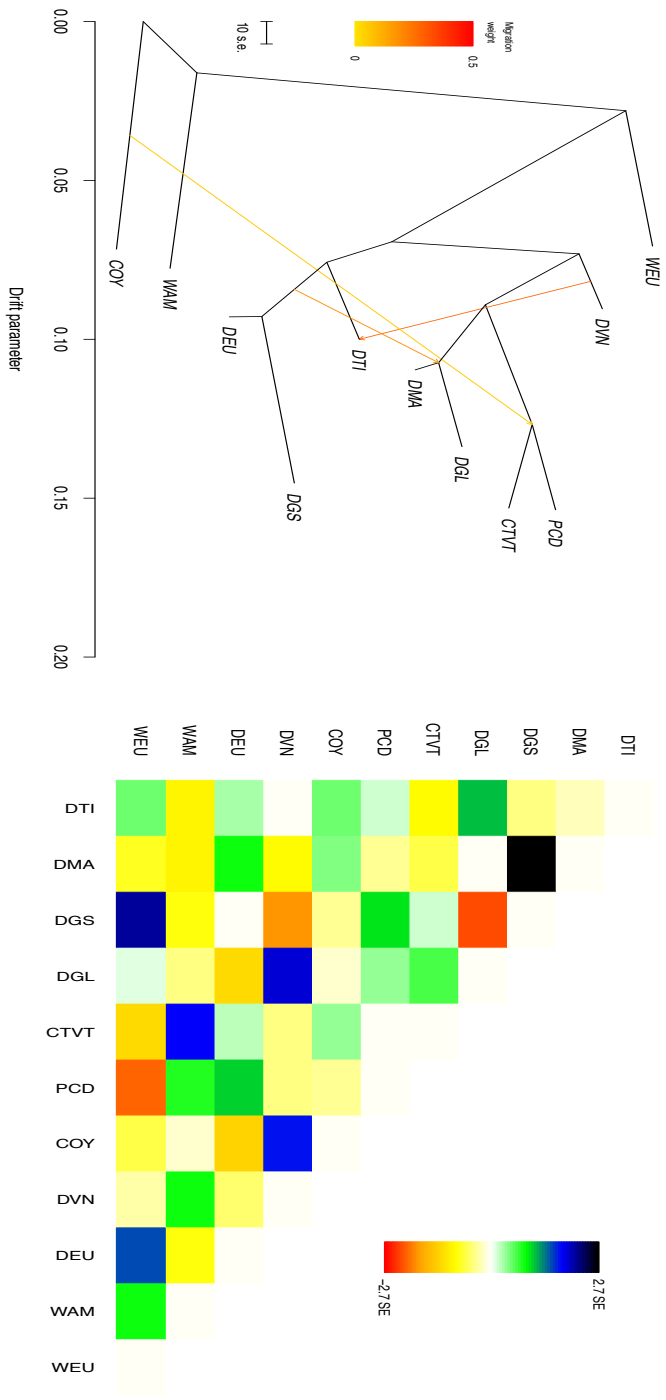


Figure S3.24 Admixture graph with three migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on transversions). *Western Eurasian dogs* - Portuguese village dogs (DEU), German Shepherd (DGS), *East Asian dogs* - Vietnamese village dogs (DVN) and Tibetan village dogs (DTI), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA) and Greenland dogs (DGL), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

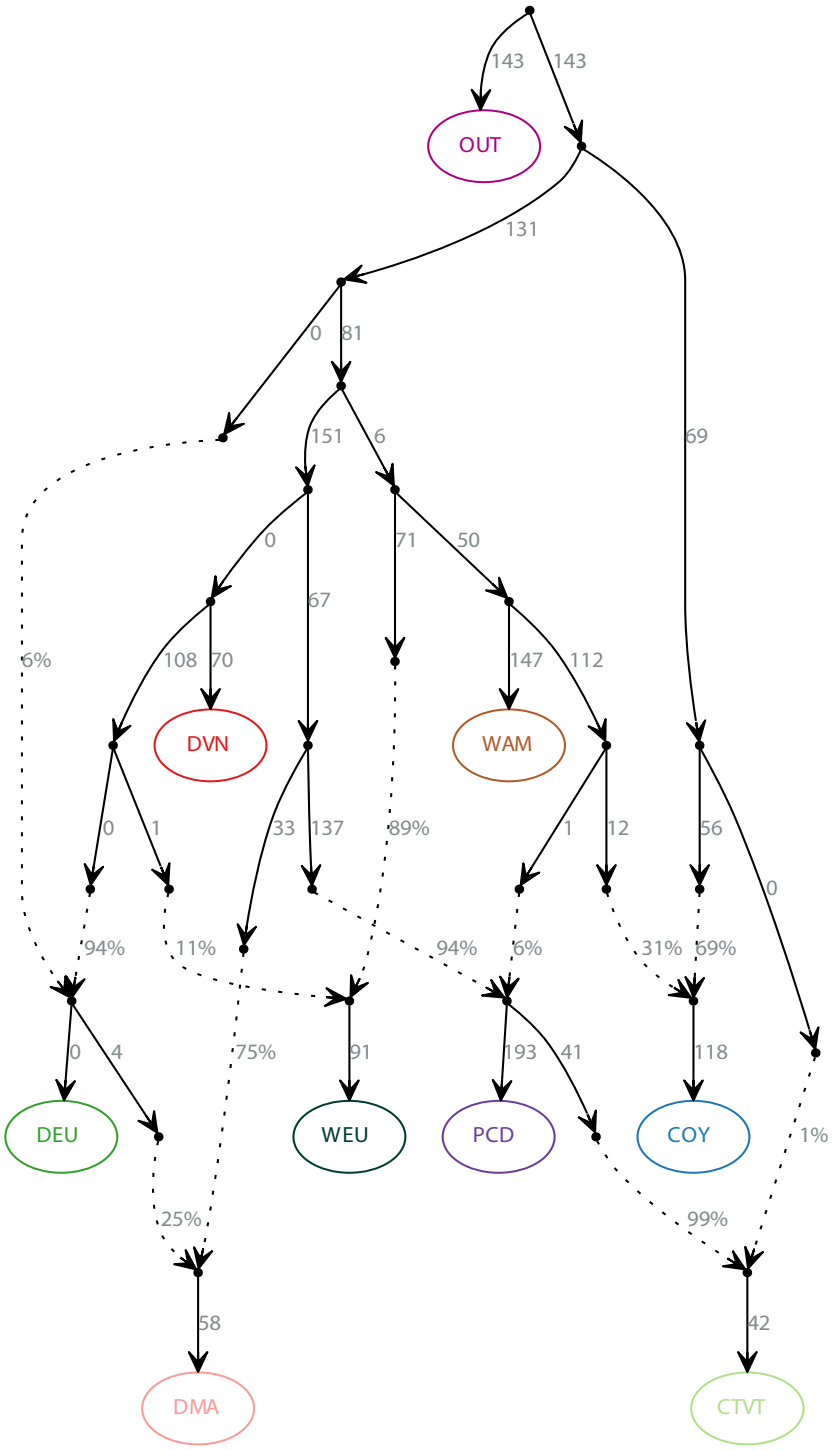


Figure S3.25 qpGraph model with admixture fractions. *Western Eurasian dogs* - Portuguese village dogs (DEU), *East Asian dogs* - Vietnamese village dogs (DVN), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

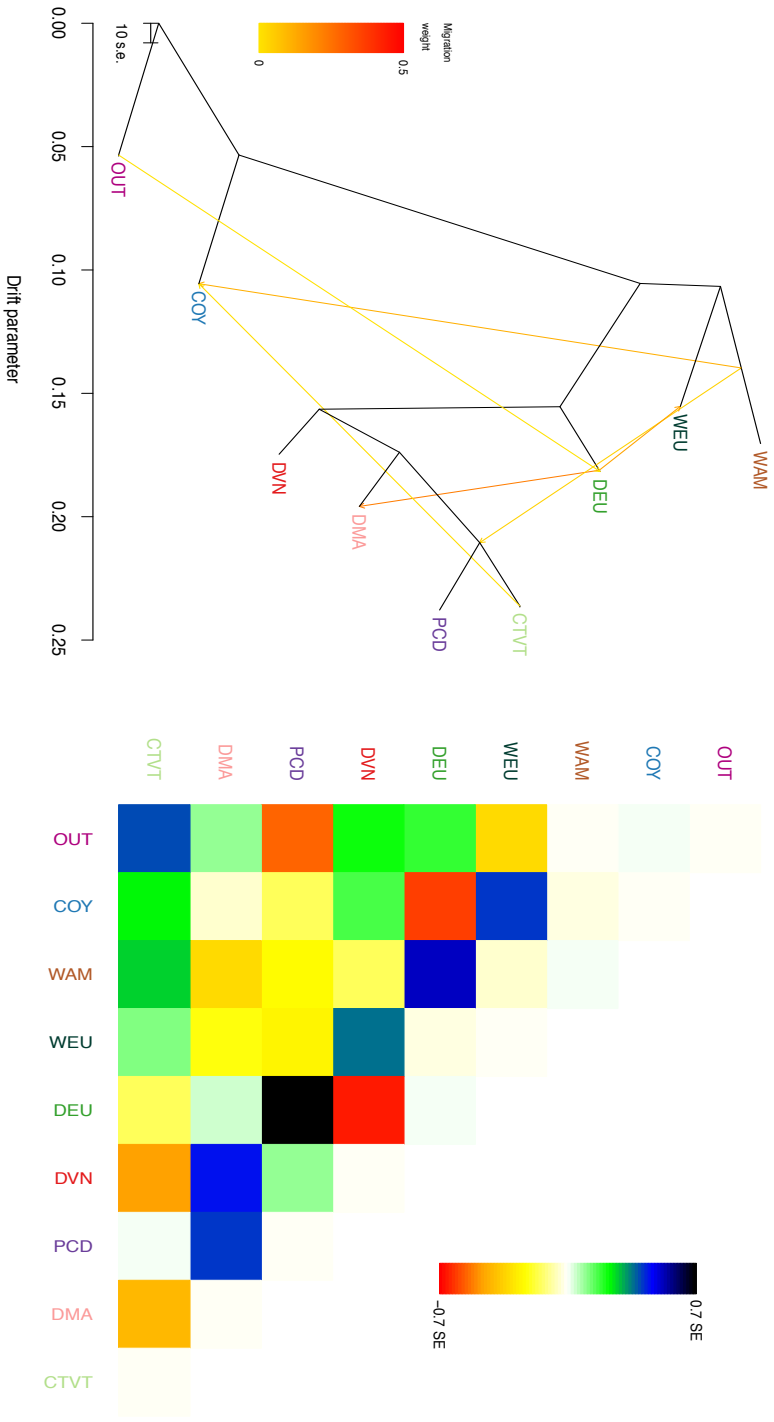


Figure S3.26 Admixture graph and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix for the same population as in Figure S3.25. *Western Eurasian dogs* - Portuguese village dogs (DEU), *East Asian dogs* - Vietnamese village dogs (DVN), *Pre-contact dogs* (PCD), including both Port au Choix (AL3194) and Weyanoke Old Town (AL3223), *Arctic dogs* - Malamute (DMA), *CTVT* - (79T and 24T), *Eurasian wolves* (WEU) from Spain and Portugal, *North American wolves* (WAM) from Yellowstone, *Coyotes* (COY) as an outgroup.

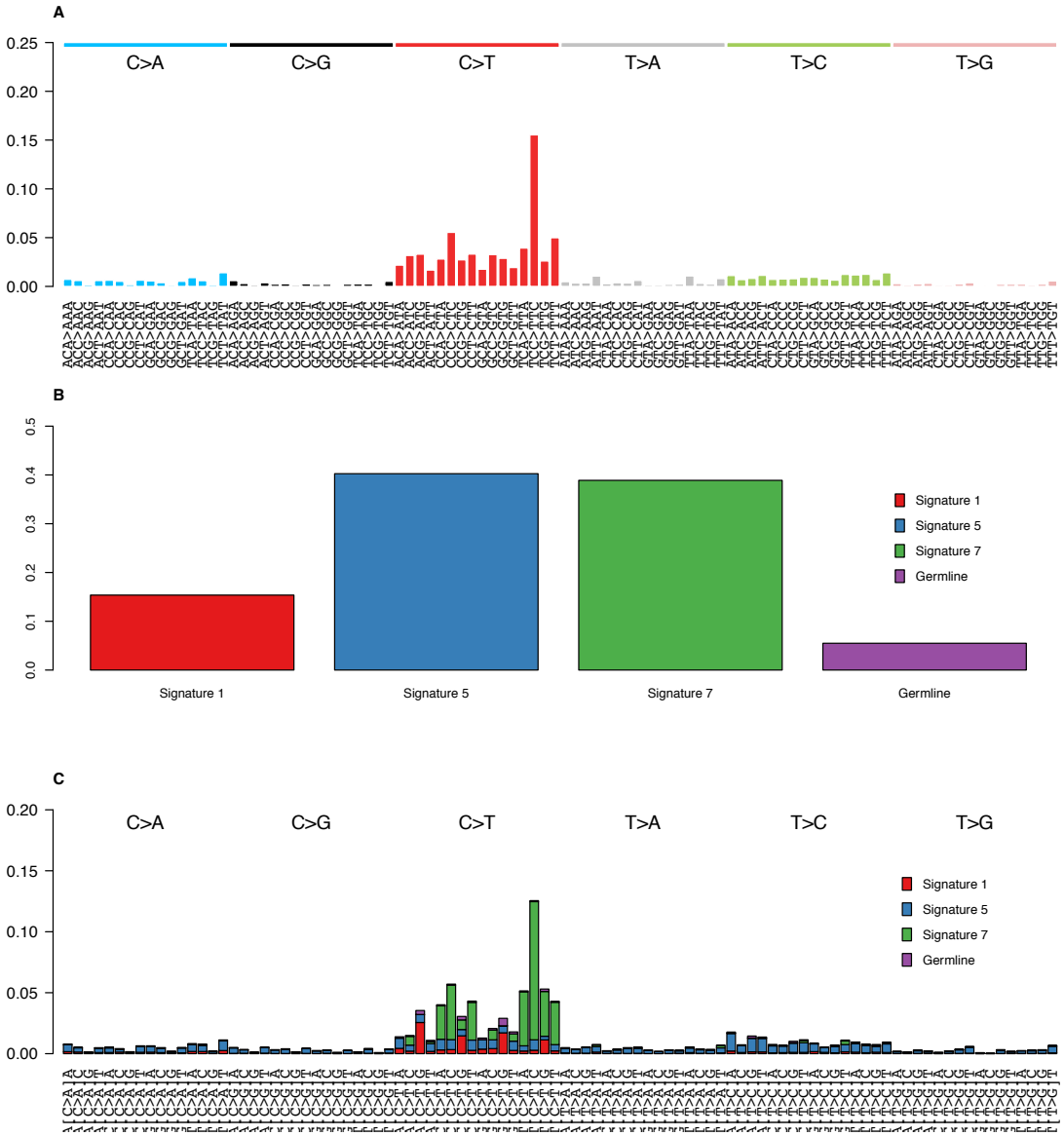


Figure S3.27 A. CTVT mutation spectrum. 1,925,779 tumour-only mutations in CTVT are displayed by mutation type (in pyrimidine context) with immediate 5' and 3' context. Each of the 96 mutation classes is displayed on the horizontal axis. Mutation proportions are displayed relative to CanFam3.1 B. Fraction of CTVT tumour-only mutations attributable to COSMIC Signatures 1, 5, 7, and the dog germline signature, as estimated using sigfit. C. Reconstruction of CTVT tumour-only spectrum using COSMIC signatures 1,5 and 7 and the dog germline signature.

3.7 Acknowledgments and Notes

Acknowledgments: We thank L. Orlando, R. K. Wayne and D. Meltzer for their valuable comments, B. M. Kemp, M. Masson and J. Chupasko for support and J. Southon (Keck Carbon Cycle Accelerator Mass Spectrometer, University of California, Irvine) for the radiocarbon date on the Port au Choix dog. We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility for providing computing time. We thank the Illinois State Museum, the Illinois State Archaeological Survey, the G. A. Black Laboratory of Archaeology at Indiana University Bloomington, the Instituto Nacional de Antropologia e Historia and the Ohio Historical Society for access to material. We thank The Rooms (Museum Division), Board Executive and Government of Newfoundland and Labrador for permission to access and sample the Port au Choix material. We are grateful to M. Ptaszynska for useful information and to S. Zhang for assistance with samples.

Funding: L.A.F.F. was supported by the Wellcome Trust (210119/Z/18/Z) and by Wolfson College (University of Oxford). L.A.F.F., J.H, A.L., A. H-B, O.L., K.M.D. 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. M.N.L. and E.P.M. were supported by the Wellcome Trust (102942/Z/13/A). A.R.P. was supported by the Max Planck Society. E.K.I.P. was supported by a Clarendon Fund Scholarship, University of Oxford. M.T.P.G was supported by European Research Council grant (ERC-2015-CoG-681396 – Extinction Genomics). A.M. was supported by the Muséum national d'Histoire naturelle. K.E.W. and R.S.M were supported by a National Science Foundation grant (BCS-1540336) and a Wenner-Gren grant. V.G. was supported by a Social Sciences and Humanities Research Council Insight Grant. V.V.P., E.Y.P., and P.A.N. were supported by Russian Science Foundation project N16-18-10265-RNF. Y.M.K was supported by a Herchel Smith Research Fellowship, S.J.C was supported by Millennia Research Limited, J.J was supported by the Santa Barbara Museum of Natural History and A.B. supported by the American Kennel Club and the NIH (R01GM103961). We thank the staff of the Danish

National High-Throughput Sequencing Centre for assistance in data generation and the Illinois State Museum Society for funding.

Author contributions: L.A.F.F, G.L. and E.P.M conceived the project and designed research; A.P., K.D. and G.L. coordinated the archaeological analyses and sampling collection efforts with input from R.P.M, C.A., A.B-H and K.E.W.; A.P., C.A, J.B, E.G., A.J.H, M-H.S.S., S.J.S, M.E., V.G., J.J, A.K.K, P.A.N, C.P.L, A.M.. T.M., K.N.M., M.O., E.Y.P, P.S, V.V.I., C.W. and V.V.P provided/collected samples; K.W.E, A.L., J.H., O.L., S.B., A.D. E.A.D., J.E., J-M. R., M-H.S.S. conducted the ancient laboratory work with input from R.P.M., G.L., L.A.F.F, E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided/collected CTVT samples; M.N.L. and Y-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F and E.K.I.P conducted the analyses of ancient data with input from S.G., A.K., A.B. and E.P.M.; L.A.F.F., G.L., E.P.M., M.N.L. and A.P., wrote the paper with input from all other authors.

Competing interests: A.D., J.E, and J-M.R. are employees of Arbor Biosciences which provided target enrichment kits used in this study. J-M.R. is also a founder of Arbor Biosciences. KD currently holds honorary Professor positions in the Departments of Archaeology at both the University of Aberdeen and Simon Fraser University. A significant portion of the research included in this paper (along with associated NERC funding) was also undertaken whilst KD was a full-time member of faculty at the University of Aberdeen. A.R.B. is founder and CSO of Embark Veterinary.

Data and materials availability: The reads for the ancient data have been deposited at the European Nucleotide Archive (ENA) with project number PRJEB22026. Reads for new CTVT genomes were deposited at the European Nucleotide Archive (ENA) with project number PRJEB22148. Mitochondrial sequences alignments, genotype files (in plink format), and phylogenetic trees were deposited on Dryad (doi:10.5061/dryad.s1k47j4).

3.8 Permission from Co-authors

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9 OCT 2018

Name(s): ADAM BOYKO

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: October 13, 2018

Name(s): P. Fedko V.V., Kasperov A.K., E. Yu. Pavlova

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/09/2018

Name(s): JEAN-MARIE ROUILLARD
JACOB ENK
ALISON DEVAULT

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 15/10-2018

Name(s): Anders J. Hansen

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/10/2018

Name(s): ANDREA STRAKOVA

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9 October 2018

Name(s): Andrew Kitchen

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 23/10/2018

Name(s): Angela Perri

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9/10 - 2018

Name(s): ANNA LINDERHOLM

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 23. 10. 2018

Name(s): ARDERN HULME-BEAMAN

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 17/10/2018

Name(s): Aurélie Manin

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: October 10, 2018

Name(s): Carlos Peraza Lope

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/10/2018

Name(s): CARLY AMEEN

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/9/18

Name(s): Chris Widga

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 15/10/2018

Name(s): Elizabeth Murchison

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: Oct, 9, 2018

Name(s): Dr. Eric Guiry

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 23.10.2018

Name(s): Eske Willerslev

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9th of October of 2018

Name(s): Evangelos Antonios Dimopoulos

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: Oct 15, 2018

Name(s): Greger Larson

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work "The evolutionary history of dogs in the Americas" as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A.,J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M.,T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L.,L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K.,A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 23.10.2018

Name(s): James Haile

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9 Oct 2018

Name(s): John R. Johnson (J.J.)

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 09/10/18

Name(s): Professor Keith Dobney

Signature(s): Keith Dobney

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: November 6, 2018

Name(s): Kelsey Noack Myers, PhD, RPA

Signature(s): Kelsey Noack Myers

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: October 9, 2018

Name(s): Kelsey E. Witt

Signature(s): Kelsey Witt Dillon

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9/10/2018

Name(s): KEVIN GORI

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 09/10/2018

Name(s): Laurent Frantz

Signature(s): Laurent Frantz

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 22/10/18

Name(s): MÁIRE NÍ LEATHLOBHAIR

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9oct2018

Name(s): MARK OMURA

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10. Octobe 2018

Name(s): Michael Meyer

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10 10 18

Name(s): Holger Stroder S. Mirg

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: Oct 9, 2018

Name(s): Morley Eldridge

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 11th October 2018

Name(s): Dr Ophélie Lebrasseur

Signature(s):

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/29/18

Name(s): Paul W Sciulli

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: October 16 2018

Name(s): Pavel A. Nikolskiy

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/22/18

Name(s): Ripan S. Malhi

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9/10/2018

Name(s): SELINA BRACE, IAN BARNES

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 15/10/2018

Name(s): SHYAM GOPALAKRISHNAN

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 10/10/2018

Name(s): Susan Crockford

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9 OCTOBER 2018

Name(s): TERRANCE J. MARTIN

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 9-10-18

Name(s): MTP GILBERT

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 09.10.2018

Name(s): Varvara Ivanova

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 15 Oct. 2018

Name(s): Vaughan Grimes

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"The evolutionary history of dogs in the Americas"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date: 22 october, 2018

Name(s): Yajaira Núñez-Cortés

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work "The evolutionary history of dogs in the Americas" as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, that Evan K. Irving-Pease's overall contribution was equal to that of the other co-first authors, and that his contribution towards the ancestry analyses of the ancient nuclear DNA was greater than that of any other co-author.

Author contributions: L.A.F.F., G.L., and E.P.M. conceived of the project and designed the research; A.R.P., K.M.D., and G.L. coordinated the archaeological analyses and sample collection efforts with input from R.S.M., C.A., A.H.-B., and K.E.W.; A.R.P., C.A., J.B., E.G., A.J.H., M.-H.S.S., S.J.C., M.E., Y.N.C., V.G., J.J., A.K.K., P.A.N., C.P.L., A.M., T.M., K.N.M., M.O., E.Y.P., P.S., V.V.I., C.W., and V.V.P. provided and/or collected samples; K.E.W., A.L., J.H., O.L., S.B., A.D., E.A.D., J.E., J.-M.R., and M.-H.S.S. conducted the ancient laboratory work with input from R.S.M., G.L., L.A.F.F., E.W., I.B., and M.T.P.G.; M.M., E.P.M., and A.S. provided and/or collected CTVT samples; M.N.L. and Y.-M.K. conducted the CTVT analyses with input from E.P.M., K.G., and L.A.F.F.; M.N.L., L.A.F.F., and E.K.I.-P. conducted the analyses of ancient data with input from S.G., A.K., A.R.B., and E.P.M.; and L.A.F.F., G.L., E.P.M., M.N.L., and A.R.P. wrote the paper with input from all other authors.

Date:
9/10/18

Name(s):
Young Mi Kwon

Signature(s):

4 Genetic Analysis of the Dwarf Shorthorn Cattle of Socotra

4.1 Statement of Authorship

Design: I designed the research, with input from of L.A.F.F and G.L.

Data: I curated the modern genome-wide sequencing data and performed all quality checks and filtering.

Analysis: I performed all the computational analyses, and interpreted the results with input from L.A.F.F. and G.L.

Manuscript: I wrote the manuscript, with input from L.A.F.F and G.L.

4.2 Authors and Affiliations

Authors

Evan K. Irving-Pease[1], Laurent A. F. Frantz[1,2], Greger Larson[1]

Affiliations

[1] *The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK.*

[2] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*

4.3 Abstract

The island of Socotra, in the Arabian Sea, was a remote trading post throughout antiquity. Multilingual inscriptions, found in coastal cave sites, testify to the island's use by sailors writing in Indian, Arabian, Ethiopian, Greek, Palmyran and Kushan languages between the 1st century BC and the 6th century AD. Since the earliest scientific expeditions to the island in the 19th century, Socotra has been famed for the uniqueness of its endemic flora. It is also home to an unusual breed of humpless, dwarf shorthorn cattle. These cattle are unlike any other found in neighbouring East Africa or the Arabian Peninsula, except in the highlands of Oman. To investigate the ancestry of these dwarf cattle, and their implications for the history of domestic cattle in the region, we sequenced five modern individuals from two populations in Socotra and the Dhofar Mountains, Oman. Using a reference panel of 793 modern samples, from 68 populations, analysis shows that Socotran and Dhofari dwarf cattle form a monophyletic clade, unique to any known breed of cattle. Admixture analyses indicates that the breed is mostly Eurasian taurine (*Bos taurus*) in origin, with a unique pattern of introgression from zebu (*Bos indicus*) and African taurine cattle.

4.4 Introduction

4.4.1 Cattle domestication

Cattle are one of the most economically important domestic animals, and the history of their domestication and rise to global dominance is deeply intertwined with intercontinental trade and environmental change. Genome-wide studies of modern domestic cattle have shown that cattle form three divergent clades: two groups of taurine cattle (*Bos taurus*), from Eurasia and Africa, and zebu or indicine cattle (*Bos indicus*), from South Asia (Decker et al., 2014; Gibbs et al., 2009).

The earliest evidence for the domestication of taurine cattle comes from the Neolithic site of Dja'de el Mughara, in the middle Euphrates Valley (Syria), dated to ~10,500 BP (Helmer et al., 2005). Archaeological evidence indicates that taurine cattle were domesticated from the Eurasian aurochs (*Bos primigenius primigenius*) (Conolly et al.,

2011), and mitochondrial evidence suggests the initial recruitment may have involved as few as 80 female aurochs (Bollongino et al., 2012). Taurine cattle subsequently migrated outwards from their centre of domestication, leaving a characteristic pattern of decreased mitochondrial (mtDNA) genetic diversity along their axes of migration (Scheu et al., 2015).

A recent study published the first whole-genome sequence of an extinct Eurasian aurochs, recovered from a 6,750-year-old British specimen (Park et al., 2015). Analysis of the genome-wide data revealed localised nuclear gene flow into the ancestors of British and Irish taurine cattle. A larger study, of 67 ancient nuclear cattle genomes from across the Near East, revealed localised aurochs introgression also occurred in the Levant and was likely a common process throughout the range expansion of domestic cattle (Verdugo et al., 2019).

The domestication of zebu cattle occurred in South Asia, approximately 8,000 years ago, from the Asian aurochs (*Bos primigenius namadicus*) (Chen et al., 2010; Utsunomiya et al., 2019), although it remains unclear whether this was a fully independent domestication or a secondary recruitment of local aurochs (Larson and Burger, 2013). Mitochondrial evidence suggests that the Indus Valley was the principal region of zebu domestication, although the origin of the second major mtDNA haplogroup is unclear (Chen et al., 2010). Zebu cattle expanded out of South Asia and into Africa, East Asia, and Europe between 4,000–1,300 BP; forming a cline of admixture with taurine cattle across Africa and Eurasia (Utsunomiya et al., 2019). The suggested origins of this intercontinental dispersal are linked to a multi-century drought, beginning around 4,200 BP (Sharifi et al., 2015). Evidence from ancient genomes, sampled across Southwest Asia, show a sudden influx of male biased zebu ancestry coinciding with this climate anomaly (Verdugo et al., 2019). Zebu cattle are generally better adapted to hot and arid climates (Hansen, 2004), and it is suggested that this dispersal was an intentional process of adaptive introgression in response to rapid climate change (Verdugo et al., 2019).

4.4.2 African cattle

African cattle form the third major clade of domestic cattle, and a hypothesised independent domestication of African aurochs (*Bos primigenius africanus*), in the Western Desert of Egypt, has long been suggested (Bradley et al., 1996; Hanotte et al., 2002a; Stock and Gifford-Gonzalez, 2013; Wendorf and Schild, 2005). However, recent analyses of modern genetic data (Decker et al., 2014; Pitt et al., 2018) and reanalysis of the Egyptian archaeological material (Brass, 2018) has largely discounted this possibility. Instead, the deep divergence between African and Eurasian taurine cattle most likely originated through admixture between Near Eastern domestic cattle and local African aurochs (Decker et al., 2014).

Taurine cattle are thought to have first arrived in North Africa from the Near East between 7,000–9,000 years ago (Brass, 2018; Edwards et al., 2017), and reached the forests of West Africa by around 4,000 years ago (MacDonald and MacDonald, 2000). The earliest evidence for zebu cattle in Africa, comes from Egyptian tomb paintings of the Twelfth Dynasty, in the second millennium BC (Hanotte et al., 2002b; Marshall, 2000). However, the main influx of zebu ancestry in Africa is thought to have arrived with the Swahili civilization, via the Horn of Africa, from the 7th century AD onwards (Mwai et al., 2015). This is supported by a prominent East–West cline of zebu admixture, with the highest proportions of zebu ancestry found in East Africa (Decker et al., 2014). Some relatively unadmixed populations of indigenous taurine cattle can still be found in West Africa (e.g. N'Dama and Lagune), where they preferentially survive over sanga and zebu breeds because of their tolerance to infection from disease carrying trypanosomes (Mwai et al., 2015; Smetko et al., 2015; Tijjani et al., 2019).

Presently, there are more than 150 recognised breeds of cattle in Africa, with three main groupings recognised: (i) humpless taurine cattle, found mostly in West Africa; (ii) sanga cattle, an ancient mixture of African taurine and zebu; and (iii) zenga, a more recent mixture of sanga and zebu. Despite the broad diversity of African cattle, all indigenous breeds have taurine mtDNA (mostly the T1 haplogroup), suggesting that both waves of zebu admixture were primarily driven by male zebu introgression (Edwards et al., 2017).

This is contrasted with Y-chromosome and autosomal markers, which show an abundance of zebu ancestry across Africa (Decker et al., 2014; Pérez-Pedal et al., 2018). Interestingly, the $Y3_B$ haplogroup, which is now found exclusively in West Africa, appears to be a relic of the first zebu migration out of South Asia (Pérez-Pedal et al., 2018).

The prevalence of zebu ancestry in Africa was further consolidated by the rinderpest panzootic of the 1890s, in which as much as 90% of African cattle were wiped out (Sunseri, 2018). Following the panzootic, zebu cattle were reintroduced from along the eastern coastline of Africa.

4.4.3 Arabian cattle

Based on rock art and faunal remains, taurine cattle may have arrived in Arabia as early as the seventh millennium BC (McCorriston and Martin, 2010), and likely reached the highlands of Yemen by the early sixth millennium BC (Boivin and Fuller, 2009; Edens and Wilkinson, 1998; McCorriston and Martin, 2010). The sporadic distribution of early cattle finds across Arabia suggest their dispersal did not occur in a single wave of migration across the peninsula (McCorriston and Martin, 2010). Based on circumstantial evidence, Arabian cattle are thought to have originated in the Near East rather than Africa, as they generally arrived alongside sheep and goats (Boivin and Fuller, 2009). Indeed, cattle might not have arrived in nearby Ethiopia until as late as ~2000 BC (Boivin and Fuller, 2009; Fattovich, 2005; Marshall, 2000).

The earliest evidence for zebu cattle in Arabia is less clear; however, they are present in coastal Eastern Arabia form around ~2000 BC, at the site of Tell Abraq (Uerpmann, 2001). The contemporaneous arrival of African crops in Gujarat, India, point to maritime trade links between Africa, Arabia and South Asia that likely facilitated the reciprocal movement of zebu cattle at this time (Boivin and Fuller, 2009). This maritime dispersal is thought to have occurred independently of the overland movement of zebu cattle (Boivin and Fuller, 2009), which was occurring in parallel across Iran and the Levant (Verdugo et al., 2019).

4.4.4 Socotra

The island of Socotra, in the Arabian Sea (east of Somalia and south of Yemen), was in the direct path of maritime trade between Africa, Arabia and South Asia (Cheung et al., 2006). The discovery in 2000 of a large corpus of multilingual inscriptions, in Hoq cave on the island's north-east coast, demonstrates the geographic extent of Socotra's role in maritime trade (Strauch, 2012). Between the 1st century BC and the 6th century AD, sailors left hundreds of inscriptions written in Indian, Arabian, Ethiopian, Greek, Palmyran and Kushan languages. Socotra is also mentioned in the *Periplus of the Erythraean Sea*, and was likely an important part of the regional incense trade; as frankincense (*Boswellia spp.*), myrrh (*Commiphora spp.*) and Dragon's blood (*Dracaena cinnabari*) resin are all produced on the island (Cheung et al., 2006). The incense was of great economic value, and used for religious, funerary, cosmetic, and medicinal purposes across South Asia, the Middle East, and the Mediterranean regions (Seland, 2014). The extent of Socotra's ancient trade links are noteworthy, as the island is currently considered one of the most isolated places in the world (Seland, 2014). Because Socotra lacks any naturally protected harbours, it was routinely cut off from the outside world by the southwestern summer monsoon (Botting, 1958; Scholte and De Geest, 2010).

Since the earliest scientific expeditions to the island in the 19th century, Socotra has been famed for the uniqueness of its endemic flora and fauna. It is also home to an unusual breed of humpless, dwarf shorthorn cattle. These cattle are unlike any other found in the neighbouring Horn of Africa or the Arabian Peninsula—except for cattle from the Dhofar Mountains, Oman (Cheung et al., 2006). Socotran cattle are an important part of the island's pastoral economy, producing a prized ghee (clarified butter) which was a well-known export product from the island (Botting, 1958; Forbes and Ogilvie-Grant, 1903; Miller, 1912; Wellsted, 1835).

The incongruous appearance of these humpless, dwarf shorthorn cattle has long been noted by European visitors to Socotra, who described them as resembling Alderney (Ravenstein, 1876; Wellsted, 1835) or Jersey cattle (Botting, 1958; Meinertzhagen,

1958)—distinctive breeds of dairy cattle from the Channel Islands. The leading hypothesis for their incongruous appearance is that Socotran cattle are a relict population of African taurine cattle, predating the arrival of zebu cattle; like those still found in West Africa (Cheung et al., 2006; Gwynne, 1967). To investigate the ancestry of these dwarf cattle, and their implications for history of domestic cattle in the region, we sequenced five modern individuals from two populations in Socotra and the Dhofar Mountains, Oman.

4.5 Materials and methods

4.5.1 Sample collection and sequencing

4.5.1.1 Low-coverage shotgun sequencing

Hair samples were acquired from Socotran (n=3) and Dhofari (n=2) cattle by Dr Kay van Damme (Senckenberg Institute, Frankfurt). DNA extraction and Illumina library preparation were performed by Dr Anna Linderholm (Texas A&M University, College Station). Single-end shotgun sequencing was performed at the Danish National High-throughput Sequencing Centre, Copenhagen.

4.5.1.2 Published data

Global reference panels for cattle, comprising Eurasian taurine, African taurine, zebu, and hybrid populations, were obtained from previously published studies. Datasets genotyped on both the Illumina BovineSNP50v1 (n=1,328) and Illumina BovineHD (n=649) arrays were downloaded from the WIDDE database (Sempéré et al., 2015).

Additional reference samples, genotyped on the Illumina BovineHD array, comprising cattle from various primitive European cattle breeds (n=144) (Upadhyay et al., 2017) and a whole-genome sequence from an ancient British aurochs (*Bos primigenius*) (Park et al., 2015) was downloaded from the Dryad database (Upadhyay et al., 2016). As were populations of indigenous Thai cattle, genotyped on the BovineSNP50v1 array (Wangkumhang et al., 2015).

Table 4.1 Summary of modern reference panels

Ref. Panel	SNP array	Populations	Samples	References
Widde50K	BovineSNP50v1	53	1,328	(Gautier et al., 2009) (Matukumalli et al., 2009) (Gautier et al., 2010)
Thai50K	BovineSNP50v1	4	28	(Wangkumhang et al., 2015)
WiddeHD	BovineHD	30	649	(Illumina, 2016)
PrimitiveHD	BovineHD	38	144	(Upadhyay et al., 2017)

4.5.2 Computational Pipeline

The computational pipeline built to perform all analyses was written using the Luigi (v. 2.8.3) (Spotify, 2019) pipeline framework, in the Python scripting language, with supporting code written in bash and R.

4.5.3 Alignment and variant calling

4.5.3.1 Alignment and pre-processing

The cattle reference genome, UMD3.1 (Zimin et al., 2009), was downloaded from Ensembl (Hubbard et al., 2002) and indexed with the `samtools faidx` command from *SAMtools* (v. 1.3.1) (Li et al., 2009) and the `CreateSequenceDictionary` command from *Picard* (v. 2.5.0) (Broad Institute, 2016).

Single-end libraries were trimmed of adapters using *Trimmomatic* (v. 0.36) (Bolger et al., 2014). Trimmed reads were aligned to the cattle reference genome using the *Burrows-Wheeler Aligner (BWA)* (v. 0.7.5) (Li and Durbin, 2009), with the `bwa mem` command (Li, 2013). Aligned records were sorted into coordinate order and converted into Binary Alignment Map (BAM) format using the `samtools sort` command. PCR duplicates were removed using the `MarkDuplicates` command from *Picard* (v. 2.5.0) (Broad Institute, 2016), by grouping duplicate reads using their 5' alignment, and only retaining

the read with the largest sum of base quality scores. Deduplicated BAM files were indexed using the `samtools index` command.

4.5.3.2 Variant calling

For the low-coverage shotgun libraries, pseudo-haploid genotypes were called for each sample individually. The pysam library (v. 0.15.0) (Heger and Jacobs, 2018), a Python wrapper around htslib (Li et al., 2009), was used to step through each position in each BAM file. For each target site, a random read was chosen from all reads that met the following conditions:

- Mapping quality ≥ 30
- Base quality ≥ 30
- Distance from read termini > 5
- Base belongs to one of the two ascertained alleles at that site

Alleles were ascertained from either the BovineSNP50v1 or BovineHD arrays. All pseudo-haploid calls which passed these conditions were converted into BED format using the *PLINK* (v. 1.90) software package (Chang et al., 2015).

Table 4.2 Sequencing depth and number of intersecting SNPs per array

Sample	Population	Mean DoC (≥ q30)	BovineSNP50	BovineHD
AL2002	SOC	0.1406	1,294	26,187
AL2003	SOC	0.0842	792	15,266
AL2004	SOC	1.6480	6,595	134,013
AL2007	OMN	0.1600	1,351	28,011
AL2008	OMN	0.0926	893	17,999

4.5.3.3 Combined datasets

The low-coverage shotgun dataset was intersected and merged with the BovineSNP50 and BovineHD reference panels using the `plink --merge-list` command. Any sites

which were polyallelic in the merged dataset were dropped from further processing. The cattle recombination map (Ma et al., 05-Nov-2015) was downloaded from the Dryad database (Ma et al., 2015), and used to annotate the combined datasets with the `plink --cm-map` command.

Table 4.3 Summary of combined modern datasets

Reference panel	Populations	Samples	SNPs	Genotyping Rate
Bovine50kMerged	59	1,361	52,870	0.97
BovineHDMerged	70	798	736,071	0.86

4.5.3.4 Genotyping rates

The reference panels were filtered for only high quality sites, using the `plink --make-bed` command and the following flags:

- Maximum missing per-sample 30% (`--mind 0.70`)
- Maximum missing per-variant 65% (`--geno 0.35`)
- Minimum minor allele frequency 1% (`--maf 0.01`)

4.5.4 Neighbour-joining tree

A neighbour-joining phylogenetic tree (Saitou and Nei, 1987) was constructed using the *APE* package (Paradis and Schliep, 2018) in R, from an identity-by-state (IBS) distance matrix made with the `plink --distance square 1-ibs` command in *PLINK*, and rooted on the branch leading to the lowland anoa.

4.5.5 Principal component analysis

Principal component analysis (PCA) was performed using *SmartPCA* (v. 13050) from the *EIGENSOFT* software package (Patterson et al., 2006), by first calculating the eigenvectors for the reference populations and then projecting the low coverage samples on top, using the least squares projection method. PCA scatter plots were plotted in R, using the ggplot2 (Wickham, 2016) package.

4.5.6 Model-based clustering

4.5.6.1 ADMIXTURE

To produce a list of SNPs with high discriminating power, sites under linkage disequilibrium were filtered out, using the `plink --indep-pairwise` and `plink --extract` workflow in *PLINK*, with flags specifying a window size of 50 SNPs, a step size of 10 SNPs and an $r^2 \leq 0.25$. Model-based clustering was performed on the resulting SNPs, for values of K ranging from 1 to 22 including cross-validation (CV), using the *ADMIXTURE* (v. 1.3.0) software package (Alexander et al., 2009). Admixture components were plotted in R, using the ggplot2 (Wickham, 2016) and tess3r (Caye et al., 2018) packages.

4.5.7 Admixture graphs

4.5.7.1 TreeMix

In order to test the topology suggested by our neighbour-joining tree, an admixture graph analyses was performed using the *TreeMix* (v. 1.13) software package (Pickrell and Pritchard, 2012), for values of M ranging from 1 to 10.

4.5.7.2 qpGraph

Admixture graphs were also fitted using *qpBrute* (Liu et al., 2019), a python package I developed for automatically fitting *qpGraph* (v. 6050) models (Patterson et al., 2012) (code available at https://github.com/ekirving/qpbrute). Given an outgroup with which to root the graph, a stepwise addition order algorithm is used for adding leaf nodes to a graph. At each step, insertion of a new node is tested at all branches of the graph, except the outgroup branch. Where a node cannot be inserted without producing f4 outliers (i.e., $|Z| \geq 3$) then all possible admixture combinations are also attempted. If a node cannot be inserted via either approach, that sub-graph is discarded. If the node is successfully inserted, the remaining nodes are recursively inserted into the graph. All possible starting node orders are attempted to ensure full coverage of the graph space.

For this analysis, only 8 cattle populations were used (i.e. GIR, NEL, JER, LAG, NDA, SHK, SOC and OMN), because the number of possible admixture graphs grows super-exponentially with each additional population. This process fitted 141,878 unique admixture graphs, of which 3,671 graphs left no f4 outliers (i.e., $|Z| < 3$). This set of graphs was filtered to retain only graphs in which SOC was admixed and JER was not admixed (based on results from the NJ tree, PCA, ADMIXTURE, and TreeMix analyses). This resulted in 156 unique admixture graphs.

4.6 Results

4.6.1 Principal component analysis

In the principal component analysis for the BovineHD array (see Figure 4.1), PC1 (19.2%) separates between the Eurasian and African taurine (*Bos taurus*) populations on the left, and the zebu (*Bos indicus*) populations on the right. The outgroup species—domestic yak (*Bos grunniens*), Gaur (*Bos gaurus*) and lowland anoa (*Bubalus depressicornis*)—cluster with the zebu cattle, due to the ascertainment bias of the BovineHD array. PC2 (2.5%) separates between Eurasian taurine in the middle to top, and African taurine at the bottom. PC2 is partially dominated by the intra-population variation of Hereford cattle (HFD), on the top left, due to the ascertainment bias of the BovineHD array. The projected Socotran (SOC) and Dhofari (OMN) populations cluster tightly together, nearest to Beefmaster cattle (BMA)—a commercial crossbreed of Eurasian taurine (Hereford and Shorthorn) and zebu (Brahman) (Decker et al., 2014).

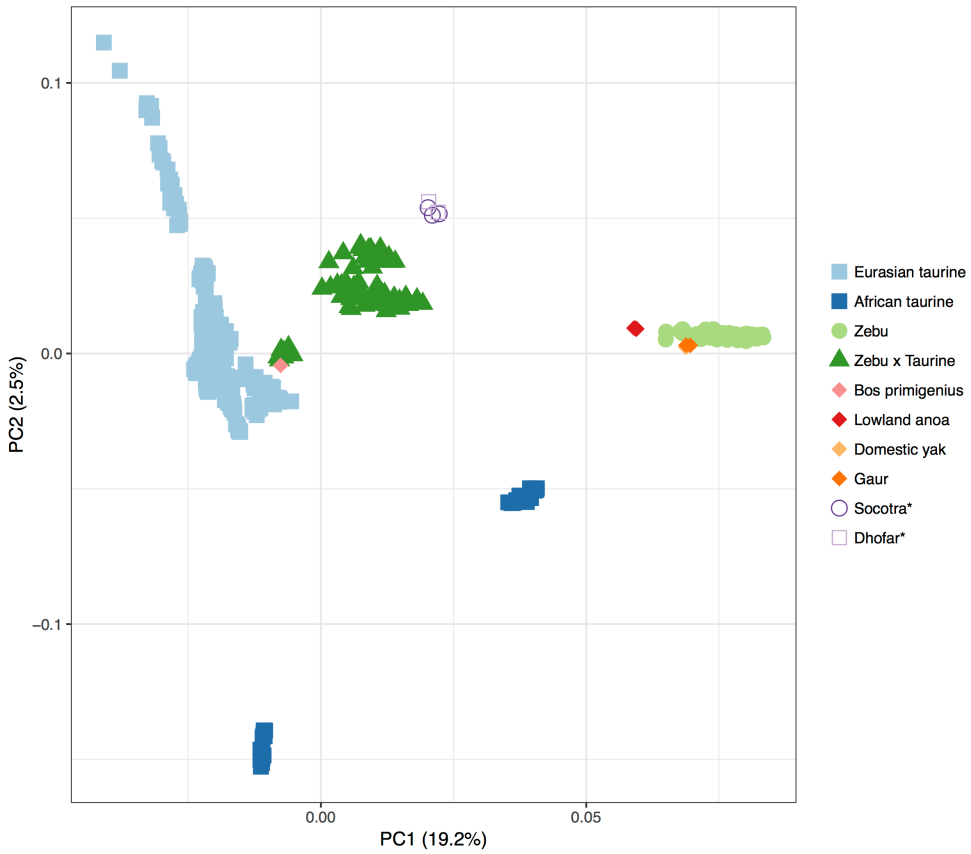


Figure 4.1 Principal Components Analysis (PC1 vs. PC2) of 798 cattle samples (from 70 populations) based on 736,071 SNPs ascertained using the BovineHD SNP array. All Socotran and Dhofari samples were projected.

In the principal component analysis for the BovineSNP50 array (see Figure 4.2), PC1 (10.8%) separates between Eurasian and African taurine (*Bos taurus*) populations on the right, and zebu (*Bos indicus*) populations on the left. PC2 (5%) separates between African taurine at the bottom, and their cline of admixture with zebu cattle (top left). The projected Socotran (SOC) and Dhofari (OMN) populations cluster together, closest to HFD and BMA.

Despite the low number of callable SNPs for four of the five SOC and OMN samples in the BovineSNP50 array, their tight clustering in the projected PCA space suggests that the low genotyping rate is not a biasing factor in this analysis.

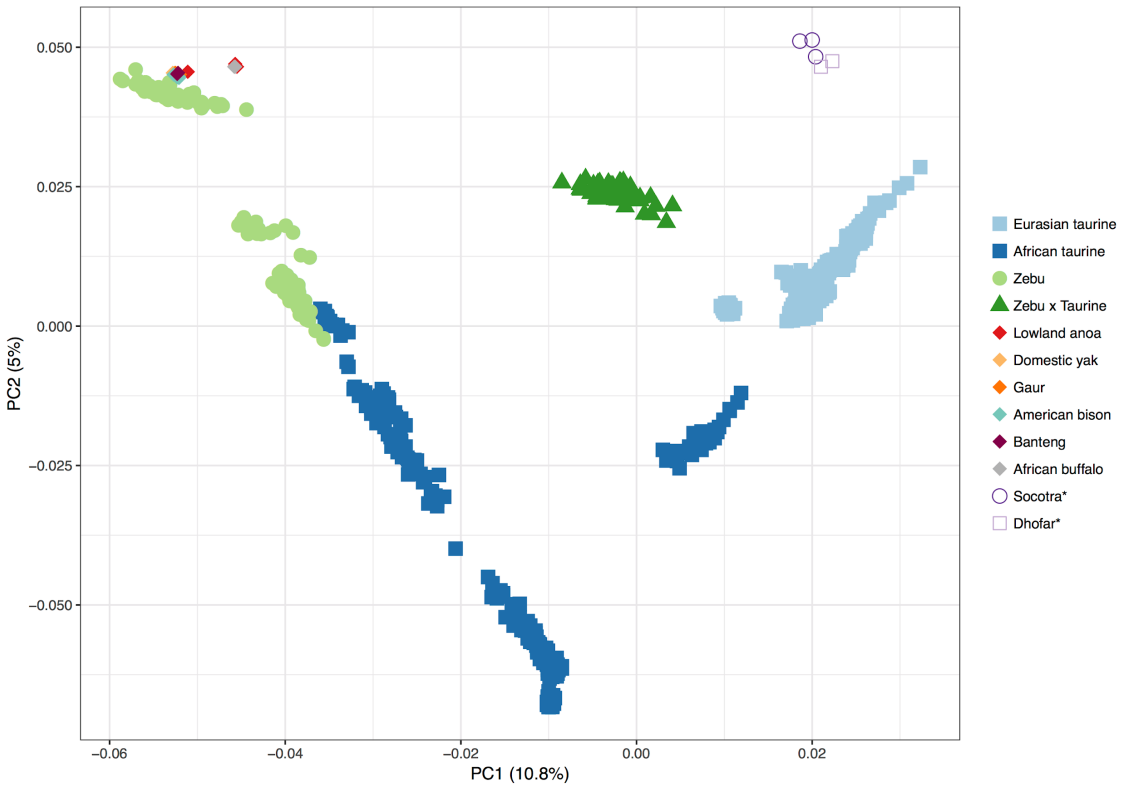


Figure 4.2 Principal Components Analysis (PC1 vs. PC2) of 1,361 cattle samples (from 59 populations) based on 52,870 SNPs ascertained using the BovineSNP50 SNP array. All Socotran and Dhofari samples were projected.

4.6.2 Neighbour-joining tree

In the neighbour-joining tree (see Figure 4.3), the Socotran (AL2002, AL2004 and AL2003) and Dhofari (AL2007 and AL2008) samples form a monophyletic clade to the exclusion of all other sampled populations; with the two Dhofari samples nested inside the variation of the three Socotran samples. The branch leading to the Socotran and Dhofari samples sits intermediate between Sheko (SHK), an Ethiopian native breed known to be an admixture of African taurine and zebu (Bahbahani et al., 2018), and Beefmaster (BMA), a commercial crossbreed of Eurasian taurine (Hereford and Shorthorn) and zebu (Brahman) (Decker et al., 2014). The long branch length leading to the Socotran and Dhofari samples is most likely an artefact of false homozygosity from the pseudo-haploid calling.

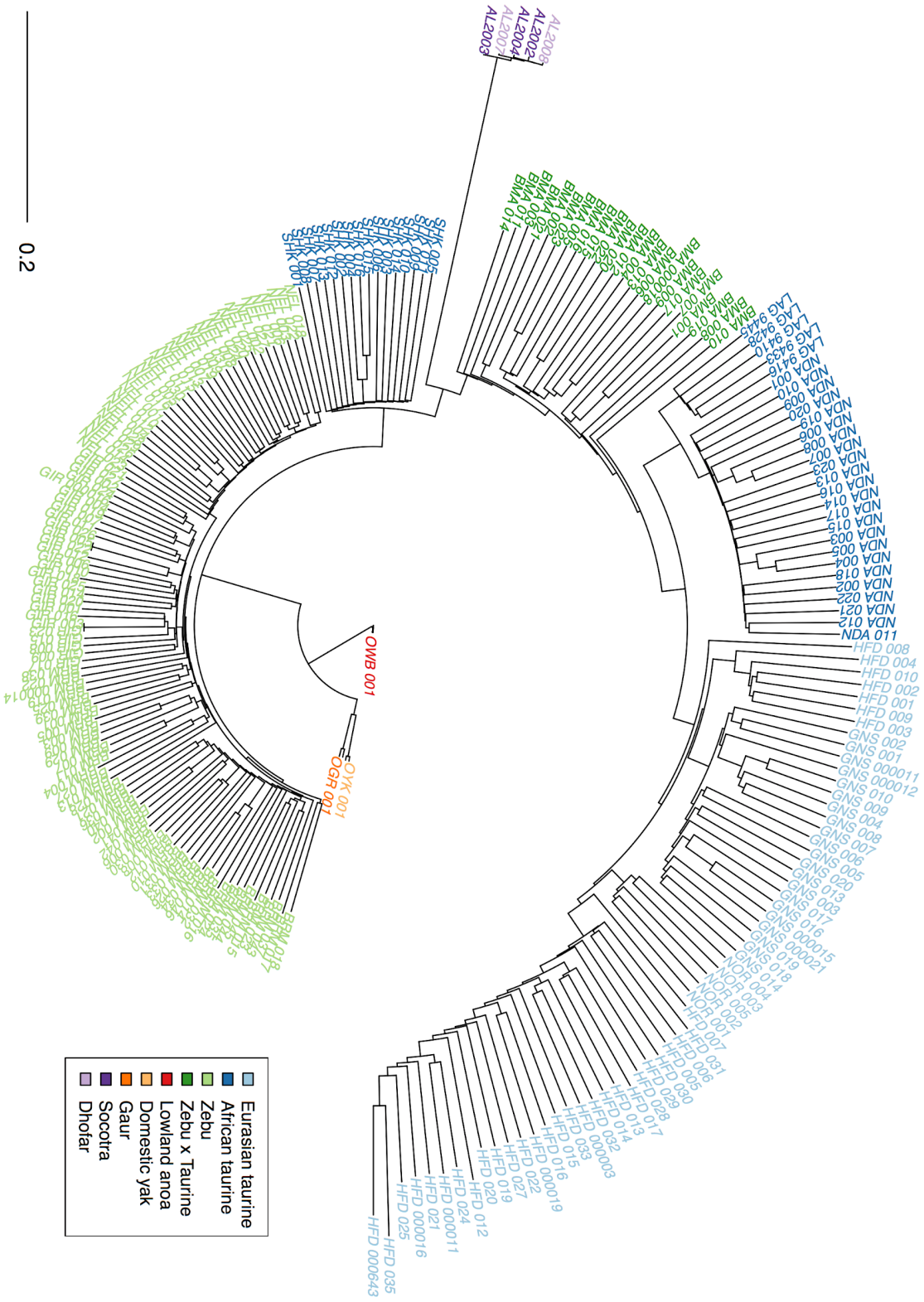


Figure 4.3 Neighbour Joining (NJ) tree built from Identity By State (IBS) matrix of 247 cattle samples (including 15 populations) based on 736,071 SNPs ascertained using the BovineHD SNP array.

4.6.3 Model-based clustering

4.6.3.1 ADMIXTURE

For the ADMIXTURE analysis (see Figure 4.5), all populations in the BovineHD array were used, except for OMN (due to insufficient SNPs) and the outgroup samples. The BovineSNP50 array was not used for model-based clustering, due to insufficient SNPs in both the SOC and OMN populations.

When K=3, the clustering recapitulates the main split between Eurasian taurine, African taurine, and zebu cattle. Both SOC and the British aurochs (AU) are modelled as an admixture of Eurasian taurine, African taurine, and zebu like ancestry clusters. SHK is modelled as an admixture of African taurine and zebu like ancestry clusters.

When K=12, the clustering shows the population structure within the non-primitive Eurasian taurine populations, and within the Eurasian taurine portion of the zebu-taurine cross populations. SOC and AU continue to be modelled as an admixture of Eurasian taurine, African taurine, and zebu like clusters, however, the Eurasian taurine cluster is further split into a Hereford (HFD) like cluster and a primitive taurine like cluster. Similarly, the zebu like cluster in SHK is split into separate Nelore (NEL) and Brahman (BRM) like clusters.

When K=20 (the best fitting K value by cross-validation error; see Figure 4.4), the clustering shows further structure within the non-primitive Eurasian taurine populations. SOC is modelled as a combination of HFD, SHK and primitive taurine like clusters.

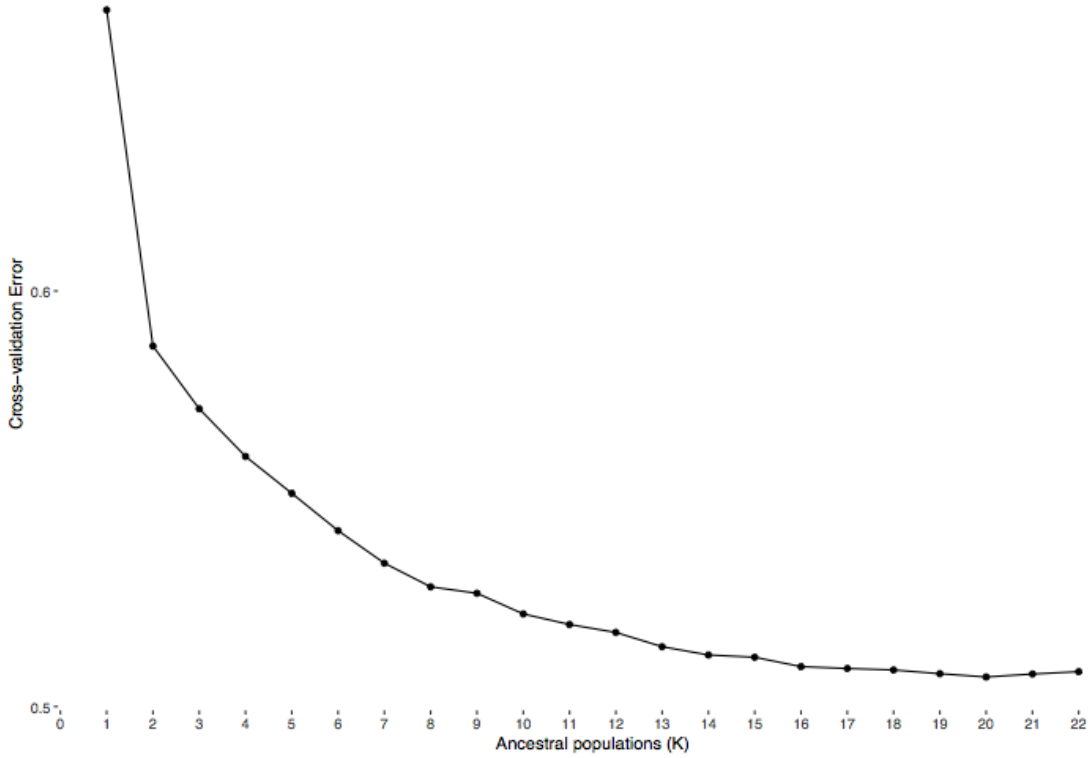


Figure 4.4 Cross-validation error for ADMIXTURE analysis, K between 1 and 22. K = 20 has the lowest cross-validation error.

To explore the geographic distribution of population ancestry, the ADMIXTURE clusters were interpolated onto a world map showing the sampling locations for each population (see Figure 4.6). At K = 6, the regional distinctiveness of the SOC and OMN populations are clearly demonstrated, where they cluster with HFD and JER despite being geographically very distant.

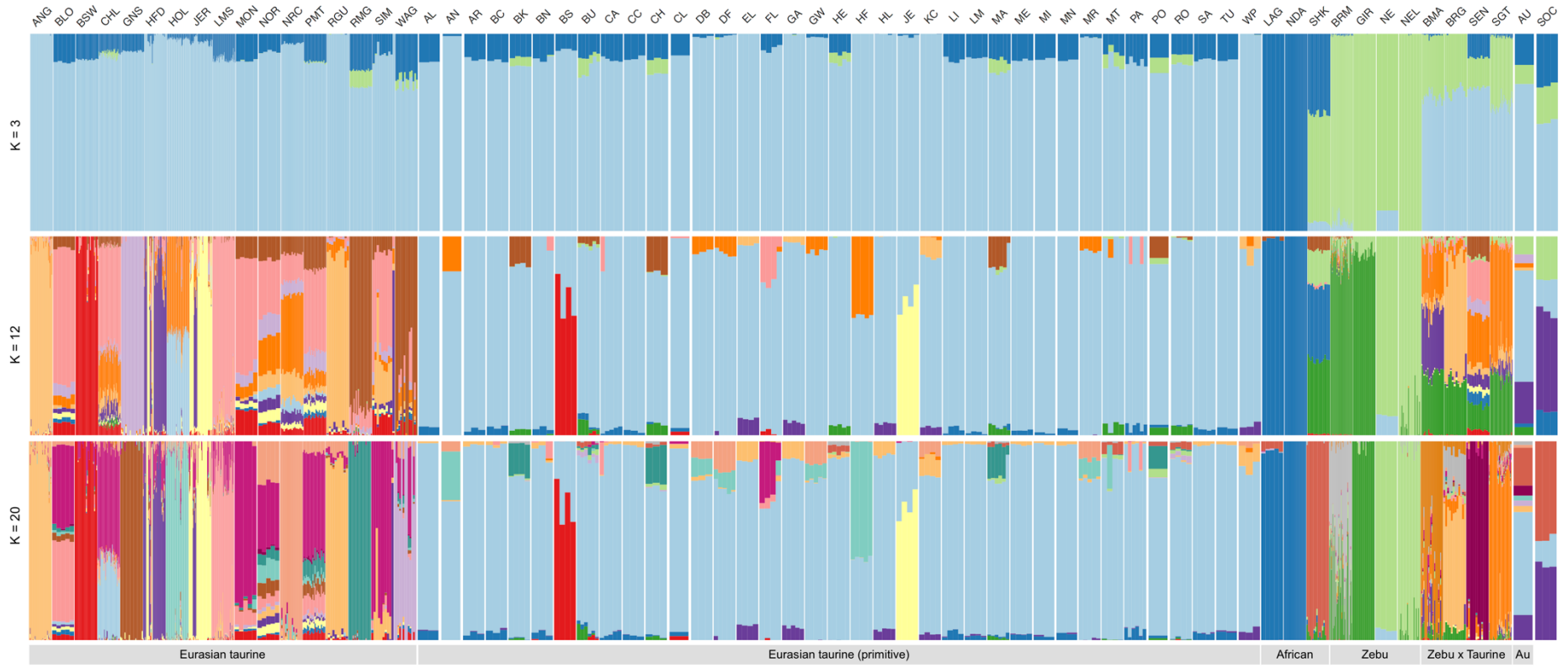


Figure 4.5 ADMIXTURE clusters, for K = 3, 12 and 20, from 788 cattle samples (including 67 populations) based on 236,614 SNPs ascertained using the BovineHD SNP array and pruned for LD. Population tiles are shows with a fixed width, regardless of sample size.

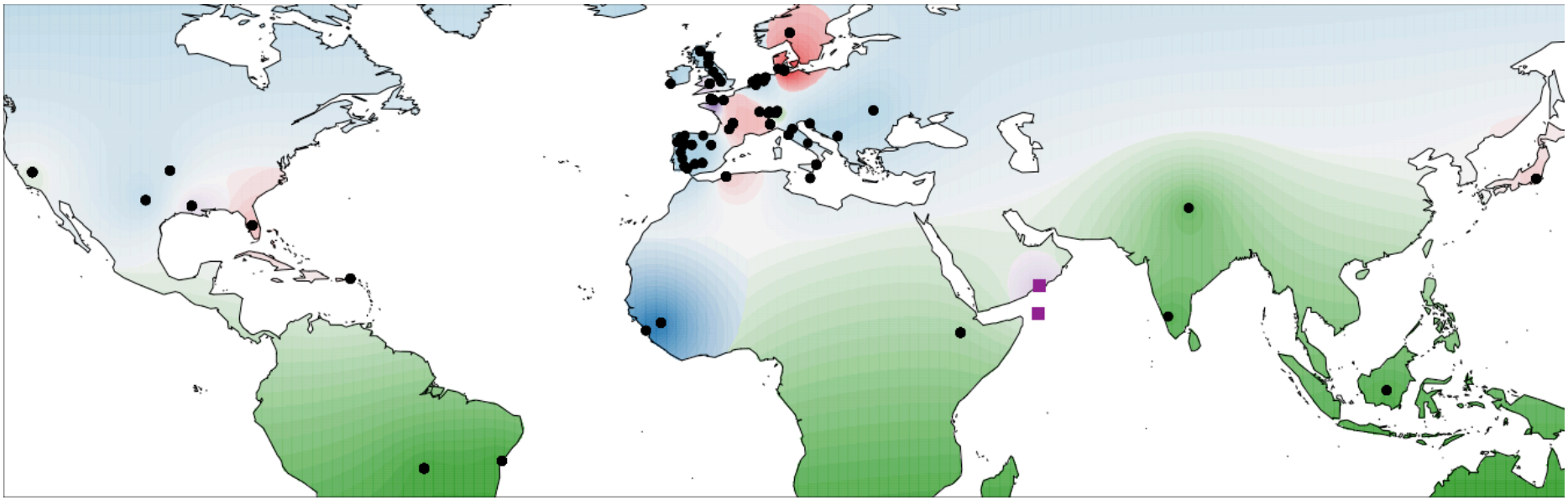


Figure 4.6 Interpolated map of ADMIXTURE clusters for K = 6, from 798 cattle samples (including 70 populations) based on 236,614 SNPs ascertained using the BovineHD SNP array and pruned for LD. Dots represent sampling locations for each population. SOC and OMN populations shown as square dots.

4.6.4 Admixture graphs

4.6.4.1 TreeMix

For the TreeMix analyses, a subset of representative populations from the BovineHD array were used. The BovineSNP50 array was not used for the TreeMix analyses, due to insufficient callable SNPs in the SOC and OMN populations.

When modelled with no migration branches (see Figure 4.7), the topology of the maximum likelihood tree recapitulates that of the neighbour-joining tree, such that the SOC and OMN populations form a monophyletic clade, and the branch leading to that clade sits intermediate between the admixed populations SHK (African taurine x zebu) and BMA (Eurasian taurine x zebu).

When modelled with one migration branch (see Figure 4.8), SHK repositions to join the zebu clade with a migration branch from the root of the African taurine populations NDA and LAG.

When modelled with two migration branches (see Figure 4.9), the SOC and OMN clade receives a migration branch from HFD, with a migration weight of almost 50%.

When modelled with three migration branches (see Figure 4.10), BMA receives a migration branch from BRM. This additional migration branch alters the balance of admixture into the SOC and OMN clade, causing the clade to reposition to be a sibling to HFD with a migration branch from near the split between taurine and zebu cattle.

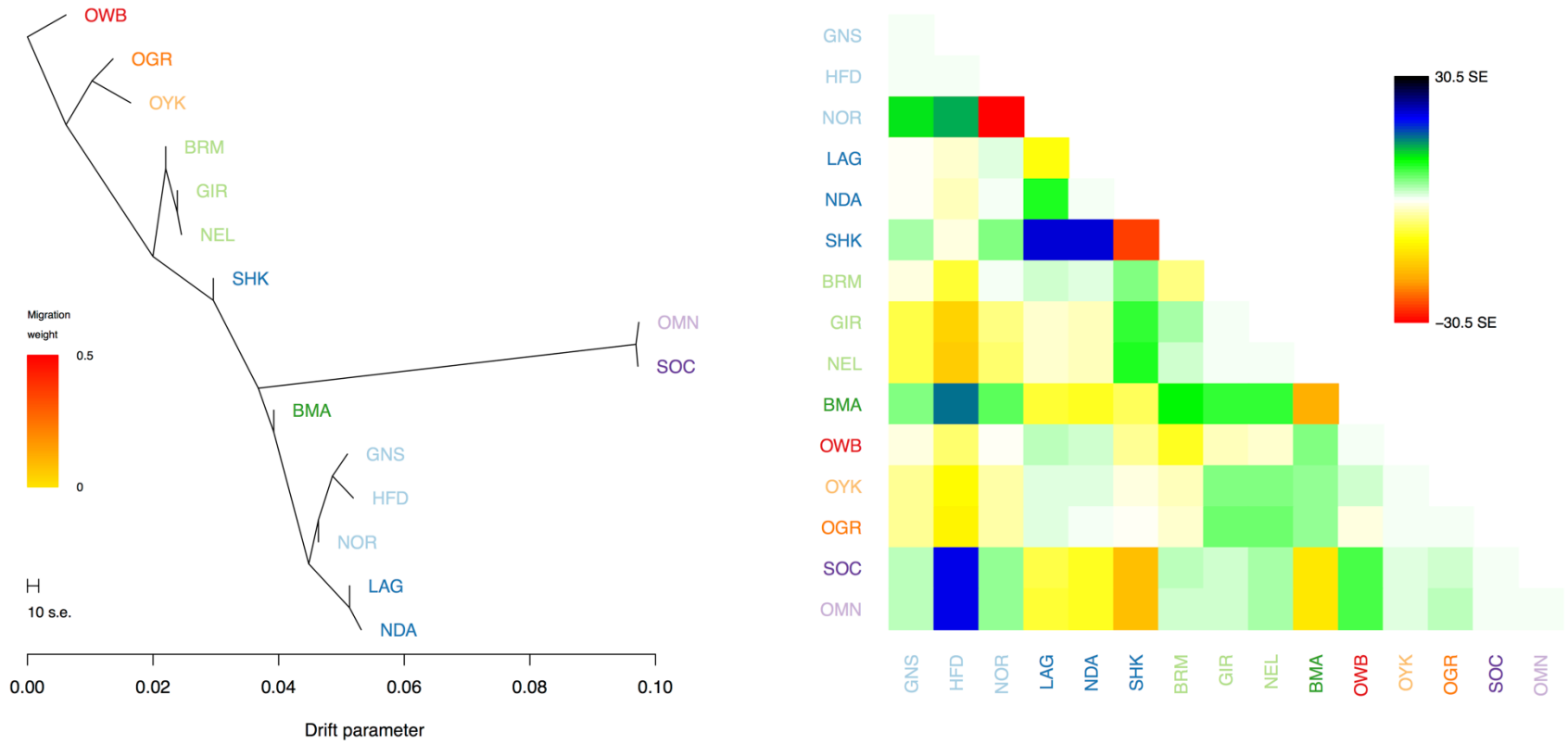


Figure 4.7 Admixture graph with no migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array).

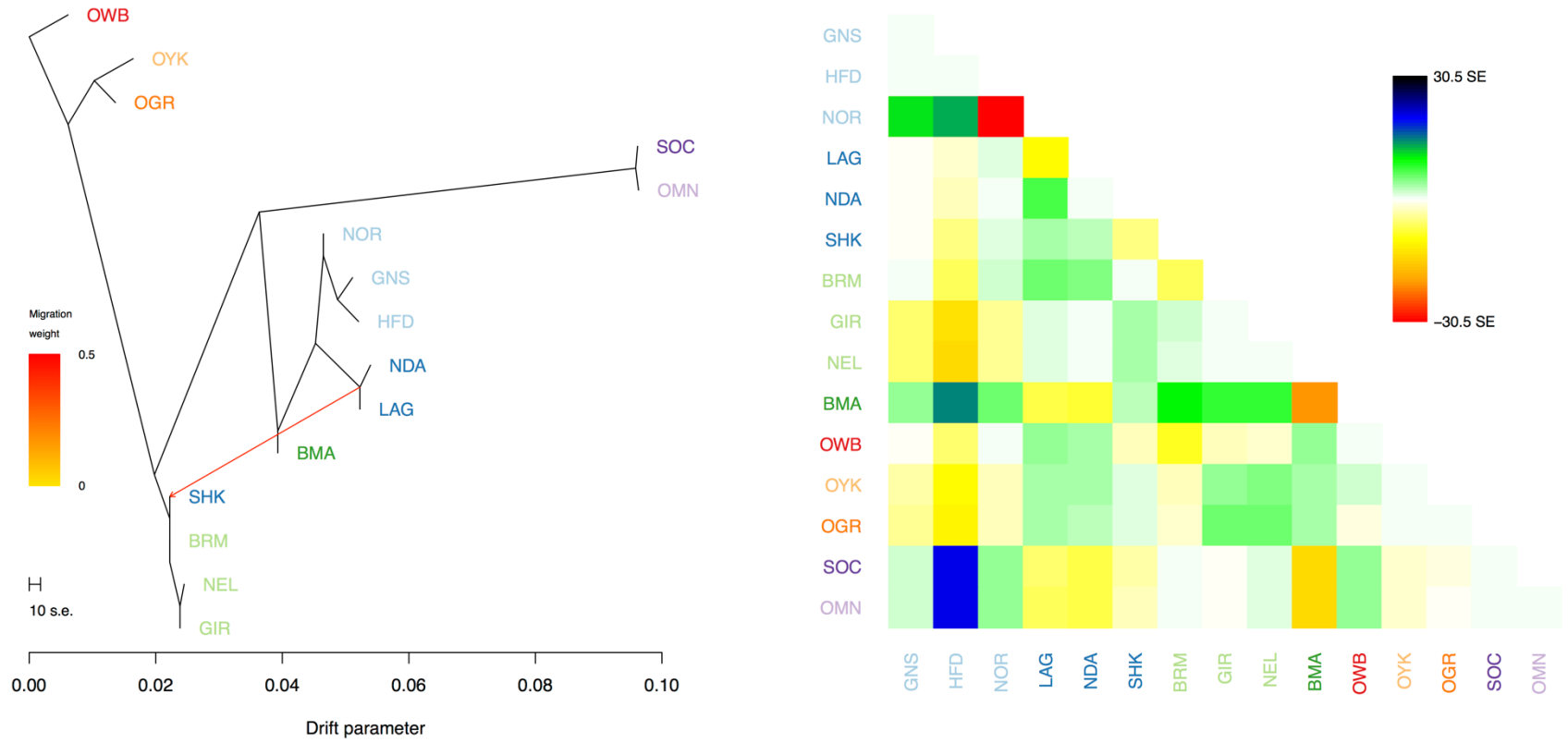


Figure 4.8 Admixture graph with one migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array).

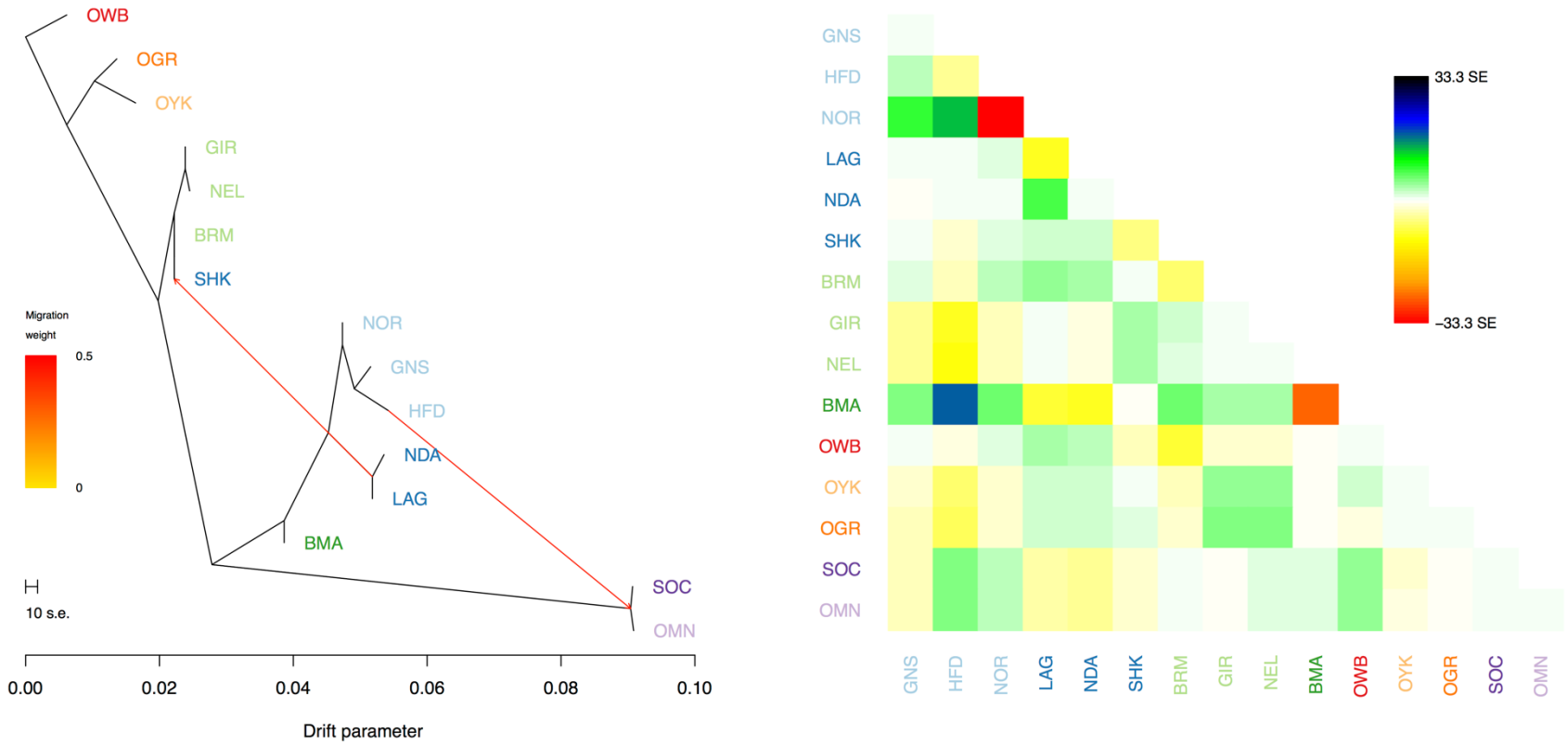


Figure 4.9 Admixture graph with two migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array).

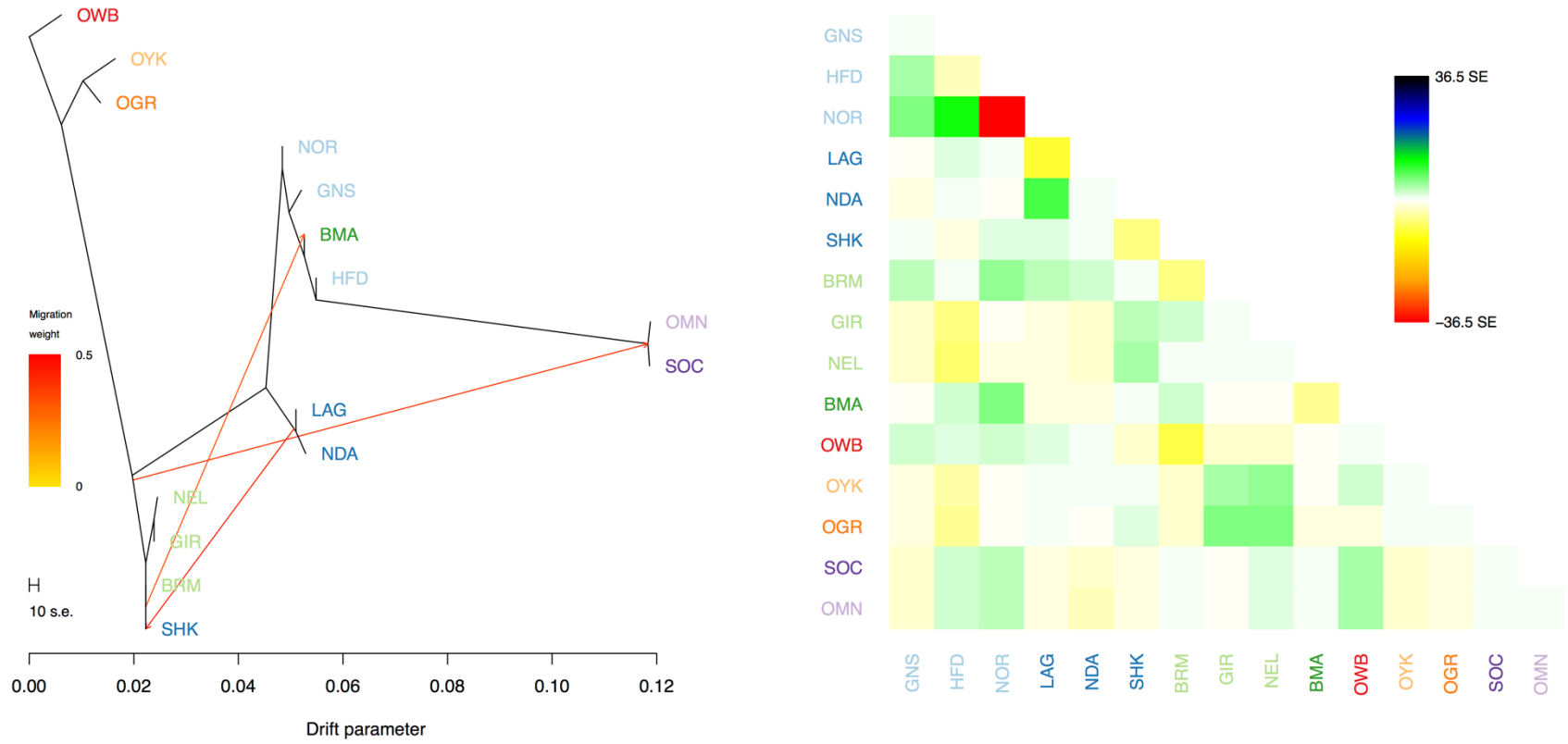


Figure 4.10 Admixture graph with three migration edges and matrix of residuals, expressed as the number of standard errors, inferred using TreeMix (based on BovineHD array).

4.6.4.2 qpGraph

For the qpGraph analyses, manual inspection of the 156 admixture graphs revealed that the graph topologies differed primarily in the arrangement of the zebu admixture into the African taurine clade. In all graphs, LAG and NDA were siblings and received admixture from the zebu clade. SHK was always a sibling to the LAG and NDA clade, and received a substantially larger portion of admixture from the zebu clade. The models differed in whether this admixture happened twice independently into the African lineages (see Figure 4.11), or consecutively into the SHK lineage (see Figure 4.12). The models also differed in the positioning of the migration branch from the zebu clade (e.g. before, between or after the split between GIR and NEL). In all graphs, SOC and OMN were siblings, and modelled as an admixture of the taurine lineage—after the split with Eurasian but basal to African—and a ghost lineage basal to the split between zebu and taurine.

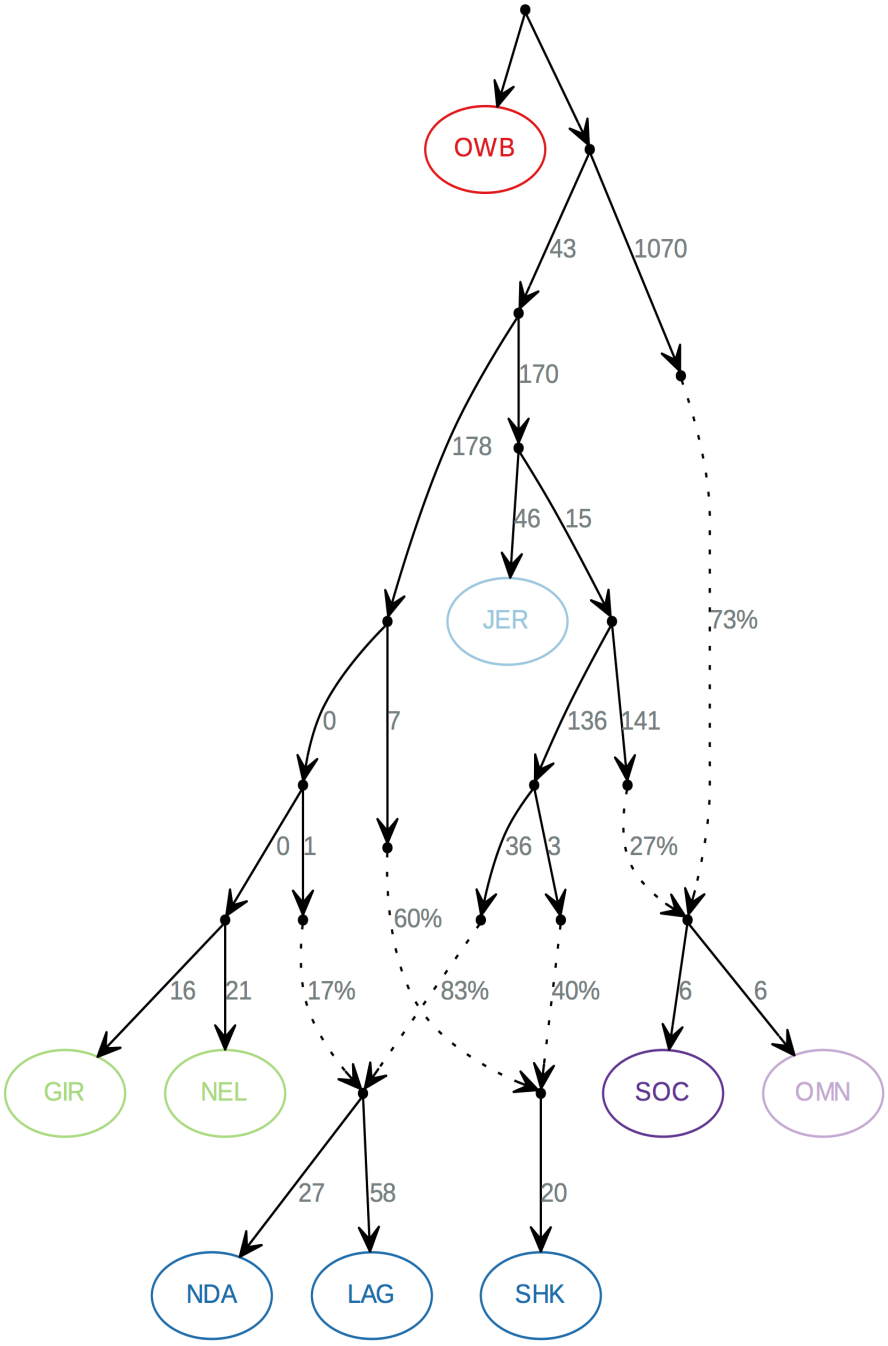


Figure 4.11 Admixture graph, inferred using qpGraph, with no f4 outliers (i.e., |Z| < 3), based on SNPs from the BovineHD array, showing two independent zebu introgressions into African taurine.

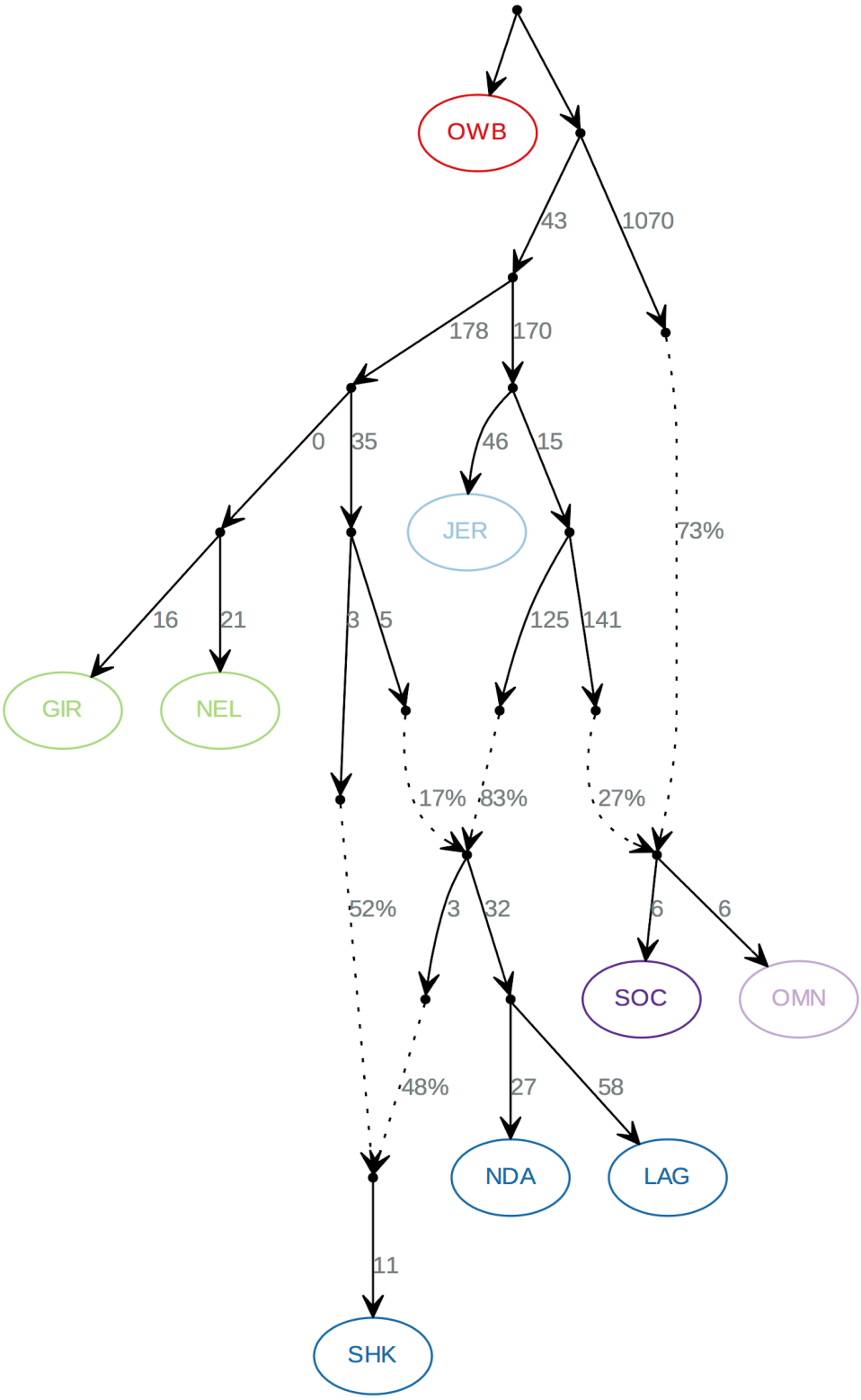


Figure 4.12 Admixture graph, inferred using qpGraph, with no f4 outliers (i.e., |Z| < 3), based on SNPs from the BovineHD array, showing two consecutive zebu introgressions into African taurine.

4.7 Discussion

The humpless, dwarf shorthorn cattle of Socotra are unlike any other cattle breed found in the neighbouring Horn of Africa or the Arabian Peninsula—except for cattle from the Dhofar Mountains, Oman (Cheung et al., 2006). Previous publications have hypothesised that Socotran cattle are a relict population of African taurine cattle, free from zebu introgression; similar to those still found in West Africa (Cheung et al., 2006; Gwynne, 1967).

From the neighbour-joining tree, the PCA, and the TreeMix analyses, it is clear that Socotran cattle are very closely related to Dhofari cattle. In the NJ tree, the nesting of the Dhofari samples within the clade of the Socotran samples suggests that they are drawn from the same geographically dispersed population. Socotra is frequently beset with droughts that affect the viability of local cattle populations, and as recently as 1999 a major drought killed off a large proportion of the cattle on the island (Cheung et al., 2006). Consequently, it seems likely that there is restocking or trade of cattle between Socotra and the Dhofar Mountains. It is already well established that cattle are regularly traded between the Horn of Africa and southern Arabia (Di Nardo et al., 2011), and therefore evidence of movement between Socotra and the Dhofar Mountains would not be surprising. Interestingly, analyses of human mitochondrial diversity in Socotra and southern Arabia genetically link the people of Socotra to those of the Dhofar Mountains (Al-Abri et al., 2012; Černý et al., 2011, 2009; Gandini et al., 2016).

The hypothesis that Socotran cattle are an unadmixed relict population of taurine cattle, free from the widespread introgression of zebu ancestry in the region, is not upheld by the analyses performed here. In the NJ tree, the branch leading to the SOC and OMN clade sits intermediate between SHK and BMA, two populations known to be admixtures of zebu and taurine cattle. This is supported by the BovineHD PCA, in which SOC and OMN are located between BMA and SHK in PC1. In PC2, they are located much closer to BMA than to SHK, suggesting that the origin of their taurine admixture is Eurasian, rather than African. In the BovineSNP50 PCA they are placed between BMA and HFD in PC1, and in PC2 they are far closer to zebu than African taurine.

It should be noted that the BovineSNP50 has a much stronger Eurasian taurine ascertainment bias than the BovineHD array (Utsunomiya et al., 2019), and therefore has less power to identify African taurine and zebu ancestry. Nevertheless, the BovineSNP50 analysis includes 9 populations of African cattle, all of which are clearly distinguishable from the Eurasian and zebu populations, and none are proximal to SOC or OMN.

In the ADMIXTURE analysis, K = 3 models SOC as an admixture of a Eurasian taurine like cluster, with smaller fractions of African taurine and zebu like clusters. At K = 12, SOC is modelled as an admixture of an HFD or JER like cluster, with varied fractions of other ancestries. These results are consistent with both the NJ tree and PCA analyses, and suggest that SOC and OMN are of primarily Eurasian taurine descent. Contrastingly, at K = 20 this changes, and SOC is modelled as an admixture of a primarily SHK like cluster, with a substantial HFD or JER like fraction. However, this African ancestry is not supported by any of the other analyses. Interestingly, the ADMIXTURE clustering for the ancient British aurochs (AU) shows a similar pattern to SOC, but with different proportions of the admixed clusters. In both cases, this may be due to small sample sizes, and a lack of other closely related populations.

The Eurasian taurine origin of SOC and OMN is further supported by the TreeMix analysis with three migration branches, in which they are a sibling to HFD with a large fraction of admixture from basal zebu, just inside the split with the taurine lineage. This unexpected topology is supported by the qpGraph analysis, where SOC and OMN are modelled as an admixture of the taurine lineage—after the split with JER but basal to African—and a ghost lineage basal to the split between zebu and taurine.

Consistent across all analyses are the findings that SOC and OMN have substantial zebu ancestry, despite their humpless phenotype. However, the source of their zebu ancestry is unclear, as is the origin of their taurine ancestry.

A recent study of 67 ancient cattle genomes from across the wider Near East revealed a sudden influx of zebu ancestry beginning around 4,200 years ago (Verdugo et al., 2019). In their PCA analysis, the resulting admixed cattle from the Near East were projected into a very similar PCA space to that of SOC and OMN. Furthermore, their qpGraph analyses suggest that the zebu introgression into these Near Eastern cattle is basal to all modern cattle (mirroring the qpGraph analyses in this study), and that modern zebu cattle have introgression from ancient taurine cattle. To resolve the nature of the zebu introgression into SOC and OMN, and to resolve the origin of their taurine ancestry, future work should consider rerunning these analyses with the ancient cattle from Verdugo et al. (2019) to see how those samples affect the inference. Additionally, larger reference panels of modern cattle are now available, including more African cattle populations genotyped on the BovineHD array, and future work should make use of these extra samples.

In summary, Socotran and Dhofari cattle form a single geographically dispersed population which is mostly Eurasian taurine in origin, with a unique pattern of admixture with zebu and African taurine cattle. However, the origin of these ancestries is unclear, and the Socotran and Dhofari samples need to be reanalysed alongside the recently published ancient Near Eastern cattle genomes to resolve the evolutionary history of this unique breed of cattle.

4.8 References

Al-Abri, A., Podgorná, E., Rose, J.I., Pereira, L., Mulligan, C.J., Silva, N.M., Bayoumi, R., Soares, P., Černý, V., 2012. Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. Am. J. Phys. Anthropol. 149, 291–298.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Bahbahani, H., Afana, A., Wragg, D., 2018. Genomic signatures of adaptive introgression and environmental adaptation in the Sheko cattle of southwest Ethiopia. PLoS One 13, e0202479.

Boivin, N., Fuller, D.Q., 2009. 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 22, 113–180.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J.-D., Thomas, M.G., 2012. Modern Taurine Cattle Descended from Small Number of Near-Eastern Founders. Mol. Biol. Evol. 29, 2101–2104.

Botting, D., 1958. The Oxford University Expedition to Socotra. Geogr. J. 124, 200–207.

Bradley, D.G., MacHugh, D.E., Cunningham, P., Loftus, R.T., 1996. Mitochondrial diversity and the origins of African and European cattle. Proc. Natl. Acad. Sci. U. S. A. 93, 5131–5135.

Brass, M., 2018. Early North African Cattle Domestication and Its Ecological Setting: A Reassessment. Journal of World Prehistory 31, 81–115.

Broad Institute, 2016. Picard Tools [WWW Document]. URL http://broadinstitute.github.io/picard/

Caye, K., Jay, F., Michel, O., François, O., 2018. Fast inference of individual admixture coefficients using geographic data. Ann. Appl. Stat. 12, 586–608.

Černý, V., Mulligan, C.J., Fernandes, V., Silva, N.M., Alshamali, F., Non, A., Harich, N., Cherni, L., Gaaied, E., Ammar, A.B., Al-Meeri, A., Pereira, L., 2011. Internal Diversification of Mitochondrial Haplogroup R0a Reveals Post-Last Glacial Maximum Demographic Expansions in South Arabia. Mol. Biol. Evol. 28, 71–78.

Černý, V., Pereira, L., Kujanová, M., Vašíková, A., Hájek, M., Morris, M., Mulligan, C.J., 2009. Out of Arabia—The settlement of Island Soqotra as revealed by mitochondrial and Y chromosome genetic diversity. Am. J. Phys. Anthropol. 138, 439–447.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

Chen, S., Lin, B.-Z., Baig, M., Mitra, B., Lopes, R.J., Santos, A.M., Magee, D.A., Azevedo, M., Tarroso, P., Sasazaki, S., Ostrowski, S., Mahgoub, O., Chaudhuri, T.K., Zhang, Y.-P., Costa, V., Royo, L.J., Goyache, F., Luikart, G., Boivin, N., Fuller, D.Q., Mannen, H., Bradley, D.G., Beja-Pereira, A., 2010. Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic. Mol. Biol. Evol. 27, 1–6.

Cheung, C., DeVantier, L., van Damme, K., 2006. Socotra: A Natural History of the Islands and Their People. Odyssey, Hong Kong.

Conolly, J., Colledge, S., Dobney, K., Vigne, J.-D., Peters, J., Stopp, B., Manning, K., Shennan, S.,

2011. 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.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Molina Alcalá, A., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Correia de Almeida Regitano, L., Yildiz, M.A., Heaton, M.P., Liu, W.-S., Lei, C.-Z., Reecy, J.M., Saif-Ur-Rehman, M., Schnabel, R.D., Taylor, J.F., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genet. 10, e1004254.

Di Nardo, A., Knowles, N.J., Paton, D.J., 2011. Combining livestock trade patterns with phylogenetics to help understand the spread of foot and mouth disease in sub-Saharan Africa, the Middle East and Southeast Asia. Rev. Sci. Tech. 30, 63–85.

Edens, C., Wilkinson, T.J., 1998. Southwest Arabia During the Holocene: Recent Archaeological Developments. Journal of World Prehistory 12, 55–119.

Edwards, C.J., MacHugh, D.E., Magee, D.A., 2017. How Modern and Ancient Genomic Analyses can Reveal Complex Domestic Histories Using Cattle as a Case Study, in: Hunter, D., Guarino, L., Spillane, C., McKeown, P.C. (Eds.), Routledge Handbook of Agricultural Biodiversity. Routledge, New York, NY : Routledge, 2017., pp. 30–44.

Fattovich, R., 2005. The archaeology of the Horn of Africa, in: Raunig, W., Wenig, S. (Eds.), Afrikas Horn: Akten Der Ersten Internationalen Littmann-Konferenz 2. Bis 5. Mai 2002 in München, Meroitica. Harrassowitz, Wiesbaden, pp. 3–29.

Forbes, H.O., Ogilvie-Grant, W.R., 1903. The natural history of Sokotra and Abd-el Kuri. Free Public Museums, HYoung, Liverpool.

Gandini, F., Achilli, A., Pala, M., Bodner, M., Brandini, S., Huber, G., Egyed, B., Ferretti, L., Gómez-Carballa, A., Salas, A., Scozzari, R., Cruciani, F., Coppa, A., Parson, W., Semino, O., Soares, P., Torroni, A., Richards, M.B., Olivieri, A., 2016. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. Sci. Rep. 6, 25472.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.-L., 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10, 550.

Gautier, M., Laloë, D., Moazami-Goudarzi, K., 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One 5. https://doi.org/10.1371/journal.pone.0013038

Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L.V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O.C., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C., Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P., Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.-Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M., Lau, L.P.L., Liu, G.E., Lynn, D.J., Panzitta, F., Dodds, K.G., 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528–532.

Gwynne, M.D., 1967. The possible origin of the dwarf cattle of Socotra. Geogr. J. 133, 39–42.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W., Rege, J.E.O., 2002a. African pastoralism: genetic imprints of origins and migrations. Science 296, 336–339.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W., Rege, J.E.O., 2002b. African Pastoralism: Genetic Imprints of Origins and Migrations. Science 296, 336–339.

Hansen, P.J., 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. Anim. Reprod. Sci. 82-83, 349–360.

Heger, A., Jacobs, K., 2018. pysam. Github. https://github.com/pysam-developers/pysam

Helmer, D., Gourichon, L., Monchot, H., Peters, J., Saña Seguí, M., 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism, in: Vigne, J.-D., Peters, J., Helmer, D. (Eds.), The First Steps of Animal Domestication: New Archaeozoological Approaches, International Council for Archaeozoology. Conference ; 9. Oxbow, Oxford, pp. 86–95.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., Clamp, M., 2002. The Ensembl genome database project. Nucleic Acids Res. 30, 38–41.

Illumina, 2016. Agrigenomics | Applying agricultural genomics to improve the food supply [WWW Document]. URL https://emea.illumina.com/areas-of-interest/agrigenomics.html

Larson, G., Burger, J., 2013. A population genetics view of animal domestication. Trends Genet. 29, 197–205.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, L., Bosse, M., Megens, H.-J., Frantz, L.A.F., Lee, Y.-L., Irving-Pease, E.K., Narayan, G., Groenen, M.A.M., Madsen, O., 2019. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. Nat. Commun. 10, 1992.

MacDonald, K.C., MacDonald, R.H., 2000. The origins and development of domesticated animals in Arid West Africa, in: Blench, R.M., MacDonald, K.C. (Eds.), The Origins and Development of African Livestock: Archaeology, Linguistics, and Ethnography. UCL Press, pp. 127–162.

Ma, L., O'Connell, J.R., Vanraden, P.M., Shen, B., Padhi, A., Sun, C., Bickhart, D.M., Cole, J.B., Null, D.J., Liu, G.E., Da, Y., Wiggans, G.R., 2015. Data from: Cattle sex-specific recombination and genetic control from a large pedigree analysis. https://doi.org/10.5061/dryad.q2q84

Ma, L., O'Connell, J.R., VanRaden, P.M., Shen, B., Padhi, A., Sun, C., Bickhart, D.M., Cole, J.B., Null, D.J., Liu, G.E., Da, Y., Wiggans, G.R., 05-Nov-2015. Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet. 11, e1005387.

Marshall, F., 2000. The origins and spread of domestic animals in East Africa, in: Blench, R.M., MacDonald, K.C. (Eds.), The Origins and Development of African Livestock: Archaeology,

Linguistics, and Ethnography. UCL Press, pp. 191–221.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, e5350.

McCorriston, J., Martin, L., 2010. Southern Arabia's Early Pastoral Population History: Some Recent Evidence, in: Petraglia, M.D., Rose, J.I. (Eds.), The Evolution of Human Populations in Arabia. Springer Netherlands, Dordrecht, pp. 237–250.

Meinertzhagen, R., 1958. Socotra Island Cattle. Geogr. J. 124, 587.

Miller, A.W., 1912. Products of the Island of Socotra. The Journal of the American Pharmaceutical Association (1912) 1, 874–877.

Mwai, O., Hanotte, O., Kwon, Y.-J., Cho, S., 2015. African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. Asian-australas. J. Anim. Sci. 28, 911–921.

Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 1, 3.

Park, S.D.E., Magee, D.A., McGettigan, P.A., Teasdale, M.D., Edwards, C.J., Lohan, A.J., Murphy, A., Braud, M., Donoghue, M.T., Liu, Y., Chamberlain, A.T., Rue-Albrecht, K., Schroeder, S., Spillane, C., Tai, S., Bradley, D.G., Sonstegard, T.S., Loftus, B.J., MacHugh, D.E., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. Genome Biol. 16, 234.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, e190.

Pérez-Pardal, L., Sánchez-Gracia, A., Álvarez, I., Traoré, A., Ferraz, J.B.S., Fernández, I., Costa, V., Chen, S., Tapio, M., Cantet, R.J.C., Patel, A., Meadow, R.H., Marshall, F.B., Beja-Pereira, A., Goyache, F., 2018. Legacies of domestication, trade and herder mobility shape extant male zebu cattle diversity in South Asia and Africa. Sci. Rep. 8, 18027.

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8, e1002967.

Pitt, D., Sevane, N., Nicolazzi, E.L., MacHugh, D.E., Park, S.D.E., Colli, L., Martinez, R., Bruford, M.W., Orozco-terWengel, P., 2018. Domestication of cattle: Two or three events? Evol. Appl. 18, R157.

Ravenstein, E.G., 1876. Sokotra Island. Geogr. Mag. 3, 119–124.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Scheu, A., Powell, A., Bollongino, R., Vigne, J.-D., Tresset, A., Çakırlar, C., Benecke, N., Burger, J., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet. 16, 54.

Scholte, P., De Geest, P., 2010. The climate of Socotra Island (Yemen): A first-time assessment of the timing of the monsoon wind reversal and its influence on precipitation and vegetation patterns. J. Arid Environ. 74, 1507–1515.

Seland, E.H., 2014. Archaeology of Trade in the Western Indian Ocean, 300 BC–AD 700. Journal of Archaeological Research 22, 367–402.

Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., Flori, L., 2015. WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. BMC Genomics 16, 940.

Sharifi, A., Pourmand, A., Canuel, E.A., Ferer-Tyler, E., Peterson, L.C., Aichner, B., Feakins, S.J., Daryaee, T., Djamali, M., Beni, A.N., Lahijani, H.A.K., Swart, P.K., 2015. Abrupt climate variability since the last deglaciation based on a high-resolution, multi-proxy peat record from NW Iran: The hand that rocked the Cradle of Civilization? Quat. Sci. Rev. 123, 215–230.

Smetko, A., Soudre, A., Silbermayr, K., Müller, S., Brem, G., Hanotte, O., Boettcher, P.J., Stella, A., Mészáros, G., Wurzinger, M., Curik, I., Müller, M., Burgstaller, J., Sölkner, J., 2015. Trypanosomosis: potential driver of selection in African cattle. Front. Genet. 6, 137.

Spotify, 2019. Luigi. Github. https://github.com/spotify/luigi

Stock, F., Gifford-Gonzalez, D., 2013. Genetics and African Cattle Domestication. African Archaeological Review 30, 51–72.

Strauch, I., 2012. Foreign Sailors on Socotra : the Inscriptions and drawings from the cave Hoq. Vergleichende Studien zu Antike und Orient ; Band 3.

Sunseri, T. (Ed.), 2018. The African Rinderpest Panzootic, 1888–1897, in: Oxford Research Encyclopedia of African History. Oxford University Press.

Tijjani, A., Utsunomiya, Y.T., Ezekwe, A.G., Nashiru, O., Hanotte, O., 2019. Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. Front. Genet. 10, 442.

Uerpmann, M., 2001. Remarks on the animal economy of Tell Abraq (Emirates of Sharjah and Umm al-Qaywayn, UAE). Proceedings of the Seminar for Arabian Studies 31, 227–233.

Upadhyay, M.R., Chen, W., Lenstra, J.A., Goderie, C.R.J., MacHugh, D.E., Park, S.D.E., Magee, D.A., Matassino, D., Ciani, F., Megens, H.-J., van Arendonk, J.A.M., Groenen, M.A.M., European Cattle Genetic Diversity Consortium, Crooijmans, R.P.M.A., 2016. Data from: Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. https://doi.org/10.5061/dryad.f2d1q

Upadhyay, M.R., Chen, W., Lenstra, J.A., Goderie, C.R.J., MacHugh, D.E., Park, S.D.E., Magee, D.A., Matassino, D., Ciani, F., Megens, H.-J., van Arendonk, J.A.M., Groenen, M.A.M., European Cattle Genetic Diversity Consortium, RPMA Crooijmans, 2017. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. Heredity 118, 169–176.

Utsunomiya, Y.T., Milanesi, M., Fortes, M.R.S., Porto-Neto, L.R., Utsunomiya, A.T.H., Silva, M.V.G.B., Garcia, J.F., Ajmone-Marsan, P., 2019. Genomic clues of the evolutionary history of Bos indicus cattle. Anim. Genet. https://doi.org/10.1111/age.12836

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., Hesse, P., Fulton, D., Sauer, E.W., Mohaseb, F.A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Vahidi, S.M.F., MacHugh, D.E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B.S., Mashkour, M., Orton, D., Horwitz, L.K., Teasdale, M.D., Bradley, D.G., 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 365, 173–176.

Wangkumhang, P., Wilantho, A., Shaw, P.J., Flori, L., Moazami-Goudarzi, K., Gautier, M., Duangjinda, M., Assawamakin, A., Tongsima, S., 2015. Genetic analysis of Thai cattle reveals

a Southeast Asian indicine ancestry. PeerJ 3, e1318.

Wellsted, J.R., 1835. Memoir on the Island of Socotra. Journal of the Royal Geographical Society of London 5, 129.

Wendorf, F., Schild, R., 2005. Are the early holocene cattle in the eastern sahara domestic or wild? Evol. Anthropol., BAR International Series 496 3, 118–128.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

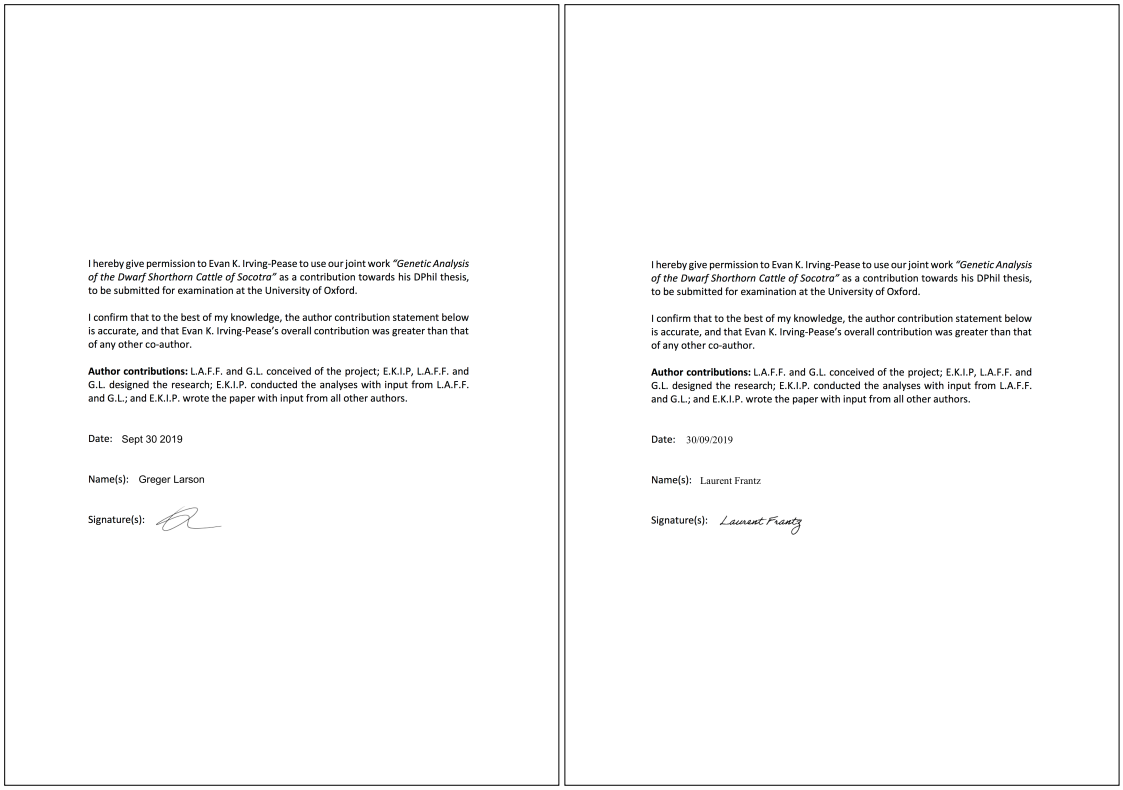
Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A., Salzberg, S.L., 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10, R42.

4.9 Acknowledgments and Notes

Acknowledgments: We thank K. van Damme for providing the hair samples from Socotra and Dhofar, A. Linderholm for performing the DNA extraction and Illumina library preparation, and A. Beja-Pereira for helpful comments about cattle in the region. We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility for providing computing time.

Author contributions: L.A.F.F. and G.L. conceived of the project; E.K.I.P, L.A.F.F. and G.L. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F. and G.L.; and E.K.I.P. wrote the paper with input from all other authors.

4.10 Permission from Co-authors



I hereby give permission to Evan K. Irving-Pease to use our joint work *"Genetic Analysis of the Dwarf Shorthorn Cattle of Socotra"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions: L.A.F.F. and G.L. conceived of the project; E.K.I.P, L.A.F.F. and G.L. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F. and G.L.; and E.K.I.P. wrote the paper with input from all other authors.

Date: Sept 30 2019

Name(s): Greger Larson

Signature(s):

I hereby give permission to Evan K. Irving-Pease to use our joint work *"Genetic Analysis of the Dwarf Shorthorn Cattle of Socotra"* as a contribution towards his DPhil thesis, to be submitted for examination at the University of Oxford.

I confirm that to the best of my knowledge, the author contribution statement below is accurate, and that Evan K. Irving-Pease's overall contribution was greater than that of any other co-author.

Author contributions: L.A.F.F. and G.L. conceived of the project; E.K.I.P, L.A.F.F. and G.L. designed the research; E.K.I.P. conducted the analyses with input from L.A.F.F. and G.L.; and E.K.I.P. wrote the paper with input from all other authors.

Date: 30/09/2019

Name(s): Laurent Frantz

Signature(s): Laurent Frantz

5 Conclusions

The domestication of plants and animals was one of the most important transformations in human prehistory. Much of the literature on animal domestication has focussed on livestock species that dominate modern global food economies—i.e., cattle, pigs, sheep, and goats. These species were among the earliest animals to be domesticated, and their significance throughout history has given rise to a predominance of teleological interpretations regarding causality. Domestication has often been posited as the result of human intentionality—a deliberate action to solve a problem, or to respond to a novel incentive. These directed models of domestication often characterise the process not solely by genetic and phenotypic characteristics, but also by population processes such as a strong bottlenecks, reproductive isolation, and directed breeding (Marshall et al., 2014). With the development of high-throughput, or so-called "next generation sequencing" (NGS) platforms, in the mid-2000s (Bentley et al., 2008; Margulies et al., 2005), a wealth of new sequence has data become available to test these assumptions.

5.1 Characterising modern patterns of diversity

Large studies of genome-wide modern DNA have characterised the global patterns of diversity and admixture seen in modern domestic populations—e.g., cattle (Decker et al., 2014; Gibbs et al., 2009), sheep (Kijas et al., 2012), goats (Brito et al., 2017; Wang et al., 2016), pigs (Ai et al., 2013), horses (McCue et al., 2012; Petersen et al., 2013; Schaefer et al., 2017), chickens (Muir et al., 2008; Stainton et al., 2017), and dogs (Shannon et al., 2015; vonHoldt et al., 2010). These large modern datasets benefit from the relative ease of sampling and low cost of data generation, however, modern populations are often a poor proxy for past diversity.

In Chapter 4 of this thesis, modern genome-wide sequence data was analysed from a rare breed of humpless, dwarf shorthorn cattle found on the remote island of Socotra. The leading hypothesis for their incongruous appearance is that Socotran cattle are a relict population of African taurine cattle, free from zebu introgression; like those still found in West Africa (Cheung et al., 2006; Gwynne, 1967). The results showed that these

regionally incongruous cattle come from a single geographically dispersed population, found on the island of Socotra and in the Dhofar Mountains, Oman. In comparison to a global reference panel of >1,000 cattle, results showed that this population is unlike any other cattle sampled in the region. Admixture analyses revealed that they are predominantly Eurasian taurine in ancestry, and contrary to expectations, have only partial African taurine ancestry. Furthermore, preliminary evidence suggests possible admixture with a basal clade of zebu cattle, previously unseen in modern populations.

Native cattle breeds are being rapidly lost around the world, as globalised food production replaces traditional pastoral lifestyles (Mwai et al., 2015). Identifying the extant patterns of diversity in regional landraces is an important first step towards preserving their unique genetic heritage.

5.2 Introgression and population turnover

The results presented in this thesis highlights how modern domestic animal genomes are often complex palimpsests—containing traces of many different ancestral populations—due to successive waves of gene flow over thousands of years. Indeed, recent ancient DNA studies have shown that reproductive isolation was much less common than previously thought, revealing complex and varied patterns of introgression among domestic pigs (Frantz et al., 2019), goats (Daly et al., 2018), cattle (Verdugo et al., 2019), horses (Fages et al., 2019), cats (Ottoni et al., 2017), and many other species. As people spread outwards from the major centres of domestication, the domestic animals that accompanied them frequently interbred with wild populations encountered along their routes of dispersal. In many cases, these initial domestic populations were later replaced by subsequent waves of migration, or their ancestry substantially diluted by protracted introgression from sympatric wild populations (e.g. Frantz et al., 2019).

In Chapter 3 of this thesis, the first whole-genome sequence data from ancient North American dogs was analysed. The results showed that pre-contact dogs were not independently domesticated in the Americas, although they do contain introgression

from North American wolves. Dogs likely arrived in the Americas accompanying humans during the peopling of the continent, from a source population of Arctic dogs originating in Eastern Siberia. Following the arrival of European colonists, pre-contact dogs were quickly replaced by European dogs, and have left almost no genetic legacy in modern populations. Indeed, no detectable traces of pre-contact ancestry was found in >100 village dogs from South America or in putatively "native" breeds (e.g., hairless dogs and Catahoulas). Surprisingly, the most closely related living organisms to these pre-contact dogs is not a dog, but is instead an 8,000-year-old contagious cancer clone, known as Canine Transmissible Venereal Tumour (CTVT). These findings were published in the journal *Science*.

5.3 Ancient DNA time-series

These results demonstrate how ancient DNA can resolve big questions about the origins of domestic populations, and reveal population turnovers or replacements that are otherwise undetectable with modern DNA alone. As the field of palaeogenomics progresses, and the sizes of the aDNA datasets increase, broad-brush studies of continental scale population movements are on the wane, and the field is beginning to address more complex and fine-grained questions, such as the strength and timing of selection for polygenic traits. Recent publications on the evolutionary history of goats (Daly et al., 2018), horses (Fages et al., 2019) and cattle (Verdugo et al., 2019) have, for the first time, produced time-series datasets of sufficient depth and density that we can now begin to directly observe fine scale evolutionary processes like selection.

Understanding the process of natural selection is one of the defining problems of evolutionary biology, and many methods have been developed to infer selection from the indirect patterns that they leave in modern genomes. By analysing wild and domestic animal genomes for segregating signatures of selection, previous studies have identified hundreds of soft sweep loci, putatively associated with domestication. These analyses have been formative in establishing an empirical model of animal domestication that is highly polygenic; involving subtle shifts in the allele frequencies of variants at hundreds of loci, each with low effect size. Very few fixed derived genetic

variants have been identified in domestic populations, and none have been associated with important functional effects. In short, empirical data has shown that animal domestication did not occur through strong selection on common variants with large effect sizes, and so-called 'domestication genes' do not exist.

Genome-wide association studies (GWAS) have also been pivotal in revealing the highly polygenic nature of quantitative traits underlying domestic phenotypes. The development of cost-effective medium and high density SNP arrays, coupled with statistical imputation techniques, have made GWAS for quantitative traits in domestic livestock increasingly commonplace. For the first time, we now have both large databases of GWAS trait associations and ancient DNA datasets of sufficient resolution to track the allele frequency changes in those variants.

In Chapter 2 of this thesis, an ancient DNA time-series analysis of selection for quantitative traits in cattle and horses was performed. Using a Bayesian modelling approach, the allele frequency trajectories of thousands of GWAS variants was reconstructed, along with estimates of the age of the allele under selection, and the selection coefficients for hundreds of polygenic traits. Technical issues with the MCMC model convergence confounded much of the analyses, however, detailed recommendations for improvements to the computational modelling software were made.

5.4 Future perspectives

5.4.1 Larger sample sizes

A decade ago, ancient DNA studies of animal domestication were focused on the retrieval of a single gene locus from few ancient samples. Now, studies involving genome-wide data from tens (Daly et al., 2018; Frantz et al., 2019; Verdugo et al., 2019) or even hundreds (Fages et al., 2019) of ancient domestic animals are being published, and in the future larger studies with thousands of samples are likely to become common.

These increases in sample size have been driven by both technological and methodological improvements. For example, the cost of sequencing has fallen dramatically, as each new generation of sequencing platform has increased overall throughput. Endogenous aDNA yields have also increased, because of an improved understanding of DNA preservation in different skeletal elements—especially petrous bones and tooth cementum (Hansen et al., 2017). Additionally, improvements in extraction protocols (e.g. Rohland et al., 2018) and single-stranded DNA library preparation (e.g. Gansauge et al., 2017) are pushing the boundaries of aDNA recovery.

5.4.2 Computational advancements

As both ancient and modern datasets grow in size, they are creating computational bottlenecks which prevents their information content from being exploited to its full potential. For example, in Chapter 2 of this thesis, 250,000 CPU hours (28.5 years) were expended on computing the MCMC chains, and only a fraction of the total GWAS associations for cattle were modelled.

Most computational methods designed for use directly on whole-genome data can only handle small sample sizes. Methods that can accommodate large sample sizes often sacrifice precision and accuracy by using computational approximations, or by operating on summary statistics. The potential of ancestral recombination graphs (ARGs) (Griffiths and Marjoram, 1996) to fully characterise the genealogical history of a set of samples, has long been recognised. Until recently, the most accurate method for inferring ARGs could only scale to a few dozen samples (Rasmussen et al., 2014). This has now changed dramatically with the publication of two new methods, which can accurately infer ARGs for thousands (Speidel et al., 2019) and hundreds of thousands (Kelleher et al., 2019) of samples.

Currently, no ARG inference methods are designed to specifically accommodate ancient DNA, however, these features are under active development. Once these new methods become available, the potential for palaeogenomics is hard to overstate. Many ancient DNA samples have both direct radiocarbon dates and GPS coordinates, which would

allow inference of population genetic parameters across both space and time. The ability of infer demography, migration and selection within an explicit temporal and geographic framework would be transformational for the study of evolutionary biology in general, and animal domestication, in particular.

5.4.3 Multidisciplinary perspectives

Larger sample sizes and better computational methods are not, in and of themselves, a panacea for all questions related the domestication of animals. Palaeogenomics can tell us much about the ancestry of domestic animal populations, and identify which genetic variants increased in frequency over time, however, these topics are often tangential to the questions of direct interest to archaeologists.

What social, ecological and economic factors were involved in the emergence of animal domestication? What subsequent processes led to the adoption of agriculture? Why did agricultural lifestyles lag behind the emergence of domestication? Are there unifying theoretical models which can explain the diversity of causal factors leading to animal domestication? How did the resulting mutualism between people and their domesticates influence human culture and our relationship to the natural word?

Palaeogenomic approaches to domestication can assist in addressing such questions, but only when genetic perspectives are integrated into a broader multidisciplinary synthesis. A widespread criticism of paleogenomic research has been its perceived scientific chauvinism, in which genetic evidence is seen to be privileged above all other sources. Animal domestication involves much more than just the phenotypic expression of genetic traits, and tracing its origins requires an understanding of the broader cultural context in which the relationship developed.

5.4.4 Detailed transects

One avenue for achieving better multidisciplinarity would be to conduct fine-scale paleogenomic studies of domestic animal remains from sites which have already been

well-studied archaeologically. For example, the site of Çatalhöyük (Turkey) was occupied for ~1,500 years during the Neolithic and Chalcolithic, and the long running excavations there have produced an exemplary zooarchaeological record. By conducting paleogenomic analyses on such a collection, observations of genetic changes over time could be jointly interpreted alongside the extensive archaeological evidence.

Another promising direction for future animal domestication research is to survey the historical and archaeological records to identify the timing of shifts in the management and manipulation of domestic species, and to establish when people first began to make use of specialised material culture (e.g. rabbit hutches, pigeon dovecotes, horse bridles, etc.) to actively manage and adapt domestic animals to a human niche. We know these interventions had a dramatic effect on the physiology, behaviour and reproductive traits of domestic animals, as efforts to force wild animals to reproduce in captivity frequently fail. By obtaining ancient DNA from dated and contextualised archaeological remains, it would be possible to establish which genes were under selection during these pivotal changes in the domestic relationship. Such data would reveal much about the way in which changes in technology, and social, economic and religious practices have shaped the genomes of our pets and livestock.

The pace of change in palaeogenomics is showing no signs of abatement, and the rapid increase in ancient sequence data holds the potential to greatly expand our understanding of both animal domestication and the underlying genomic architecture of complex traits. However, the greatest potential for new insights are to be had when palaeogenomic data is interpreted in conjunction with historical and archaeological evidence.

5.5 References

Ai, H., Huang, L., Ren, J., 2013. Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. PLoS One 8, e56001.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara E Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Brito, L.F., Kijas, J.W., Ventura, R.V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A., Feng, Z., Jafarikia, M., Schenkel, F.S., 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genomics 18, 229.

Cheung, C., DeVantier, L., van Damme, K., 2006. Socotra: A Natural History of the Islands and Their People. Odyssey, Hong Kong.

Daly, K.G., Delser, P.M., Mullin, V.E., Scheu, A., Mattiangeli, V., Teasdale, M.D., Hare, A.J., Burger, J., Verdugo, M.P., Collins, M.J., Kehati, R., Erek, C.M., Bar-Oz, G., Pompanon, F., Cumer, T., Çakırlar, C., Mohaseb, A.F., Decruyenaere, D., Davoudi, H., Çevik, Ö., Rollefson, G., Vigne, J.-D., Khazaeli, R., Fathi, H., Doost, S.B., Sorkhani, R.R., Vahdati, A.A., Sauer, E.W., Kharanaghi, H.A., Maziar, S., Gasparian, B., Pinhasi, R., Martin, L., Orton, D., Arbuckle, B.S., Benecke, N., Manica, A., Horwitz, L.K., Mashkour, M., Bradley, D.G., 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science 361, 85–88.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Regitano, L.C. de A., Yildiz, M.A., Heaton, M.P., Liu, W.-S., Lei, C.-Z., Reecy, J.M., Saif-Ur-Rehman, M., Schnabel, R.D., Taylor,

J.F., 2014. Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet. 10, e1004254.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., Alfarhan, A.H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J.H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., Berrocal-Rangel, L., Biglari, F., Boessenkool, S., Boldgiv, B., Brem, G., Brown, D., Burger, J., Crubézy, E., Daugnora, L., Davoudi, H., de Barros Damgaard, P., de Los Ángeles de Chorro Y de Villa-Ceballos, M., Deschler-Erb, S., Detry, C., Dill, N., do Mar Oom, M., Dohr, A., Ellingvåg, S., Erdenebaatar, D., Fathi, H., Felkel, S., Fernández-Rodríguez, C., García-Viñas, E., Germonpré, M., Granado, J.D., Hallsson, J.H., Hemmer, H., Hofreiter, M., Kasparov, A., Khasanov, M., Khazaeli, R., Kosintsev, P., Kristiansen, K., Kubatbek, T., Kuderna, L., Kuznetsov, P., Laleh, H., Leonard, J.A., Lhuillier, J., Liesau von Lettow-Vorbeck, C., Logvin, A., Lõugas, L., Ludwig, A., Luis, C., Arruda, A.M., Marques-Bonet, T., Matoso Silva, R., Merz, V., Mijiddorj, E., Miller, B.K., Monchalov, O., Mohaseb, F.A., Morales, A., Nieto-Espinet, A., Nistelberger, H., Onar, V., Pálsdóttir, A.H., Pitulko, V., Pitskhelauri, K., Pruvost, M., Rajic Sikanjic, P., Rapan Papeša, A., Roslyakova, N., Sardari, A., Sauer, E., Schafberg, R., Scheu, A., Schibler, J., Schlumbaum, A., Serrand, N., Serres-Armero, A., Shapiro, B., Sheikhi Seno, S., Shevnina, I., Shidrang, S., Southon, J., Star, B., Sykes, N., Taheri, K., Taylor, W., Teegen, W.-R., Trbojević Vukičević, T., Trixl, S., Tumen, D., Undrakhbold, S., Usmanova, E., Vahdati, A., Valenzuela-Lamas, S., Viegas, C., Wallner, B., Weinstock, J., Zaibert, V., Clavel, B., Lepetz, S., Mashkour, M., Helgason, A., Stefánsson, K., Barrey, E., Willerslev, E., Outram, A.K., Librado, P., Orlando, L., 2019. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell 177, 1419–1435.e31.

Frantz, L.A.F., Haile, J., Lin, A.T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V.E., Daly, K.G., Battista, V.M., Price, M., Gron, K.J., Alexandri, P., Arbogast, R.-M., Arbuckle, B., Bălăşescu, A., Barnett, R., Bartosiewicz, L., Baryshnikov, G., Bonsall, C., Borić, D., Boroneanţ, A., Bulatović, J., Çakirlar, C., Carretero, J.-M., Chapman, J., Church, M., Crooijmans, R., De Cupere, B., Detry, C., Dimitrijevic, V., Dumitraşcu, V., du Plessis, L., Edwards, C.J., Erek, C.M., Erim-Özdoğan, A., Ervynck, A., Fulgione, D., Gligor, M., Götherström, A., Gourichon, L., Groenen, M.A.M., Helmer, D., Hongo, H., Horwitz, L.K., Irving-Pease, E.K., Lebrasseur, O., Lesur, J., Malone, C., Manaseryan, N., Marciniak, A., Martlew, H., Mashkour, M., Matthews, R., Matuzeviciute, G.M., Maziar, S., Meijaard, E., McGovern, T., Megens, H.-J., Miller, R., Mohaseb, A.F., Orschiedt, J., Orton, D., Papathanasiou, A., Pearson, M.P., Pinhasi, R., Radmanović, D., Ricaut, F.-X., Richards, M., Sabin, R., Sarti, L., Schier, W., Sheikhi, S., Stephan, E., Stewart, J.R., Stoddart, S., Tagliacozzo, A., Tasić, N., Trantalidou, K., Tresset, A., Valdiosera, C., van den Hurk, Y., Van Poucke, S., Vigne, J.-D., Yanevich, A., Zeeb-Lanz, A., Triantafyllidis, A., Gilbert, M.T.P., Schibler, J., Rowley-Conwy, P., Zeder, M., Peters, J., Cucchi, T., Bradley, D.G., Dobney, K., Burger, J., Evin, A., Girdland-Flink, L., Larson, G., 2019. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. Proc. Natl. Acad. Sci. U. S. A. 116, 17231–17238.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., Meyer, M., 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. 45, e79.

Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L.V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O.C., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C.,

Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P., Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.-Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M., Lau, L.P.L., Liu, G.E., Lynn, D.J., Panzitta, F., Dodds, K.G., 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528–532.

Griffiths, R.C., Marjoram, P., 1996. Ancestral inference from samples of DNA sequences with recombination. J. Comput. Biol. 3, 479–502.

Gwynne, M.D., 1967. The possible origin of the dwarf cattle of Socotra. Geogr. J. 133, 39–42.

Hansen, H.B., Damgaard, P.B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., Allentoft, M.E., 2017. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PLoS One 12, e0170940.

Kelleher, J., Wong, Y., Wohns, A.W., Fadil, C., Albers, P.K., McVean, G., 2019. Inferring whole-genome histories in large population datasets. Nat. Genet. 51, 1330–1338.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R.P., Cristobal, M.S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B., Consortium, O.M. of T.I.S.G., 2012. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biol. 10, e1001258.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Marshall, F.B., Dobney, K., Denham, T., Capriles, J.M., 2014. Evaluating the roles of directed breeding and gene flow in animal domestication. Proc. Natl. Acad. Sci. U. S. A. 111, 6153–6158.

McCue, M.E., Bannasch, D.L., Petersen, J.L., Gurr, J., Bailey, E., Binns, M.M., Distl, O., Guérin, G., Hasegawa, T., Hill, E.W., Leeb, T., Lindgren, G., Penedo, M.C.T., Røed, K.H., Ryder, O.A., Swinburne, J.E., Tozaki, T., Valberg, S.J., Vaudin, M., Lindblad-Toh, K., Wade, C.M., Mickelson, J.R., 2012. A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. PLoS Genet. 8, e1002451.

Muir, W.M., Wong, G.K.-S., Zhang, Y., Wang, J., Groenen, M.A.M., Crooijmans, R.P.M.A., Megens, H.-J., Zhang, H., Okimoto, R., Vereijken, A., Jungerius, A., Albers, G.A.A., Lawley, C.T., Delany, M.E., MacEachern, S., Cheng, H.H., 2008. 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. 105, 17312–17317.

Mwai, O., Hanotte, O., Kwon, Y.-J., Cho, S., 2015. African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. Asian-australas. J. Anim. Sci. 28, 911–921.

Ottoni, C., Van Neer, W., Cupere, B.D., Daligault, J., Guimaraes, S., Peters, J., Spassov, N., Prendergast, M.E., Boivin, N., Morales-Muñiz, A., Bălăşescu, A., Becker, C., Benecke, N., Boroneant, A., Buitenhuis, H., Chahoud, J., Crowther, A., Llorente, L., Manaseryan, N., Monchot, H., Onar, V., Osypińska, M., Putelat, O., Morales, E.M.Q., Studer, J., Wierer, U., Decorte, R., Grange, T., Geigl, E.-M., 2017. The palaeogenetics of cat dispersal in the ancient world. Nature Ecology & Evolution 1, 0139.

Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., Machado, A. da C., Distl, O., Felicetti, M., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C.T., Piercy, R.J., Raekallio, M., Rieder, S., Røed, K.H., Silvestrelli, M., Swinburne, J., Tozaki, T., Vaudin, M., Wade, C.M., McCue, M.E., 2013. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. PLoS One 8, e54997.

Rasmussen, M.D., Hubisz, M.J., Gronau, I., Siepel, A., 2014. Genome-wide inference of ancestral recombination graphs. PLoS Genet. 10, e1004342.

Rohland, N., Glocke, I., Aximu-Petri, A., Meyer, M., 2018. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. 13, 2447–2461.

Schaefer, R.J., Schubert, M., Bailey, E., Bannasch, D.L., Barrey, E., Bar-Gal, G.K., Brem, G., Brooks, S.A., Distl, O., Fries, R., Finno, C.J., Gerber, V., Haase, B., Jagannathan, V., Kalbfleisch, T., Leeb, T., Lindgren, G., Lopes, M.S., Mach, N., da Câmara Machado, A., MacLeod, J.N., McCoy, A., Metzger, J., Penedo, C., Polani, S., Rieder, S., Tammen, I., Tetens, J., Thaller, G., Verini-Supplizi, A., Wade, C.M., Wallner, B., Orlando, L., Mickelson, J.R., McCue, M.E., 2017. Developing a 670k genotyping array to tag \textasciitilde2M SNPs across 24 horse breeds. BMC Genomics 18, 565.

Shannon, L.M., Boyko, R.H., Castelhano, M., Corey, E., Hayward, J.J., McLean, C., White, M.E., Said, M.A., Anita, B.A., Bondjengo, N.I., Calero, J., Galov, A., Hedimbi, M., Imam, B., Khalap, R., Lally, D., Masta, A., Oliveira, K.C., Pérez, L., Randall, J., Tam, N.M., Trujillo-Cornejo, F.J., Valeriano, C., Sutter, N.B., Todhunter, R.J., Bustamante, C.D., Boyko, A.R., 2015. Genetic structure in village dogs reveals a Central Asian domestication origin. Proc. Natl. Acad. Sci. U. S. A. 112, 13639–13644.

Speidel, L., Forest, M., Shi, S., Myers, S.R., 2019. A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. 51, 1321–1329.

Stainton, J. j., Charlesworth, B., Haley, C. s., Kranis, A., Watson, K., Wiener, P., 2017. Use of high-density SNP data to identify patterns of diversity and signatures of selection in broiler chickens. J. Anim. Breed. Genet. 134, 87–97.

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., Hesse, P., Fulton, D., Sauer, E.W., Mohaseb, F.A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Vahidi, S.M.F., MacHugh, D.E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B.S., Mashkour, M., Orton, D., Horwitz, L.K., Teasdale, M.D., Bradley, D.G., 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 365, 173–176.

vonHoldt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J.C., Mosher, D.S., Spady, T.C., Elkahloun, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C.,

Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P.G., Qian, Z., Huang, W., Ding, Z.-L., Zhang, Y.-P., Bustamante, C.D., Ostrander, E.A., Novembre, J., Wayne, R.K., 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature 464, 898–902.

Wang, X., Liu, J., Zhou, G., Guo, J., Yan, H., Niu, Y., Li, Y., Yuan, C., Geng, R., Lan, X., An, X., Tian, X., Zhou, H., Song, J., Jiang, Y., Chen, Y., 2016. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci. Rep. 6. https://doi.org/10.1038/srep38932

6 Appendix

In addition to the research presented in the main body of this thesis, I also contributed to other research projects throughout my doctoral studies. The following are a selection of the published co-authored papers I have contributed to.

A. Rabbits and the Specious Origins of Domestication

In this paper, based on my MSc project, I led the historical research, performed all the computational analyses, and wrote the manuscript.

Trends in Ecology & Evolution

CellPress
REVIEWS

examples, and the interspecific competition processes raised by Doherty and Driscoll, highlight the crucial role of biotic interactions in species recovery efforts.

Too often, conservation efforts are focused on species' abiotic requirements for persistence, and the complexities of the multiple interacting processes shaping species occurrence and responses to threats are underappreciated. The NRH provides a framework to consider species declines and conservation management in terms of the biotic and abiotic processes influencing the realized niche of declined species. The NRH aims to improve opportunities for drawing on ecological theory for applied conservation research and management. In the face of the emerging extinction crisis, the NRH can facilitate new insights into the causes of species decline, barriers to recovery, and options for innovative management solutions.

[1]Fenner School of Environment and Society, The Australian National University, Canberra, Australia
[2]National Environmental Science Programme, Threatened Species Recovery Hub, Australia

*Correspondence:
ben.scheele@anu.edu.au (B.C. Scheele).

https://doi.org/10.1016/j.tree.2017.12.001

References
1. Scheele, B.C. *et al.* (2017) Niche contractions in declining species: mechanisms and consequences. *Trends Ecol. Evol.* 32, 346–355
2. Doherty, T.S. and Driscoll, D.A. (2017) Competition in the historical niche: a response to Scheele *et al.*. *Trends Ecol. Evol.*
3. Fraser, C.I. *et al.* (2015) Priority effects can lead to underestimation of dispersal and invasion potential. *Biol. Invasions* 17, 1–8
4. Waters, J.M. *et al.* (2013) Founder takes all: density-dependent processes structure biodiversity. *Trends Ecol. Evol.* 28, 78–85
5. Hinton, J.W. *et al.* (2013) Red wolf (*Canis rufus*) recovery: a review with suggestions for future research. *Animals* 3, 722–744
6. Perring, M.P. *et al.* (2015) Advances in restoration ecology: rising to the challenges of the coming decades. *Ecosphere* 6, 1–25
7. Thomas, J.A. *et al.* (2009) Successful conservation of a threatened *Maculinea* butterfly. *Science* 325, 80–83

Forum

Rabbits and the Specious Origins of Domestication

Evan K. Irving-Pease,[1] Laurent A.F. Frantz,[1,2] Naomi Sykes,[3] Cécile Callou,[4] and Greger Larson[1,*]

Rabbits are commonly thought to have been domesticated in ~AD600 by French monks. Using historical and archaeological records, and genetic methods, we demonstrate that this is a misconception and the general inability to date domestication stems from both methodological biases and the lack of appreciation of domestication as a continuum.

Traditional archaeological approaches for inferring the origins of domestic taxa have recently been complemented by the application of genetic methods, though the two techniques have often produced widely discordant estimates [1]. The lack of consilience between these approaches has frustrated efforts to understand the origins of domestic plants and animals. More generally, the wide variation in reported dates raises questions about what aspects of domestication are being dated.

Most efforts to establish the timing of domestication have focused on the late Pleistocene and early Holocene when the first animals were domesticated [1]. To better assess the lack of methodological consilience, we investigated European rabbits (*Oryctolagus cuniculus*). This species is ideal since they were domesticated in historic times from a geographically restricted source population (on the Iberian Peninsula and southwest France), and are present in archaeological faunal records inside and outside their indigenous distribution. The well-resolved geographic origin and the presence of an extant wild progenitor population also allowed for the application of population genetic methods to model the timing of their domestication.

The Historical Record

The earliest documentary records of rabbits were authored by Romans who encountered the species in the Iberian Peninsula. Varro, writing in the 1st century BC, gave instructions to his wife to keep rabbits alongside hares in her *leporarium* (the Roman precursor to medieval warrens) and to fatten them in hutches before slaughter (*De Re Rustica*, 3.12). Nachstein, however, argued that this did not lead to domestication since the Roman practice of actively hunting rabbits within *leporaria* would select against tameness, and that because rabbits continued to breed underground they escaped direct animal husbandry [2].

A recent study [3] reported that rabbit domestication was initiated by French monks in ~AD600 as the result of an edict by Pope Gregory the Great that allowed Christians to consume newborn or foetal rabbits (*laurices*) during Lent, since they were not considered meat. The idea that rabbits were a popular source of protein during Lent can be traced to Nachtsheim [2] and Zeuner [4], both of whom miscited a late 6th century Latin manuscript by St Gregory of Tours [5]. Though laurices were first described by Pliny the Elder in the 1st century AD as a most delicate food (*Naturalis Historia*, 8.55), there is no evidence that they were not considered meat. This fallacy, along with their wrongly assumed popularity during Lent, resulted directly

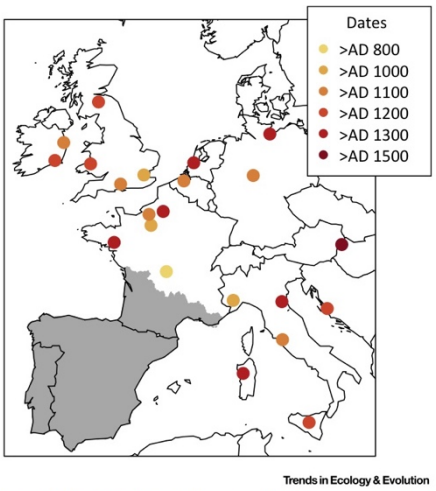


Figure 1. Map of the Medieval Dispersal of Rabbits across Western Europe. The grey region depicts the approximate natural range of the European rabbit. Coloured dots indicate the earliest historically or archaeologically documented appearance of rabbits in those regions. Adapted from [8].

from the miscitation (see Supplemental Information online). Lastly, this popular narrative also mistakenly conflates Pope Gregory the Great and Saint Gregory of Tours, two contemporaneous but unrelated individuals.

The Archaeological Record

Archaeological evidence demonstrates that rabbits were extensively exploited during the Epipaleolithic, Mesolithic, and early Neolithic in the Iberian Peninsula and southwest France (e.g., [6]). Besides a few isolated cases of rabbits appearing on Mediterranean islands ~2500 years ago [7], they were intentionally transported across Europe only during the Middle Ages when they were considered a high-status food (Figure 1) [8]. Though the expansion is historically well-attested, identifying and dating it archaeologically has been difficult owing to site recovery biases and the intrusion of rabbits into archaeological stratigraphies [8].

In addition, transported rabbits were largely indistinguishable from their wild counterparts. In fact, skeletal changes do not appear until the 18th century [8], almost 2000 years after the earliest historical account of their exploitation in captivity. The first appearance of skeletal morphological changes distinguishing wild from domestic populations instead coincides with the earliest evidence for rabbits as pets [8].

The Genetic Perspective

Genetic approaches to domestication can reveal the time depth of the most recent common ancestor of wild and domestic taxa. The conversion of molecular time estimates into calendar years

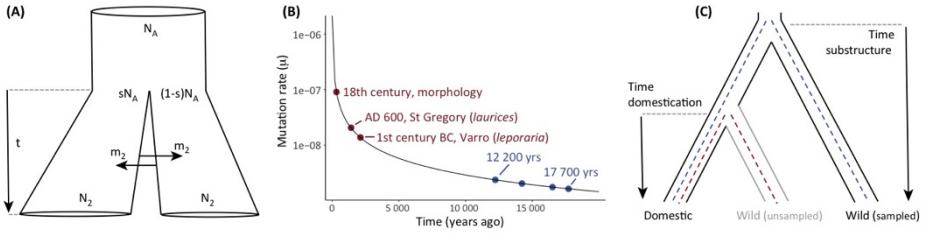


Figure 2. Demographic Modelling of Rabbit Domestication. An illustration of modelling results of the evolutionary history of rabbits based on genomic data from wild French and domestic rabbits [11], using diffusion approximations for demographic inference (*∂a∂i*) [12] with an isolation–migration (IM) model (see Supplemental Information online). (A) A schematic of the IM model where t is the time elapsed since the two populations separated, s is the bottleneck ratio (the proportion of the wild population that underwent domestication), m_1 and m_2 are migration rates (i.e., the amount of gene flow between the two populations) and N_1 and N_2 are effective population sizes. Inferring split times requires a mutation rate (μ) and a generation time to convert results into calendar years. (B) Time versus mutation rate. Blue dots represent inferred calendar year split times using five published estimates for μ. Red dots represent suggested rabbit domestication dates based on different criteria. Even when armed with an accurate mutation rate, estimating the time of domestication would require sampling the wild population from which domestic rabbits arose [see (C)]. The dates obtained by sampling other wild populations are consistent with events (e.g., deglaciation) that induced the substructure in wild rabbits.

requires a robust mutation rate, and for rabbits, four separate published rates vary by up to 45% (from 1.62×10^{-9} to 2.35×10^{-9}). As a result, analyses of rabbit genomic data suggest that wild French and domestic rabbit possibly split between 12 200 years and 17 700 years ago (Figure 2; see Supplemental Information online), though these estimated mutation rates are derived from imprecise fossil calibrations. When compared with estimates derived from more sophisticated methods, these rates are an order of magnitude faster than human rate estimates and up to three times slower than rates in domestic mice (see Supplemental Information online).

By applying these mutation rates at face value, the divergence estimates are more consistent with the Last Glacial Maximum than with domestication (Figure 2). This could be the case since population substructure is a feature of rabbit evolutionary history [3], and it is possible that by making use of wild populations not descended from those involved in the domestication process, the resulting split times can significantly predate the origins of domestication (Figure 2). Regardless, the wide range of intraspecific variation in mutation rates generally (see Supplemental Information online) and the lack of clear population divergences during the domestication process suggest that molecular dating approaches to domestication should be treated with caution.

Domestication as a Process, Not an Event

Rabbits are amongst the most recently domesticated animals, yet none of the three aforementioned methods can satisfactorily identify the rabbit's temporal origins. The historical record does not support the narrative built upon it since there was no papal edict, no dispensation to eat *laurices*, and no historical or archaeological evidence that the practice was commonplace. The archaeological evidence records skeletal morphological changes coinciding with modern pet-keeping, and the shifts in distribution sometimes post-date the historical evidence. Lastly, genetic approaches are complicated by large mutation rate uncertainty, population substructure, and the lack of clear separation between populations during domestication.

Discrepancies also result from *a priori* definitions of domestication. For instance, rabbit domestication may be concomitant with the earliest record of penning in Roman *leporaria* in the 1st century BC, with *laurice* consumption in the Middle Ages [2] or with the appearance of morphological changes distinguishing wild from domestic in the 18th century [8] (Figure 2). Archaeologists also commonly use the translocation of a species outside its native range as circumstantial evidence for the process of domestication. For rabbits, this is complicated by the fact that there is no evidence that the rabbits dispersed across Europe in the Middle Ages were domestic.

The willingness of scholars across broad disciplinary boundaries to accept the erroneous story of *laurices* in ~AD600 reveals how frequently the domestication process is misconstrued as a discrete event. Instead, the combination of the methodological and semantic factors highlighted in this study suggests that a precise domestication date does not exist. The domestication of rabbits, like other animals, was the result of a continuous, dynamic process that reflects gradual shifts in the nature and intensity of the relationship between humans and other species [9].

To obtain a satisfying rabbit domestication narrative, we need to view domestication and its associated biological changes as a process that occurs along a continuum [9,10]. Timing domestication should therefore focus on questions related to the numerous changes in the way humans interacted with domesticates, how those relationships varied in time and space, the relative intentionality of human actions and the genetic and morphological effects on the taxa in question. For example, rabbits were hunted during the Paleolithic, deliberately transported to Mediterranean islands, consumed as foetuses, housed in Roman *leporaria*, kept in Medieval pillow mounds and warrens, forced to reproduce in hutches, and only recently bred for morphological novelties as pets. No single one of these activities can be classified as the domestication threshold but collectively, they formed the processes by which rabbits became domesticated. Investigating domestication from a perspective that makes systematic use of multiple lines of evidence and emphasises the entirety of the process will result in a far more sophisticated appreciation of the origins of our pets and livestock.

Acknowledgements

L.A.F.F. and G.L. were supported by a European Research Council grant (ERC-2013-StG-337574-UNDEAD) and Natural Environmental Research Council grants (NE/K005243/1 and NE/K003259/1). L.A.F.F. was supported by a Junior Research Fellowship (Wolfson College, University of Oxford). E.K.I-P. was supported by a Clarendon Fund Scholarship, University of Oxford.

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tree.2017.12.009.

[1]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
[2]Department of Organismal Biology, Queen Mary University of London, London, UK
[3]Department of Archaeology, University of Nottingham, Nottingham, NG7 2RD, UK
[4]Muséum national d'Histoire naturelle, UMR 7209 du CNRS, Archéozoologie, Archéobotanique: sociétés, pratiques et environnements, case postale 55, 55 rue Buffon, 75005 Paris, France

*Correspondence:
greger.larson@arch.ox.ac.uk (G. Larson).
https://doi.org/10.1016/j.tree.2017.12.009

References

1. Ho, S. and Larson, G. (2006) Molecular clocks: when times are a-changin'. *Trends Genet.* 22, 79–83
2. Nachtsheim, H. (1936) *Vom Wildtier zum Haustier*, Alfred Metzner (in German)
3. Carneiro, M. *et al.* (2011) The genetic structure of domestic rabbits. *Mol. Biol. Evol.* 28, 1801–1816
4. Zeuner, F.E. (1963) The small rodents. In *A History of Domesticated Animals*, pp. 409–416, Harper & Row
5. Gregory-Bishop of Tours (1969) *History of the Franks*, W.W Norton & Company
6. Sana, I. *et al.* (2013) Domestication of animals in the Iberian Peninsula. In *The Origins and Spread of Domestic Animals in Southwest Asia and Europe* (Colledge, S., ed.), Left Coast Press, pp. 195
7. Quintana, J. *et al.* (2016) Primera datació d'un mamífer no autòcton (*Oryctolagus cuniculus* [Linnaeus, 1758]) (Mammalia: Lagomorpha) del jaciment holocènic del Pas d'en Revull (barranc d'Algendar, Ferreries). *Rev. de Menorca. Tom* 95, 185–200 (in French)
8. Callou, C. (2003) *De la Garenne au Clapier: Étude Archéozoologique du Lapin en Europe Occidentale*, Publications scientifiques du Muséum Paris (in French)
9. Vigne, J.-D. (2011) The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *C. R. Biol.* 334, 171–181
10. Zeder, M.A. (2012) The domestication of animals. *J. Anthropol. Res.* 68, 161–190
11. Carneiro, M. *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074–1079
12. Gutenkunst, R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695

B. Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events

In this paper, I led the spatial analysis of the mitochondrial and microsatellite data from the babirusa (*Babyrousa spp.*), Sulawesi warty pig (*Sus celebensis*) and anoa (*Bubalus spp.*); see Figure 4.

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Frantz LAF *et al.* 2018 Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events. *Proc. R. Soc. B* **285**: 20172566. http://dx.doi.org/10.1098/rspb.2017.2566

Received: 30 November 2017
Accepted: 16 March 2018

Subject Category:
Evolution

Subject Areas:
genetics, evolution

Keywords:
biogeography, evolution, geology, Wallacea

Authors for correspondence:
Laurent A. F. Frantz
e-mail: laurent.frantz@qmul.ac.uk
Greger Larson
e-mail: greger.larson@arch.ox.ac.uk

†Contributed equally to this study.
¶Present address: Pertamina University, Jl. Teuku Nyak Arief, Kawasan Simprug, Kebayoran Lama, Jakarta Selatan 12220, Indonesia.
‖Deceased.
‡Co-supervised the study.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4040687.

THE ROYAL SOCIETY PUBLISHING

Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events

Laurent A. F. Frantz[1,2,†], Anna Rudzinski[3,†], Abang Mansyursyah Surya Nugraha[4,¶,†], Allowen Evin[5,6,†], James Burton[7,8,†], Ardern Hulme-Beaman[2,6], Anna Linderholm[2,9], Ross Barnett[2,10], Rodrigo Vega[11], Evan K. Irving-Pease[2], James Haile[2,10], Richard Allen[2], Kristin Leus[12,13], Jill Shephard[14,15], Mia Hillyer[14,16], Sarah Gillemot[14], Jeroen van den Hurk[14], Sharron Ogle[17], Cristina Atofanei[11], Mark G. Thomas[3], Friederike Johansson[18], Abdul Haris Mustari[19], John Williams[20], Kusdiantoro Mohamad[21], Chandramaya Siska Damayanti[21], Ita Djuwita Wiryadi[‖], Dagmar Obbles[22], Stephano Mona[23,24], Hally Day[25], Muhammad Yasin[25], Stefan Meker[26], Jimmy A. McGuire[27], Ben J. Evans[28], Thomas von Rintelen[29], Simon Y. W. Ho[30], Jeremy B. Searle[31], Andrew C. Kitchener[32,33], Alastair A. Macdonald[7,‡], Darren J. Shaw[7,‡], Robert Hall[4,‡], Peter Galbusera[14,‡] and Greger Larson[2,‡]

[1]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[2]The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford OX1 3QY, UK
[3]Research Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
[4]SE Asia Research Group, Department of Earth Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK
[5]Institut des Sciences de l'Evolution, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, Cedex 05, France
[6]Department of Archaeology, Classics and Egyptology, University of Liverpool, 12-14 Abercromby Square, Liverpool L69 7WZ, UK
[7]Royal (Dick) School of Veterinary Studies & The Roslin Institute, University of Edinburgh, Easter Bush Campus, Roslin, Edinburgh EH25 9RG, UK
[8]IUCN SSC Asian Wild Cattle Specialist Group and Chester Zoo, Cedar House, Caughall Road, Upton by Chester, Chester CH2 1LH, UK
[9]Department of Anthropology, Texas A&M University, College Station, TX 77843-4352, USA
[10]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark
[11]Ecology Research Group, Section of Life Sciences, School of Human and Life Sciences, Canterbury Christ Church University, North Holmes Road, Canterbury CT1 1QU, Kent, UK
[12]Copenhagen Zoo, IUCN SSC Conservation Breeding Specialist Group—Europe, Roskildevej 38, Postboks 7, 2000 Frederiksberg, Denmark
[13]European Association of Zoos and Aquaria, PO Box 20164, 1000 HD Amsterdam, The Netherlands
[14]Centre for Research and Conservation (CRC), Royal Zoological Society of Antwerp, Koningin Astridplein 20-26, 2018 Antwerp, Belgium
[15]Environment and Conservation Sciences, School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia 6150, Australia
[16]Molecular Systematics Unit/Terrestrial Zoology, Western Australian Museum, Welshpool, Western Australia, Australia
[17]Edinburgh Medical School: BMTO, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK
[18]Gothenburg Natural History Museum, Box 7283, 402 35 Gothenburg, Sweden
[19]Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry, Bogor Agricultural University, PO Box 168, Bogor 16001, Indonesia

[20]Davies Research Centre, School of Animal and Veterinary Sciences, Faculty of Sciences, University of Adelaide, Roseworthy, Southern Australia 5371, Australia
[21]Faculty of Veterinary Medicine, Bogor Agricultural University, Jalan Agatis, IPB Campus, Darmaga, Bogor 16680, Indonesia
[22]Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, Ch. Deberiotstraat 32, 3000 Leuven, Belgium
[23]Institut de Systématique, Évolution, Biodiversité, ISYEB - UMR 7205 - CNRS, MNHN, UPMC, EPHE, Ecole Pratique des Hautes Etudes, 16 rue Buffon, CP39, 75005 Paris, France
[24]EPHE, PSL Research University, Paris, France
[25]No affiliation
[26]Department of Zoology, State Museum of Natural History Stuttgart, Rosenstein 1, 70191 Stuttgart, Germany
[27]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA
[28]Department of Biology, McMaster University, Hamilton, Ontario, Canada
[29]Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
[30]School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales 2006, Australia
[31]Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY 14853, USA
[32]Department of Natural Sciences, Chambers Street, National Museums Scotland, Edinburgh EH1 1JF, UK
[33]Institute of Geography, School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

LAFF, 0000-0001-8030-3885; AE, 0000-0003-4515-1649; AH-B, 0000-0001-8130-9648; DJS, 0000-0003-2016-1541; GL, 0000-0002-4092-0392

The high degree of endemism on Sulawesi has previously been suggested to have vicariant origins, dating back to 40 Ma. Recent studies, however, suggest that much of Sulawesi's fauna assembled over the last 15 Myr. Here, we test the hypothesis that more recent uplift of previously submerged portions of land on Sulawesi promoted diversification and that much of its faunal assemblage is much younger than the island itself. To do so, we combined palaeogeographical reconstructions with genetic and morphometric datasets derived from Sulawesi's three largest mammals: the babirusa, anoa and Sulawesi warty pig. Our results indicate that although these species most likely colonized the area that is now Sulawesi at different times (14 Ma to 2–3 Ma), they experienced an almost synchronous expansion from the central part of the island. Geological reconstructions indicate that this area was above sea level for most of the last 4 Myr, unlike most parts of the island. We conclude that emergence of land on Sulawesi (approx. 1–2 Myr) may have allowed species to expand synchronously. Altogether, our results indicate that the establishment of the highly endemic faunal assemblage on Sulawesi was driven by geological events over the last few million years.

1. Introduction

Alfred Russel Wallace was the first to document the 'anomalous' biogeographic region in Island Southeast Asia (ISEA) now known as Wallacea [1,2]. This biodiversity hotspot [3] is bounded by Wallace's Line in the west and Lydekker's Line in the east [4]. It consists of numerous islands in the Indonesian Archipelago, all of which boast a high degree of endemism. For example, on Sulawesi, the largest island in the region, at least 61 of the 63 non-volant mammalian species are endemic [5] and this figure is likely to be an underestimate.

The geological origins of Wallacea are as complex as its biogeography. Until recently, Sulawesi had been regarded as the product of multiple collisions of continental fragments from the Late Cretaceous [6–9]. This assumption has been challenged and a recent reinterpretation suggests instead that the island began to form as the result of continental collisions during the Cretaceous, which were then followed by Eocene rifting of the Makassar Strait. This process led to the isolation of small land areas in western Sulawesi from Sundaland. In the Early Miocene (approx. 23 Ma), a collision between the Sula Spur (a promontory of the Australian continent) and north Sulawesi led to uplift and emergence of land [10–12]. Later tectonic movements led to the present-day configuration of islands between Borneo and Australia [13,14].

A previous interpretation, involving the assembly of multiple terranes by collision, was used to suggest that Sulawesi's peculiar species richness resulted from vicariance and amalgamation over long geological time periods [10,15,16]. However, recent molecular-clock analyses suggest that a dispersal, starting in the Middle Miocene (approx. 15 Ma) from both Sunda and Sahul, is a more plausible explanation [17–19]. These conclusions suggest a limited potential for animal dispersal to Sulawesi prior to approximately 15 Ma. Rapid tectonic changes, coupled with the dramatic sea-level fluctuations over the past 5 Myr [20] might also have affected land availability and influenced patterns of species dispersal to Sulawesi, intra-island species expansion and speciation.

The hypothesis of a recent increase in land area [19] can be tested by comparing the population histories of multiple species on the island. Analyses of genetic and morphometric variability can be used to infer the timing and trajectories of dispersal, and the geographical and temporal origins of expansion. For example, if land area had increased from a single smaller island, extant species now living on Sulawesi would all have expanded from the same area. In addition, under this assumption, within the same geographical region their respective diversifications would be expected to have been roughly simultaneous.

Here, we focus on three large mammals endemic to Sulawesi: the babirusa (*Babyrousa* spp.), the Sulawesi warty pig (SWP, *Sus celebensis*) and the anoa, a dwarf buffalo (*Bubalus* spp.). The babirusa is a suid characterized by wrinkled skin and two extraordinary curved upper canine tusks displayed by males [21–23]. It represents a 'ghost lineage', because there are no closely related extant species outside Sulawesi (e.g. African suids are more closely related to all other Asian suids than to the babirusa) and the babirusa is unknown in the fossil record outside Sulawesi [24]. Three extant species of babirusa (distributed primarily in the interior of Sulawesi and on surrounding islands [21–23] have been described: *Babyrousa babyrussa* (Buru and Sulu Islands), *Babyrousa celebensis* (mainland Sulawesi) and *Babyrousa togeanensis* (Togian Island) [25].

The anoa is an endemic 'miniature buffalo' related to indigenous bovids in the Philippines and East Asia [26,27]. It stands approximately 1 m tall, weighs 150–200 kg and mostly inhabits pristine rainforest [28]. Although the subgenus *Anoa* has been divided into two species, the lowland anoa (*Bubalus depressicornis*) and the highland anoa (*Bubalus quarlesi*) [29], this classification is still contentious [27]. In contrast with anoa and babirusa, the SWP occupies a wide range of habitats, from swamps to rainforests. This species is closely related to the Eurasian wild pig (*Sus scrofa*), from which it

diverged during the Early Pleistocene (approx. 2 Ma) [24,30]. The SWP has been found on numerous islands throughout ISEA, probably as the result of human-mediated dispersal [31]. As its name implies, male SWPs develop facial warts. These cultural icons (e.g. SWP/babirusa and anoa are represented in the oldest prehistoric cave paintings [32,33]) have undergone recent and significant population reduction and range contraction due to overhunting and conversion of natural habitat for agricultural use.

Here, we establish when Sulawesi gained its modern shape and size, including connectivity between its constituent peninsulae, and assessed the impact of island formation on the evolution of Sulawesi's biodiversity. To do so, we used new reconstructions of the island's palaeogeography that allowed us to interpret the distribution of land and sea over the last 8 Myr at 1 Myr intervals. To determine the timings of diversification of the three largest endemic mammals on the island, we generated and analysed genetic and/or morphometric data from a total of 1289 samples of the SWP, anoa and babirusa obtained from museums, zoos and wild populations (456, 520 and 313 samples, respectively; electronic supplementary material, table S1). More specifically, we measured a total of 356 teeth from 227 specimens (357 babirusa and 191 SWP) using a geometric morphometric approach. In addition, we sequenced mitochondrial loci (cytb and/or control region) from 142 anoas, 213 babirusa and 230 SWP. Lastly, we typed 13 microsatellite loci from 163 anoa, 14 loci from 238 SWP and 13 from 182 babirusa (see the electronic supplementary material for more information). Although these taxa have been divided into multiple species (see taxonomic notes in the electronic supplementary material), for the purpose of this study, we treated SWP, anoa and babirusa as single taxonomic units.

2. Results and discussion

(a) Contemporaneous divergence

We generated mitochondrial DNA (mtDNA) sequences and/or microsatellite data from 230 SWPs, 155 anoas and 213 babirusa sampled across Sulawesi and the neighbouring islands (electronic supplementary material, figure S1 and table S1). Using a molecular-clock analysis, we inferred the time to the most recent common ancestor (TMRCA) of each species. The estimates from this method represent coalescence times, which provide a reflection of the crown age of each taxon. The closer relationship between babirusa and SWP (approx. 13 Ma) [34], compared with the divergence of either species from the anoa (approx. 58 Ma) [35] allowed us to align sequences from babirusa and SWP alongside one another, and jointly infer their relative TMRCAs. Separate analyses were performed for the anoa. The inferred TMRCA of SWP was 2.19 Myr (95% credibility interval (CI) 1.19–3.41 Myr; electronic supplementary material, figure S2) and of babirusa was 2.49 Myr (95% CI 1.33–3.61 Myr) (figure 1; electronic supplementary material, figure S2). The inferred TMRCA of anoa was younger (1.06 Myr; figure 1; electronic supplementary material, figure S3), though its 95% CI (0.81–1.96 Myr) overlapped substantially with the TMRCAs of the other two species.

The relatively recent divergence between babirusa and SWP also allowed us to compare their TMRCAs using identical microsatellite loci. To do so, we computed the average square distance (ASD) [36,37] between every pair of individuals

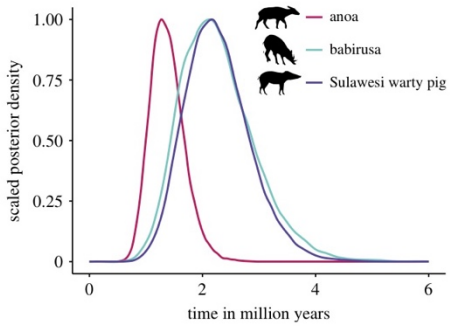


Figure 1. Time to the most recent common ancestor (TMRCA) for three mammal species on Sulawesi. Posterior densities of the TMRCA estimates for anoa, babirusa and Sulawesi warty pig.

within each species at the same 13 microsatellite loci. Although such an analysis might be affected by population structure (see below), we found that the distributions of ASD values were not significantly different between these two species (Wilcoxon signed-rank test, $p = 0.492$). This is consistent with the mitochondrial evidence for the nearly identical TMRCAs in the two species.

Recent molecular analyses have indicated that babirusa may have colonized Wallacea as early as 13 Ma, whereas SWP and anoas appear to have only colonized Sulawesi within the last 2–4 Myr [17,30,32,34]. An early dispersal of babirusa to Sulawesi (Late Palaeogene) has also been suggested on the basis of palaeontological evidence [19]. In addition, our data corroborate previous studies in indicating that both SWP and babirusa are monophyletic with respect to their most closely related taxa on neighbouring islands (e.g. Borneo), which is consistent with only one colonization of Sulawesi (electronic supplementary material, figure S4–S6) [30].

We then examined whether patterns of morphological diversity in these taxa are consistent with the molecular date estimates. To do so, we obtained measurements of 356 second and third lower molars (M2 and M3) from 95 babirusa and 132 SWPs. SWP and babirusa do not overlap morphologically (figure 2*a*), and we were thus able to assign each specimen to its correct species with success rates of 94.3% (CI: 92.7%–95.5%, distribution of leave-one-out cross validation of a discriminant analysis based on a balanced sample design) [38] and 94.7% (CI: 93.8%–96.7%) based on their M2 and M3, respectively. Our results also indicate that babirusa did not accumulate more tooth shape variation within Sulawesi (Fligner–Killeen test $\chi^2 = 1.04$, $p = 0.3$ for M2, $\chi^2 = 3.45$, $p = 0.06$ for M3). The data instead suggest that SWP has greater variance in the size of its M3 ($\chi^2 = 4.52$, $p = 0.03$, but not in the size of the M2, $\chi^2 = 3.44$, $p = 0.06$), and that the population from west central Sulawesi has an overall smaller tooth size than the two populations from northwest and northeast Sulawesi (figure 2*b*; electronic supplementary material, table S2). While these results may result from different selective constraints, they indicate that babirusa did not accumulate greater morphological variation in tooth shape than did the SWP, despite arriving on Sulawesi up to 10 Myr earlier.

Altogether our analyses suggest that although the three species are believed to have colonized the island at different

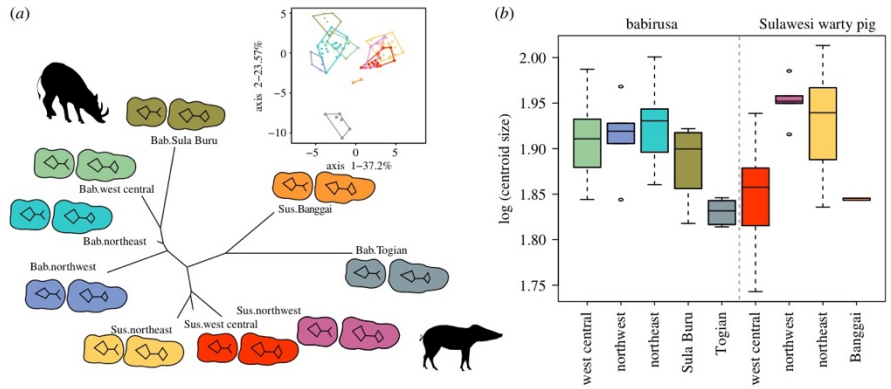


Figure 2. Population morphological variation inferred from geometric morphometric data. (*a*) Neighbour-joining network based on Mahalanobis distances measured from second and third lower molar shapes and visualization of population mean shape. Bab, babirusa; Sus, Sulawesi warty pig. (*b*) Variation of third molar size per population (log centroid size).

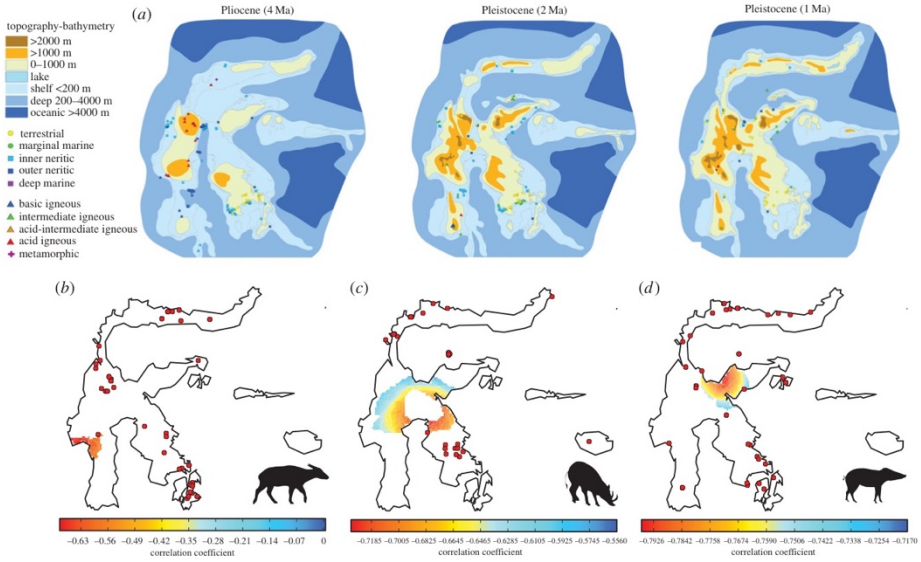


Figure 3. Geological maps of Sulawesi and the geographical origin of expansion. (*a*) Reconstruction of Sulawesi over the last 5 Myr (adapted from [39]) and potential origin of expansion of (*b*) anoa, (*c*) babirusa and (*d*) SWP. Red dots represent the location of the samples used for this analysis. Low correlation values (between distance and extrapolated genetic diversity; see electronic supplementary material) represent most likely origin of expansion.

times, their similar degrees of morphological diversity and their nearly synchronous TMRCAs raise the possibility that they (and possibly other species) responded to a common mechanism that triggered their contemporaneous diversification.

(b) Past land availability correlates with the expansion origins

Increasing land area may have promoted a simultaneous diversification and range expansion in babirusa, SWPs and anoas. To test this hypothesis, we used a new reconstruction that depicts land area in the Sulawesi region through time using information from the geological record. The reconstructions in 1 Myr increments (figure 3*a*; electronic supplementary material, figure S7) [39] support a scenario in which most of Sulawesi was submerged until the Late Pliocene to Early Pleistocene (2–3 Ma). Large-scale uplifts over the last 2–3 Myr would have rapidly and significantly increased land area, making it possible for non-volant species to expand their ranges.

To further assess whether these Plio-Pleistocene uplifts were responsible for a synchronous expansion, we inferred the most likely geographical origin of expansion using microsatellite

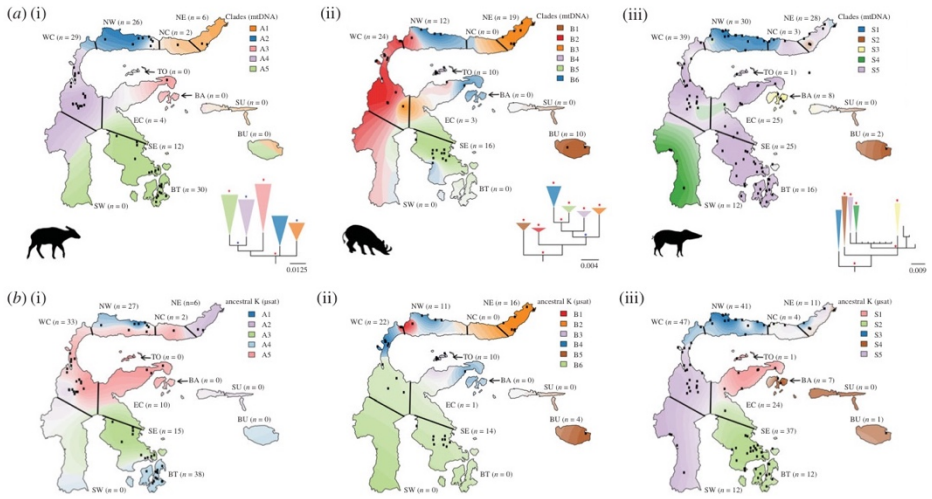


Figure 4. Population structure and geographical patterning of three mammal species on Sulawesi inferred from mitochondrial and microsatellite DNA. (*a*) A tessellated projection of sample haplogroups in each region of endemism and phylogeny of (i) anoa, (ii) babirusa and (iii) Sulawesi warty pig. Each region is labelled with the number of samples used for the projection. The projection extends over regions with no samples (e.g. the southwest peninsula for babirusa and anoa) and the population membership affinities for these regions are, therefore, unreliable. Red and blue stars on the phylogenetic trees correspond to posterior probabilities greater than 0.9 and 0.7, respectively. (*b*) Tessellated projection of the STRUCTURE analysis, using the microsatellite data, for (i) anoa, (ii) babirusa and (iii) Sulawesi warty pig. The best *K*-value for each species was used ($K = 5$ for anoa; $K = 6$ for babirusa; $K = 5$ for Sulawesi warty pig; electronic supplementary material, figure S8). NE, northeast; NC, north central; NW, northwest; TO, Togian; BA, Banggai Archipelago; EC, east central; WC, west central; SU, Sula; BU, Buru; SE, southeast; SW, southwest; BT, Buton.

data under a model of spatial loss of diversity with distance from expansion origin (electronic supplementary material). These estimates were obtained independently of, and uninformed by, either the geological reconstructions or modern phylogeographical boundaries inferred from other species. We deduced that the most likely origin for both SWP and babirusa was in the east central region of Sulawesi (figure 3*c*,*d*), and the most likely origin of anoa was in the west central region (figure 3*b*).

The origins of the population expansions of both SWP and babirusa occurred in an area of Sulawesi that only emerged during the Late Pliocene to Early Pleistocene (figure 3*a*; electronic supplementary material, figure S7). On the other hand, the anoa's most likely origin of diversification lies in a region that was submerged until the Pleistocene, consistent with palaeontological evidence [32] and with the slightly more recent TMRCA inferred for this species (figure 1). Thus, for all three species, the inferred geographical origins of their range expansions match the land availability derived from our geological reconstruction of Sulawesi.

(c) Geological history of past land isolation correlates with zones of endemism

Previous studies have identified endemic zones that are common to macaques, toads [18,40], tarsiers [41–44] and lizards [45]. We tested whether the same areas of endemism are linked to the population structure in our three species by generating a phylogenetic tree for each species using mtDNA and defined five to six haplogroups per species based on well-supported clades (figure 4*a–c*; electronic supplementary material, figure S4–6). We found that haplogroup proportions were significantly different between previously defined areas of endemism in all three species (Pearson's χ^2-test; $p < 0.001$), suggesting population substructure.

We also used STRUCTURE [46] to infer population structure from microsatellite data. The optimum numbers of populations (K) were 5, 6 and 5 for anoa, babirusa and SWP, respectively (electronic supplementary material, figure S8; figure 4*b*). Plotting the proportion of membership of each sample onto a map revealed a strong correspondence with the previously described zones of endemism (figure 4*b*). Using an analysis of molecular variance (AMOVA), we found that these areas of endemism explained approximately 17%, 27% and 5% of the variance in allele frequencies in anoa, babirusa and SWP, respectively (electronic supplementary material, table S5). Populations of babirusa and SWP in these zones of endemism were also strongly morphologically differentiated (figure 2).

Altogether, these data and analyses indicate that, despite some differences, the zones of endemism identified in tarsiers, macaques, toads and lizards [18,40–45,47] are largely consistent with the population structure and morphological differentiation in the three species studied here. This is particularly striking for the north arm of Sulawesi (NW, NC and NE in figure 4), where we identify two highly differentiated populations (reflected in both mtDNA and nuclear datasets) in all three taxa. This pattern could result from either adaptation to local environments or from isolation due to the particular geological history associated with the northern arm. Geological reconstructions (figure 3*a*) indicate that although land was present in this region during the past

4 Myr, it was often isolated from the rest of Sulawesi until the mid-Pleistocene. Thus, the combined geological and biological evidence presented here indicate that the high degree of divergence observed in the northern-arm populations in a multitude of species (e.g. three ungulates, macaques and tarsiers) might have been shaped by isolation from the rest of the island until the last 1 million years (figure 3*a*).

(d) Recent and contemporary land isolation also affected morphological evolution including dwarfism

Similar isolation is likely to have influenced the populations inhabiting the smaller islands adjacent to Sulawesi, including the Banggai Archipelago, Buru, Togian and Sula Islands. Interestingly, our geometric morphometric analyses demonstrated that these island populations of SWP and babirusa are the most morphologically divergent (figure 2*a*). For example, the insular populations from the Togian Islands (babirusa) and the Banggai Archipelago (SWP) were found to have much smaller tooth sizes than their counterparts on the mainland (figure 2*b*).

The significant morphometric divergences between populations on various islands are consistent with the genetic differentiation between babirusa/SWP on Togian, Sula and Buru (figure 4; electronic supplementary material, figure S9 and figure S10) and between island populations of SWP on Banggai Archipelago, Buton and Buru (figure 4; electronic supplementary material, figure S9 and figure S10).

Together, these results show that while suture zones between tectonic fragments are consistent with genetic and morphometric differentiation within Sulawesi, isolation on remote islands is likely to have had a much greater effect on morphological distinctiveness. Rapid evolution on islands has been described in many species (e.g. [48]), including in pigs [49] where island populations are known to have smaller tooth sizes than their mainland counterparts [50,51].

(e) Demographic history

Isolation of subpopulations across Sulawesi might also be linked to recent anthropogenic disturbances, especially for anoa and babirusa that occupy pristine forest or swamps [21,28]. In order to assess the impact of recent anthropogenic changes on the three species, we inferred their demographic history using approximate Bayesian computation (ABC). We fitted various demographic models to the genetic data (combining both mtDNA and microsatellite data; electronic supplementary material; figure S11). The best-supported demographic model involved a long-term expansion followed by a recent bottleneck in all three species (electronic supplementary material, table S3), corroborating the results of recent analyses of the SWP genome [30].

While our ABC analysis had insufficient power to retrieve the time of expansion (electronic supplementary material, table S4), it provided relatively narrow estimates of the current effective population sizes (figure 5; electronic supplementary material, table S4). We inferred a larger effective population size in SWP (83 021; 95% CI 46 287–161 457) than in babirusa (30 895; 95% CI 17 522–54 954) or anoa (27 504; 95% CI 13 680–54 056). Sulawesi warty pig occupies a wide range of habitats, including agricultural areas [52]. Thus, this species is likely to be less affected by continuing deforestation than babirusa or anoa, which is typically restricted to less disturbed forest and swamps [21,26]. Phylogenetic analyses of microsatellite data indicate more geographical structuring in babirusa and anoa than in SWP (electronic supplementary material, figure S12 and table S5). Altogether, these results are consistent with species-specific responses to habitat loss.

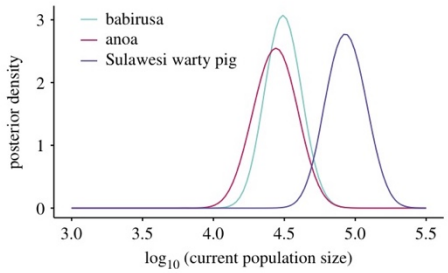


Figure 5. Posterior distribution of the current population size (Ne) of each species as inferred via ABC.

3. Conclusion

Our results indicate that, while the different geological components of Sulawesi were assembled at about 23 Ma, the island only acquired its distinctive modern form in the last few million years. By 3 Ma there was a large single island at its modern centre, but the complete connection between the arms was established more recently. The increasing land area associated with Plio-Pleistocene tectonic activity is likely to have provided the opportunity for a synchronous expansion in the three endemic mammal species in this study, as well as numerous other species. Interestingly, both our Pleistocene geological reconstruction and our proposed origins of expansion in the centre of the island closely resemble maps inferred from a study of tarsier species distribution on Sulawesi [42].

Furthermore, the recent emergence of connections between Sulawesi's arms coincides with a faunal turnover on the island and the extinction of multiple species. The geological reconstruction, and in particular, the recent elimination of the marine barrier at the Tempe depression separating the southwest and central regions, fits well with suggested replacement in tarsier species that occurred in the last approximately 1 million years [41]. The dispersal of our three species from the central region of Sulawesi may therefore have played a role in other local extinctions, such as the extinct suid known from southwest Sulawesi, *Celebochoerus*.

Sulawesi's development by emergence and coalescence of islands had a significant impact on the population structure and intraspecific morphological differentiation of Sulawesi's three largest mammals and many other endemic taxa. Thus, while most of Sulawesi's extant fauna arrived relatively recently, the more ancient geological history of the island (collision of multiple fragments) might have also affected patterns of endemism. Many aspects of Sulawesi's interconnected natural and geological histories remain unresolved. Integrative approaches that combine biological and geological datasets are therefore essential for reconstructing a comprehensive evolutionary history of Wallace's most anomalous island.

Data accessibility. All datasets, including microsatellites, mitochondrial, morphometric and meta data, are available on Dryad (http://

dx.doi.org/10.5061/dryad.dv322) [53]. The mitochondrial data are also available on GeneBank (accession MH021990–MH022712).

Authors' contributions. L.A.F.F., J.B., P.G., D.J.S., R.H., A.C.K. A.A.M., and G.L. conceived the study. L.A.F.F. and G.L. wrote the paper with input from all authors. L.A.F.F., S.Y.W.H., A.R., A.M.S.N., A.E., J.B., A.H-B., A.L., G.L., P.G., D.J.S. and E.K.I.-P. analysed the data. All other authors provided samples, data and analytical tools.

Competing interests. The authors have no competing interests.

Acknowledgements. We thank Joshua Schraiber and Erik Meijaard for valuable comments. We also thank the National Museums of Scotland for logistic support, and the Negaunee Foundation for their continued support of a curatorial preparator. We are also indebted to the Indonesian Ministry of Forestry, Jakarta (PHKA), Sulawesi's Provincial Forestry Departments (BKSDA); the Indonesian Institute of Science (LIPI); Museum of Zoology, Research Center for Biology, Cibinong (LIPI); and the project's long-standing Indonesian sponsor, Ir. Harayanto MS, Bogor Agricultural University (IPB) for sample collection/permission.

Funding. L.A.F.F., J.H., A.L., A.H.-B. and G.L. were supported by a European Research Council grant (ERC-2013-StG-337574-UNDEAD) and Natural Environmental Research Council grants (NE/K005243/1 and NE/K003259/1). L.A.F.F. was supported by a Junior Research Fellowship (Wolfson College, University of Oxford) and a Wellcome Trust grant (210119/Z/18/Z). P.G., S.G., J.v.d.H., C.A. and D.O. were supported by Flemish government structural funding. A.R. was supported by a Marie Curie Initial Training Network (BEAN—Bridging the European and Anatolian Neolithic, GA no. 289966) awarded to M.G.T. M.G.T. is supported by a Wellcome Trust Senior Research Fellowship (GA no. 100719/Z/12/Z). B.J.E. was supported by the Natural Science and Engineering Research Council of Canada. This work received additional support from the University of Edinburgh Development Trust, the Roslin Institute, the Balloch Trust and the Stichting Dierentuin Helpen (Consortium of Dutch Zoos). Additional support was also provided by the Rufford Small Grant, Royal Geographical Society, London, the Royal Zoological Society of Scotland and the University of Edinburgh Birrell-Gray Travel Award.

References

1. Wallace AR. 1863 On the physical geography of the malay archipelago. *J. R. Geogr. Soc. Lond.* **33**, 217–234. (doi:10.2307/1798448)
2. Wallace AR. 2012 *Island life*. Cambridge, UK: Cambridge University Press.
3. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858. (doi:10.1038/35002501)
4. Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR, von Rintelen T. 2011 Biogeography of the Indo-Australian Archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226. (doi:10.1146/annurev-ecolsys-102710-145001)
5. Musser GG. 1987 The mammals of Sulawesi. In *Biogeographical evolution of the Malay Archipelago* (ed. TC Whitmore), pp. 73–93. Oxford, UK: Oxford University Press.
6. Smith RB, Silver EA. 1991 Geology of a Miocene collision complex, Buton, eastern Indonesia. *Geol. Soc. Am. Bull.* **103**, 660–678. (doi:10.1130/0016-7606(1991)103<0660:GOAMCC>2.3.CO;2)
7. Hall R. 1996 Reconstructing Cenozoic SE Asia. In *Tectonic evolution of SE Asia* (eds R Hall, DJ Blundell), pp. 153–184. London, UK: Geological Society of London.
8. Hamilton W. 1979 *Tectonics of the Indonesian region*. Geological Survey Professional Paper 1078. Washington, DC: US Government Printing Office.
9. Davidson JW. 1991 The geology and prospectivity of Buton Island, Southeast Sulawesi, Indonesia. In *Proceedings, 20th annual convention, Jakarta*, vol. I, pp. 209–234. Jakarta, Indonesia: Indonesian Petroleum Association.
10. Moss SJ, Wilson MEJ. 1998 Biogeographic implications from the Tertiary palaeogeographic evolution of Sulawesi and Borneo. In *Biogeography and geological evolution of SE Asia* (eds R Hall, JD Holloway), pp. 133–163. Leiden, The Netherlands: Backhuys Publishers.
11. Hall R. 1998 The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In *Biogeography and geological evolution of SE Asia* (eds R Hall, JD Holloway), pp. 99–131. Leiden, The Netherlands: Backhuys Publishers.
12. Hall R. 2013 The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J. Limnol.* **72**, 1–17. (doi:10.4081/jlimnol.2013.s2.e1)
13. Spakman W, Hall R. 2010 Surface deformation and slab–mantle interaction during Banda arc subduction rollback. *Nat. Geosci.* **3**, 562–566. (doi:10.1038/ngeo917)
14. Hall R. 2011 Australia-SE Asia collision: plate tectonics and crustal flow. In *The SE Asian gateway: history and tectonics of Australia-Asia collision* (eds R Hall, MA Cottam, MEJ Wilson), pp. 75–109. London, UK: Geological Society of London.
15. Michaux B. 1996 The origin of southwest Sulawesi and other Indonesian terranes: a biological view. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **122**, 167–183. (doi:10.1016/0031-0182(95)00100-X)
16. de Boer AJ, Duffels JP. 1996 Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **124**, 153–177. (doi:10.1016/0031-0182(96)00007-7)
17. Stelbrink B, Albrecht C, Hall R, von Rintelen T. 2012 the biogeography of sulawesi revisited: is there evidence for a vicariant origin of taxa on wallace's 'Anomalous Island'? *Evolution* **66**, 2252–2271. (doi:10.1111/j.1558-5646.2012.01588.x)
18. Evans BJ, Supriatna J, Andayani N, Setiadi MI, Cannatella DC, Melnick DJ. 2003 Monkeys and toads define areas of endemism on Sulawesi. *Evolution* **57**, 1436–1443. (doi:10.1111/j.0014-3820.2003.tb00350.x)
19. van den Bergh GD, de Vos J, Sondaar PY. 2001 The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **171**, 385–408. (doi:10.1016/S0031-0182(01)00255-3)
20. Miller KG *et al.* 2005 The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298. (doi:10.1126/science.1116412)
21. Macdonald AA, Burton JA, Leus K. 2008 *Babyrousa babyrussa*. The IUCN Red List of Threatened Species. (doi:10.2305/IUCN.UK.2008.RLTS.T2461A9441445.en)
22. Macdonald A, Leus K, Masaaki I, Burton J. 2016 *Babyrousa togeanensis*. The IUCN Red List of Threatened Species. (doi:10.2305/iucn.uk.2016-1.rlts.t136472a44143172.en)
23. Leus, K., Macdonald, A., Burton, J. & Rejeki, I. 2016 *Babyrousa celebensis*. The IUCN Red List of Threatened Species. (doi:10.2305/iucn.uk.2016-1.rlts.t136446a44142964.en)
24. Frantz L, Meijaard E, Gongora J, Haile J, Groenen MAM, Larson G. 2016 The evolution of Suidae. *Ann. Rev. Anim. Biosci.* **4**, 61–85. (doi:10.1146/annurev-animal-021815-111155)
25. Meijaard E, Groves C. 2002 Upgrading three subspecies of babirusa (*Babyrousa* sp.) to full species level. *Asian Wild Pig News* **2**, 33–39.
26. Semiadi G, Mannullang B, Burton JA, Schreiber A, Mustari AH. 2008 *Bubalus depressicornis*. The IUCN Red List of Threatened Species. (doi:10.2305/IUCN.UK.2008.RLTS.T3126A9611738.en)
27. Burton JA, Hedges S, Mustari AH. 2005 The taxonomic status, distribution and conservation of the lowland anoa *Bubalus depressicornis* and mountain anoa *Bubalus quarlesi*. *Mamm. Rev.* **35**, 25–50. (doi:10.1111/j.1365-2907.2005.00048.x)
28. Burton JA, Wheeler P, Mustari A. 2016 *Bubalus depressicornis*. The IUCN Red List of Threatened Species. (doi:10.2305/IUCN.UK.2016-2.RLTS.T3126A46364222.en)
29. Groves CP. 1969 Systematics of the anoa (Mammalia, Bovidae). *Beaufortia* **17**, 1–12.
30. Frantz LA *et al.* 2013 Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol.* **14**, R107. (doi:10.1186/gb-2013-14-9-r107)

31. Groves CP. 1984 Of mice and men and pigs in the Indo-Australian Archipelago. *Canberra Anthropol.* **7**, 1–19. (doi:10.1080/03149098409508559)
32. Rozzi R. 2017 A new extinct dwarfed buffalo from Sulawesi and the evolution of the subgenus Anoa: an interdisciplinary perspective. *Quat. Sci. Rev.* **157**, 188–205. (doi:10.1016/j.quascirev.2016.12.011)
33. Aubert M *et al.* 2014 Pleistocene cave art from Sulawesi, Indonesia. *Nature* **514**, 223–227. (doi:10.1038/nature13422)
34. Gongora J *et al.* 2011 Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zool. Scr.* **40**, 327–335. (doi:10.1111/j.1463-6409.2011.00480.x)
35. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012 Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B.* **279**, 3491–3500. (doi:10.1098/rspb.2012.0683)
36. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995 An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463–471.
37. Sun JX, Mullikin JC, Patterson N, Reich DE. 2009 Microsatellites are molecular clocks that support accurate inferences about history. *Mol. Biol. Evol.* **26**, 1017–1027. (doi:10.1093/molbev/msp025)
38. 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)
39. Nugraha AMS, Hall R. 2016 Late Cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **490**, 191–209. (doi:10.1016/j.palaeo.2017.10.033)
40. Evans BJ, Supriatna J, Andayani N, Melnick DJ. 2003 Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* **57**, 1931–1946. (doi:10.1111/j.0014-3820.2003.tb00599.x)
41. Driller C, Merker S, Perwitasari-Farajallah D, Sinaga W, Anggraeni N, Zischler H. 2015 Stop and go—waves of tarsier dispersal mirror the genesis of Sulawesi Island. *PLoS ONE* **10**, e0141212. (doi:10.1371/journal.pone.0141212)
42. Merker S, Driller C, Perwitasari-Farajallah D, Pamungkas J, Zischler H. 2009 Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. *Proc. Natl Acad. Sci. USA* **106**, 8459–8464. (doi:10.1073/pnas.0900319106)
43. Burton JA, Nietsch A. 2010 Geographical variation in duet songs of Sulawesi Tarsiers: evidence for new cryptic species in South and Southeast Sulawesi. *Int. J. Primatol.* **31**, 1123–1146. (doi:10.1007/s10764-010-9449-8)
44. Shekelle M, Meier R, Wahyu I, Ting N. 2010 Molecular phylogenetics and chronometrics of Tarsiidae based on 12S mtDNA haplotypes: evidence for Miocene origins of crown tarsiers and numerous species within the Sulawesian clade. *Int. J. Primatol.* **31**, 1083–1106. (doi:10.1007/s10764-010-9457-8)
45. McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, Orange DI, Lemos-Espinal J, Riddle BR, Jaeger JR. 2007 Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* **61**, 2879–2897. (doi:10.1111/j.1558-5646.2007.00239.x)
46. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
47. Evans BJ, McGuire JA, Brown RM, Andayani N, Supriatna J. 2008 A coalescent framework for comparing alternative models of population structure with genetic data: evolution of Celebes toads. *Biol. Lett.* **4**, 430–433. (doi:10.1098/rsbl.2008.0166)
48. van der Geer A. 2010 *Evolution of island mammals: adaptation and extinction of placental mammals on islands*. New York, NY: Wiley-Blackwell.
49. Evin A, Dobney K, Schafberg R, Owen J, Vidarsdottir US, Larson G, Cucchi T. 2015 Phenotype and animal domestication: A study of dental variation between domestic, wild, captive, hybrid and insular *Sus scrofa*. *BMC Evol. Biol.* **15**, 6. (doi:10.1186/s12862-014-0269-x)
50. Kruska D, Röhrs M. 1974 Comparative–quantitative investigations on brains of feral pigs from the Galapagos Islands and of European domestic pigs. *Z. Anat. Entwicklungsgesch.* **144**, 61–73. (doi:10.1007/BF00518633)
51. McIntosh GH, Pointon A. 1981 The Kangaroo Island strain of pig in biomedical research. *Aust. Vet. J.* **57**, 182–185. (doi:10.1111/j.1751-0813.1981.tb00505.x)
52. Burton JA, Macdonald AA. 2008 *Sus celebensis*. The IUCN Red List of Threatened Species. (doi:10.2305/IUCN.UK.2008.RLTS.T41773A10559537.en)
53. Frantz LAF *et al.* 2018 Data from: Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events. Dryad Digital Repository. (https://doi.org/10.5061/dryad.dv322)

C. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion

In this paper, I lead the qpGraph analysis of the pygmy hog data, and developed a Python package for automating the fitting of qpGraph models and comparison of the fitted models by computing their Bayes factors with the R package *admixturegraph*.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-10017-2 OPEN

Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion

Langqing Liu [1], Mirte Bosse [1], Hendrik-Jan Megens[1], Laurent A.F. Frantz[2,3], Young-Lim Lee[1], Evan K. Irving-Pease [3], Goutam Narayan [4,5], Martien A.M. Groenen [1] & Ole Madsen[1]

Wild boar (*Sus scrofa*) drastically colonized mainland Eurasia and North Africa, most likely from East Asia during the Plio-Pleistocene (2–1Mya). In recent studies, based on genome-wide information, it was hypothesized that wild boar did not replace the species it encountered, but instead exchanged genetic materials with them through admixture. The highly endangered pygmy hog (*Porcula salvania)* is the only suid species in mainland Eurasia known to have outlived this expansion, and therefore provides a unique opportunity to test this hybridization hypothesis. Analyses of pygmy hog genomes indicate that despite large phylogenetic divergence (~2 My), wild boar and pygmy hog did indeed interbreed as the former expanded across Eurasia. In addition, we also assess the taxonomic placement of the donor of another introgression, pertaining to a now-extinct species with a deep phylogenetic placement in the *Suidae* tree. Altogether, our analyses indicate that the rapid spread of wild boar was facilitated by inter-specific/inter-generic admixtures.

[1] Animal Breeding and Genomics, Wageningen University & Research, 6708PB Wageningen, the Netherlands. [2] School of Biological and Chemical Sciences, Queen Mary University of London, E1 4NS, London, United Kingdom. [3] Palaeogenomics and Bioarcheology Research Network, Research Laboratory for Archeology and History of Art, University of Oxford, Oxford OX1 3QY, United Kingdom. [4] Durrell Wildlife Conservation Trust, Les Augrès Manor, Jersey JE3 5BP Channel Islands, United Kingdom. [5] Pygmy Hog Conservation Programme, EcoSystems-India, Indira Nagar, Basistha, Guwahati, Assam 781029, India. Correspondence and requests for materials should be addressed to L.L. (email: langqing.liu@wur.nl) or to O.M. (email: ole.maden@wur.nl)

The expansion of species into novel habitats can have tremendous impacts on the native fauna. If the native fauna contains species that are closely related to the invasive population, hybridization may also threaten the integrity and survival of native species[1]. Observation of admixture in expanding populations has led to speculation whether admixture has an important role in driving the success of those populations[2–4]. Although the old-world pigs, the *Suidae*, are distributed throughout Africa and Eurasia, only two species are recognized across mainland Eurasia: wild boar (*Sus scrofa*) and pygmy hog (*Porcula salvania*)[5–9].

This, however, was not always the case and extensive fossils records suggest that Eurasia hosted a highly diverse set of *Suidae* species that originated during the Miocene[10–13](Fig. 1a). In middle Miocene, the first isolated lineage to split from those early *Suidae* was *Babyrousa,* which forms an ancient lineage endemic to the island of Sulawesi (Fig. 1b)[14,15]. Along with the global cooling during the late Miocene[16], a new subfamily, the *Suinae*, emerged in the fossil record[17,18] and replaced almost all other subfamilies of *Suidae* present at that time[19,20] (Fig. 1c). The *Suinae* later diversified into multiple tribes[17,18]. This was followed by a divergence of the African *Suidae* and the Eurasian *Sus* genus, at around the Miocene/Pliocene boundary (Fig. 1d). Shortly thereafter, the divergence within the *Sus* genus started during the early Pliocene (Fig. 1f). Several *Sus* species on the islands of southeast Asia (ISEA) evolved during the early/mid Pliocene[21]. Relatively high levels of species diversity were likely maintained across Eurasia, until the early Pleistocene, when wild boar expanded out of East Asia into almost every type of ecosystem across the old-world. This expansion was highly efficient mirroring the great human expansion during the late Pleistocene[22], and previous study have suggested that it is the reason for the disappearance of most suid species across Eurasia[23,24] (Fig. 1g, h). With this, the layout for the modern *Suidae* species became settled.

As the only reminiscence of the once highly diverse Pliocene suid fauna, the pygmy hog represents not only an important species to conserve but also the key to our understanding of the expansion of wild boar. Indeed, previous studies have suggested that the rapid and highly efficient expansion of wild boar was facilitated by interspecific adaptive gene-flow[21,25]. Under this scenario, wild boar absorbed rather than replaced species, a process paralleled to some extent in humans[26–28]. Although pygmy hog is now highly endangered and restricted in a small corridor of high grassland at the southern foothills of the Himalaya[29–31], it was far more widespread in the past[32] (Fig. 1e). According to middle Pleistocene fossil remains from southwest China, the geographical range of pygmy hog and wild boar did

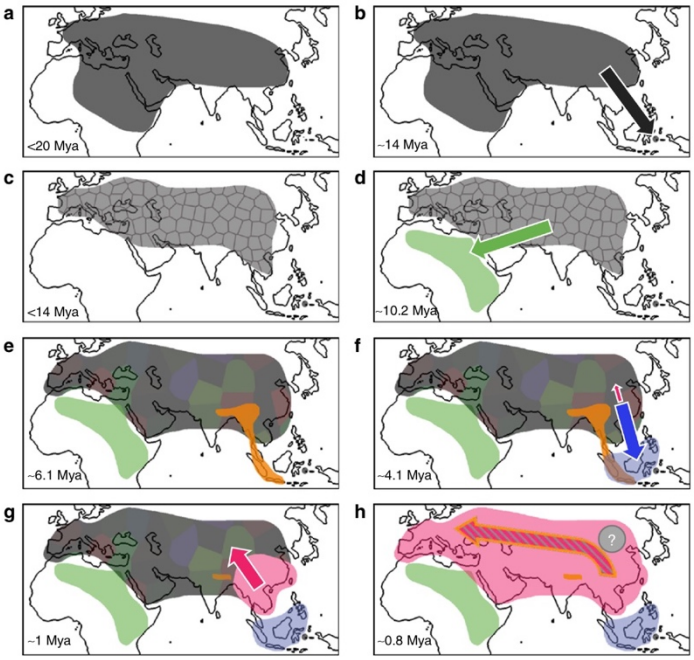


Fig. 1 A series of schematic models depicting the geographic evolution of *Suidae* species over the past 20 Mya. **a** Emergence of Suidae across Eurasian and Northern Africa. **b** The black arrow depicts the hypothesized trajectory of *Babyrussa* migrating to ISEA. **c** Emergence of *Suinae* (gray collage pattern) that replaced other *Suidae*. **d** The green arrow indicates the diversification of the ancestor of Sub-Saharan suids. **e** Eurasian *Suinae* split into multiple genera (multi-color collage pattern), including pygmy hog (orange shade). **f** The blue arrow depicts the migrations of *Sus* to ISEA. The red arrow indicates emergence of *Sus scrofa*. **g**, **h**. The spread of *Sus scrofa* from southern Asia to Europe and replacement of all Suinae species except pygmy hog. During the replacement, Sus scrofa populations introgressed at least three times with one of these Suinae species (ghost linage), pygmy hog and ISEA *Sus*, respectively. The colors correspond to those used in Fig. 2 and represent the cluster on the tree to which the samples belong

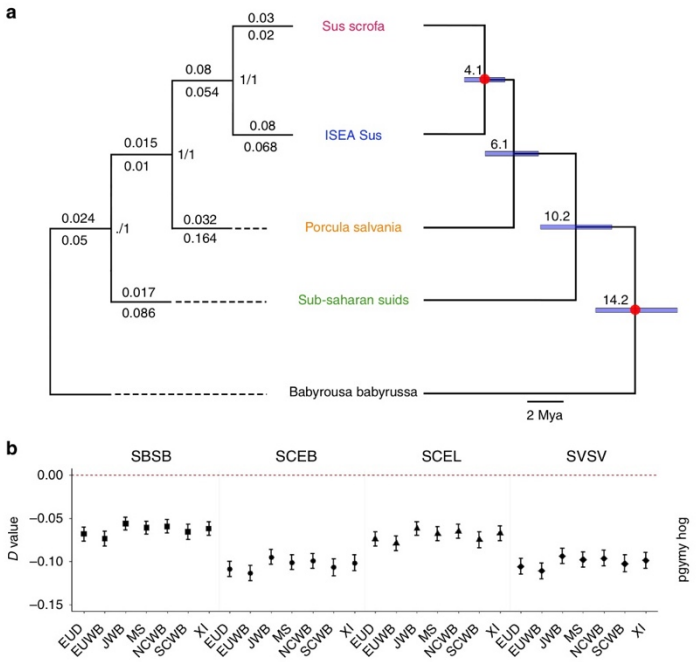


Fig. 2 Phylogenetic relationships and divergence of the *Suidae* species used in the current study and admixture event between pygmy hog and *Sus scrofa*. **a** The tree on the left is the ML tree of the Suidae family based on consensus and concatenation methods. Branch length of the consensus analysis above branches, concatenation below branches. Node labels show bootstrap values of the concatenation analysis and concordance factors of the consensus analysis, respectively. The tree on the right is the time tree of divergence. Node labels show age in million years. Blue bars indicate 95% confidence interval and red dots show the calibration points (See Supplementary material Figs. 1–3 for full trees). **b** A diagram depicting the excess derived allele sharing when comparing sister taxa and outgroups. Each column contains the fraction of excess allele sharing by a taxon (up/down) with the pygmy hog/outgroup compared with its sister taxon (up/down). We computed D-statistics of the form D (X, Y, Pygmy hog, warthog). Error bars correspond to three standard errors. (SBSB = *Sus barbatus*, SCEB = *Sus cebrifons*, SCEL = *Sus celebensis*, SVSV = *Sus verrucosus*, EUD = European domesticated pig, EUWB = European wild boar, JWB = Japanese wild boar, MS = Meishan, NCWB = Northern China wild boar, SCWB = Southern China wild boar, XI = Xiang)

overlap[33,34], implying that the temporal and geographical proximity of pygmy hog and wild boar could have resulted in hybridization. Therefore, pygmy hog provides an excellent comparative framework to study the evolutionary processes that occurred during a fast and extensive radiation. We analyzed six genomes of pygmy hog in combination with genomes of related suid species and found strong support for an important role of inter-species hybridization during range expansion. The results suggest that wild boar hybridized with pygmy hog and a now-extinct suid species during the rapid spread across Eurasia and North Africa.

Results

Phylogenetic relationships and divergence of the *Suidae* species. We sequenced the genomes of six pygmy hogs and one *Babyrousa babyrussa* and analyzed these with the genome sequences of 31 individuals, in total, representing 10 of the extant *Suidae* species (See Material and methods and Supplementary data 1 for details). We first assessed the phylogenetic relationship of these species. The concatenation and consensus methods resulted in the same main topology (Fig. 2a, Supplementary Figs. 1, 2). The phylogenetics analyses clearly show that the most basal split within the *Suinae* are sub-Saharan suids followed by a highly supported split of pygmy hogs (BS = 100 in supermatrix and CF = 1 in supertree) from all *Sus* species. To compare our phylogenetics results to an earlier study using fragments of mitochondrial DNA[35], we also carried out a Bayesian phylogenetic analysis using complete mitochondrial genomes (Supplementary Fig. 3). The resulting topology is consistent with previous studies confirming pygmy hog as basal to *Sus*. Thus, both the genome-wide autosomal phylogenetic analysis and complete mitochondrial genome analysis support, with very high confidence, that pygmy hog is a monophyletic sister taxon of *Sus*.

We selected autosomal genomic loci supporting the main topology to obtain the basal divergence between the studied taxa (Fig. 2a, Supplementary Fig. 4). Molecular clock analyses indicated that the divergence between Sub-Saharan suids and Eurasian *Suidae* (pygmy hog and *Sus*) took place shortly after the divergence of *Babyrousa babyrussa* ~ 10.2 Mya (95% highest posterior density (HPD) = 12.7–7.9). Pygmy hog separated from the common ancestor with *Sus* during the early Pliocene, ~6.1 Mya (95% HPD = 7.8–4.2) and the Eurasian wild boar split from

other *Sus* species during the early Pliocene ~ 4.1 Mya (95% HPD = 5.5–2.7). (Fig. 2a, Supplementary Fig. 4, Supplementary Note)

Admixture between pygmy hog and wild boar. In order to test whether temporal and geographical proximity of closely related species could have resulted in hybridization, we looked for interspecific admixture signal within our genome-wide data. Several studies have reported that in diverged species sex-linked markers may show evolutionary histories incongruent to other sex-linked and/or autosomal markers[36–40]. Thus, we analyzed autosomes and the X chromosome separately. To investigate whether any of the sequenced pygmy hogs showed evidence for autosomal introgression from ancestors of present-day *Sus* species, or vice-versa, the pygmy hog was compared with representatives of eleven *Sus* populations using *D*-statistics. We found a significant overrepresentation of derived alleles between the pygmy hog and mainland Eurasian wild boar at autosomal chromosomes (Fig. 2b, Supplementary data 2), indicative of admixture. This signal of admixture was also supported by TreeMix analysis. The best-fitting model suggests an ancient admixture between ancestral wild boar and pygmy hog (Supplementary Fig. 5).

To further examine this autosomal genome-wide pattern of admixture between pygmy hog and wild boar, we combined *D*-statistics and fd to infer regions of introgression. We also calculated DNA sequence divergence (d_{xy}) for each window to reduce false-positive signals[41]. This resulted in 636 putative introgression intervals between pygmy hog and wild boar from Europe, North China and South China, of which 427 (67.1%) are shared within wild boars (Supplementary Fig. 6, Supplementary Note). This suggests the admixture occurred before the divergence within Eurasian wild boar, further sustaining the evidence of an ancestral gene-flow between pygmy hog and the common ancestor of wild boar.

Wild boar harbors genetic introgression from a ghost linage. We next addressed evidence of admixture from the X chromosome starting with reconstructing a ML tree for the X chromosome. This tree displays a different topology compared with our main phylogenetic tree (Supplementary Fig. 8). Previous study has reported two distinct haplotypes in European pigs and South Asian pigs, and proposed that this might be derived from a now-extinct Suid (ghost) lineage[40]. To investigate the existence of genealogical discordance, we carried out a sliding-window *D*-statistic and a machine-learning based detection of local phylogenetic incongruent regions[42]. Both approaches supported that there is a ~ 40.6 Mbp (46.2–86.8 Mbp) region on the X chromosome, where pygmy hog clusters with ISEA *Sus* and South China pigs, whereas the European pigs and North China wild boars appear to be basal to this cluster. (Fig. 3, Supplementary data 3, Supplementary Figs. 9–15, Supplementary Note). In addition, an ambiguous pattern was also observed in northern Chinese domestic pigs, where the region from 46.2 to 57.1 Mbp support a clustering of northern Chinese domestic pigs with European pigs/North Chinese wild boars, whereas from 57.1 to 86.8 Mbp support northern Chinese domestic pigs clustering with southern Chinese pigs (Fig. 3, Supplementary Fig. 13). The signature of the introgression regions was also supported by ML trees (Supplementary Fig. 16). Taken together, our results show that within this genomic region, sequences of European/North Chinese pigs have an ancient origin. With the inclusion of pygmy hog genome, we could locate this ghost lineage to be older than the split of pygmy hogs but younger than the split to the Sub-Saharan suids. So far, there is no molecular or fossil evidence for this ancient linage, which was probably extinct long time ago. We further looked for evidence of this introgression in the autosomes. Comparison between different wild boar populations further identified autosomal regions supporting the X-chromosome introgression signal (Supplementary Figs. 10–15). The amount, length and magnitude of ghost introgression in autosomes are similar among the wild boar populations (Supplementary Figs. 17 & 18), which suggests that this hybridization likely occurred early within the evolutionary history of wild boar.

It has been reported that the region around the centromere on the X chromosome in pigs has an extremely low recombination rate[43,44]. Also, as a consequence of the global reduction in effective population size of wild boar in the past ~1 My to the end of the Last Glacial Maximum[9], wild boar went through processes like incomplete lineage sorting (ILS) and positive natural selection. The joint effect likely resulted in distinct distribution of ancient introgressed haplotype between European/northern Chinese and southern Chinese wild boar populations[40]. Finally, the genealogical discordance on the X chromosome became fixed in pigs from different regions. Recombination rates are highly variable on the autosomes and this hybridization probably happened at least 1 Mya. The long period of recombination, in addition with post-divergence gene-flow between wild boar population[9], highly truncated autosomal haplotype blocks. This would lead to many very short introgression segments scattered over the autosomes, which is what we observed in our analyses (Supplementary Fig. 18).

For the male individuals, we also reconstructed the phylogeny based on the non-recombining part of the Y-chromosome (4.8~43.5 Mb), which resulted in a topology consistent with our main phylogenetic tree (Supplementary Fig. 19).

Demographical modeling of the Suidae evolutionary history. We then used an automated qpGraph approach[45,46] to evaluate the fit of various admixture graphs to our data (see Material and methods). We then estimated the marginal likelihood of 37 models that left no f4 outliers using a MCMC approach[47] (Supplementary Fig. 21). According to these models, both pygmy hog and ISEA *Sus* contributed to wild boar. Interestingly, these models suggest that the ISEA *Sus* only contributed to south Chinese wild boar. This discrepancy is likely the result of the near simultaneous divergence of all three wild boar lineages, as well as the fact that a population closer to the South Chinese wild boar potentially also contributed to ISEA *Sus*[21]. The best admixture graph, however, is slightly different from those obtained from *TreeMix* and phylogenetic analyses, as it finds a signal of ISEA *Sus* admixture into pygmy hog population. This suggests that pygmy hog actually interbred with an unsampled admixing/separating population, which was intermediate between ISEA *Sus* and *Sus scrofa*. However, our parsimonious model may be too simple to reflect the complexity of the reticulate admixture in these populations and to disentangle the ancestral variants between ISEA *Sus* and *Sus scrofa*. Altogether, this analysis validates the existence of gene-flow between wild boar and ISEA *Sus* as well as between pygmy hog and wild boar. Yet, it also suggests that this admixture could have been bidirectional in which case the south Chinese wild boar population forms the best proxy for these events.

To coalesce the known demographic implications by far, we further fitted various fitted various gene-flow models, which are based on our priori assumption of *Suidae* systematic, to a phylogenetic scenario using G-phocs. We separated admixture branches into each of the *Sus scrofa* populations, to better account for their variable levels of basal admixture. (Supplementary Fig. 22, Supplementary Note). In support of our *D*-statistic findings, high probability of gene-flow between the common

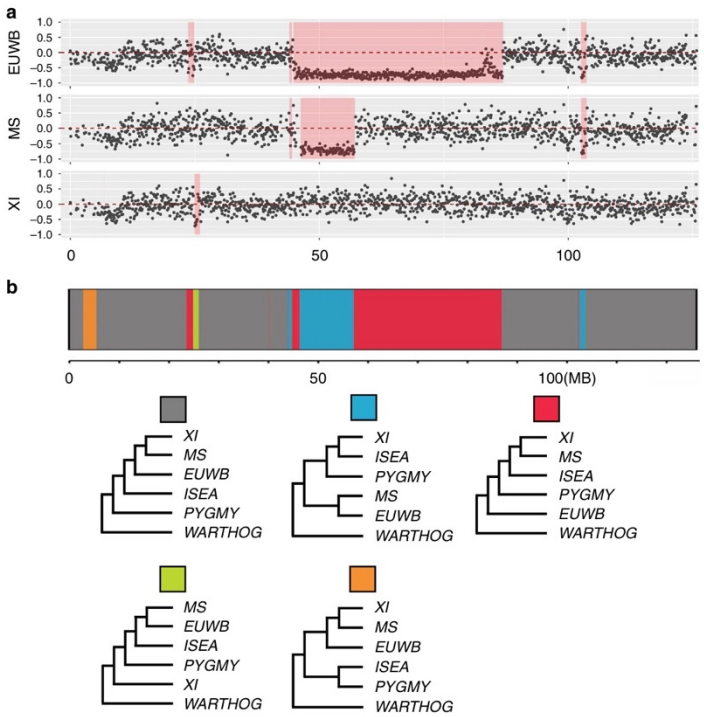


Fig. 3 Genealogical discordance in Chromosome X. **a** The D-statistic for testing introgression for 100-kb windows on chromosome X for tree topology (((ISEA, X), pygmy hog), warthog) (X = EUWB, MS, and XI). Excesses of BABA in 46–86 Mb indicated higher genetic similarity between ISEA and pygmy hog. See supplementary Figs. 9–14 for D-statistic testing introgression for 100-kb windows on autosome chromosomes. **b** SAGUARO plot illustrating the distribution of the five frequent rooted local topologies over the X chromosome. The red unrooted topologies support introgression from a missing linage into European/North China wild boar. Note that the blue unrooted topology includes the discrepant haplotype within Chinese population (see text for further explanation). ISEA = Islands of Southeast Asia Sus; MS = Meishan; XI = Xiang; EUWB = European wild boar; See Supplementary data 3 for further details and supplementary Figure 8 for results for autosomal chromosomes

ancestor of wild boar and pygmy hog was inferred. In our full model, signatures of admixture in the ISEA population were also examined and significant gene-flow between the ISEA *Sus* and Asian wild boar was found in agreement with previous analyses[21]. Furthermore, a model assuming a basal ghost population was applied and confirmed a post-speciation gene-flow between the common ancestor of wild boar and the ghost population (Supplementary data 4, Supplementary Fig. 23, Supplementary Note).

Testing for ILS. It is well known that incomplete ancestral linage sorting can inflate admixture signals. Therefore, we used two methods to test for ILS in our data. First, we calculated calculate the maximum length of a shared haplotype by pygmy hog and wild boar owing to ILS (probability < 0.005). With the deep divergence split between pygmy hog and wild boar (at least 4.2 My), the estimated length of a shared haplotype owing to ILS is extremely small (< 688 bp) and significantly different from the window size used in our sliding-window *D*-statistic analysis (100 kb). Furthermore, in our Saguaro analysis, we also filter out all the segment having alternative topology with a length ≤ 688 bp.

Second, we also follow the approach described in Huerta-Sánchez's[27] to assess the probability of ILS. We simulated 10,000 loci with length of 100 kb under the model described in Supplementary Note, and calculated *D*-statistics with the same quadruplets we used in the sliding-window analysis. All results from the simulations resulted in $P < 0.001$ against ILS for all comparisons (Supplementary Fig. 25). Thus, we conclude that it is unlikely that ILS have contributed significantly to our observed introgression signal.

Identification and functional analysis of introgressed genes. The different introgression signals that we observe in *Sus scrofa* could have played an important role in its successful expansion. We therefore accessed the functional annotation of the genes overlapping introgressed regions. However, given our low-coverage unphased genomic data sets and limited sample size, our ability to reveal ultra-short introgression segments broken down by long-term recombination is limited. With this in mind,

we undertook a functional annotation analysis for candidate introgressed genes. For the introgressed pygmy hog genes in wild boars (384 genes), enrichment for GO terms related to the sensory perception of taste, olfactory pathways and participating in glycolysis and fatty-acid metabolism was observed (Supplementary Fig. 26, Supplementary data 5). This finding is in agreement with the knowledge that smelling, taste, and energy metabolism pathway do have specific roles in adaptive capacity to environment[48–50]. However, it should be noted that especially olfactory genes are prone to copy number variation making them consistent enriched in such analysis. The Ghost introgression genes (104 genes) are related to more broad GO terms including neurogenesis, immune response and TCA cycle (Supplementary data 5). Of the Ghost introgressed genes, the POR gene is of most interest as it is involved in Vitamin D metabolism[51], which could potentially boost the fitness during the expansion of wild boar from sun-belt region (Southern Asia) to short day-length region (Northern Asia and European).

Discussion

Our analyses reveal the phylogeny and diversification times of the *Suidae* family (Fig. 2). The demographic analysis suggests at least three independent events of inter-species gene-flow during *Suidae* evolution—the most notable from an ancient and now-extinct linage (Fig. 3 and Supplementary Fig. 23). Combined, these results allow us to dramatically refine the evolutionary history of the *Suidae* family. After the Sub-Saharan suids evolved during the late Miocene (~ 10.2 Mya, Fig. 1d), the divergence between pygmy hog and *Sus* took place around the Miocene/Pliocene boundary (~ 6.1 Mya, Fig. 1e), followed by the emergence of ISEA *Sus* and wild boar (Fig. 1f). At ~ 1 Mya, populations of wild boar from Asia started to spread and reach Europe ~ 0.8 Mya (Fig. 1g, h). During this migration, wild boar colonized Eurasia and efficiently replaced all but one of the local species along the way[23,24], with pygmy hog as the only survivor. Moreover, during this the expansion, despite long divergences (~ 2 My between wild boar and pygmy hog, even longer for the ghost linage), wild boar hybridized with both pygmy hogs and an extinct, more divergent, *Suinae* species (Fig. 1h). The frequent climatic fluctuations during the Pleistocene led to alternating warm and cold periods (ices ages)[52,53], which likely resulted in multiple rounds of north-south directed migration[54]. Although expanding instantly, wild boar had greater chances of encountering and temporal co-existing with local species, enabling possible inter-species hybridization. Although our knowledge of the impact of admixture on the fitness of expanding populations is still limited, it is likely that changes in the genetic architecture that arise from admixture will generate heterosis that could boost adaptation to local niches (i.e., high grassland for pygmy hog, high altitude for ghost lineage).

Here, we have shown that an effective replacement of species is accompanied by consistently absorption of part of gene pool of the local related species. This suggests that admixture may play a role as an evolutionary biological driving force in successful range expansion and provides pertinent evolutionary hypothesis on the model of massive species replacement. With the booming development of paleogenomics technology, several case studies have directly verified ancient gene-flow between genomes of extinct species and extent recipient species[27,55]. Future studies, where the genome of fauna from early/middle Pleistocene remain is retrieved, will probably further refine the *Sus scrofa* expansion from Asia to Europe. Overall, the demographic history of pig species not only demonstrates how explosive and invasive range expansion can be, but also reminds us of the ubiquity of inter-species hybridization during speciation.

Methods

Sampling, genome sequencing, alignment, and SNP calling. The pygmy hog used for this research consists of three individuals sampled from the wild and three individuals from captivity. Whole-genome Illumina PE 100 bp re-sequencing was performed at SciGenom Laboratories in Chennai, India on these six pygmy hog samples. The *Babyrousa babyrussa* was sampled from Copenhagen Zoo. Libraries of ~ 300 bp fragments were prepared using Illumina paired-end kits (Illumina, San Diego, CA) and 100 bp paired-end sequenced with Illumina HiSeq. A selection of published genome from other *Suinae* species was included (Supplementary data 1). All these samples were also sequenced with the Illumina sequence technology. The whole-genome sequencing data were trimmed using sickle (https://github.com/najoshi/sickle) with default parameters. The trimmed reads were aligned to the Sscrofa 11.1 reference genome using the Burrows-Wheeler Aligner (BWA 0.7.5a)[56]. Local re-alignment was performed using GATK v3.6.0 RealignmentTargetCreator and IndelRealigner and variants were called using GATK UnifiedGenotyper[57], with the –stand_call_conf option set to 50, the –stand_emit_conf option set to 20, and the −dcov option set to 200. Variants with a read-depth between 0.5 and 2.0 times of the average sample genome coverage were selected and stored in variant calling format. We identified the sex of all individuals by calculating the ratio of read-depth on X chromosome and the autosomes. For the individuals whose molecular sex are male, we filtered out variants in the non-PAR regions, which are heterozygous and with a coverage larger than the average read-depth in autosomes. We do not have any explicit pedigree for the *Babyrousa Babyrussa* sample. To avoid potential biases caused by recent interbreeding, we did decide to used warthog as the outgroup in all the analyses related to introgression. Pygmy hog samples were collected within the Pygmy Hog Conservation Programme in Assam, India in accordance with ethical and legal regulations in India. The *Babyrousa babyrussa* was sampled from a dead individual in Copenhagen Zoo in accordance with ethical and legal regulations in Denmark (The Animal Experimentation Act LBK no. 253, March 8th 2013). This study was ethically approved by the European Research Council under the European Community's 256 Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement no. 249894".

Phylogenetic analysis. Phylogenetic trees on autosomes were construct based on the maximum-likelihood (ML) method as implemented in RAxML 8.2.3[58] using the best-fitting model of substitutions, identified by jModelTest2[59] on 100 random subsets of 1 Mbp. In order to eliminate possible bias stemming from alignment and genotyping errors, we only used autosome one-to-one orthologous gene coding sequences (CDS)[60] between pig and cow for this analysis. A list of one-to-one orthologous genes (between cow and pig) and coordinates of corresponding one-to-one CDS region were extracted from ENSEMBL with biomart[61]. Finally, we got 486,203 CDS regions from 18,313 genes. The total number of SNPs in the one-to-one gene regions was 2571419. We used both supermatrix and supertree techniques[62], using *Babyrousa babyrussa* as an outgroup. In the supermatrix approach, the concatenated CDS alignment was analyzed under best fitting substitution model (GTR + Γ + I) with 100 bootstrap replications. In the supertree approach, pig-to-cow orthologous genes with CDS alignments longer than 1000 bp were used. Individual gene trees were inferred separately under GTR + Γ + I substitution model with 100 rapid bootstraps. All gene trees with an average bootstrap value above 40 were combined into a consensus tree using the software ASTRAL-II[63]. To assess support for particular clades in the supertree analysis, we calculated concordance factors in DensiTree[64].

RAxML 8.2.3 was used to reconstructed ML phylogenetic trees on the whole X chromosome and on the two regions, which have anomalous phylogenetic relationship in the SAGUARO analysis. For mitochondria DNA analysis, we used a Bayesian Markov Chain Monte Carlo simulation (MCMC) to estimate the most likely phylogenetic trees with MrBayes 3.2.3[65], using the best fitting model of substitutions, identified by JModelTest2. The length of the MCMC was set to 10,000,000. The parameter estimates and consensus trees resulting from 10 MrBayes runs were recorded and compared. The best-supported phylogenetic consensus tree was summarized with TRACER v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) discarding the first 10% as burn-in. All trees were depicted using the software FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Molecular clock analyses. We estimated divergence times using an approximate likelihood method as implemented in MCMCtree[66], with an independent relaxed-clock and birth-death sampling. To overcome difficulties arising from computational efficiency and admixture, we only used CDS with pig-to-cow ortholog filter. We fitted an GTR + Γ4 model to each genomic bin and estimated a mean mutation rate by fitting a strict clock to each fragment setting a root age at 20 Mya, which represent the earliest *Tayasuidae* fossil. This mean rate was used to adjust the prior on the mutation rate (rgene) modeled by a gamma distribution as G (1,241). The BDS and sigma2 values were set at 7 5 1 and G (1,10) respectively. We ran 100,000 (25% burn in) MCMC samples for fossil calibration reported previously. We used a float prior and a maximum bound age, with a scale parameter of $c = 2$. Allowing the MCMC to explore a wide range of time for the divergence between *Babyrussa* and *Suinae* and calibrate the time later than the split of "new world" peccaries (tU = 2 [20 Mya], $p = 0.1$, $c = 2$). For MRCA of sub-Saharan African suid and *Sus*, we used the same fossil calibration as in Frantz et al. 2013 (tL = 0.55 [5.5 Mya], $p = 0.9$, $c = 0.5$). For MRCA of *Sus*, we used a minimum bound at 2 Ma [tL = 0.2 [2

Mya], $p = 0.1$, $c = 0.5$] to represents the earliest appearance of *Sus* in the fossil record of Island Southeast Asia.

Detecting gene-flow among *Suidae*. We integrated Patterson's *D*-statistic to examine the phylogenetic distribution of derived alleles at loci that display either an ABBA or BABA allelic configuration across the chromosomes among *Suidae* using warthog as an outgroup. For admixture estimation, we assigned 18 autosomes and selected a block size of 5 Mb to calculate the standard errors on *D*-statistics using Admixtools[45]. We also identified candidate introgression loci using *D*-statistic and fd in slide window[67]. To avoid D returning inflated values in small genomic regions[68], we set the window size as 100-Kb and summarized the results in Venn diagrams.

SAGUARO. Phylogenetic relationships of genomic regions may differ from the species tree due to incomplete linage sorting and introgression. To test whether this is the case in our analysis, and to detect breakpoints between genomic segments supporting different local topologies, we used the machine-learning approach implemented in SAGUARO[42]. We first ran SAGUARO with six representative individuals for the overview of whole genome (See Supplementary data 3). Then, to better estimate length of phylogenetic incongruent regions, we performed the same approach but using the quadruplets in sliding-window *D*-statistic analysis. We constrained SAGUARO to use only nucleotide positions with no missing data and ran with 20 iterations and 500 neurons.

Fitting models of population history. We used qpGraph[45] to fit admixture graphs to six populations representing European wild boar, North Chinese wild boar, South Chinese wild boar, ISEA, pygmy hog, and warthog as the outgroup. We filter the data set using following criteria: SNPs with at least 10 Kb distance from one another, no more than 10% missing data. This resulted in 361,837 SNPs. To explore the space of all possible admixture graphs, we used a heuristic search algorithm first described in Leathlobhair et al.[46] (code available at https://github.com/ekirving/qpbrute). Given an outgroup with which to root the graph, a stepwise addition order algorithm was used for adding leaf nodes to the graph. At each step, insertion of a new node was tested at all branches of the graph, except the outgroup branch. Where a node could not be inserted without producing f4 outliers (i.e., $|Z| \geq 3$) then all possible admixture combinations were also attempted. If a node could not be inserted via either approach, that sub-graph was discarded. If the node was successfully inserted, the remaining nodes were recursively inserted into that graph. All possible starting node orders were attempted to ensure full coverage of the graph space. We fitted 2444 unique admixture graphs for these six populations and recorded the 37 graphs that left no f4 outliers (i.e., $|Z| < 3$). We then used the MCMC algorithm implemented in the ADMIXTUREGRAPH R package[47] to compute the marginal likelihood of these 37 models and their Bayes Factors (BF). We ran two independent replications, each with two million iterations, five heated chains, a burn in of 50%, and no thinning. Convergence was assessed by calculating the potential scale reduction factor for the model likelihoods using the CODA R package[69]. We found one particularly well supported model (1d9676e) which, when compared with all others, had $K < 119$ (Supplementary Figs. 20–21). We also note that among the remaining models there are small differences in the admixture topologies, however, most models support gene-flow between ISEA *Sus* and Chinese pigs, as well as between the pygmy hog and basal wild boar (Supplementary Fig. 21).

We further carried out demographic analysis based on the G-PhoCS, applied to 10,000 loci of 1 kb of length chosen via a series of filter to obtain putatively neutral loci. We filtered out exons of protein coding genes and 10 kilobases (kb) flanking them on each side, as well as conserved noncoding elements and 100 bases on each side of these elements. We selected 1 kb loci located at least 30 kb apart. We identified a collection of 11,274 loci that followed these criteria, and random selected 10,000 loci for the G-PhoCS. Multiple sequence alignments for these loci were extracted using sequence data from the all individual genomes. We conditioned inference on the population phylogeny based upon the neighbor-joining tree constructed with MEGA[70] on the basis of the IBS distance matrix data of neutral loci used in G-PhoCS analysis calculated by PLINK 1.9[71] (Supplementary Fig. 24). The prior distributions over model parameters was defined by a product of Gamma distributions with $\alpha = 1$ and $\beta = 10{,}000$ for population size and divergence time scaled by mutation rate, and $\alpha = 0.002$ and $\beta = 0.00001$ for the migration rates. Markov Chain was run for 100,000 burn-in iterations, after which parameter values were sampled for 200,000 iterations every 10 iterations, resulting in a total of 20,001 samples from the approximate posterior. Convergence was inspected manually for each run (effective sample size for all parameters > 200). We converted probabilities into rates using the formula $p = 1 - e^{-m}$ (where p is the probability of gene-flow and m is the total migration rate)[72]. We checked for convergence between runs using Tracer v1.7[73].

Finally, we also used Treemix v1.12[74] to test models of possible admixture for Babyrussa, warthog, pygmy hog, ISEA, European pigs, Northern China pigs, and Southern China pigs. Windows with 500 consecutive SNPs were used to account for the non-independence of SNPs located in close vicinity. Migrations from m0 to m10 were tested, with five replicates per m to assess consistency.

Probability of introgression fragment from shared ancestral lineage. Based on the equation described in Huerta-Sánchez et al.[27], we calculate the probability of a haplotype shared by pygmy hog and wild boar as a result of ancestral ILS. In brief, let k be the introgressed haplotype length of the two species' branches since divergence. The expected length of a shared ancestral sequence is $L = 1/(r \times t)$, where r is the recombination rate and t is branch lengths of pygmy hog and wild boar since divergence. The probability of a length of at least k is 1-GammaCDF(k, shape = 2, $L = 1/L$), where GammaCDF is the Gamma distribution function. The lower estimate of 4.2 My of the Sus–pygmy hog was used as branch length and an assumed generation time of 5 years. The recombination rate was set to 0.8 cM/Mb[75].

Another approach to assess probability of ILS is comparing *D*-statistics between populations under simulations of demographic model. However, it requires a very detailed and precise demographic model to obtain a better assessment. The historical demographic information of pygmy hog and the ancestral population of Suinae species are still deficient. Here, we can only fit in a simplified model. Inaccuracy of the effective ancestral population size and bottleneck event may lead to over-/underestimation of ILS. With this in mind, we compared *D*-statistics with the same quadruplet as we used in the sliding-window steps under simulations of a simple demographic model with no gene-flow. (see Supplementary Note for details). All simulations resulted in $P < 0.001$ against ILS for all comparisons (see Supplementary Fig 25 and Supplementary Note).

Functional annotation of genes in introgressed regions. We applied a functional annotation analysis using PANTHER v.11[76] on the candidate introgressed genes. Genes from the pygmy hog/Sus scrofa introgression and the Sus scrofa/ghost lineage introgression were analyzed separately. Gene-enrichment analyses were performed using clusterProfiler[77]. False discovery rate was performed to adjust *P* values using the Benjamini and Hochberg method. A *P* value of < 0.05 was used as the cutoff criterion.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary Information files, or from the corresponding author upon request. Raw reads of pygmy hog and *Babyrousa babyrussa* have been deposited in the European Nucleotide Archive (ENA) under accession PRJEB30129. Sequences for the sequenced *Sus scrofa* and other species have been deposited on the EBI Sequence Read Archive under accession number ERP001813.

Received: 9 October 2018 Accepted: 15 April 2019
Published online: 30 April 2019

References

1. Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83–109 (1996).
2. Nuijten, R. J. M. et al. The use of genomics in conservation management of the endangered visayan warty pig (Sus cebifrons). *Int. J. Genomics* **2016**, 1–9 (2016).
3. Kolbe, J. J., Larson, A., Losos, J. B. & de Queiroz, K. Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biol. Lett.* **4**, 434–437 (2008).
4. Lawson Handley, L.-J. et al. Ecological genetics of invasive alien species. *BioControl* **56**, 409–428 (2011).
5. Munz, E. D. Psychotherapie in der Psychiatrie. *Nervenheilkunde* **36**, 800–805 (2017).
6. Groves, C. *Ancestor for the pigs: taxonomy and phylogeny of the genus Sus. Technical Bulletin by the Department of Prehistory, Research School of Pacific Studies*, Australian National University 3, (Department of Prehistory, Research School of Pacific Studies, Australian National University, 1981).
7. Hardjasasmita, H. S. Taxonomy and phylogeny of the Suidae (Mammalia) in Indonesia. *Scripta Geol.* **85**, 1–68 (1987).
8. Oliver, W. L. R. Taxonomy and conservation of Asian wild pigs. *Asian Wild Pig News* **1**, 3–5 (2001).
9. Groenen, M. A. M. et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012).
10. Pickford, M. Revision of the Miocene Suidae of the Indian subcontinent. *Muenchner Geowiss. Abh.* **12**, 1–92 (1988).
11. Andrews, P. Cainozoic paleontological sites in Western Kenya. *J. Hum. Evol.* **17**, 273 (1988).
12. Pickford Senut, B., Hadoto, D., M. Geology and Palaeobiology of the Albertine Rift Valley, Uganda - Zaire - Volume I: Geology. 24, (Cifeg, France, 1993).
13. Made, Jvander Biometrical trends in the Tetraconodontinae, a subfamily of pigs. *Trans. R. Soc. Edinb. Earth Sci.* **89**, 199–225 (1999).
14. Frantz, L. et al. The evolution of suidae. *Annu. Rev. Anim. Biosci.* **4**, 61–85 (2016).

15. Frantz, L. A. F. et al. Synchronous diversification of sulawesi's iconic artiodactyls driven by recent geological events. *Proc. R. Soc. B Biol. Sci.* **285**, 20172566 (2018).
16. Zachos, J., Pagani, H., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
17. Roth, J. & Wagner, J. A. Die fossilen Knochenüberreste von Pikermi in Griechenland: Gemeinschaftlich bestimmt u. beschrieben nach d. Materialien, welche durch die von dem Erstgenannten im Winter 1852/3 dortselbst vorgenommenen Ausgrabungen erlangt wurden. 7, (Verlag d. Akad., 1854).
18. Geraads, D., Spassov, N. & Garevski, R. New specimens of *Propotamochoerus* (Suidae, Mammalia) from the late Miocene of the Balkans. *N. Jb. für Geol. Paläontol. Abh.* **248**, 103–113 (2008).
19. Orliac, M. J., Pierre-Olivier, A. & Ducrocq, S. Phylogenetic relationships of the Suidae (Mammalia, Cetartiodactyla): New insights on the relationships within Suoidea. *Zool. Scripta* **39**, 315–330 (2010).
20. van der Made, J., Morales, J. & Montoya, P. Late Miocene turnover in the Spanish mammal record in relation to palaeoclimate and themessinian salinity crisis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **238**, 228–246 (2006).
21. Frantz, L. A. F. et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol.* **14**, R107 (2013).
22. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
23. Fistani, A. B. Sus scrofa priscus (Goldfuss, de Serres) (Mammalia, Artiodactyla, Suidae) from the Middle Pleistocene layers of Gajtan 1 site, southeast of Shkoder (North Albania). *Ann. Paliontologie* **82**, 177–229 (1996).
24. Guérin, C. & Faure, M. The wild boar (Sus scrofa priscus) from the post-Villafranchian lower Pleistocene of Untermassfeld. *Das. Pleistozän von. Unter. bei Meining.* **1**, 375–384 (1997).
25. Frantz, L. A. F., Madsen, O., Megens, H. J., Groenen, M. A. M. & Lohse, K. Testing models of speciation from genome sequences: Divergence and asymmetric admixture in Island South-East Asian Sus species during the Plio-Pleistocene climatic fluctuations. *Mol. Ecol.* **23**, 5566–5574 (2014).
26. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
27. Huerta-Sánchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
28. Slon, V. et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113–116 (2018).
29. Breeding, C., Of, R. & Endangered, C. CRITICALLY ENDANGERED PYGMY HOG (Porcula salvania) (2013).
30. PHCP. Conservation Strategy and Action Plan for Pygmy Hog in Assam EcoSystems-India. (2008).
31. von Koenigswald, G. H. R. Fossil Pygmy Suidae from Java and China. *Proc. Ser. B* **66**, 192–197 (1963).
32. Pickford, M. Suids from the Pleistocene of Naungkwe Taung, Kayin State, Myanmar. *Paleontol. Res.* **16**, 307–317 (2013).
33. Han D F. Artiodactyla fossils from Liucheng *Gigantopithecus* cave in Guangxi (in Chinese). Memoirs of Institute of Vertebrate Paleontology and Paleoanthropology, Academia Sinica, No.18. Beijing: Science Press, 1987. 135–208.
34. Han, D.F. Quaternary mammalian fossils from Bijiashan, Luizhou, Guangxi. *Vert. PalAs.* **13**, 250–256 (1975).
35. Funk, S. M. et al. The pygmy hog is a unique genus: 19th century taxonomists got it right first time round. *Mol. Phylogenet. Evol.* **45**, 427–436 (2007).
36. Yannic, G., Dubey, S., Hausser, J. & Basset, P. Additional data for nuclear DNA give new insights into the phylogenetic position of Sorex granarius within the Sorex araneus group. *Mol. Phylogenet. Evol.* **57**, 1062–1071 (2010).
37. Nakagome, S., Pecon-Slattery, J. & Masuda, R. Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. *Mol. Biol. Evol.* **25**, 1344–1356 (2008).
38. Ropiquet, A. & Hassanin, A. Hybrid origin of the Pliocene ancestor of wild goats. *Mol. Phylogenet. Evol.* **41**, 395–404 (2006).
39. Chan, K. M. A. & Levin, S. A. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**, 720–729 (2005). http://pygmyhog.org/wp-content/uploads/2014/01/Conservation-breeding-and-reintroduction-of-pygmy-hog_2013.pdf
40. Ai, H. et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
41. Smith, J. & Kronforst, M. R. Do Heliconius butterfly species exchange mimicry alleles? *Biol. Lett.* **9**, 20130503–20130503 (2013).
42. Zamani, N. et al. Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics* **14**, 347 (2013).
43. Ma, J. et al. Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. *BMC Genomics* **11**, 13 (2010).
44. Fernández, A. I. et al. Recombination of the porcine X chromosome: a high density linkage map. *BMC Genet.* **15**, 148 (2014).
45. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
46. Leathlobhair, M. N. et al. The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).
47. Leppälä, K., Nielsen, S. V. & Mailund, T. Admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017).
48. Kishida, T., Thewissen, J., Hayakawa, T., Imai, H. & Agata, K. Aquatic adaptation and the evolution of smell and taste in whales. (2015). https://doi.org/10.1186/s40851-014-0002-z.
49. Paudel, Y. et al. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics.* **14**, 449 (2013).
50. Chevin, L.-M., Lande, R. & Mace, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive Theory. *PLoS Biol.* **8**, e1000357 (2010).
51. Larson-Meyer, D. E. et al. Sun exposure in pigs increases the vitamin D nutritional quality of pork. *PLoS ONE* **12**, e0187877 (2017).
52. Hansen, J., Sato, M., Russell, G. & Kharecha, P. Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**, 20120294 (2013).
53. Guo, Z. T., Peng, S. Z., Hao, Q. Z., Biscaye, P. E. & Liu, T. S. Origin of the miocene - Pliocene Red-Earth formation at Xifeng in northern China and implications for paleoenvironments. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 11–26 (2001).
54. Seebacher, F. & Post, E. Climate change impacts on animal migration. *Clim. Change Responses* **2**, 5 (2015).
55. Barlow, A. et al. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
57. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
58. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
59. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. JModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
60. Frantz, L. A. F. *Speciation and Domestication in Suiformes: a Genomic Perspective.* (Wageningen University, 2015).
61. Kinsella, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011**, bar030–bar030 (2011).
62. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).
63. Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44–i52 (2015).
64. Bouckaert, R. R. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372–1373 (2010).
65. Ronquist, F. et al. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
66. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
67. Rosenzweig, B. K., Pease, J. B., Besansky, N. J. & Hahn, M. W. Powerful methods for detecting introgressed regions from population genomic data. *Mol. Ecol.* **25**, 2387–2397 (2016).
68. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
69. Best, N. G. & Cowles, M. K. CODA: convergence diagnosis and output analysis software for Gibbs sampling output. *MRC Biostat. Unit. Cambridge University* **6**, 7–11 (1997).
70. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
71. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
72. VonHoldt, B. M. et al. Whole-genome sequence analysis shows that two endemic species of North American Wolf are admixtures of the coyote and gray Wolf. *Sci. Adv.* **2**, e1501714–e1501714 (2016).
73. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **00**, 1–3 (2018).
74. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
75. Tortereau, F. et al. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* **13**, 586 (2012).

76. Westbury, M. V. et al. Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the Brown Hyena. *Mol. Biol. Evol.* **35**, 1225–1237 (2018).
77. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).

Acknowledgements

L.L. has received financial support from the China Scholarship Council (Grant No. 201707720055). This project was financially supported by a European Research Council grant (ERC-2009-AdG: 249894). Sample of *Babyrousa babyrussa* was kindly provided by professor Dr. M. Fredholm, University of Copenhagen.

Author contributions

M.A.M.G., O.M., and L.L. designed the study. M.A.M.G., O.M., and H.-J.M. initially conceived the project; G.N. provided the pygmy hog samples; L.L. analyzed the data; Y.-L.L. and L.L. performed the phylogenetic analyses; E.K.I.-P. preformed the qpgraph analyses; L.L., M.B., H.-J.M., and O.M. discussed the results; L.L. wrote the manuscript; M.B., H.-J.M., L.A.F.F., M.A.M.G., and O.M provided valuable suggestion and comments to improve the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10017-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks Mingzhou Li and other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

In this paper, I assisted in the spatial modelling of the ancient and modern pig data (see Figure 1), and assisted in the phylogenetic analyses of the selective sweep loci.

PNAS

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[j], 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[j], 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

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

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.

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.

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

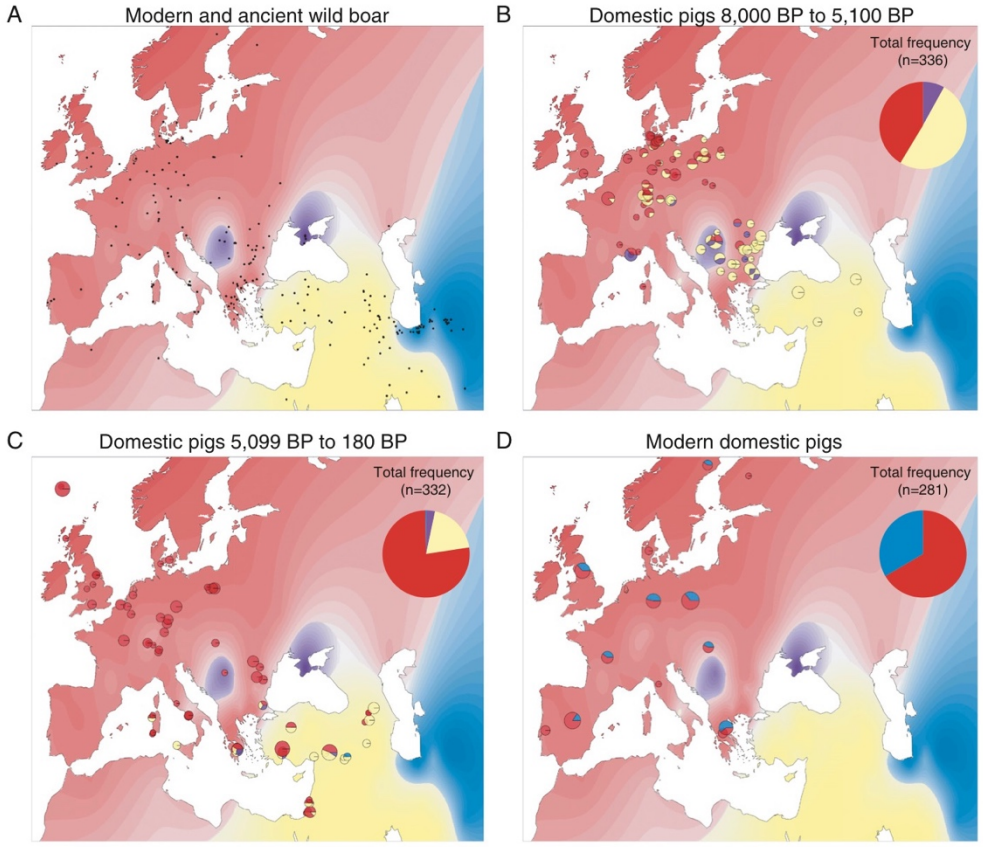


Fig. 1. (*A*) Map representing the distribution of East Asian (blue), Near Eastern (including haplogroups mt-Y1 and mt-ArmT; yellow), European (including haplogroups mt-Italian, mt-A, mt-C, and mt-Y2; red), and Y2 (purple) haplogroups in wild boars. Black dots represent the locations of 696 modern and 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.

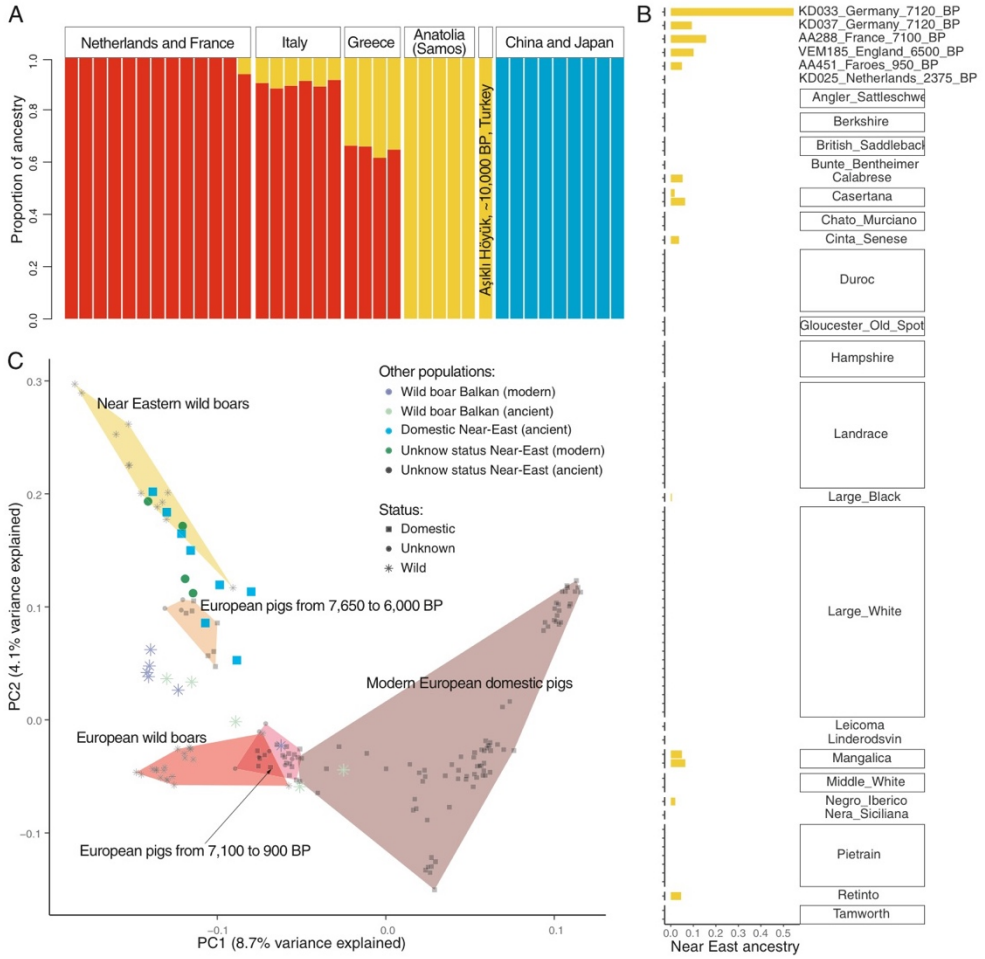


Fig. 2. (*A*) Bar plots representing the proportion of ancestry from Europe (red), the Near East (yellow), and East Asia (blue) in Eurasian wild boar genomes. (*B*) Bar plots depicting the proportion of Near Eastern ancestry in modern and ancient European domestic pigs. (*C*) PCA (excluding East Asian 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).