Functional adaptation in Toll-like receptors of wild and domesticated animals



Steven Robert Fiddaman

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Abstract

Functional adaptation in Toll-like receptors of wild and domesticated animals

Steven R Fiddaman, *Doctor of Philosophy*, Trinity 2021

Animals sense their pathogens through detection of conserved pathogen-associated molecules by pattern-recognition receptors (PRRs). The best characterised family of PRRs is the Toll-like receptors. TLR subfamilies retain specificity for particular agonist classes over great evolutionary distances. This does not, however, mean that TLRs are evolutionarily static. TLRs, along with other pathogen-facing immune genes, frequently undergo directional selection to maintain leverage in the perpetual arms race between hosts and pathogens.

In this thesis, I argue that TLRs are critical determinants of disease susceptibility and resistance, and are therefore under strong selective pressures. To support this, I provide evidence of adaptive evolution operating over multiple nested timescales. For instance, I show that, even in the few thousand years since the domestication of chickens, TLRs have undergone adaptation, likely to better deal with higher pathogen loads in denser, commercialised populations. At the other extreme (hundreds of millions of years), I document the repurposing of the CpG-DNA-recognising TLR21 between fishes and birds, arguing that this compensated for the loss of the unrelated CpG-recognising TLR9 in the birds and reptiles.

To contextualise and reinforce my bioinformatic findings, I provide evidence from laboratory experiments to confirm changes in TLR function. For instance, I demonstrate that two positively selected sites in Gentoo penguin *TLR5* are sufficient to change the responsiveness of the receptor to its agonist, flagellin. Furthermore, I demonstrate that the ancestor to the *Eudyptes* (crested) penguins had a cryptic pseudogene for *TLR15* that was then overtly pseudogenised several independent times in the genus. Functional evidence was crucial in both of these conclusions, and neither

could have been deduced from bioinformatic methods alone.

Taken together, this body of work provides several lines of evidence that TLRs are key targets for natural selection across multiple evolutionary scales, and that functional evidence plays an important role in interpreting bioinformatic predictions. In addition to contributing to the fundamental understanding of how hosts cope with changing pathogen landscapes, this work provides important information for conservation and sustainable food production.

Acknowledgements

Firstly, I would like to thank Adrian Smith for co-supervising this project. The work presented in this thesis is a fraction of the projects we have undertaken in the last decade, which has been possible primarily through a freedom to investigate new areas and go where the science takes us. I thank Adrian for allowing me this freedom, for welcoming me back into the fold after time away, and for our continued friendship. I am also grateful for the valuable guidance of Laurent Frantz, my second co-supervisor. I thank Laurent for his insightful, critical appraisal of my work, for always finding the time to discuss a topic, and for believing in me enough to offer me a post-doc.

I would also like to acknowledge receipt of funding from the Biotechnology and Biological Sciences Research Council.

I also extend my thanks to my collaborators, particularly Hila Levy, Simon Spiro, Michal Vinkler and Tom Hart, with whom I have shared many useful discussions about this work.

I thank my lab colleagues, particularly Patrik, Rob, Stefan and Stephen who were always on hand to help out with overnight Maxiprep cultures.

I would like to thank my family, particularly my parents, Elaine and Paul, and grandparents, Sylvia, David and Elizabeth, for their ceaseless love and support. To my brother, Mark, and Freyja and Talitha: thank you for always reminding me that my work is interesting, giving me continued motivation.

I would like to give my heartfelt thanks to Jen. Having two children during a DPhil was never going to be easy, especially latterly during a pandemic. My completion of this thesis is in no small part down to the sacrifices you have made, both personally and professionally, for which I am eternally grateful. Finally, I would like to thank my children, Mia and Jacob. You are too young to understand TLRs, but your curiosity for the world around you and the adventures we have inspire me to carry on.

Statement of Contribution and Associated Publications

The work presented in this thesis includes research that has been published in a peer-reviewed journal and manuscripts that are currently in preparation and/or review. Below, I present a summary of my contributions to each chapter, and those of others. An asterisk or dagger in an author list denotes equal contribution.

Chapter 2

Hila Levy, **Steven R Fiddaman***, Juliana A Vianna, Daly Noll, Gemma V Clucas, Jasmine K H Sidhu, Michael J Polito, Charles A Bost, Richard A Phillips, Sarah Crofts, Gary D Miller, Pierre Pistorius, Francesco Bonnadonna, Céline Le Bohec, Andrés Barbosa, Phil Trathan, Andrea Raya Rey, Laurent A F Frantz, Tom Hart, Adrian L Smith, Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird, Molecular Biology and Evolution, Volume 37, Issue 6, June 2020, Pages 1708–1726, doi: 10.1093/molbev/msaa040*

I conducted PCRs and sequencing of *TLR5* from all samples in conjunction with JKHS—an undergraduate student—and assisted in sequencing of *TLR7*. HL carried out all remaining PCRs and sequencing. HL carried out all population genetic analyses. I performed all positive selection analyses and identification of functional sites. I performed all cloning and functional work. I generated all figures and wrote the paper with HL. ALS, LAFF and TH supervised the project and provided editorial assistance. All other authors provided samples and/or sequence data.

Chapter 3

***Steven R Fiddaman**, Michal Vinkler, Simon G Spiro, Hila Levy, Christopher A Emerling, Amy C Boyd, Evangelos A. Dimopoulos, Juliana A Vianna, Theresa L Cole, Hailin Pan, Miaoquan Fang, Guojie Zhang, Tom Hart, Laurent AF Frantz,*

Adrian L Smith, Adaptation and cryptic pseudogenization in penguin Toll-like receptors, Molecular Biology and Evolution, 2021, msab354, doi: 10.1093/molbev/msab354

I conceived the project and conducted all analyses except for homology modelling, which was done by MV. I wrote the paper and made all figures. LAFF, ALS and TH supervised the project and provided editorial assistance. All other authors provided samples or sequence data.

Chapter 4

Theresa L. Cole†, Chengran Zhou†, Miaoquan Fang†, Hailin Pan, Daniel T. Ksepka*, Daniel Thomas*,* ***Steven R. Fiddaman*****, Christopher Emerling*, Xupeng Bi, Qi Fang, Martin R. Ellegaard, Adrian L. Smith, Tracy Heath, Alan J. D. Tennyson, Pablo García Borboroglu, Jamie R. Wood, Charles-André Bost, Peter Hadden, Misha Vorobyev, Jason Turuwhenua, Yves Cherel, Thomas Mattern, Tom Hart, Mikkel-Holger S. Sinding, Lara D. Shepherd, Richard A. Phillips, Petra Quillfeldt, Juan F. Masello, Stefanie Grosser, Juan L. Bouzat, Peter Ryan, David Thomas, Ursula Ellenberg, Peter Dann, Gary Miller, P. Dee Boersma, M. Thomas P. Gilbert, De-Xing Zhang, Guojie Zhang, A comprehensive genomic history of the evolution of penguins. In review (Nature Communications).*

This chapter forms part of a larger paper that is about to be submitted. While the work contained within the chapter is entirely my own (with supervision from ALS), all other authors contributed to other parts of the final manuscript. My site-based selection analyses are being used as a basis for other parts of the paper concerning physiological adaptation in penguins.

Chapter 5

Steven R Fiddaman*, Adrian L Smith, Laurent AF Frantz. Diversity and evolution in chicken Toll-like receptors. Manuscript in preparation.*

LAFF curated accession numbers for publicly available sequences, remapped reads to Galgal6, generated VCFs, and oversaw the project. I conducted all downstream analyses, wrote the manuscript, and made all figures. ALS provided editorial assistance.

Chapter 6

Steven R Fiddaman *and Adrian L Smith, Functional repurposing of TLR21 following the loss of TLR9 in birds and reptiles. Manuscript in preparation.*

ALS and I came up with the concept for the paper. I conducted all analyses and wrote the paper with guidance from ALS.

Chapter 7

Adrian L Smith, Sarah C Hill and I came up with the concept for the chapter. I conducted Nanopore sequencing alongside SCH. I extracted all DNA for whole genome sequencing and conducted all analyses. SCH collected all samples for previously conducted work.

Additional information

Formal statements of contributions for published/submitted chapters can also be found at the end of the relevant chapter. Additional publications to which I have contributed over the course of my DPhil are given in Appendix G.

Contents

1 Introduction **1**

1.1 Early evolution of the Toll-like receptors 2

1.2 TLR expansion and contraction in the invertebrates 3

1.3 Phylogenetic relationships between TLRs 4

1.4 Agonist specificities of TLRs 5

1.5 TLR signalling and interaction with the adaptive immune system . . 10

1.6 Subcellular localisation and regulation of TLRs 14

1.7 Involvement of TLRs in disease 16

1.8 Diversity and selection in TLRs 17

1.9 Other pattern-recognition receptors 18

1.10 TLRs in nonmodel species: the value of the comparative approach . . 20

1.11 Theory and tools of positive selection analysis 21

1.12 Structure of the thesis 23

2 Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird **27**

2.1 Abstract 27

2.2 Introduction 28

2.3 Materials and methods 31

2.3.1 Sample collection 31

2.3.2 DNA extraction 32

2.3.3 TLR genotyping 33

2.3.4 mtDNA genotyping . . . 34
2.3.5 Population-level analyses . . . 35
2.3.6 Population divergence and phylogeography . . . 37
2.3.7 Selection analyses . . . 37
2.3.8 *In silico* prediction of polymorphism functional consequences . 38
2.3.9 Functional analysis of *TLR5* genotype expression in CRISPR-Cas9 edited HEK-Blue cells . . . 38
2.3.10 Data availability . . . 39
2.4 Results . . . 40
2.4.1 Amplification of TLR genes in the gentoo penguin . . . 40
2.4.2 Diversity indices and population differentiation . . . 40
2.4.2.1 Mitochondrial hypervariable region . . . 40
2.4.2.2 *TLR4* . . . 41
2.4.2.3 *TLR5* . . . 44
2.4.2.4 *TLR7* . . . 46
2.4.3 Effect of population size on diversity . . . 47
2.4.4 Isolation by distance . . . 47
2.4.5 Analysis of positive selection . . . 47
2.4.6 Functional analysis of selected TLR5 residues . . . 50
2.5 Discussion . . . 55
2.5.1 Patterns of diversification and selection . . . 55
2.5.2 Functional polymorphisms in *TLR5* support positive selection 58
2.5.3 Potential drivers of selection . . . 60
2.5.4 Concluding remarks . . . 63

3 Adaptation and cryptic pseudogenization in penguin Toll-like receptors 66

3.1 Abstract . . . 66
3.2 Introduction . . . 67

3.3 Materials and methods . . . 69
3.3.1 Sequence retrieval . . . 69
3.3.2 Alignments and polymorphism identification . . . 70
3.3.3 Positive selection analysis . . . 71
3.3.4 Homology modelling . . . 71
3.3.5 Phasing of TLR15 haplotypes . . . 72
3.3.6 Fungal lysate preparation . . . 72
3.3.7 Cloning of TLR15 constructs . . . 72
3.3.8 Transient transfection and stimulation . . . 73
3.3.9 Immunoblotting . . . 73
3.4 Results . . . 74
3.4.1 Toll-like receptor and pseudogene identification . . . 74
3.4.2 Patterns of diversity in penguin Toll-like receptors . . . 75
3.4.3 Positive selection in penguin Toll-like receptors . . . 75
3.4.4 Pseudogenization of *TLR15* . . . 84
3.4.5 Functional analysis of full-length TLR15 . . . 89
3.4.6 Loss-of-function mutation . . . 90
3.5 Discussion . . . 90
3.5.1 Penguin Toll-like receptors have undergone diversifying selection at key functional sites . . . 91
3.5.2 Pseudogenization of *TLR5* in the Snares Crested penguin . . . 92
3.5.3 Cryptic pseudogenization in *Eudyptes* spp. *TLR15* . . . 93
3.6 Conclusions . . . 97

4 A genome-wide analysis reveals extensive positive selection at immune loci in penguins 99

4.1 Abstract . . . 99
4.2 Introduction . . . 100
4.3 Methods . . . 102

4.3.1 Protein-coding gene annotation . . . 102
4.3.2 Gene function annotation and orthologue alignment . . . 102
4.3.3 Selection analysis . . . 103
4.3.4 Immune gene identification . . . 103
4.3.5 Homology modelling . . . 103
4.4 Results and Discussion . . . 104
4.4.1 Positively selected immune genes . . . 104
4.4.1.1 Toll-like receptors and adaptors . . . 104
4.4.1.2 *IFIT5* . . . 106
4.4.1.3 *CD81* . . . 107
4.4.1.4 Transferrin . . . 108
4.4.1.5 *PIGR* . . . 108
4.4.1.6 Other positively selected genes . . . 109
4.4.2 Conclusions . . . 109

5 A comprehensive description of diversity in chicken Toll-like receptors 112

5.1 Abstract . . . 112
5.2 Introduction . . . 113
5.3 Methods . . . 115
5.3.1 Data gathering and assembly pipeline . . . 115
5.3.2 Variant discovery pipeline . . . 115
5.3.3 Phasing TLR loci . . . 116
5.3.4 Phylogenetic reconstruction . . . 116
5.3.5 Variant identification . . . 117
5.3.6 Neutrality and diversity statistics . . . 117
5.3.7 Genome-wide Tajima's D comparison . . . 117
5.3.8 Positive selection analyses . . . 117
5.4 Results . . . 118

5.4.1 Dataset description . . . 118
5.4.2 Diversity in chicken Toll-like receptors . . . 118
5.4.3 Neutrality and selective sweep analysis . . . 128
5.4.4 Positive selection analysis . . . 130
5.4.5 Phylogenetic reconstruction . . . 132
5.4.6 Variant analysis . . . 135
5.5 Discussion . . . 146
5.5.1 Future directions and conclusions . . . 150

6 Evolutionary repurposing of TLR21 as a broad-spectrum CpG-DNA recogniser in the birds and reptiles 152

6.1 Abstract . . . 152
6.2 Introduction . . . 153
6.3 Materials and Methods . . . 155
6.3.1 Sequence retrieval . . . 155
6.3.2 Maximum-likelihood tree . . . 156
6.3.3 Positive selection analyses . . . 156
6.3.4 Homology modelling . . . 156
6.3.5 TLR21 cloning strategy . . . 157
6.3.5.1 RNA extraction . . . 157
6.3.5.2 cDNA synthesis . . . 158
6.3.5.3 Vector and insert PCRs . . . 158
6.3.5.4 Gibson assembly/enzyme-free assembly . . . 161
6.3.5.5 Transformation and colony screening . . . 161
6.3.5.6 Endo-free maxipreps . . . 162
6.3.6 Transfection and stimulation . . . 163
6.3.7 Immunoblotting . . . 164
6.4 Results . . . 165
6.4.1 *TLR21* is distinct from *TLR13* . . . 165

6.4.2 Features of *TLR21* 166
6.4.3 Molecular cloning and expression of TLR21 172
6.4.4 Functional testing of cloned TLR21 constructs 174
6.4.5 Signatures of positive selection in avian/reptilian *TLR21* . . . 178
6.5 Discussion 183
6.5.1 *TLR21* structure and phylogeny 183
6.5.2 CpG preference 184
6.5.3 Signatures of selection 187
6.5.4 Conclusions 188

7 Characterisation of key anti-influenza effectors in the highly susceptible mute swan (*Cygnus olor*) 189

7.1 Abstract 189
7.2 Introduction 190
7.3 Materials and methods 193
7.3.1 Study population 193
7.3.2 PCR of *RIG-I*, *TLR3* and *TLR7* 193
7.3.3 MinION sequencing of PCR products 196
7.3.3.1 PCR cleanup using SPRI beads 196
7.3.3.2 PCR pooling 196
7.3.3.3 End-prep and barcode ligation 196
7.3.3.4 Barcode adaptor ligation 198
7.3.3.5 Flow cell loading, sequencing and data processing . . 198
7.3.4 BUSCO analysis of swan genome 198
7.3.5 Genomic DNA extractions 198
7.3.6 Whole-genome sequencing 199
7.3.7 Read mapping 199
7.3.8 Candidate gene identification 201
7.4 Results 205

7.4.1 Pilot experiment: sequencing of *TLR3*, *TLR7* and *RIG-I* . . . 205
7.4.2 Completeness of published mute swan genome 206
7.4.3 Preparation of genomic DNA for resequencing genomes 207
7.4.4 Candidate gene finding in the mute swan genome 208
7.4.5 Read depth of candidate genes 212
7.5 Discussion 216

8 General Discussion **220**
8.1 Key results for each chapter 220
8.1.1 Chapters 2, 3, and 4 220
8.1.2 Chapter 5 223
8.1.3 Chapter 6 224
8.1.4 Chapter 7 225
8.2 Integration of bioinformatic and functional methods 226
8.3 Limitations of functional assays 228
8.4 Toll-like receptor evolution at different scales 231
8.5 Wildlife disease in a changing climate 233
8.6 Future directions for the field 235
8.6.1 Sequencing costs and technology 235
8.6.2 Large-scale patterns of host and pathogen diversity 236
8.6.3 Taking the comparative approach 237
8.7 Data availability 237
8.8 Conclusions 238

A Appendix to Chapter 2 **239**
A.1 CRISPR-Cas9 editing of endogenous *TLR5* in HEK-Blue Null1 cells . 239
A.2 Tables 243
A.2.1 Notes on Table A.3, A.4, A.5 and A.6 252
A.3 Figures 261

B Appendix to Chapter 3 **264**

B.1 Figures 264

B.2 Tables 270

B.3 Coverage depth of *TLR15* relative to all known coding sequences . . 276

B.3.1 Background 276

B.3.2 Methods 276

B.3.3 Results 278

B.3.4 Interpretation 278

B.4 Identification of putative loss-of-function mutations in *Eudyptes* TLR15 278

B.4.1 Alignment of intact *Eudyptes* TLR15 haplotypes with other birds 278

B.4.2 Variant conservation across vertebrate genomes 279

B.4.3 Change-of-function analysis 281

B.4.4 Conclusions 282

C Appendix to Chapter 4 **283**

C.1 Homology modelling 283

D Appendix to Chapter 5 **286**

E Appendix to Chapter 6 **330**

F Appendix to Chapter 7 **340**

G Additional publications **343**

G.1 Additional publications and manuscripts in preparation 343

G.1.1 Published manuscripts and book chapter 343

G.1.2 Manuscript in review 345

G.1.3 Pre-submission manuscript 345

Bibliography **346**

List of Figures

1.1 Simplified TLR signalling pathways 12

2.1 Locations and sample sizes for sampled Gentoo penguin colonies . . . 30
2.2 Haplotype diversity for HVR1, *TLR4*, *TLR5* and *TLR7* 42
2.3 Minimum spanning haplotype networks for HVR1, *TLR4*, *TLR5* and *TLR7* 43
2.4 Haplotype diversity across Gentoo penguin sample populations 45
2.5 Pairwise F_{ST} comparisons for *TLR4*, *TLR5* and *TLR7* 52
2.6 Positions of polymorphic and positively selected sites in *Pygoscelis papua TLR4*, *5*, and *7* 53
2.7 Distribution and functional differences of two selected amino acid residues in TLR5 54

3.1 Number of polymorphisms per TLR. 76
3.2 Locations of positively selected sites in penguin Toll-like receptors. . . 78
3.3 Alignments of positively selected sites 79
3.4 Locations of positively selected sites on penguin Toll-like receptor ectodomains, relative to sites of known function. 83
3.5 *TLR15* sequences used in pseudogene analysis by location and species 85
3.6 *TLR15* appears to have been pseudogenized repeatedly in *Eudyptes* penguins. 86
3.7 Full-length *Eudyptes* consensus *TLR15* is non-functional, and is therefore a cryptic pseudogene. 87

4.1 Numbers and proportions of positively selected genes by functional category . . . 105
4.2 Homology model of penguin *IFIT5* showing positively selected sites and RNA-binding cleft . . . 107
4.3 Homology model of penguin *CD81* showing positively selected sites and putative HCV E2 glycoprotein-binding site . . . 108

5.1 Map showing sampling locations of feral/backyard chickens and related wild junglefowl species across Asia and Africa . . . 119
5.2 Depth of coverage for chicken and junglefowl genomes . . . 121
5.3 Within-population nucleotide diversity of chicken TLRs . . . 122
5.4 Within-population nucleotide diversity of chicken TLRs, stratified by chicken subpopulations . . . 123
5.5 Between-population nucleotide diversity of chicken TLRs . . . 126
5.6 Between-population nucleotide diversity of chicken TLRs including chicken subpopulations . . . 127
5.7 Values of Tajima's D for TLRs compared to the rest of the chicken genome . . . 128
5.8 Selective sweep analysis of chicken TLRs . . . 129
5.9 Positions of positively selected sites in chicken Toll-like receptors . . . 131
5.10 Phylogeny of a representative TLR in chickens and junglefowl . . . 133
5.11 Phylogeny of a representative TLR in chickens and junglefowl including ancient sequences . . . 135
5.12 Positions of variants discovered in chicken and junglefowl TLR sequences 137
5.12 Positions of variants discovered in chicken and junglefowl TLR sequences (*continued*) . . . 138
5.12 Positions of variants discovered in chicken and junglefowl TLR sequences (*continued*) . . . 139

6.1 Phylogenetic relationship between *TLR11*, *TLR12*, *TLR13* and *TLR21* 167

6.2 Conserved features of the TLR21 protein 168
6.3 Identification of poly-glutamic acid repeats in the TIR domain of avian TLR21 170
6.4 Direct sequencing of zebrafish *TLR21*-containing plasmid 171
6.5 Western blot analysis of cloned TLR21 construct expression 175
6.6 Specificity of TLR21 from a fish and an amphibian to a panel of CpG-ODN agonists 176
6.7 Specificity of TLR21 from birds and reptiles to a panel of CpG-ODN agonists 177
6.8 Strong positive selection in avian/reptilian *TLR21* 180
6.9 Test branches in *TLR21* phylogeny for branch-site model in PAML . 181
6.10 Homology model of the TLR21 ectodomain showing positively selected sites. 182

7.1 BUSCO analysis of completeness of the publicly available mute swan genome. 207
7.2 Testing three commercially available genomic DNA extraction kits. . 208
7.3 Extraction of genomic DNA from additional birds for whole-genome sequencing. 209
7.4 Read depth over candidate genes and their parental contig. 214
7.5 Read depth over the putative Z chromosome in the mute swan. 216

A.1 T7EI mismatch cleavage assay for detection of edited *TLR5* 240
A.2 Stimulation of putative *TLR5*$^{-/-}$ CRISPR-Cas9-edited cells 242
A.3 Conservation of Gentoo penguin TLR5 transmembrane domain region surrounding selected site 262
A.4 Positions of positively selected sites in penguin TLR4 and TLR5 ectodomains 263

B.1 TLR1B homology model showing positively selected sites. 265

B.2 TLR2B homology model showing positively selected sites. 266
B.3 TLR4 homology model showing positively selected sites. 267
B.4 TLR5 homology model showing positively selected sites. 268
B.5 TLR7 homology model showing positively selected sites. 269
B.6 Pairwise depth of coverage of *TLR15* and all other known coding sequences in the genomes of three penguin species. 277
B.7 Per-nucleotide conservation scores for *TLR15* in chicken compared to other vertebrate genomes 280

List of Tables

1.1 The main microbe-associated molecular patterns of TLRs 7

1.2 The main damage-associated molecular patterns of TLRs 8

5.1 Metadata associated with ancient samples 116

5.2 Diversity and neutrality statistics for chicken TLRs 120

5.3 Comparison of diversity statistics between studies 120

5.4 Comparison of haplotype diversity between studies 120

5.5 Distribution of high-impact TLR variants in chickens and related junglefowl 141

6.1 Primer sequences and PCR conditions used for generating inserts, linearised vector and for sequencing. 159

6.2 CpG-ODN sequences used for stimulation of TLR21. 165

6.3 Cloning of *TLR21* from fish, amphibians, reptiles and birds 173

6.4 Variants in cloned chicken *TLR21* compared to known SNPs 174

7.1 PCR primer sequences and conditions for amplifying *TLR3*, *TLR7* and *RIG-I* genes 195

7.2 Recipe for end-prep of PCRs. 197

7.3 Recipe for barcoding end-prepped PCRs. 197

7.4 Barcodes and associated metadata for birds included in the pilot experiment. 197

7.5 Metadata for swan whole-genome samples. 200

7.6 Calculated read depth of swan whole genomes 201

7.7 Identification of candidate anti-influenza genes. 203
7.7 Identification of candidate anti-influenza genes. 204
7.8 Single-nucleotide polymorphisms identified in pilot experiment candidate genes within the swan population. 206
7.9 Candidate gene analysis results. 210
7.9 Candidate gene analysis results. 211

A.1 Polymorphisms and diversity indices for fourteen colonies of Gentoo penguin at three Toll-like receptor loci and the mitochondrial hypervariable control region. 244
A.2 List of primers used for PCR amplification and sequencing of Gentoo penguin TLR genes, along with reaction conditions. 247
A.3 *TLR4* pairwise F_{ST} values and pairwise ϕ_{ST} values 248
A.4 *TLR5* pairwise F_{ST} values and pairwise ϕ_{ST} values 249
A.5 *TLR7* pairwise F_{ST} values and pairwise ϕ_{ST} values 250
A.6 HVRI pairwise F_{ST} values and pairwise ϕ_{ST} values 251
A.7 Geographic distances between sampling locations. 253
A.8 Mantel Test results for isolation by distance. 254
A.9 Log-likelihood comparisons for PAML analysis of *TLR4* 254
A.10 Bayes Empirical Bayes analysis of *P. papua TLR4* sites under positive selection (M2a) 254
A.11 Bayes Empirical Bayes analysis of *P. papua TLR4* sites under positive selection (M8) 255
A.12 Log-likelihood comparisons for PAML analysis of *TLR5* 255
A.13 Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection (M2a) 255
A.14 Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection (M8) 255

A.15 Predicted functional consequences of amino acid substitutions using homology-based methods. 256
A.16 Proximity of selected sites to selected and functional sites in other taxa. 257
A.17 Coordinates and sample sizes of sampling locations. 258
A.18 Census population size (number of breeding pairs) at each location using best available data. 259
A.19 Diversity statistics of Gentoo penguin *vs.* avian Toll-like receptors from published studies 260
A.20 Spearman's rank correlations calculated for haplotype diversity (H_d) *vs.* census population size. 261

B.1 Results from site models comparisons (M2/M2a and M7/M8) in the codeml program in PAML 270
B.2 Positively selected sites in penguin TLRs. 272
B.3 Analysis of positively selected sites in penguin TLRs using MEME and FUBAR 274
B.4 Sampling locations for *Eudyptes* spp. *TLR15* sequences 275
B.5 Amino acid and nucleotide positions of positions which are conserved among penguins and other birds, but are distinct in the *Eudyptes*. . . 279
B.6 Provean predictions for changes in TLR15 protein function 281
B.7 SIFT predictions for changes in TLR15 protein function 281

C.1 PAML results for all positively selected immune loci 284

D.1 Metadata of chicken genomes used in population genetic analysis of Toll-like receptors 286
D.2 HyPhy results using FEL. 309
D.3 HyPhy results using FUBAR. 315
D.4 HyPhy results using MEME. 321
D.5 HyPhy results using SLAC. 325

E.1 Accession numbers for *TLR11*, *TLR12*, *TLR13* and *TLR21* sequences 331

E.2 Results from PAML branch-site analysis of *TLR21* 339

F.1 Variants in swan candidate genes with respect to the swan goose (*Anser cygnoides*). 340

Chapter 1

Introduction

All animals have a capability to sense pathogens and detection represents the first step in mounting a defensive immune response. However, inappropriate responses to normal host molecules can be devastating and so the immune system requires the ability to distinguish between self and non-self. Pattern-recognition receptors (PRRs) overcome this issue by specifically responding to molecular motifs that signify either infection or tissue damage. Microbe-associated molecular patterns (MAMPs) are specific to microbial life and are not found (in any abundance) in the host. MAMPs are diverse and include protein, carbohydrate, nucleic acid, and lipoprotein agonists. In addition, PRRs have evolved to respond to host molecules associated with tissue damage (which can also act as a proxy for infection, as well as sterile damage), termed damage-associated molecular patterns (DAMPs). Responding to these molecular patterns, PRRs initiate inflammation and orchestrate the early stages of the adaptive immune response.

The Toll-like receptors (TLRs) are the largest family of PRRs. TLRs are type I transmembrane proteins with an extracellular leucine-rich repeat domain which binds agonists, a single transmembrane domain, and a Toll/interleukin-1 receptor (TIR) homology domain, which is responsible for transducing the activation signals through an intracellular signalling cascade by interacting with adaptor proteins. While TLR families typically retain specificity for an agonist class, there may be significant between-

species (and even within-species) variability in TLR responses to agonists derived from different pathogens. Unlike the receptor repertoire of the adaptive immune system, which is somatically generated, innate immune receptor diversity is germline-encoded and may therefore be under natural selection to adapt to different pathogen threats.

This thesis documents several examples of pathogen-induced adaptation in the innate immune system in diverse contexts. To introduce the topic, the following sections provide an overview of Toll-like receptor evolution and function. I conclude the introductory chapter by presenting a structural overview of the thesis.

1.1 Early evolution of the Toll-like receptors

> Das war ja toll!
>
> (German for "That was amazing!")
>
> *Christiane Nüsslein-Volhard*[1]

The *Toll* gene was originally identified in *Drosophila melanogaster* as having a role in dorso-ventral patterning of the embryo [15, 235], and also for immune responses against fungi and Gram-positive bacteria in the adult fly [349, 350]. The origin of Toll-like receptors, however, was much earlier in the evolution of metazoans. The prototypical Toll-like receptor comprises three domains: (i) an ectodomain containing variable numbers of leucine-rich repeat (LRR) motifs, responsible for binding pathogen- or damaged-host-derived molecules, (ii) a single-span transmembrane domain, and (iii) an intracellular Toll/interleukin-1 receptor (TIR) domain, which interacts with adapter proteins to initiate downstream signalling [76]. TLRs can be broadly categorised based on the number of cysteine clusters located in the ectodomain [355]. Single cysteine-cluster TLRs (sccTLRs) have one cluster of cysteine residues located at the C-terminal end of the LRRs. Multiple cysteine-cluster TLRs (mccTLRs) also

[1] upon seeing a strange *Drosophila* larval phenotype caused by a mutation in the gene later named "Toll."

have additional cysteine clusters within the LRR, often at the N-terminus of the LRR (LRRNT; [355]).

The timing of the appearance of both mccTLRs and sccTLRs is debated. Brennan and Gilmore (2018) argue that mccTLRs emerged in the phylum Cnidaria and sccTLRs arose secondarily in the Mollusca [76], while Liu *et al.* (2020) contest that both mccTLRs and sccTLRs may have arisen contemporaneously in the Cnidaria [374]. Prototypical TLRs appear to be absent in the phylum Porifera, although the sponge *Amphimedon queenslandica* is known to have at least two TIR domain-containing proteins [199, 248, 553], along with another LRR domain-containing protein [199]. Several of the TLR-to-NF-κB (Nuclear Factor kappa-light-chain-enhancer of activated B cells) pathway components are, however, present in the Porifera, and the early developmental tissue-specific expression of these components may indicate a role in development [199, 76]. It seems likely that the origin of the prototypical TLR was tied to the fusion of LRR-only and TIR-only proteins [76]. All vertebrate TLRs are sccTLRs, suggesting that mccTLRs were lost prior to the emergence of the vertebrates [355, 374].

1.2 TLR expansion and contraction in the invertebrates

Vertebrates are fairly consistent in the number of TLRs in their repertoire ($\sim$10; [374]). The invertebrate TLR repertoire, however, is somewhat less conserved, and significant expansions and contractions of TLRs have been reported in different phyla. The purple sea urchin (*Strongylocentrotus purpuratus*, phylum: Echinodermata) has undergone a remarkable expansion of TLR genes (104-253 TLRs, depending on the analysis; [83, 250, 574]) that are generally more self-similar than they are to other animals, suggesting the expansion is sea urchin-specific [250, 574]. Similarly, the amphioxus (*Branchiostoma floridae*, phylum: Chordata) is reported to have as many

as 71 TLRs [262], although this has been contested with a more recent estimate of between 19-39 TLRs [574]. Several other expansions of the TLR family have been reported in invertebrates; the polychaete *Capitella teleta* (phylum: Annelida) has 105 TLRs [136], the owl limpet (*Lottia gigantea*, phylum: Mollusca) encodes ∼60 TLRs [83], and the blue mussel (*Mytilus edulis*, phylum: Mollusca) expresses 27 TLRs [470], although the number encoded in the genome (but not detectable through transcriptomic analysis) may be somewhat higher [83]. Conversely, other invertebrate taxa have reportedly undergone contraction of TLR diversity. For instance, the nematodes *Caenorhabditis elegans* and *C. briggsae* (phylum: Nematoda) each only have a single TLR [374]. The drivers behind large expansions or contractions of TLR numbers are currently unknown, although it is interesting to point out that TLR expansion typically (though not always) goes hand-in-hand with expansion of other pattern recognition receptor families (such as the NOD-like receptors; [83]). Since both of these families have immune function, it seems likely that the driver behind mutual expansions is pathogen-driven, rather than coincidental.

1.3 Phylogenetic relationships between TLRs

Despite TLRs being almost universal in metazoans, orthologous relationships of TLRs are difficult to track between invertebrates and vertebrates [374]. This implies that TLRs found within the vertebrates are vertebrate-specific, having arisen around the time of the emergence of vertebrates and diversified into several families. The exception to this general rule appears to be *TLR3*[2], which clusters with TLRs from echinoderms, hemichordates and cephalochordates in phylogenetic analyses [374, 574] suggesting this TLR may represent a true invertebrate-vertebrate ortholog. *TLR3* may therefore represent an evolutionarily ancient sensor of double-stranded RNA that is

indispensable in extant vertebrates.

Vertebrate TLRs fall into six major subfamilies (the TLR1/2/6/10 family, the TLR3 family, the TLR4 family, the TLR5 family, the TLR7/8/9 family, and the TLR11/12/13/21 family; [493]) although more recent analyses including additional sequences indicate that the avian- and reptilian-specific *TLR15* may form its own family [374], and that the TLR11/12 family should be considered distinct from the TLR13/21 family because of their different extracellular domain architecture [618]. Therefore, there may be as many as eight subfamilies of vertebrate TLRs. Among the TLR families, the TLR1 family appears to have undergone the most extensive diversification, with *TLR14*, *TLR18*, *TLR25*, and *TLR27* being acquired in the fishes [433], *TLR6* and *TLR10* in mammals [264], and tandem duplications of *TLR1* and *TLR2* in birds (*TLR1A*/*TLR1B* and *TLR2A*/*TLR2B*; [264, 603]). This contrasts with the single-TLR families TLR3, TLR4, TLR5, and TLR7 which are almost invariably present as single family members [374, 493, 603]. Although considered a single family member, *TLR7* is known to undergo frequent and independent duplication in terminal branches, particularly in birds [121, 221, 374, 491, 603]. Pseudogenisations of TLRs are also relatively common, as indicated by the loss of *TLR11-13* in humans [493], the loss of *TLR8* and *TLR9* in birds [469, 579], the loss of *TLR21* in mammals [82, 301], and the frequent pseudogenisation of *TLR5* in terminal branches of avian phylogeny [10, 36, 603].

1.4 Agonist specificities of TLRs

In a prescient insight long before the discovery of TLRs, Charles A. Janeway, Jr. predicted that the innate immune system responds to conserved microbial patterns via pattern-recognition receptors, as a means of identifying 'non-self' [270]. Toll-like

[2]Throughout this thesis, italics are used to denote gene names, while non-italicised text refers to protein products. Thus *TLR1* refers to the Toll-like receptor 1 gene, and TLR1 refers to its protein product. Discussions of TLR families (e.g., the TLR7/8/9 family) use non-italicised text. Since positive selection acts at the level of the nucleotide sequence, statements referring to positive selected sites use italicised gene names, whereas any discussion of the functional importance of the site (e.g., to agonist-binding) refers to the protein, and therefore uses non-italicised text.

receptors have evolved to recognise a wealth of molecular patterns that are specific to microbial pathogens. These microbe-associated molecular patterns (MAMPs) include proteins, lipopeptides, bacterial cell wall components, polysaccharides, and nucleic acids, and are summarised in Table 1.1. Furthermore, as predicted by Polly Matzinger in 1994 prior to the discovery of TLRs [396], the immune system can respond to 'danger' signals, i.e., molecules that indicate host tissue damage. These damage-associated molecular patterns (DAMPs) are numerous, and include extracellular matrix components, plasma membrane components, nuclear and cytosolic proteins, heat shock proteins, and nucleic acids [159]. Some of the best-characterised DAMPs are shown in Table 1.2.

Table 1.1: The main microbe-associated molecular patterns of TLRs

TLR	PAMP	Reference
TLR1/2	Triacylated lipopeptides	[288, 571]
TLR2/6	Diacylated lipopeptides	[288, 570]
TLR2	Peptidoglycan	[155]
	Lipoteichoic acid	[528]
TLR3	dsRNA	[68]
TLR4	Lipopolysaccharide (lipid A)	[86]
	F protein (Syncytial virus)	[331]
	Mannuronic acid	[181]
	Teichuronic acid	[651]
	Flavolipin	[209]
	Mannan	[501]
	NS1 protein (Dengue)	[412]
	Paclitaxel	[297]
TLR5	Flagellin	[659]
TLR7	ssRNA	[385]
TLR8	ssRNA	[186]
TLR9	Unmethylated CpG DNA	[106]
TLR10	dsRNA	[345]
TLR11	Profilin	[656]
TLR12	Profilin	[317]
TLR13	Ribosomal RNA	[449, 537]
TLR15	Fungal component	[72]
	Protease cleavage	[138]
TLR21	Unmethylated CpG DNA	[82]

Table 1.2: The main damage-associated molecular patterns of TLRs

Structural class	TLR	DAMP	Reference
Proteins, peptides	TLR1	β-defensin-3	[192]
	TLR2	HSP60, 70, Gp96	[28, 519, 597]
	TLR2	HMGB1	[314, 461, 462]
	TLR2	HMGB1-nucleosome complexes	[400]
	TLR2	Surfactant protein A, D	[418, 447]
	TLR2	Eosinophil-derived neurotoxin	[648]
	TLR2	Antiphospholipid antibodies	[521]
	TLR4	HMGB1	[314, 461, 462]
	TLR4	Fibronectin EDA	[448]
	TLR4	Fibrinogen	[546]
	TLR4	Surfactant protein A, D	[418, 447]
	TLR4	B-defensin-2	[57]
	TLR4	HSP60, 70, 72, 22, Gp96	[28, 444, 497, 519, 596, 597, 633]
	TLR4	S100A8 (MRP8)	[182, 609]
	TLR4	S100A9 (MRP14)	[182, 609]
	TLR4	Neutrophil elastase	[143]
	TLR4	Antiphospholipid antibodies	[417, 471]
	TLR4	Lactoferrin	[127]
	TLR7	Antiphospholipid antibodies	[267]
	TLR8	Antiphospholipid antibodies	[150, 267]
Fatty acids, lipoproteins	TLR2	Serum amyloid A	[99, 243]
	TLR4	Serum amyloid A	[257, 517]
	TLR4	Oxidised LDL	[35, 408, 644]
	TLR4	Saturated fatty acids	[523, 536]
Proteoglycans, glycosaminoglycans	TLR2	Biglycan	[522]
	TLR2	Versican	[308]

Cont. next page

Structural class	TLR	DAMP	Reference
	TLR2	Hyaluronic acid fragments	[580]
	TLR4	Biglycan	[522]
	TLR4	Heparan sulfate fragments	[276]
	TLR4	Hyaluronic acid fragments	[580]
Nucleic acids	TLR3	mRNA	[289]
	TLR7	ssRNA	[610]
	TLR8	ssRNA	[610]
	TLR9	IgG-chromatin complexes	[341]

TLRs generally retain specificity for the same agonist class across evolutionary time [493]. There are some exceptions, however, such as mouse TLR8 not recognising ssRNA ligands that are potent activators of human TLR8 [247, 376]. Within the same TLR family, TLRs tend to recognise agonists from the same structural class. For instance, the TLR1/2 family usually respond to lipopeptides, and the TLR7/8/9 family respond to nucleic acids (Table 1.1). TLR orthologs in different species often evolve subtly different agonist specificities, presumably in response to pathogen-induced selection. Examples of this include changes in CpG-DNA motif-binding specificity in human vs. mouse TLR9 [50] and preferences for flagellins derived from different bacterial species in TLR5 [169]. An additional example of extensive changes in CpG-DNA specificity in TLR21 is presented in Chapter 5 of this thesis. Upon binding an agonist, most TLRs for functional homodimers which allows recruitment of intracellular adapter proteins. However, TLR2 is unusual in that it forms functional heterodimers with TLR1 or TLR6. This acts to increase the breadth of agonists seen by TLR2; the TLR2/1 heterodimer recognises triacylated lipopeptides [571], while the TLR2/6 heterodimer recognises diacylated lipopeptides [570].

1.5 TLR signalling and interaction with the adaptive immune system

TLR signalling to downstream transduction pathways is dependent on receptor dimersation. Some TLRs (such as TLR8) exist as pre-formed homodimers which then undergo conformational changes to bring the C-terminal TIR domains into close proximity and allow for adaptor molecule recruitment [573]. The dimerisation and downstream signalling of other TLRs is dependent on ligand binding [69]. Most TLRs signal via homodimerisation, but some can form heterodimers with a different TLR, notably the TLR1/2 and TLR2/6 heterodimers [273], TLR4/6 heterodimers [560], and TLR11/12 heterodimers [16]. Upon ligand binding, the intracellular TIR do-

mains recruit adaptor molecules to transduce downstream signals. A brief overview of TLR signalling is provided in the following section and in Figure 1.1.

Two major pathways of TLR signalling have been described, initiated by two distinct adaptor molecules. Most TLRs are dependent on signalling via myeloid differentiation primary response protein (MyD88), either directly, or via a bridging molecule—MyD88 adaptor-like (MAL, also known as TIRAP; [175, 261, 646]). However, TLR3 is exclusively dependent on an alternative adaptor molecule, TIR-domain-containing adapter-inducing interferon-β (TRIF) [645], with which it interacts directly. TLR4 can signal via both the MyD88- and TRIF-dependent pathways [534], but interacts with TRIF via an additional adaptor, TRIF-related adaptor molecule (TRAM; [176]).

After MyD88 recruitment, a complex of MyD88 and IRAK1 and IRAK4 forms, known as the Myddosome [369]. IRAK1, activated by IRAK4, undergoes autophosphorylation [320] and is released from the Myddosome. IRAK1 then binds TRAF6, which recruits the ubiquitinases UBC13 and UEV1A to polyubiquitinate both TRAF6 itself and the TAK1 protein kinase complex. TAK1 forms a complex with TAB1, TAB2 and TAB3, and TAK1 is activated [4, 97], possibly by ubiquitination or phosphorylation. From TAK1, two pathways are activated, NF-κB and the mitogen-activated protein kinase (MAPK) family member (ERK, JNK and p38) pathway, which leads to the activation of the proinflammatory AP-1 family of transcription factors [8, 296]. The TRIF-dependent pathway can also activate NF-κB- and AP-1-dependent responses [298], but also interacts with TRAF3 to initiate a distinct signalling pathway. TRAF3 induces IRF3 phosphorylation via TRK1, IKKi and NEMO recruitment. Phosphorylated IRF3 forms a dimer and is translocated to the nucleus where it up-regulates expression of type I interferon (IFN) genes [8, 296].

Stimulation of the innate immune system via PRRs leads to the production of innate immune effectors. These can include cytokines such as IL-1, IL-6 and TNF-a which can act to promote features of local inflammation (i.e. heat, redness, swelling) by inducing vascular and cellular changes. Furthermore, the production of directly

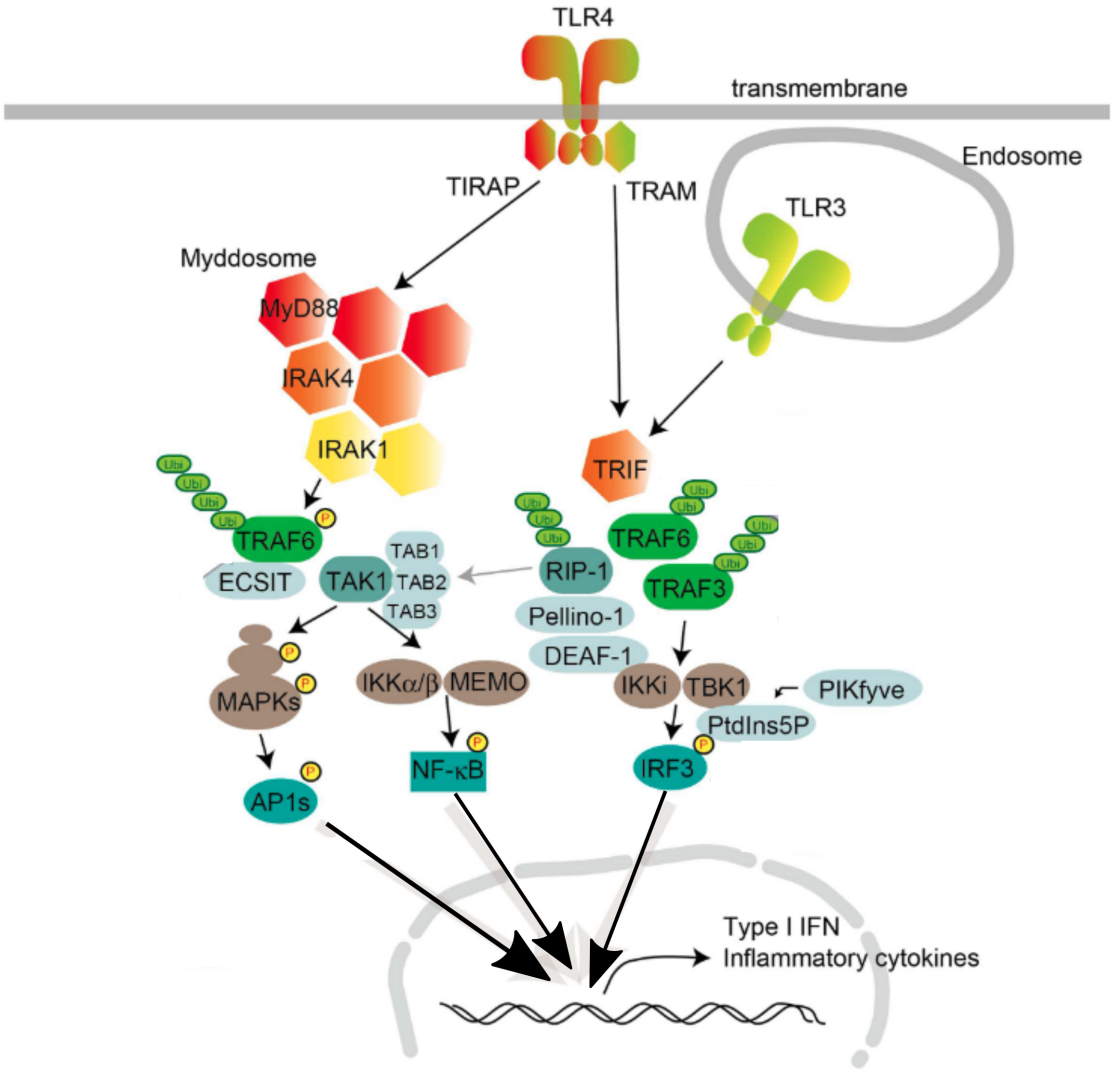


Figure 1.1: A simplified representation of TLR signalling pathways. The diagram was taken from [298] and modified. The two main TLR signalling pathways are illustrated (MyD88-dependent and MyD88-independent (i.e., TRIF-dependent). TLR3 is located within the endosomes and exclusively signals via the TRIF-dependent pathway, while TLR4 is (primarily) located at the cell surface and signals via the MyD88-dependent pathway. TLR4 can also signal via the TRIF-dependent pathway, and is the only TLR to signal via both pathways. All other TLRs exclusively use the MyD88-dependent pathway. The two pathways result in the activation of transcription factors including AP1, NF-κB, and IRF3, which lead to the up-regulation of pro-inflammatory cytokines and type I interferons. Further details of the intermediate steps in the signalling pathways can be found in the main text.

antimicrobial effectors such as antimicrobial peptides and nitric oxide (NO) can initiate direct pathogen killing. The two main families of antimicrobial peptides, defensins and cathelicidins, act to disrupt bacterial cell membranes and cause cell puncturing. Stimulation of viral-recognising TLRs such as TLR3 or TLR7 can also upregulate the production of type I IFNs. Expression of type I IFN acts on cells in an autocrine or paracrine manner to stimulate expression of hundreds of additional interferon-stimulated genes (ISGs), which collectively act to interfere with viral replication and induce adaptive immunity.

The interactions between innate and adaptive immunity, as well as the multiplicity of downstream adaptive immune responses, are complex and will only be dealt with briefly in the following section. TLRs are considered to be a bridge between innate and adaptive immunity [429]. Specifically, specialised antigen-presenting cells such as dendritic cells (DCs) express TLRs and undergo maturation upon stimulation with MAMPs. Mature DCs upregulate costimulatory molecules that are required for priming and activating T cells. Communication between the matured DC and the T cell takes the form of four signals [109]. Firstly, the DC presents a linearised peptide antigen on MHC class I or II to the T cell via the T cell receptor (peptide-MHC complexes are also promoted by TLR-induced endosomal acidification [60, 96]). This signal alone is insufficient to activate T cells, and a lack of a second signal may render the T cell refractory to future stimulation. The second signal is a co-stimulation and can be achieved through a combination of several mediators, including CD80, CD86, CD28 and CTLA-4 [51, 96]. The purpose of the second signal is to give positive or negative confirmation to the T cell as to whether it should respond to the antigen. The third signal determines the polarisation a T cell should adopt to best deal with the detected pathogen, in very simple terms either a T-helper (Th) 1, Th2, or regulatory (Treg) response. Th1 cells secrete cytokines, including IFN-γ and tumour necrosis factor (TNF)-β, and work to activate macrophages and stimulate cytotoxic CD8+ T cells, termed cell-mediated immunity. Th2 cells secrete IL-4, IL-5 and IL-13 and are effective stimulators of B cells to initiate antibody production,

therefore inducing humoral immunity. The final signal affects T cell tissue localisation: DCs induce the expression of homing receptors in T cells that will direct them to the tissues in which the antigen was first discovered, increasing the opportunity of the reactive T cell to identify its target. TLRs appear to exert considerable influence of the earliest stages of the adaptive immune response. TLR signalling can promote the establishment of a Th1, Th2 or Treg response [451], and may be influenced by the dose of agonist received. For instance, high doses of LPS stimulating TLR4 may induce Th1 polarisation, while low-dose LPS may stimulate a Th2 response [158]. TLRs are therefore more than gatekeepers to the adaptive immune response—signalling via activated TLRs can not only induce adaptive immunity, but shape the specific adaptive immune responses and tailor the response to the specific threat.

1.6 Subcellular localisation and regulation of TLRs

TLRs are functional in two subcellular compartments, depending on the receptor. Receptors involved in lipid, protein, and carbohydrate detection (e.g. TLR1, TLR2, TLR4, TLR5, TLR6, TLR15) typically embed in the plasma membrane and functional at the cell surface, while receptors involved in the detection of nucleic acids (e.g. TLR3, TLR7, TLR8, TLR9, TLR21) are functional in the endolysosomal system. Nucleic acid-recognising TLRs are sequestered in the endoplasmic reticulum and are continuously trafficked to the endolysosomal system at low levels to survey MAMPs, whereas cell surface TLRs are continuously sent to the plasma membrane and are not sequestered in the ER [346].

The additional compartmentalisation of nucleic acid-recognising TLRs is not coincidental. Nucleic acids are unusual among MAMPs in that they are not unique to microbes: all hosts possess nucleic acid which, under certain circumstances, can lead to aberrant TLR activation [48, 395]. One well-documented example is the involve-

ment of nucleic acid-recognising TLRs in the autoimmune pathology of systemic lupus erythematosus (SLE; [56, 104, 105, 334, 473, 563]). Under normal circumstances, the compartmentalisation of nucleic acid-recognising TLRs is likely to be a regulatory mechanism to limit their contact with host nucleic acids [48]. Additional regulatory features are also found in nucleic acid-recognising TLRs, such as the requirement to undergo acidification in order to function (as demonstrated by the use of the endosomal acidification inhibitors chloroquine and bafilomycin A [82, 229, 301]), and receptors having to undergo proteolytic cleavage in order to function [166, 459]. Intracellular TLRs undergo trafficking to the endosomes from the endoplasmic reticulum by means of a chaperone protein, UNC93B1 [566]. UNC93B1 is clearly crucial to this process, as exemplified by human patients lacking functional UNC93B1 developing Herpes simplex virus type 1 (HSV-1) encephalitis [91]. Interestingly, UNC93B1 is now thought to be essential for maintenance of UNC93B1-dependent TLR expression, and functions even when physically restricted to the endoplasmic reticulum [463].

Extracellular TLRs have fewer regulatory steps to become functional compared to intracellular TLRs, implying their ligands are more specific to pathogens than nucleic acids. However, as indicated in Table 1.2, there are considerable numbers of host-derived damage-associated molecular patterns that are detected by cell surface TLRs. Dysregulation of cell surface TLR function can lead to autoimmune pathology, similar to nucleic acid-recognising TLRs. Cell surface TLRs have been implicated in rheumatic disorders, systemic lupus erythematosus, and multiple sclerosis, among other diseases [381]. As such, tight regulation of TLR activity is of paramount importance to avoid pathology. One common feature of TLR regulation is tolerance, whereby cells become refractory to continued stimulation with the same agonist [240, 435], which can be achieved by negative regulators of TLR signalling such as SIGIRR or Tollip [336], or via transcriptional regulation mediated by the three MAPK components: ERK, JNK and p38 [465]. A complex regulatory network ensures, under normal circumstances, that the inflammatory response is proportionate and controllable, protecting the host from potentially fatal systemic inflammation.

1.7 Involvement of TLRs in disease

The importance of TLRs to health is best exemplified by situations where a non-functional or otherwise altered TLR is associated with disease. Clearly, the majority of these instances are documented in humans, but there is no reason to believe that humans would not represent a general case across all animals. Aberrant TLR function can affect the host in two ways: (i) sterile pathologies associated with immune dysregulation and autoimmunity (e.g., allergy, arthritis, and rheumatic disorders [278, 401], and (ii) by altered susceptibility to a variety of bacterial, viral, and parasitic diseases.

There are several well-documented examples of TLR polymorphism being associated with increased susceptibility to infectious disease. For instance, the TLR1 I602S variant in humans is associated with increased risk of tuberculosis (*Mycobacterium tuberculosis* infection; [390]), mechanistically due to aberrant receptor trafficking to the cell surface [275]. The TLR4 D299G variant is associated with increased risk of mortality from invasive meningococcal disease [170], and is also associated with low birth weights and maternal anaemia during pregnancy of mothers infected with *Plasmodium falciparum* malaria [411]. However, under some circumstances, the TLR4 D299G may be protective: individuals with the polymorphism are less likely to suffer the systemic consequences of inflammation following the inhalation of lipopolysaccharide [406]. This is also reminiscent of TLR4-null mice being resistant to the lethal effects of endotoxic shock [486]. Opposing susceptibility and resistance SNP phenotypes appear to be common in TLRs. A further example is the common R392STOP variant in TLR5, which is associated with susceptibility to pneumonia caused by *Legionella pneumophila* [236], but also associated with protection against the autoimmune disease systemic lupus erythematosus (SLE; [237]). The apparent pleiotropy of TLR polymorphisms may indicate that balancing selection acts to maintain such variants in populations. Evidence from the bank vole (*Myodes glareolus*) indicates that TLR alleles are maintained in the population by balancing selection, possibly in response to the blood pathogen *Bartonella* [313]. It could be that pleitropic SNPs such as

D299G are actively maintained in human populations, and, although associated with disease under some circumstances, may provide host benefits under others. The fact that TLR polymorphisms can lead to important host phenotypes also indicates that they are likely targets for natural selection, which is considered in the next section.

1.8 Diversity and selection in TLRs

Immune loci are frequently identified as being hotspots for positive selection in the genome [47, 323, 504, 512, 526, 542]. Moreover, genes whose products directly interact with pathogens are found to be particularly enriched in positive selection [156, 160, 161, 512, 624]. This is likely because hosts and pathogens are engaged in an evolutionary arms race to detect pathogens, and avoid detection, respectively. Studies focusing on TLRs have often identified them as targets for positive selection in several taxa [33, 46, 220, 358, 373, 535, 603, 608, 636, 637]. Both diversity and selection in TLRs is domain-specific: the ectodomain, which directly interacts with pathogen-derived molecules, is typically more diverse and under stronger positive selection than the intracellular TIR domain, which interacts with host molecules and is typically less diverse and under purifying selection [220, 374, 603, 637]. Diversity and selection are also dependent on whether the TLR is extracellular (typically recognising protein, lipid, or carbohydrate) or intra-endosomal (nucleic acid-recognising). Extracellular TLRs are typically more diverse and under stronger positive selection compared to the intra-endosomal TLRs [46, 358, 374, 603, 637]. This phenomenon is frequently ascribed to nucleic acid ligands being 'less complex' PAMPs than the non-nucleic-acid PAMPs [46], although in reality it is more likely that these PAMPs are simply less diverse due to selective constraints on the genomes of viruses and bacteria. In addition to coding sequences, non-coding regulatory sequences are increasingly being recognised as targets for positive selection, particularly in the human TLR1/6/10 cluster [46, 133, 222, 332].

TLRs may also evolve at different rates in different taxa. For instance, chicken

TLRs have several times the diversity and numbers of selected sites found within human TLRs [565], despite the genome-wide rate of evolution being lower in birds than mammals [666]. The proportions of positively selected sites in avian TLRs are also significantly elevated compared to amphibian TLRs [374]. Birds therefore appear to have rates of TLR evolution that far exceed predictions from the rate of genome evolution. Since analyses of diversity and positive selection of TLRs form a major part of the data presented in this thesis, continued discussion of this topic can be found in relevant data chapters, although an introduction to the theory and methods of detecting positive selection can be found later in this Chapter.

1.9 Other pattern-recognition receptors

Although TLRs are the major focus of this thesis, it is helpful to consider TLRs as part of a wider family of pattern-recognition receptors (PRRs) that together form the pathogen-sensing capability of the innate immune system. In addition to the TLRs, there are three other major classes of PRR: the soluble or transmembrane C-type lectin receptors (CLRs), and cytoplasmic retinoic acid-inducible gene (RIG)-I-like receptors (RLRs), and NOD-like receptors (NLRs) [569].

RIG-I, melanoma differentiation-associated gene 5 (*MDA5*) and *LGP2* are cytoplasmic detectors of double-stranded (ds)RNA [658], either directly recognising dsRNA viral genomes or the dsRNA intermediate of single-stranded (ss)RNA genome replication [569]. RIG-I and MDA5 have complementary roles in dsRNA recognition. RIG-I typically recognises short dsRNA (up to 1 kbp), while MDA5 recognises long dsRNA (more than 2 kb), and MDA5 agonists can become RIG-I agonists through enzymatic degradation [293]. LGP2 lacks a caspase activation and recruitment domain (CARD), which is responsible for downstream signal transduction in RIG-I and MDA5 by interaction with mitochondrial antiviral-signalling protein (MAVS) [503]. LGP2 functions as a positive regulator of RIG-I and MDA5, possibly by processing RNA to remove proteins from viral ribonucleoprotein complexes or unravelling RNA

secondary structures [520]. RIG-I and MDA5 are key components of the anti-viral response, as indicated by a lack of type I IFN production in *RIG-I*$^{-/-}$ and *MDA5*$^{-/-}$ double-knockout mice [382], and the susceptibility of chickens (which lack *RIG-I*) to highly pathogenic avian influenza infection [42].

CLRs are transmembrane or soluble receptors which specialise in recognising carbohydrate patterns of pathogens. CLRs are divided into two groups: the group I CLRs (mannose-binding receptors) and group II CLRs (asialoglycoprotein receptor family members including dectin 1) [201]. The detection of different carbohydrate structures allows CLRs to identify most classes of pathogens. For instance, mannose is typical of viruses, fungi and mycobacteria, while fucose is associated with certain bacteria and helminths [201]. Syk is a tyrosine kinase that is activated by CLRs containing an ITAM domain (or via ITAM-containing adaptor proteins; [569]). In response to *Candida* infection, dectin-1 and dectin-2 signal via the Syk kinase and CARD9 to upregulate T cell production of IL-17 and thereby drive differentiation of Th17 T cell subsets [494, 598].

NLRs are cytoplasmic PRRs that recognise a variety of endogenous and exogenous stimuli [569]. NLRs have a central NOD domain which is required for formation of a signalling complex known as the nodosome [576], a variable N-terminal domain with distinct functions in different NLRs, and a leucine-rich repeat-containing C-terminal domain responsible for ligand binding [673]. Some NLRs (NLRP1, NLRP3, NLRP6, NLRP7, NLRP12, NLRC4 and NAIP) function in inflammasomes, multimeric protein complexes that promote inflammatory responses [673]. Other NLRs (NOD1, NOD2, NLRP10, NLRX1, NLRC5 and CIITA) initiate inflammation via activation of NF-κB, MAPKs and IRF pathways [673]. NOD1 and NOD2—two of the best characterised NLRs—are activated primarily by peptidoglycan from Gram-positive bacteria [94]. Following agonist binding, NOD1/2 interact with RIPK2 via homotypic CARD-CARD domain interactions [428]. Downstream signal transduction, similar to TLRs, involves TRAFs and culminates in the activation of NF-κB which translocate to the nucleus as dimers and initiate proinflammatory gene expression [428].

1.10 TLRs in nonmodel species: the value of the comparative approach

TLRs, their agonists and signalling pathways were rapidly elucidated in the mouse and human in the first decade after their discovery at the end of the 20th century. While there is undoubtedly knowledge still to be gained in these systems, much work is focused on fine-scale detail, and our knowledge of TLRs in non-model species lags significantly further behind. However, much can be learned from TLR function in non-model animals.

For instance, while the conventionally accepted TLR paradigm involves the binding of an agonist (either a DAMP or MAMP) to elicit TLR activation, TLR15 (found only in birds and reptiles) demonstrates an altogether different mode of activation: cleavage of the receptor by microbial proteases to yield a signalling TLR [138]. The activation of TLR15 illustrates the role of TLRs as sensors of danger in the fullest sense, irrespective of canonical agonist-induced signalling.

Furthermore, much can be learned about the reliance on TLR signalling from the highly unusual case of the Atlantic cod (*Gadus morhua*), which has experienced loss of the MHC class II pathway and undergone concomitant expansion of MHC class I genes and TLRs [392, 556, 555]. Based on current knowledge, TLR repertoire in the Atlantic cod is unparalleled among the vertebrates—cod have, among others, at least five *TLR25*, at least seven *TLR8* and at least eight *TLR22*, and there is evidence of diversifying selection within these expanded families [550]. This demonstrates great innovation and plasticity within the immune system, and is likely to be linked to the increased reliance on the innate immune system in a species that has lost MHC class II.

A final benefit of studying non-model species is the identification of putative functional protein regions through patterns of recurrent positive selection in different taxa. There are several studies considering TLR diversity and adaptation in wild animals (a number of which are discussed in the data chapters). While the finding of posi-

tively selected sites in any one of these studies is unremarkable, patterns of recurrent selection at specific TLR residues across multiple taxa, from birds to mammals, are beginning to emerge [603]. Such observations are powerful in indicating putative functional sites, and would not have been possible without taking a comparative approach to TLR biology.

1.11 Theory and tools of positive selection analysis

Detecting adaptation in TLRs at the sequence level is a key part of this thesis. In the next section, I will briefly outline the theory and tools that are used in this field.

In order to understand how adaptive processes operate in the genome, we must also be able to confidently identify instances of positive selection. The relative contributions of genetic drift and natural selection to the patterns of nucleotide diversity within and between species is debated [205, 311]. On one side, a body of theory argues that a substantial proportion of genetic variation within a population has phenotypic consequences related to fitness and reproductive potential, and will therefore be subject to natural selection [205]. This contrasts to the neutral theory of molecular evolution, which posits that most nucleotide changes are either neutral or slightly deleterious in consequence, and that it is stochastic effects of genetic drift in populations of finite size that determine allele frequency [309, 312, 445].

Neutrality can be tested using metrics such as Tajima's D [567], calculated as the difference between two estimations of the population mutation rate (or genetic diversity): the mean number of pairwise nucleotide differences (π), and the number of segregating sites (θ_{w}). In a neutrally evolving population, both measures of diversity should be approximately equal, and the expected value of Tajima's D is zero. Significant deviations from zero indicate a departure from neutral evolution. Positive values of Tajima's D signify a lack of rare alleles, which could be due to an active process such as balancing selection, or a population contraction. In contrast, negative values of Tajima's D indicate an excess of rare alleles, which could be an indicator of

selection at the locus, or evidence of a population expansion. Since selective forces and population demographic history can both lead to similar patterns of nucleotide variation, these processes are difficult to disentangle. A key difference is that demography is anticipated to affect the entire genome, while selective forces are generally locus-specific. Therefore, while at the level of a single locus, deviations from neutrality are difficult to infer, these conclusions are bolstered by considering many loci that are all subject to the same demographic influences [59]. In Chapter 5 of this thesis, for instance, I consider values of Tajima's D in chicken Toll-like receptors in the context of the whole coding genome.

For protein-coding genes, an alternative method for detecting deviation from neutral evolution has been developed. In a coding sequence, a nucleotide polymorphism can either be synonymous (whereby the encoded amino acid remains the same) or non-synonymous (the amino acid is changed). In general, synonymous polymorphisms are anticipated to be of no consequence (although there are clear instances where this is not the case, such as codon usage bias [380]). Conversely, non-synonymous polymorphisms are expected to cause changes in protein function (although the degree of functional change is dependent on the conservation of amino acid properties between the variants [647]). Calculating ratios of the rates of non-synonymous (dN) to synonymous (dS) polymorphisms allows us to infer whether the gene evolving neutrally, or is under purifying or positive selection. When $dN/dS < 1$, rates of synonymous polymorphism exceed non-synonymous polymorphism, which is interpreted as purifying selection for maintenance of protein function. When synonymous polymorphisms are as common as non-synonymous polymorphisms ($dN/dS \approx 1$), the gene is evolving neutrally. In the final case ($dN/dS > 1$), non-synonymous polymorphisms are elevated, which is evidence for change in protein function. Since the vast majority of proteins have at least some element of conservation of structure or key function, dN/dS is unlikely to be elevated across the entire coding sequence. Therefore, tests for positive selection using dN/dS typically consider individual codons as selectable units, which allows positively selected sites to be found within otherwise highly conserved genes.

Methods using dN/dS to infer positive selection are broadly split into two categories: (i) those in which selection is tested along specified phylogenetic branches (branch or branch-site models), or (ii) those in which selection is tested between sites in a sequence alignment (site models). Both of these methods are used at various points in this thesis. For instance, where I consider whether immune genes have been selected within the penguins, without any predictions about which branches within penguin phylogeny were selected, I use site models (Chapters 2, 3, and 4). In contrast, where I hypothesise specifically that *TLR21* has been selected in the avian/reptilian lineage (Chapter 6), I use branch or branch-site models.

Several tools that implement tests of dN/dS to detect positive selection have been published. Two of the most widely used packages of programs are Phylogenetic Analysis by Maximum Likelihood (PAML) [652] and the suite of Bayesian and maximum-likelihood methods implemented online on the Datamonkey server (http://www.datamonkey.org; last accessed July 2021) and on the command line as the 'hypothesis testing using phylogenies' (HyPhy) tools [476]. Avoiding false positives using these methods relies heavily on data quality; errors in sequencing, assembly and alignment, as well as non-codon-based variation (e.g. large insertions) are all sources of erroneously detected positive selection. In light of this, I manually check all alignments, particularly in the genome-wide analysis of Chapter 4, where orthologous genes and alignments were all curated using automated methods. Despite this, there is still some debate surrounding rates of false positives using different methods [438, 653]. To mitigate this issue, I typically use multiple methods of detecting positive selection and highlight the positively selected sites that are concordant between different methods.

1.12 Structure of the thesis

In this thesis, I integrate bioinformatic and laboratory methods to characterise innate immune system (primarily Toll-like receptor) diversity, selective influences, and func-

tion. A wide range of animal species and populations is considered, from wild Gentoo penguins to domestic chickens, and from mute swans to crocodiles. My objective was to identify and characterise functional variation that was directly linked (wherever possible) to changes in the immune phenotype, which in turn likely influences the host-pathogen interaction.

Chapters 2, 3, and 4 consider innate immune adaptation in the penguins (order: Sphenisciformes). Penguins were once thought to have evolved in the polar Antarctic climate [37], which, being inhospitable to many forms of life, was presumed to have rendered penguins 'immunologically naïve' because of their reduced exposure to pathogens [37, 218]. However, more recent evidence suggests that penguins evolved in Australia and New Zealand during a relatively mild period [118, 604], which is inconsistent with a low pathogen pressure. Despite this, penguins are susceptible to a wide range of pathogens, likely exacerbated by changes in global pathogen profiles due to climate change, human movement, and the widespread captivity of penguins around the world.

Chapter 2 uses Sanger sequencing to characterise the nucleotide diversity and adaptation in three Toll-like receptors (*TLR4*, *TLR5* and *TLR7*) in the Gentoo penguin (*Pygoscelis papua*). Gentoo penguins were an ideal study population for characterising immune diversity, since at least four genetically delineated clades of the Gentoo penguin exist, and there is little or no admixture between clades. Putative selected sites in *TLR5* were subjected to functional analysis to reveal phenotypic consequences. Chapter 3 takes a broader look at TLR adaptation across all penguins using reference genomes. I also consider the unusual scenario of a cryptic pseudogene TLR which has subsequently undergone multiple independent overt pseudogenisations across a genus of penguins. Chapter 4 takes a further step back, and considers immune adaptation in penguins from the perspective of an unbiased, whole-genome analysis. In this chapter, I discuss evidence of likely host-pathogen interactions that mark the genomes of penguins.

Chapter 5 investigates extensive TLR diversity in the domestic chicken (*Gallus*

gallus domesticus), along with closely related wild junglefowl species (*Gallus* spp.). Chicken TLRs have previously been reported to display far more diversity than those of other animals, including humans [565]. I use a comprehensive dataset comprising >1600 genomes from commercial chickens, feral/backyard birds, fancy breeds, all five subspecies of red junglefowl (*Gallus gallus* ssp.), and all three congeneric junglefowl (green junglefowl, *Gallus varius*; grey junglefowl, *Gallus sonneratii*; and Ceylon junglefowl, *Gallus lafayettii*). The dataset yields a rich database of SNPs, pseudogenes, and putative functionally important selected sites that far exceeds the current known diversity of chicken TLRs in online variant databases.

Chapter 6 considers how the agonist binding specificities of TLRs can be remodelled over extensive evolutionary distances. *TLR21* is a CpG-DNA-recognising TLR that is present in fishes, amphibians, reptiles and birds, but has been lost in mammals. *TLR9*, an unrelated but functionally analogous TLR, also binds CpG-DNA motifs and is present in fishes, amphibians and mammals, but has been lost in birds and most reptiles. The phylogenetic distribution of the two CpG-recognising TLRs implies that fishes and amphibians have both *TLR9* and *TLR21*, birds and reptiles have only *TLR21*, and mammals have only *TLR9*. Interestingly, in fishes, the two receptors have non-redundant, and even cooperative, roles in CpG recognition [657], with each displaying a preference for a different range of CpG-DNA motifs. In fishes, TLR9 is a generalist CpG-recogniser, responding to a broad range of different motifs, while TLR21 is a fastidious CpG-recogniser, recognising only a narrow subset of the different CpG motifs. In stark contrast, the chicken TLR21 responds robustly to a broad range of different CpG motifs [107]. I hypothesised that the loss of the *TLR9* in the avian/reptilian lineage drove, or was driven by, the broadening of the TLR21 CpG-recognition profile. Using bioinformatic and functional methods, I analysed *TLR21* sequences for evidence of positive selection, which could be indicative of functional change. I then used functional methods to establish the ligand-binding preferences of cloned TLR21 receptors from amphibians and reptiles which represent the 'missing links' in the trajectory between fastidious to generalist CpG-recogniser.

Chapter 7 describes an investigation into influenza susceptibility in the mute swan (*Cygnus olor*). Swans appear to be highly susceptible to infection with highly pathogenic avian influenza (HPAI). As part of influenza monitoring efforts, influenza-positive swan carcasses are disproportionately overrepresented compared to other waterfowl. While this could partly be attributed to the swan being a conspicuous, large bird that is easy to identify, this appears only to be part of the answer. During experimental infection, swans are much more likely to succumb to HPAI infection compared to geese or ducks, and exhibit significantly elevated mortality during outbreaks in closely monitored settings, such as at the Abbotsbury Swannery (Dorset, UK). While adaptive immunity clearly plays a role in influenza susceptibility in swans and other waterfowl (older birds with previous exposure to influenza are much less likely to die), innate immunity may determine the high sensitivity of juvenile swans to influenza infection. Indeed, in another highly susceptible bird—the domestic chicken—a crucial anti-influenza innate immune component is absent [42]. In this chapter, I take a candidate gene approach, using MinION and whole-genome sequencing, to identify key mediators of the anti-influenza response in the newly sequenced swan genome.

Chapter 8 is a synthesis of major results and emergent themes from the data chapters where I discuss strengths and weaknesses of the work presented in the thesis.

Since the methods used in each chapter are generally distinct, relevant materials and methods are summarised within each chapter, rather than being presented as a separate chapter.

Chapter 2

Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird

2.1 Abstract

Over evolutionary time, pathogen challenge shapes the immune phenotype of the host to better respond to an incipient threat. The extent and direction of this selection pressure depend on the local pathogen composition, which is in turn determined by biotic and abiotic features of the environment. However, little is known about adaptation to local pathogen threats in wild animals. The Gentoo penguin (*Pygoscelis papua*) is a species complex that lends itself to the study of immune adaptation because of its circumpolar distribution over a large latitudinal range, with little or no admixture between different clades. In this study, we examine the diversity in a key family of innate immune genes—the Toll-like receptors (TLRs)—across the range of the Gentoo penguin. The three TLRs that we investigated present varying levels of diversity, with *TLR4* and *TLR5* greatly exceeding the diversity of *TLR7*. We present

evidence of positive selection in *TLR4* and *TLR5*, which points to pathogen-driven adaptation to the local pathogen milieu. Finally, we demonstrate that two positively selected co-segregating sites in *TLR5* are sufficient to alter the responsiveness of the receptor to its bacterial ligand, flagellin. Taken together, these results suggest that Gentoo penguins have experienced distinct pathogen-driven selection pressures in different environments, which may be important given the role of the Gentoo penguin as a sentinel species in some of the world's most rapidly changing environments.

2.2 Introduction

All organisms are challenged by pathogens in their surrounding environments, but it is clear that the pathogen pressure can vary by location. Similar to free-living metazoans, a latitudinal species-richness gradient has been identified in several parasitic and pathogenic taxa, which may be driven by temperature and other abiotic and biotic factors [148, 225, 499]. Given this gradient in pathogen pressure, it follows that natural selection on the host will favor distinct immune phenotypes in different environments, as suggested by major histocompatibility complex II genetic diversity patterns in Humboldt penguins associated with higher pathogen diversity in lower latitudes [515, 516]. In our study, we sought to test the hypothesis that pathogen-driven selection can drive distinct patterns of host immune system genotype and phenotype, using the Gentoo penguin (*Pygoscelis papua* ssp.) as a model species complex.

The Gentoo penguin complex [114, 605] is ideally suited for investigating pathogen-driven selection on the immune system. Firstly, it has a circumpolar range, spanning the largest latitudinal range of any penguin species, between 46 °S and 66 °S, with breeding colonies in most of the Southern Ocean's sub-Antarctic islands, as well as the islands off Tierra del Fuego in South America, South Georgia, the Scotia Arc, and the Western Antarctic Peninsula [561]. Secondly, population monitoring of the species has shown it to be growing at the southern end of its range [388], with highly fluctuating changes over time in colonies in the South Atlantic and Indian

Oceans [353, 587]. Thirdly, the Gentoo penguin is a highly philopatric seabird known to remain close to its breeding colonies year round [111, 256, 588, 635], limiting gene flow across breeding regions [114, 357, 605].

Furthermore, across its range, Gentoo penguins overlap with (and occasionally co-occur in mixed colonies with) King (*Aptenodytes patagonicus*), Magellanic (*Spheniscus magellanicus*), Macaroni (*Eudyptes chrysolophus*), and Southern Rockhopper penguins (*Eudyptes chrysocome*) in sub-Antarctic colonies, as well as congeneric Adélie (*P. adeliae*) and Chinstrap penguins (*P. antarcticus*) in its Antarctic range. Gentoo penguin colonies are also frequented by a number of flying birds with vast ranges, including albatrosses and petrels (Order: Procellariiformes), as well as predator-scavengers like skuas (genus: *Stercorarius*) and sheathbills (genus: *Chionis*) that could introduce and/or spread novel avian pathogens. Levels of human interaction also vary across the range, from permanent settlements with livestock present near colonies in the Falkland/Malvinas Islands, to seasonal or year-round scientific research stations in sub-Antarctic and Antarctic colonies, and an increasing presence of Antarctic tourism. Differences in sympatric interactions with other species across the range of the Gentoo penguin are likely to result in different pathogen challenges and therefore different selective pressures.

To investigate genetic diversity across the immune system, many immunogenetic studies on penguins have focused on the major histocompatibility complex [64, 315, 516, 589]. Increasingly, however, the Toll-like receptors (TLRs) are recognized as important monogenic determinants of disease-resistance phenotypes, and are therefore important operands for natural selection [220]. Toll-like receptors are the best-studied family of pattern-recognition receptors in the vertebrate innate immune system, representing the front line of detection of pathogen challenge [295]. TLRs respond to highly conserved microbe-associated molecular patterns (MAMPs) that are structurally conserved in large groups of pathogens. Upon binding of a MAMP, TLRs undergo dimerization and initiate an intracellular signaling cascade that culminates in the production of anti-pathogen effector molecules [7, 69].

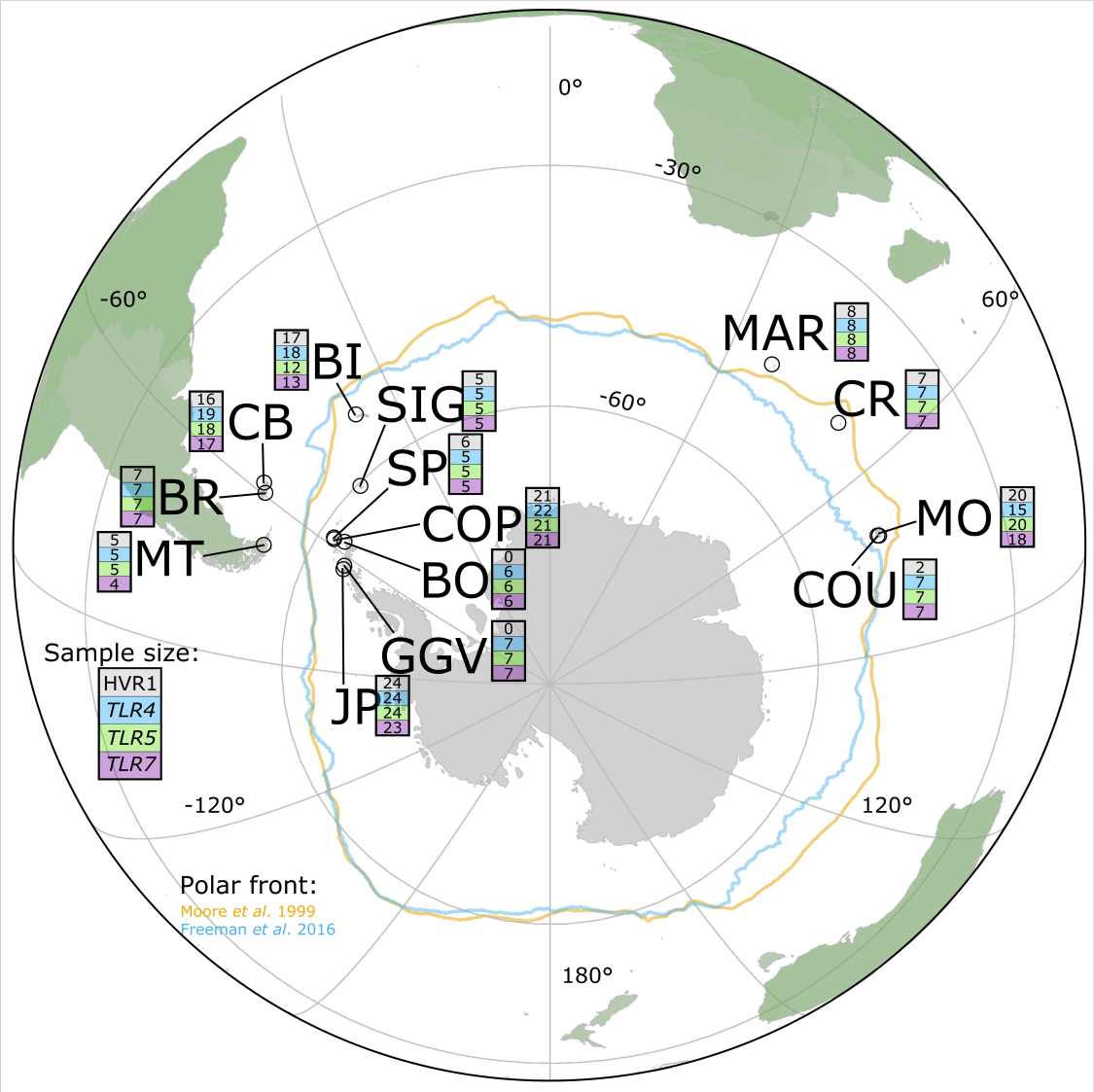


Figure 2.1: Locations and per-locus sample sizes for each sampled Gentoo penguin colony in the Southern Ocean and Antarctica. Solid-coloured lines represent the reported position of the Polar Front, based on the analyses of Freeman et al (2016) [188] and Moore et al (1999) [414]. Depending on the analysis, the Kerguelen and Crozet Islands can lie just north or just south of the Polar Front [188, 414]. CR, Crozet Island; MAR, Marion Island; COU, Courbet Peninsula, Kerguelen; MO, Pointe du Morne, Kerguelen; CB, Cow Bay, Falkland/Malvinas Islands; BR, Bull Roads, Falkland/Malvinas Islands; BI, Bird Island, South Georgia; MT, Martillo Island, Tierra del Fuego; SIG, Signy Island, South Orkney Islands; COP, Copacabana (Admiralty Bay), King George Island, South Shetland Islands; SP, Stranger Point, King George Island, South Shetland Islands; BO, Bernardo O'Higgins Base, Western Antarctic Peninsula; GGV, Gabriel González Videla Base, Western Antarctic Peninsula; JP, Jougla Point, Western Antarctic Peninsula.

Vertebrates have six major families of TLRs which are typically conserved across evolutionary time to retain specificity for a particular MAMP or family of MAMPs. In most avian species, there are ten recognized TLRs [71, 81, 493]. Of these, TLR4 and TLR5 respond to the bacterial agonists lipopolysaccharide and flagellin, respectively, whereas TLR7 responds to single-stranded RNA of viruses in the endosomal compartment [102, 202, 385].

To investigate TLR diversity across the range of the Gentoo penguin complex, we sequenced the full coding sequences of *TLR4*, *TLR5*, and *TLR7,* as opposed to targeted portions of certain exons as in previous studies [129]. These three genes represent bacterial- and viral-sensing TLRs that are present in almost all vertebrates. Samples (n = 155) were obtained from a broad geographic range across the range of the species (Figure 2.1), representing the largest geospatial scale of any immunogenetics study outside of humans. We describe patterns of diversity in TLRs that have a clear spatial component, and provide evidence that some of the diversity in *TLR4* and *TLR5* is driven by positive selection between different locations. We also demonstrate that two of the positively selected residues in *TLR5* yield a phenotypic difference in the response of the receptor to flagellin, providing further evidence that Gentoo penguins have experienced differential pathogen-driven selection pressures in different environments.

2.3 Materials and methods

2.3.1 Sample collection

This study used 155 blood samples from Gentoo penguins, previously obtained in the framework of other projects. Samples were collected between 1999 and 2017 at the 14 sites shown in Figure 2.1 (details in Appendix Table A.17). To take blood, penguins were held with the flippers restrained and the head placed under the arm of the handler, or they were wrapped in cushioned material covering the head and

preventing movement, to minimize stress during handling [348]. A second handler took up to 1 ml blood from the brachial, intertarsal, or jugular vein using a 25- or 23-G needle and 1 ml syringe or capillary, after cleaning the area with an alcohol swab. Total restraint time was generally 2–3 min. The animal was then released at the edge of the colony and observed to ensure it returned to its normal behavior. Blood was stored in 95% ethanol or Queen's lysis buffer at -20 °C for transport at room temperature and subsequent storage at -20 °C upon arrival. All blood samples were imported under the appropriate animal by-product import licenses.

Sampling was conducted under permits from each site's territorial government or governing agency. These permits for animal handling were issued following independent institutional ethical review of the sampling protocols, in accordance with Scientific Committee for Antarctic Research (SCAR) guidelines.

2.3.2 DNA extraction

DNA for samples from MO, CB, BI, COP, and JP was extracted from blood samples using QIAGEN DNeasy Blood and Tissue kits. The digestion step was modified to include 40 μl proteinase K and extended to 3 h for blood samples. Details of the modifications made to the protocols for tissue samples are available in [660, 661]. All these samples were treated with 1 μl Riboshredder (Epicentre) to reduce RNA contamination and DNA was visualized on a 1% agarose gel to confirm high-molecular weight DNA was present. DNA concentration and purity were measured on a Qubit and Nanodrop (Thermofisher Scientific), respectively. These samples are stored at the University of Oxford for future analysis. DNA from all other sampling sites was isolated using a modified salt protocol [12], with details in [605], stored at the Pontificia Universidad Católica de Chile for future analysis.

2.3.3 TLR genotyping

Primers for *TLR4*, *5*, and *7* coding regions were designed using the primer design feature based on Primer3 2.3.7 [593] in Geneious v.11.0 (Biomatters, http://www.geneious.com). Reference coding sequences for primer design were derived from the congeneric Adélie penguin (*P. adeliae*) reference genome (GenBank accession JMFP01000000) and unpublished Gentoo penguin genomic data. *TLR4* amplifications for samples from MO, CB, BI, COP, and JP were conducted in 12 μl volumes (9 μl Qiagen Taq PCR Master Mix, 2 μl 10 μM forward and reverse primer mix, and 1 μl of template DNA diluted 1:100). *TLR5* and *TLR7* amplifications for these samples were conducted in 25 μl volumes containing 5 μl 5× Phusion High Fidelity (HF) Buffer (New England Biolabs, UK), 0.5 μl 10 mM dNTPs, 1.25 μl of 10 μM forward primer, 1.25 μl of 10 μM reverse primer, 2 μl of template DNA diluted 1:100, 0.25 μl of Phusion Hot Start Flex DNA Polymerase (New England Biolabs, UK), and 14.75 μl nuclease-free water. One GC-rich region in *TLR7* required the use of Phusion GC buffer, rather than HF buffer for amplification. PCR products were visualized on a 1% agarose gel stained with SYBR Safe. The primers and PCR reaction conditions are fully detailed in Appendix Table A.2. PCR products were sequenced using Macrogen Europe's EZ-Seq (*TLR4*) or Eco-Seq (*TLR5* and *TLR7*) services (http://www.macrogen.com, Netherlands) for purification and Sanger sequencing, using the same PCR primers for sequencing.

For all remaining sampling sites, which underwent whole-genome sequencing, a total of 100 ng of genomic DNA was fragmented to an average of 350 bp to construct paired-end libraries using the Illumina TruSeq Nano kit with the included indexed adapter and barcode. A total of six PCR cycles were used for enrichment, purified with sample purification beads, quantified using a Qubit fluorometer, and then sequenced to $\approx 20\times$ coverage with 150 bp paired-end reads using an Illumina HiSeq X platform at MedGenome.

Sequences for each TLR coding region were assembled, edited, and aligned us-

ing Geneious v.11.0. For *TLR4*, which has multiple exons along the coding region, exon sequences were extracted and concatenated for further analysis. Heterozygous sites and single-nucleotide polymorphisms were detected by visually examining chromatograms. In cases of doubt, resequencing was accomplished so that only high-quality reads from multiple sequencing runs were called as single-nucleotide polymorphisms. All heterozygous sites also had homozygous individuals within the data set, and each gene had at least one haplotype homozygous across the full length of the gene. All alleles were verified using a combination of multiple independent Sanger sequencing runs and, where available, the whole-genome sequencing data. The International Union of Pure and Applied Chemistry (IUPAC) code for degenerate nucleotides was used to label heterozygous positions.

2.3.4 mtDNA genotyping

For mitochondrial DNA, the hypervariable region of the mitochondrial control region (HVR1), also known as Domain I, was amplified using the primers GPPAIR3F and GPPAIR3R [113] for samples from MO, CB, BI, COP, and JP. Amplifications were conducted in 25 μl volumes according to the manufacturer's instructions, using Phusion Hot Start Flex DNA Polymerase (New England Biolabs, UK) and 2 μl of template DNA diluted 1:100. Amplifications involved a two-step PCR, with an initial cycle of 98 °C for 30 s, 40 cycles of 98 °C for 10 s and 72 °C for 20 s, followed by a 10-min extension at 72 °C. PCR products were visualized on a 1% agarose gel stained with SYBR Safe. PCR fragments were purified using an ethanol/sodium acetate precipitation, and sequencing was performed using the Applied Biosystems BigDye Terminator v3.1 sequencing kit (Applied Biosystems) with the same PCR primers as sequencing primers.

Published mtDNA HVR1 sequences for the samples from BI (GenBank accessions KJ646314-KJ646330, n = 16) and COP (KJ646361-KJ646382,n = 21) were included in the analysis for the relevant individuals. mtDNA data were not available for the

individuals from two Antarctic sites. Although other individuals from those sites have sequence data available, we only included data from individuals sequenced for both TLRs and HVR1.

Individual mtDNA fragments from all remaining sites (CR, MAR, COU, BR, MT, SIG, SP) were amplified using the primers tRNAGlu and AH530 [496]. These reactions included 0.4 μM of each primer, 1.5 mM 1$\times$ of PCR reaction buffer, $MgCl_2$, 200 μM of each dNTP, and 1 U of Taq Polymerase Platinum (Invitrogen) in a two-phase touchdown program [321]: (i) 10 min at 95 °C, and 11 cycles of 95 °C for 15 s; a touchdown with an annealing temperature of 60-50 °C for 30 s, with one cycle per 1 °C interval, and 72 °C for 45 s; (ii) 35 amplification cycles at 95 °C for 15 s, 50 °C for 30 s, and 72 °C for 45 s; and a final extension period of 30 min at 72 °C. The purification of these PCR products and sequencing was carried out by Macrogen using an ABI PRISM 3730XL.

Only overlapping segments of the HVR1 sequences common to all samples were used by aligning and editing with Geneious v11.0. Consensus sequences for the resulting 273-bp region of interest were extracted for analysis.

2.3.5 Population-level analyses

Haplotypes were inferred for each of the diploid TLR loci using the PHASE algorithm [558] implemented in DnaSP 6.12.10 [506] with 10,000 iterations and 1,000 burn-in iterations. Phased haplotype data were used as input to determine standard genetic diversity measures of each population, including number of polymorphic sites, haplotypes, haplotype diversity, and nucleotide diversity, using DnaSP 6.12.10 and Arlequin v3.5.1.3 [167]. Minimum spanning haplotype networks were constructed and visualized using PopART 1.7 [40, 347] for each locus. DnaSP was also used to identify synonymous and nonsynonymous polymorphic sites and frequencies. Arlequin was used to calculate Tajima's D [567], and Fu's Fs [191]. FSTAT v.2.9.3 [212] was used to calculate allelic richness, taking into account differences in sample size.

Because TLR nucleotide diversity has been observed to have a correlation to population size in some island bird populations [206], we evaluated the relationship between Gentoo penguin census population sizes (N_c) and haplotype diversity (H_d) at each locus. Gentoo penguins are philopatric but also show evidence of admixture within island groups and adjacent coastlines [357, 605]. For this reason, the census population sizes selected for the analysis were numbers of breeding pairs from the most recent available surveys of each archipelago or region (in the case of the South Shetland Islands and Western Antarctic Peninsula). The N_c size and source survey reference are available in Appendix Table A.18. Spearman's rank correlations and P-values were calculated for each diversity-size comparison.

For population differentiation comparisons, Arlequin was used to calculate pairwise F_{ST} distances based on haplotype frequencies [630], and pairwise Φ_{ST} for the TLR and mtDNA sequences [168]. FindModel [478] was used to find the best fit substitution model for use in Arlequin. Φ_{ST} calculations for the TLR loci were obtained using the Tamura and Nei [572] substitution model, whereas mtDNA Φ_{ST} analysis was carried out using the Kimura 2-Parameter model [310] with a gamma of 0.27. AMOVA was used to compute hierarchical F-statistics, with 10,000 permutations, to evaluate likely patterns of genetic structure, seeking to identify the population grouping that maximized the among-group variation (F_{CT}) and minimized the variation among colonies within groups (F_{SC}) [168]. Significance of overall and pairwise genetic distances was computed using 1,000,000 permutations. We used the SGoF+ method [90] within the Myriads software [89] to correct for multiple tests.

To test for isolation-by-distance, shortest distances by sea (during summer ice extent), in km, were computed between each sampling location using Google Earth v7.3.2.5776 (Appendix Table A.7), which were then related to pairwise F_{ST} in a Mantel test, implemented in Arlequin.

2.3.6 Population divergence and phylogeography

Phylogenetic reconstruction and estimates of divergence time were carried out using BEAST 2.5.2 [70]. The evolutionary model for mtDNA analysis was selected using jModelTest v. 2.1.10 [135], testing 88 candidate models and selecting the best fit using the Bayes Information Criterion. All 88 models were within the 100% confidence interval, with HKY+G selected for further divergence analyses [234]. A total of 15 Adélie penguin (*P. adeliae*) and 15 Chinstrap penguin (*P. antarcticus*) mitochondrial HVR1 GenBank sequences [113], aligned and cropped to the equivalent size of the Gentoo sequences to avoid bias, were included in the analysis as outgroups.

The most recent common ancestor prior was set for the *Pygoscelis* genus at 7.6 Ma [562], derived from the fossil calibration for *P. grandis* [615], with a normal distribution, and SD (δ) of 1.3 Ma. A strict molecular clock with a starting prior of 1.0 and a Yule speciation process for branching rates, with uniform priors for birth and clock rates of 1.0 was applied. Four independent runs of 30 million MCMC chains were performed, logging parameters every 3,000 steps. The four independent runs were then combined using LogCombiner v.2.5.2 and assessed for convergence within Tracer v.1.7.1 [488]. All parameters converged with ESS values $>$ 6,000. A maximum clade credibility tree was then generated using TreeAnnotator v.2.5.2 (part of the BEAST software distribution) and visualized in FigTree v.1.4.3.

2.3.7 Selection analyses

Phylogenetic inference was carried out on phased sequence data for each TLR locus using RAxML-NG using a GTR substitution matrix [324]. To detect selection, maximum likelihood analysis of ratios of nonsynonymous to synonymous nucleotide substitutions (dN/dS; ω) was performed with the codeml package of programs in PAML v. 4.9 [654, 652]. Various models were fitted to the multiple alignments: M1a (neutral model; two site classes: $0 < \omega_0 < 1$ and $\omega_1 = 1$); M2a (positive selection; three site classes: $0 < 0 < 1$, $\omega_1 = 1$, and $\omega_2 > 1$); M7 (neutral model; values of ω

fit to a beta distribution where $\omega > 1$ disallowed); M8 (positive selection; similar to M7 but with an additional codon class of $\omega > 1$); and M8a (neutral model; similar to M8 but with a fixed codon class at $\omega = 1$). Likelihood ratio tests were performed on pairs of models to assess whether models allowing positively selected codons gave a significantly better fit to the data than neutral models (model comparisons were M1a vs. M2a, M7 vs. M8, and M8a vs. M8). In situations where the null hypothesis of neutral codon evolution could be rejected ($P < 0.05$), the posterior probability of codons under selection in M2a and M8 was inferred using the BEB algorithm [655].

2.3.8 *In silico* prediction of polymorphism functional consequences

Physicochemical distances between amino acid variants were assessed using distance matrices provided by several authors [162, 214, 410, 549, 594] (Appendix Table A.15). Predicted functional consequences of amino acid substitutions were assessed using the homology-based tools SIFT [432] and PolyPhen-2 [2], using online servers (https://sift.bii.a-star.edu.sg/ and http://genetics.bwh.harvard.edu/pph2/, both accessed December 2019). Transmembrane domain positions were predicted using Phobius [281] (http://phobius.sbc.su.se/, accessed December 2019).

2.3.9 Functional analysis of *TLR5* genotype expression in CRISPR-Cas9 edited HEK-Blue cells

In order to functionally assess positively selected *TLR5* polymorphisms *in vitro*, two full-length *TLR5* sequences were synthesized including the two signaling domain polymorphisms that had the strongest signature of selection (GBlocks, IDT). Synthetic genes were cloned using the Gibson assembly method [204] into the p3×FLAG-CMV-14 expression vector (Sigma) which incorporates a 3×-FLAG sequence on the C-terminus of the expressed construct. Insert-containing vector was purified using the ZymoPURE II plasmid Maxiprep with the optional endotoxin-removal step (Zymo).

Both constructs were transiently expressed using TransIT-2020 (Mirus Bio) in custom HEK-Blue Null1 cells (InvivoGen) that had undergone genome editing using the CRISPR-Cas9 technique to disrupt endogenous human *TLR5* (additional information in Appendix A). Cells expressing Gentoo penguin *TLR5* constructs were challenged with *Salmonella* Typhimurium-derived flagellin (FLA-ST; InvivoGen) at 100 ng/ml and incubated for 24 h. Cell supernatants were harvested and NF-κ B activity was assessed by measuring the absorbance at 405 nm on a FLUOstar Omega microplate reader (BMG Labtech) following the addition of p-nitrophenyl phosphate substrate, according to the manufacturer's instructions (SIGMAFAST, Sigma). Expression levels were monitored by subjecting cell lysates to a direct anti-FLAG ELISA. Cell lysates were harvested in ice-cold RIPA buffer (ThermoFisher) and proteins were immobilized on high-bind ELISA plates (VWR) using coating buffer (BioLegend) overnight at 4 °C. Wells were blocked using StartingBlock (PBS) blocking buffer (ThermoFisher) for 1 h and then incubated with mouse monoclonal anti-FLAG M2 antibody (Sigma, 1:1,000) at 37 °C for 1 h, followed by incubation with goat anti-mouse IgG-HRP (ThermoFisher, 1:10,000) for 1 h at 37 °C. Reactive protein amount was then assessed by the addition of 3,3′,5,5′-tetramethylbenzidine substrate and measurement of absorbance at 650 nm. Expression data were then used to normalize signaling data. Statistical differences between means were determined by a two-tailed Student's t-test, and statistical significance was considered to be $P < 0.05$. Transfections were carried out in three independent wells per condition, and the experiment was conducted on at least three independent occasions.

2.3.10 Data availability

All sequence data generated in this study were deposited in GenBank under accession numbers MN394222–MN394376 (*TLR4*), MN313018–MN313169 (*TLR5*), MN312870–MN313017 (*TLR7*), and MN566362–MN566421 (mitochondrial control region, HVR1).

2.4 Results

2.4.1 Amplification of TLR genes in the gentoo penguin

Through successful amplification by PCR or whole-genome sequencing, we were able to confirm that Gentoo penguins have clear homologs of *TLR4*, *TLR5*, and *TLR7*, finding no evidence of gene loss or pseudogenization, as has been reported in other avian lineages for *TLR5* [603], and in African penguins for *TLR7* [129].

The length of the *P. papua TLR4* coding sequence matches the longest reported length (2,550 bp/849 aa) in other bird species. For *TLR5*, there is a start codon that yields an open reading frame (ORF) in line with the length of previously published *TLR5* sequences (2,589 bp/862 aa; [603]), but the ORF continues upstream of the putative start codon, yielding a complete ORF that is 2,643 bp/880 aa, which is 54 bp longer than other reported avian *TLR5* sequences. Both the longer and shorter ORFs respond to flagellin in our *in vitro* system (data not shown), suggesting both could be functional *in vivo*. The length of the *TLR7* coding exon, at 3,126 bp/1,042 aa, falls within the reported range of avian coding sequence lengths.

2.4.2 Diversity indices and population differentiation

2.4.2.1 Mitochondrial hypervariable region

All colonies with more than two sampled individuals presented high levels of mitochondrial hypervariable region 1 (HVR1) haplotype diversity (H_d = 0.60–1.00; Figure 2.2 and Appendix Table A.1) as obtained in DnaSP 6.12.10 [506]. Differentiation between colonies at this locus was significant and in line with previous data for this species (Appendix A.6), with four clearly differentiated clades obtained through population-level analyses in Arlequin v3.5.1.3 [167]: (i) a southern clade consisting of colonies South of the Polar Front in South Georgia, the South Orkneys, the South Shetlands, and Western Antarctic Peninsula; (ii) a South American/Falklands/Malvinas clade; (iii) a Kerguelen clade; and (iv) a North Indian Ocean

clade (Marion and Crozet Islands). Our BEAST 2 phylogenetic analysis of HVR1 showed support for a division within the Gentoo penguin complex occurring ∼3.36 Ma (1.72–4.88 Ma), when the North Indian Ocean clade diverges from all others. The Kerguelen lineage diverges from the Atlantic lineages 2.36 Ma (1.14–3.51 Ma), and the populations North and South of the Polar Front within the Atlantic Ocean appear to diverge 1.19 Ma (0.51–1.75 Ma; Figure 2.3). Within each clade, no clear site-specific mtDNA structure was noted.

2.4.2.2 *TLR4*

For *TLR4*, 13 polymorphic sites were identified, with a total of 21 distinct phased haplotypes coding for 9 unique protein variants (Appendix Table A.1). Indian Ocean colonies (CR, MAR, COU, and MO) presented the highest levels of haplotype diversity (H_d = 0.66–0.90), whereas the South American colony in Tierra del Fuego (MT) and the Falkland/Malvinas Islands (CB and BR) had very low diversity (H_d = 0.00–0.05; Figure 2.2). Among Gentoo penguin colonies south of the Polar Front, diversity was highest in the central colonies of SIG, COP, SP, and BO (H_d = 0.51–0.63), decreasing sharply in BI in South Georgia (Hd = 0.00), as well as southward down the Western Antarctic Peninsula (GGV and JP; H_d = 0.08–0.26).

Comparing pairwise genetic distances (F_{ST} and Φ_{ST}) for *TLR4* in Arlequin (Figure 2.5 and Appendix Table A.3), the Indian Ocean populations of Crozet and Marion differed significantly ($F_{ST} > 0.3$, $P < 0.01$) or near-significantly (corrected $P \approx 0.01$) from all other colonies in the Atlantic. Crozet and Marion did not differ significantly from each other (F_{ST} = 0.001, P = 0.40), whereas some haplotypes were shared between CR/MAR and the Kerguelen Island colonies of COU and MO, yielding some nonsignificant pairings among the four Indian Ocean colonies. In the Atlantic, one haplotype, also present in the Indian Ocean in lower frequencies, dominated across all the colonies (Figure 2.4). The central sites in the Scotia Arc (COP/SP on King George Island, Signy Island, and BO on the northern end of the Western Antarctic Peninsula) exhibited greater diversity than other Atlantic sites and contained private

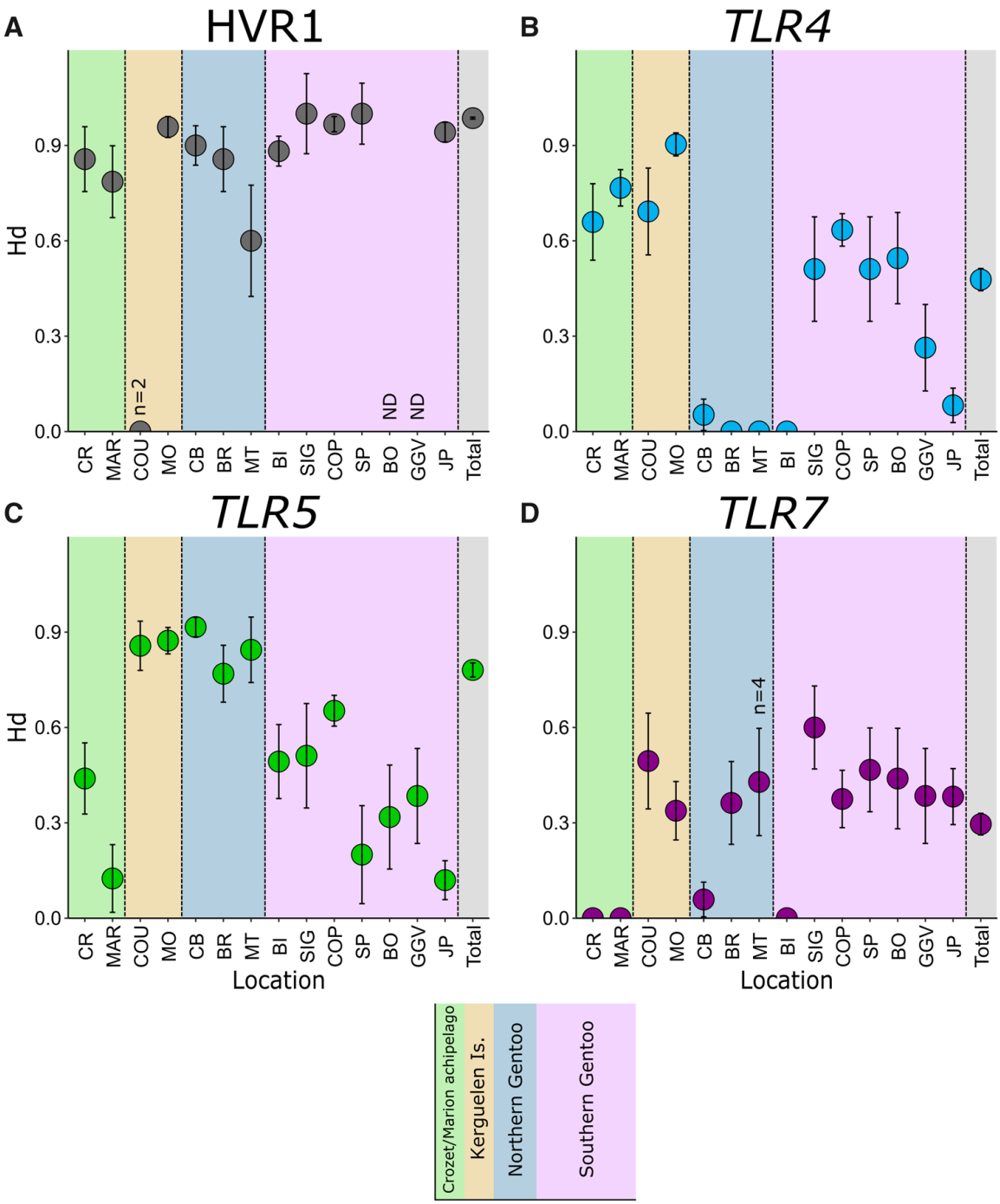


Figure 2.2: Haplotype diversity (H_d) for the hypervariable mitochondrial control region (HVR1; A) and three TLR loci (B-D), for all sampling sites for which data were available. All sites had $n \geqslant 5$ unless otherwise indicated (ND, no data). Location abbreviations are the same as in previous figure.

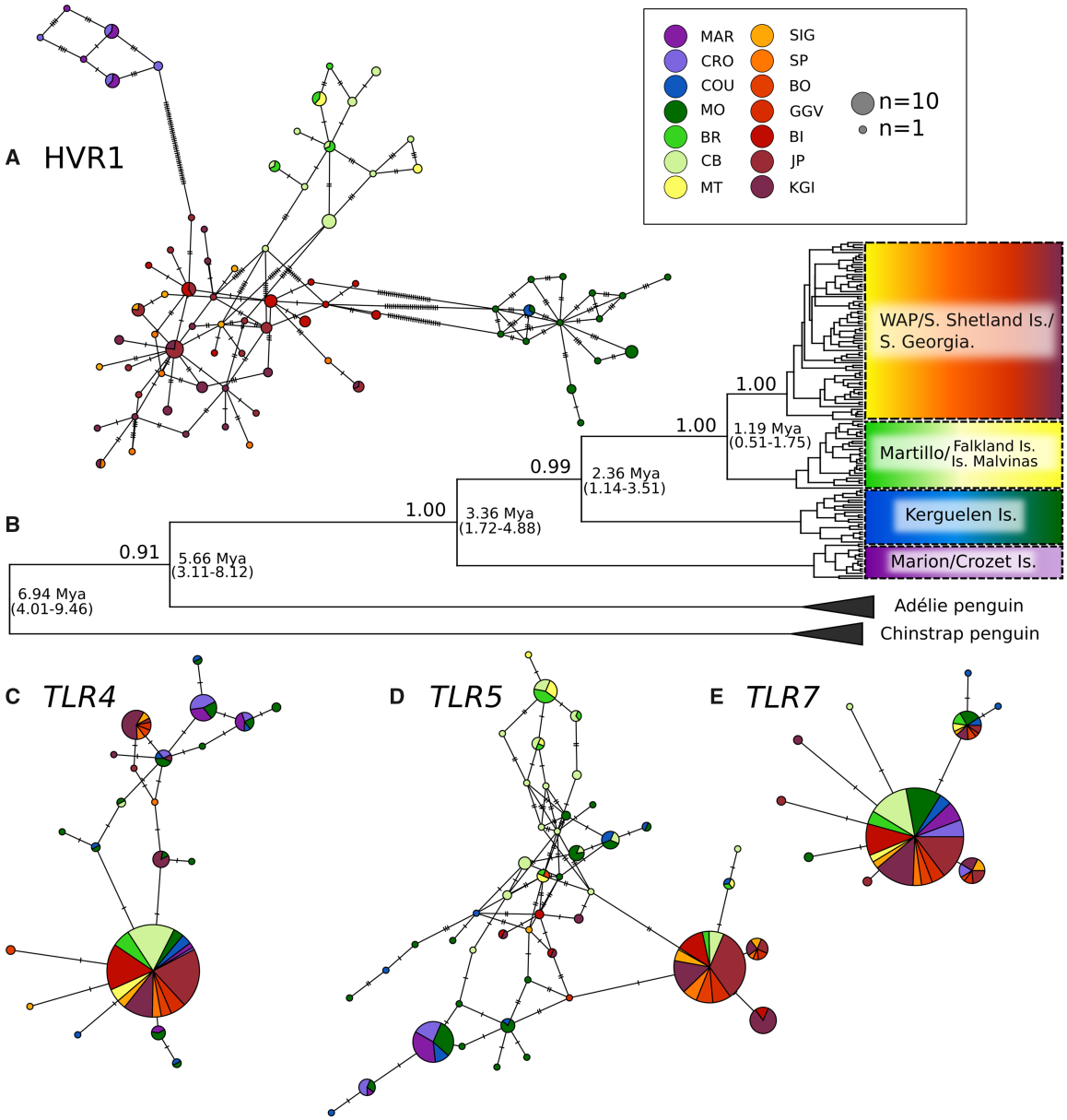


Figure 2.3: Minimum spanning haplotype networks for mtDNA HVR1 (A), *TLR4* (C), *TLR5* (D), and *TLR7* (E) for Gentoo penguin colonies, along with mtDNA HVR1 maximum clade credibility tree (B), using congeneric penguin species as outgroups. Location abbreviations are the same as in previous figures. For minimum spanning haplotype networks, pie charts represent single haplotypes, whereas segment size refers to the contribution of individual sampled sites to the proportion of overall haplotype frequency. Size of pie charts reflects the number of individual birds with the observed haplotype. Dashes on connecting lines each denote one nucleotide change.

alleles. The overall pattern of significance for F_{ST} and Φ_{ST} comparisons can be seen in Figure 2.5. Not surprisingly, colonies within the same island group that are in close proximity to each other (COU/MO, CB/BR, COP/SP) did not differ significantly, despite variations in sample sizes. Hierarchical population structure was detected for *TLR4* across Gentoo penguin colonies using Arlequin (analysis of molecular variance, AMOVA; global $F_{ST} = 0.32$, $P < 0.0001$). The proportion of variation resulting from differences among groups was 24.01% ($F_{CT} = 0.24$, $P = 0.003$) when colonies were placed into four groups, coinciding with the four mtDNA clades: (i) Marion/Crozet archipelagos, (ii) Kerguelen Is., (iii) Falkland/Malvinas and Tierra del Fuego, and (iv) south of the Polar Front in the Scotia Arc and Antarctic Peninsula. However, F_{CT} increased to 28.19% ($P = 0.007$) when Bird Island was grouped separately from other Southern Gentoo penguin colonies, a pattern also suggested in genomic-level analyses [114].

2.4.2.3 *TLR5*

TLR5 was the most diverse TLR locus analyzed. Twenty polymorphic sites were identified, with a total of 46 distinct phased haplotypes coding for 32 unique protein variants (Appendix Table A.1). Five Gentoo penguin colonies north of the Polar Front (COU/MO in Kerguelen, CB/BR in Falklands/Malvinas, and MT in Tierra del Fuego) exhibited the highest diversity measures (6–17 haplotypes, $H_d = 0.77$–0.92, Figure 2.2), despite differences in sample size. This is unexpected given that Martillo Island is the smallest known population of this species ($N_c = 12$ breeding pairs [203]), yet still maintained high diversity (five unique protein variants in a sample of $n = 5$) at this locus. Interestingly, Crozet and Marion Island colonies exhibited substantially lower genetic variation at this locus with only two haplotypes ($H_d = 0.13$–0.44), though these were shared with the Kerguelen colonies ($H_d = 0.86$–0.87; Figure 2.4). Southern colonies exhibited moderate diversity ($H_d = 0.51$–0.65), with the exception of SP in the South Shetland Islands and the Western Antarctic Peninsula at the edge of the range, with only two or three haplotypes ($H_d = 0.12$–0.38). Strikingly,

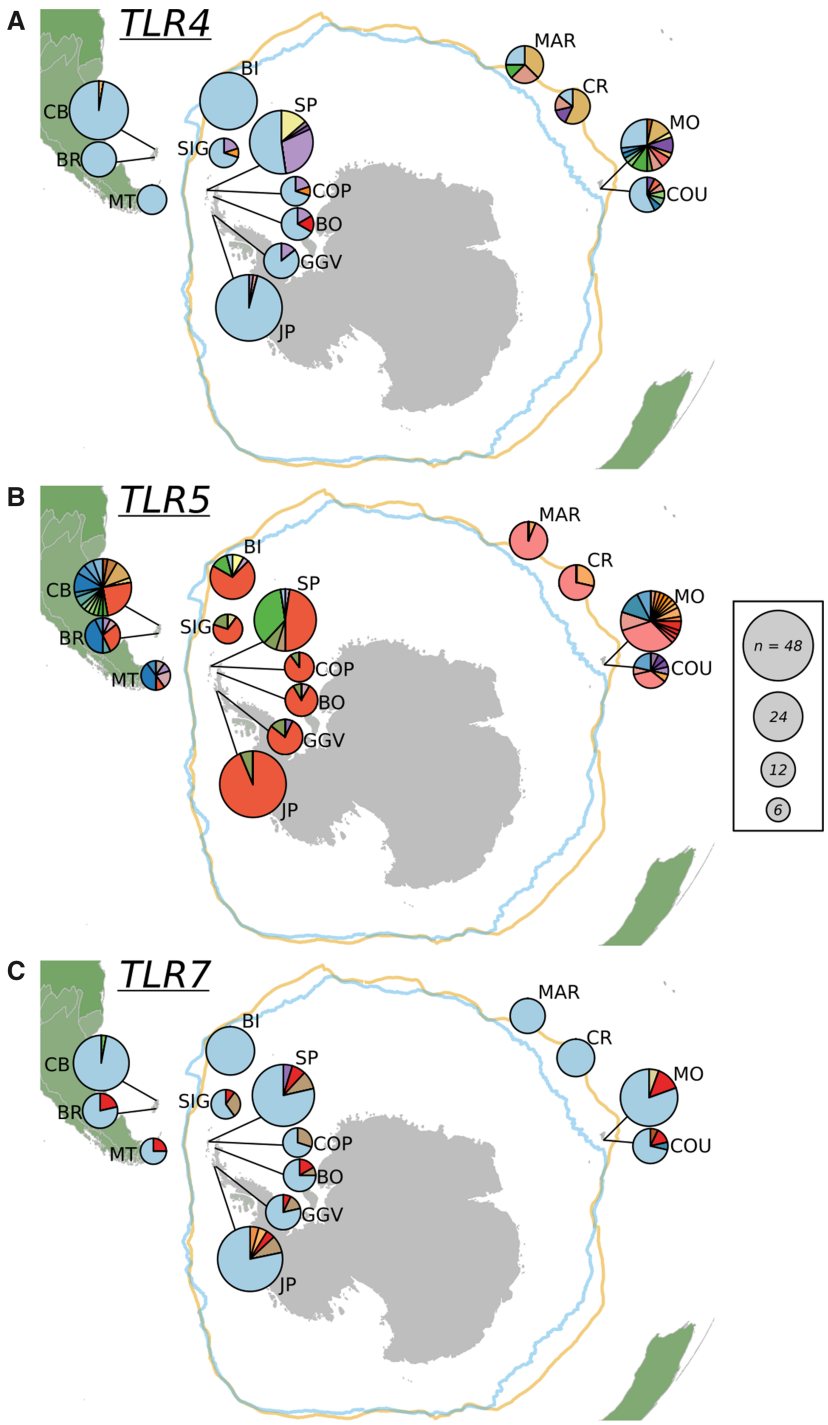


Figure 2.4: Haplotype diversity across Gentoo penguin sample populations for (A) *TLR4*, (B) *TLR5*, and (C) *TLR7*. For each locus, different colours represent unique haplotypes and each segment size reflects the proportion of birds in each location with that haplotype. Overall size of the pie chart reflects the number of birds sampled in each location. Location abbreviations are the same as in previous figures.

Atlantic colonies south of the Polar Front were dominated by one haplotype found less frequently (10–28%) in the northern Atlantic colonies and completely absent from all Indian Ocean colonies. The second-most prevalent haplotype in southern colonies was completely absent from all northern colonies.

Pairwise F_{ST} and Φ_{ST} values obtained in Arlequin revealed significant clustering by clade in terms of genetic distance at the *TLR5* locus (Figure 2.5 and Appendix Table A.4). Within three of the four clades mentioned earlier, there were no significant differences (CR/MAR, COU/MO, and CB/BR/MT). The only significantly differentiated within-clade pairs lay in Gentoo penguin colonies south of the Polar Front, where Jougla Point differed significantly from COP (highest diversity among Southern Gentoo penguins) and BI (at the opposite geographic edge of this clade's range). All other colonies differed significantly from all colonies outside their clade (F_{ST} range 0.1-0.88, $P < 0.01$), with CR/COU being near-significant ($F_{ST} = 0.13$, $P = 0.02$). This was reflected in AMOVA results, where a four-clade grouping presented a proportion of variation resulting from differences among groups of 32.70% ($F_{CT} = 0.327$, $P < 0.0001$), and isolating Jougla Point increased the F_{CT} to 56.56% ($P < 0.0001$).

2.4.2.4 *TLR7*

TLR7 was the least diverse TLR locus, with nine polymorphic sites, ten phased haplotypes, and eight unique protein variants present across the study colonies (Figure 2.2 and Appendix Table A.1). One haplotype was predominant in all colonies (frequency of 70-100%) and seven of the ten haplotypes were private alleles, only found in single colonies (Figure 2.4). In the Indian Ocean, Kerguelen colonies (COU/MO) had relatively more diversity ($H_d = 0.34$-0.50) than the Crozet and Marion Island colonies ($H_d = 0.00$), which had only one haplotype. In the Atlantic, South Georgia (BI) had only one haplotype ($H_d = 0.00$), whereas all northern Atlantic colonies (CB/BR/MT) and SP in the South Shetland Islands presented two haplotypes ($H_d = 0.06$-0.47). Other southern colonies contained three to four haplotypes ($H_d =$

0.38-0.60), whereas the southernmost colony at Jougla Point (JP) on the Western Antarctic Peninsula exhibited the most unique haplotypes (H = 5; H_d = 0.38) and contained two private alleles. Unsurprisingly, only 1 of 91 pairwise comparisons between colonies were significant in terms of F_{ST} and 3/91 for Φ_{ST} ($P < 0.01$), with no pattern to this differentiation (Figure 2.5 and Appendix Table A.5). Different AMOVA groupings yielded an among-group variation (F_{CT}) no higher than 2.85% (four-clade grouping), further highlighting the lack of structure in this locus.

2.4.3 Effect of population size on diversity

Census population size was not significantly correlated to TLR haplotype diversity (*TLR4*, P = 0.067; *TLR5*, P = 0.75; *TLR7*, P = 0.64).

2.4.4 Isolation by distance

Significant isolation by distance, using shortest distances by sea between colonies in a Mantel's test, was detected in both *TLR4* (r = 0.515, P = 0.001) and *TLR5* (r = 0.593, P = 0.001; Appendix Table A.8). mtDNA HVR1 was less strongly correlated to isolation by distance (r = 0.312), though marginally significant (P = 0.015). On the other hand, the lack of diversity and structure in the *TLR7* locus yielded no significant correlation across the range of Gentoo penguin colonies sampled.

2.4.5 Analysis of positive selection

We investigated the *P. papua TLR4*, *TLR5*, and *TLR7* genes for evidence of positive selection, which could be an indicator of adaptation to local pathogen environments across the natural range of the species. Neutrality tests (Tajima's D and Fu's Fs) did not yield observable patterns of significant deviation from neutrality across the full length of these genes (Appendix Table A.1). Using a codon-specific approach, the site models in the codeml package of programs in PAML v4.9 [654, 652] were employed to test for signatures of positive selection in *P. papua* TLR loci. Nonsynonymous

sites observed and analyzed are graphically depicted in their relative positions on the proteins in Figure 2.6.

As expected, the majority of codons (*TLR4* 98.8%; *TLR5*, 98.4%) were predicted to be under purifying selection with the ratio of nonsynonymous to synonymous substitutions (dN/dS) being <1. Interestingly, 1.2% (*TLR4*) or 1.6% (*TLR5*) of codons in the alignment were found to be positively selected using M2a, and similar frequencies, 1.2% (*TLR4*) or 1.7% (*TLR5*), were found using M8. For *TLR7*, 99.9% of sites were found to be under purifying selection, whereas the remaining 0.1% was predicted to be under neutral selection. We investigated whether models that permit positive selection were a significantly better fit to the multiple alignments than models where dN/dS $\not> 1$ by performing likelihood ratio tests between pairs of models. For *TLR4* and *TLR5*, all model comparisons (M1a vs. M2a, M7 vs. M8, and M8a vs. M8) significantly favored the positive selection model compared with the neutral model (*TLR4*, P ≤ 0.017; *TLR5*, P $\leq 7.2\times10^{-23}$ for all comparisons), indicating that *P. papua TLR4* and *TLR5* have likely undergone positive selection. In contrast, for *TLR7*, the data were not a significantly better fit to the positive selection model compared with the neutral model ($P = 1$), therefore, the null hypothesis of codons being negatively and neutrally selected was not rejected.

For *TLR4* and *TLR5*, we then used the Bayes Empirical Bayes (BEB) algorithm to infer the posterior probability that a particular codon has experienced positive selection. For *TLR4*, three codons were predicted to have undergone positive selection at posterior probability of >0.90 under model M2a (Appendix Tables A.9-A.11). All but one of the *TLR4* polymorphic codons (12, 82, 236, 316, and 445) were located in the extracellular (leucine-rich repeat, LRR) domain, of which two were positively selected (12 and 236). The final positively selected codon (659) was located in the transmembrane domain (Figure 2.6). Of the three selected sites, one (12 V/A) is a relatively nonconservative change (see Appendix Table A.15 for amino acid distance metrics). One site in *TLR4* (12) has previously been found to be under positive selection in birds [603], whereas the remaining two (236 and 659) are novel selected

sites in penguins (Appendix Table A.16).

For TLR5, there were 13 amino acid variants, of which nine were predicted to have undergone positive selection at posterior probability of >0.90 in both M2a and M8 (Appendix Tables A.12-A.14). Of these, four were located in the extracellular domain (10, 285, 442, and 535), one was in the transmembrane domain (667), and four were located in the TIR (intracellular) domain (698, 747, 788, and 845; Figure 2.6). Two sites (442 and 698) have previously been reported as being positively selected in other birds [220, 603], whereas the remaining seven sites are novel selected sites in penguins. Interestingly, one site in the TLR5 extracellular domain (285) is adjacent to two residues known to be important for flagellin binding in Interface B [551, 659], and so could be important for ligand preference (Appendix Table A.16 and Appendix Figure A.4). Four positively selected sites in TLR5 were nonconservative changes (10C/Y, 442A/V, 667S/F, and 788S/C). It is interesting to note that the nonconservative TLR5 667F/S polymorphism is located in the transmembrane domain, a region that is typically constrained by the physicochemical requirement to embed in the cell membrane. Furthermore, the homology-based methods of amino acid substitution consequence prediction SIFT [432] and PolyPhen-2 [2] both predict the S667F change to be of high functional consequence (SIFT score = 0.01; PolyPhen-2 score = 0.998; Appendix Table A.15). Transmembrane integrity is predicted to remain intact, despite the nonconservative polymorphism, and the Phobius tool [281] predicts the transmembrane domain is unchanged by the polymorphism. Although the vast majority (57/64, 89% of available sequences) of avian TLR5 has a serine in this position in the transmembrane domain, only one other bird—the Northern Fulmar (Fulmarus glacialis)—has a phenylalanine in this position (Appendix Figure A.3), indicating there is likely to have been a functional consequence to the S667F change in the Gentoo penguin.

2.4.6 Functional analysis of selected TLR5 residues

Although *in silico* methods can be useful indicators of protein residues under selection, functional study is the only means to isolate the selected phenotype and its relevance. In order to assess whether key selected sites identified in the positive selection analyses have functional consequences, we developed an *in vitro* assay using transient expression of TLRs in a reporter cell line.

Since extracellular domain polymorphisms in TLRs are likely to give rise to preferences in ligand type [169], we focused on TMD/TIR domain polymorphisms that are likely to give rise to differential signaling in response to the same agonist. Given that five *TLR5* polymorphisms were located in the TMD/TIR domain, we tested the two polymorphisms with the highest posterior probability of selection from the PAML analysis—residues 667 (TMD) and 845 (TIR). Polymorphisms at these positions segregate well with geographical location: birds from Crozet and Marion were all homozygous for the 667S/845I haplotype, whereas 87.3% ($n = 71$) of birds from colonies South of the Polar Front were homozygous for the derived haplotype, 667F/845V (Figure 2.7). In the Kerguelen Islands, the ancestral 667S/845I haplotype predominated, but variants at both positions were present at lower frequencies. South American/Falklands/Malvinas colonies were the most diverse at the positions of interest, where 46.7% (n = 14) of birds were heterozygous at one or both of the sites. Overall, however, polymorphisms at these loci tended to co-segregate: 72.4% (n = 110) of birds were homozygote for either the ancestral (667S/845I) or derived (667F/845V) haplotypes.

Given the strong tendency for the alleles at these positions to co-segregate at the extremes of the range of *P. papua*, and the fact that these were the TIR/TMD polymorphisms with the highest likelihood of positive selection, the functional consequences of altering both residues together were investigated. FLAG-tagged constructs of both of the TLR5 TIR/TMD variants with the same LRR domain were transiently expressed in *TLR5*$^{-/-}$ HEK-Blue Null1 NF-κB reporter cells (additional information

in Appendix A) and protein expression levels were normalized using an anti-FLAG ELISA. Cells expressing either constructs were then treated with *Salmonella enterica* serovar Typhimurium-derived flagellin, or PBS control, and the NF-κB response was measured. Cells expressing either construct responded to flagellin, demonstrating that the *P. papua* TIR domain interacts efficiently with human adapter molecules. Interestingly, there was a marked enhancement ($\sim$1.5-fold, P = 0.04) of the flagellin response in the variant that was predominantly found in the Southern Gentoo penguin clade compared with the variant found in the Indian Ocean clades, suggesting that the derived haplotype 667F/845V has enhanced signaling capability compared with the ancestral genotype 667S/845I (Figure 2.7). These data provide further evidence that *P. papua TLR5* has undergone positive selection for different immune capabilities during the expansion of the species below the Polar Front toward the Antarctic Peninsula.

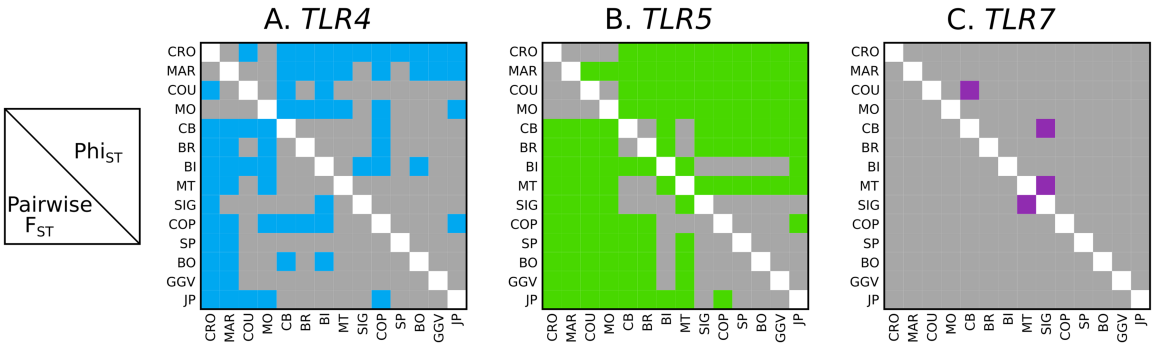


Figure 2.5: Visualization of pairings of Gentoo penguin breeding colonies, where pairwise F_{ST} values (below diagonal) or Φ_{ST} values (above diagonal) with significance $P < 0.01$ after correction for multiple tests using SGOF+ [90] are shown in colour for (A) *TLR4*, (B) *TLR5*, and (C) *TLR7*. Location abbreviations are the same as in previous figures.

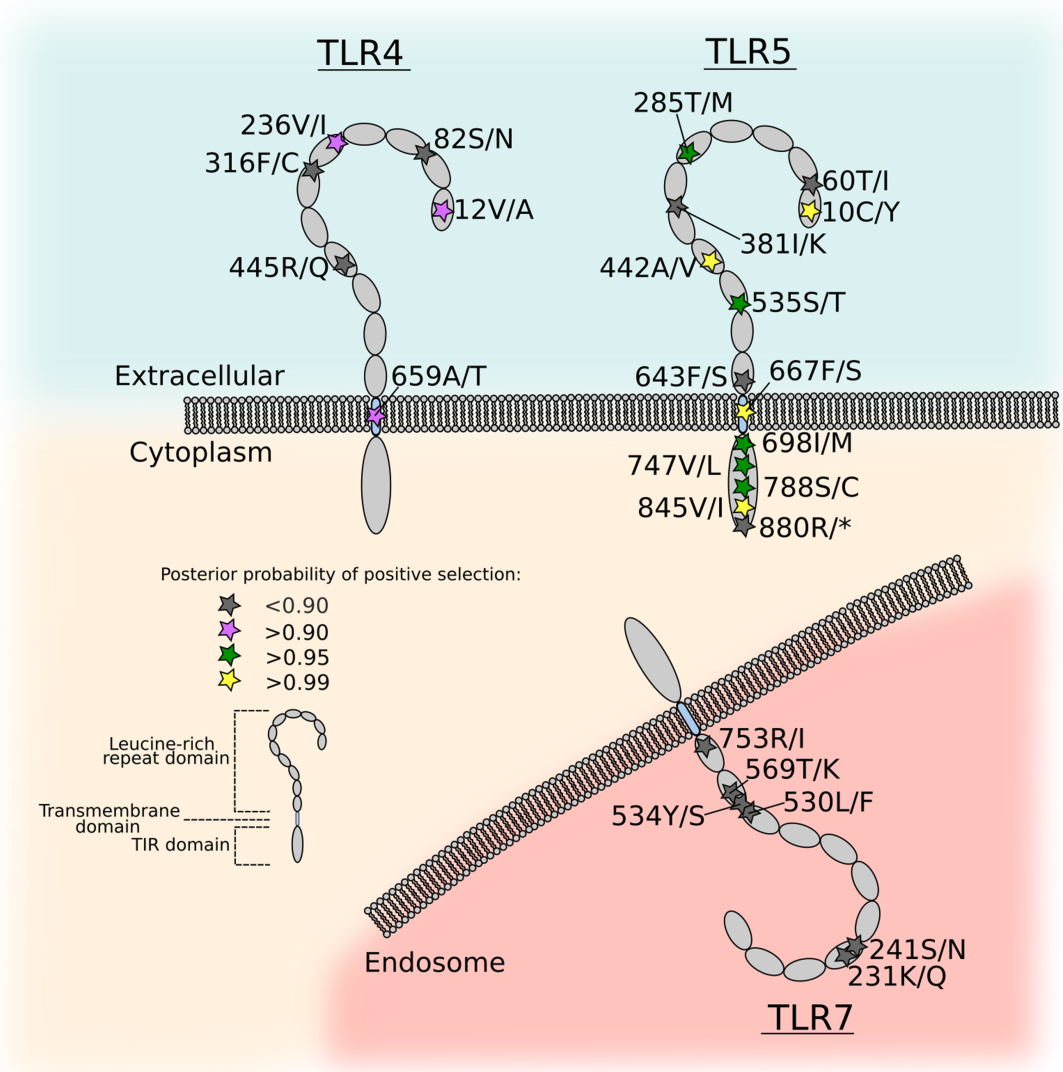


Figure 2.6: Positions of polymorphic and positively selected sites in *Pygoscelis papua* *TLR4*, *5*, and *7*. Schematic diagrams of TLRs in the extracellular (*TLR4* and *5*) and endosomal (*TLR7*) compartments show positions of amino acid variants resulting from nonsynonymous nucleotide substitutions. Variant positions are marked with stars and are coloured according to the likelihood of positive selection as determined by the M2a model in the codeml program in PAML.

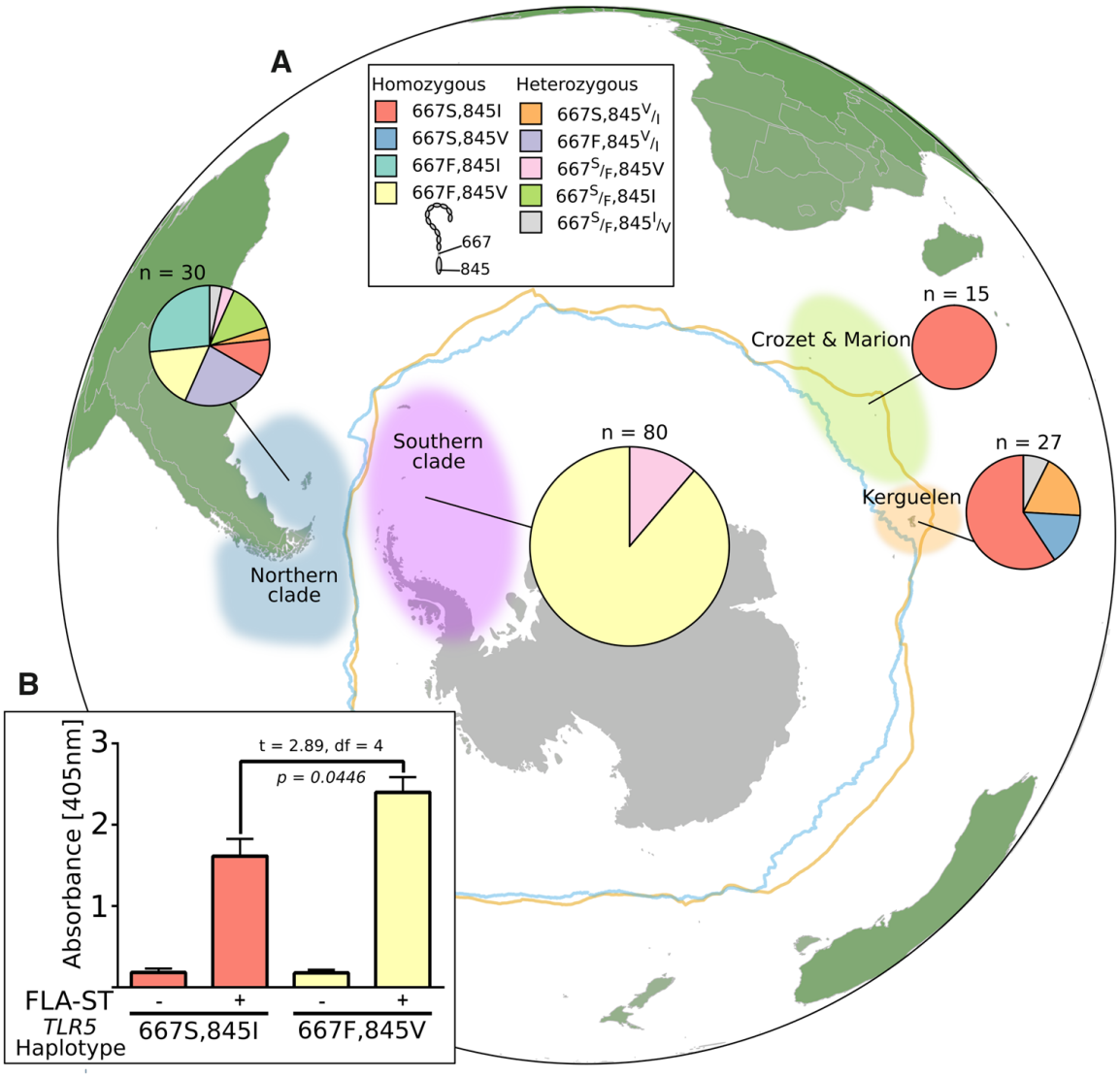


Figure 2.7: Distribution and functional differences of two selected amino acid residues in TLR5. (A) Mapped areas represent the four clades of *Pygoscelis papua* as determined by HVR1 analysis. Pie charts represent the allele proportions at the two selected residues in TLR5, and chart area is proportional to the number of samples. (B) Bar chart displays NF-κB response of the two genotypes following stimulation with FLA-ST.

2.5 Discussion

All vertebrates are subject to challenge by a plethora of pathogens that can exert strong selective pressures on host populations. The innate immune system, and in particular the TLRs, is responsible for both recognizing and responding to a pathogen threat by inducing inflammation and priming the adaptive immune response. To investigate TLR adaptation in Gentoo penguins, we sequenced the entire coding regions of *TLR4*, *TLR5*, and *TLR7*, which recognize products from bacterial and viral pathogens. Multiple individuals were sequenced from colonies at the extremes of the species' range (∼8,000 km between the most distant colonies), providing an extensive geospatial component to our analysis. We found spatially associated patterns of diversity in the TLRs, although greater diversity was observed in *TLR4* and *TLR5* compared with *TLR7*. Furthermore, clear evidence of positive selection in both *TLR4* and *TLR5* was identified, which was further reinforced by the demonstration that two of the TLR5 TMD/TIR domain polymorphisms are sufficient to alter the magnitude of responsiveness to flagellin. To our knowledge, no other TLR study outside of humans has supported predictions of positive selection with confirmation of functional polymorphism.

2.5.1 Patterns of diversification and selection

Most studies of TLR genetic diversity in wild populations have investigated small, bottlenecked, and/or endangered avian populations [129, 130, 211, 219, 221, 232]. In these populations, drift, rather than selection, is suspected to have been the main driver of sampled diversity due to recent bottlenecks or pronounced founder effects. Several studies have documented TLR diversity in domesticated animals such as pigs [134], cows [437], and chickens [565], but large-scale studies on whole-gene TLR variation in a wild population are lacking. Although study design can dramatically affect the diversity detected in different studies, it is noteworthy that with Gentoo penguins *TLR5* exhibited higher diversity than *TLR4*, whereas with domestic chicken

breeds (n = 110, 25 breeds; [565]) and grey partridge (n = 10; [607]), *TLR4* was more diverse than *TLR5* (Appendix Table A.19). The diversity of *TLR7* (compared with *TLR4* and/or *TLR5*) was relatively low in Gentoo penguins, domestic chicken breeds, and grey partridges (Appendix Table A.19).

To provide an internal reference for TLR diversity, we sequenced the mitochondrial HVR1 as a neutral marker in the same individuals. In line with previous studies, we found evidence of at least four deeply divergent clades in *P. papua* based on HVR1 sequence [114, 466, 605]. These more recent analyses support a revision, first proposed by de Dinechin et al. (2012) [137], of the previously accepted two subspecies model that was based on morphological characteristics [561]. These four clades are likely to have much greater divergence (millions of years) than what would be expected at the intraspecific level. We found evidence of differentiation according to this underlying population structure in *TLR4* and *TLR5* which further supports the argument for taxonomic revision of the species, with particular focus on the classification of colonies in South Georgia and the Indian Ocean. Conversely, *TLR7* was highly conserved across the species range and is clearly not subject to the same selection pressures as *TLR4* and *TLR5*. Overall, our study highlights that the genetic differentiation across Gentoo penguin clades is not just driven by drift but by clear population-specific adaptations to the environment.

Diversity has been widely reported to vary between different families of TLRs, particularly comparing extracellular and intracellular TLRs. Some authors have reported that TLRs that respond to viral ligands are more likely to be under purifying selection, at least in mammals [46, 313, 619, 637], although the pattern may not be consistent in birds, with *TLR3* displaying the greatest number of nonsynonymous variants of the four TLRs tested in different chicken breeds [565], and *TLR7* diversity in the house finch (*Carpodacus mexicanus*) far exceeding that of *TLR4* and *TLR5* [10]. Consistent with the pattern observed in mammals (and also the lesser kestrel, *Falco naumanni*; [10]), we observed overall nucleotide diversity measurements that were several times higher for the extracellular/bacterial TLRs 4 and 5 (*TLR4*: 6.1×10^{-4} $\pm$

0.5×10^{-4}; *TLR5*: $14.8\times10^{-4} \pm 0.6\times10^{-4}$) compared with the intracellular/viral *TLR7* ($1.0\times10^{-4} \pm 0.1\times10^{-4}$), indicating strong purifying selection for maintenance of function in *TLR7*. In addition, we found no evidence of any codons under selection in *TLR7*, compared with three and nine sites in *TLR4* and *TLR5* respectively, similar to the pattern of positively selected residues reported in several avian species [10]).

Within TLR sequences, levels of variation are not uniformly distributed across the domains of the receptor. TLRs are type I integral membrane glycoproteins with highly conserved architecture across large phylogenetic distances [69]. Typical TLR structure comprises an N-terminal extracellular (or intraluminal for intracellular TLRs) LRR domain for ligand binding, a single transmembrane helix, and a C-terminal cytoplasmic signaling (Toll/interleukin-1 receptor, TIR) domain interacting with intracellular adapter proteins [55]. The LRR domain directly binds microbe-derived ligands in all known vertebrate TLRs, with the exception of TLR4 recognition of lipopolysaccharide via an accessory molecule, myeloid differentiation factor, MD-2 [460]. Since the LRR domain represents the interface between host and pathogen, and pathogens exhibit variable MAMPs to evade detection [14], there is often an excess of diversity in the LRR domain compared with the TIR domain [565, 603]. In contrast, TIR domains interact with adapter proteins such as MyD88 (myeloid differentiation primary-response protein 88) which are shared between several TLRs, although a MyD88-independent pathway also facilitates TLR3 and TLR4 signaling [6]. Unsurprisingly, TIR domains were found to be much more highly conserved than their corresponding extracellular domains in a study of 366 vertebrate TLRs from 96 species with the exception of *TLR10* [407]. Within species, the same trend is evident: in a study of *TLR3*, *4*, *5*, and *7* diversity across domestic chicken breeds, only 3 of the 46 nonsynonymous polymorphisms (two in ch*TLR3* and one in ch*TLR7*) were located in the TIR domain [565].

In line with previous evidence for the asymmetric distribution of polymorphisms in TLR domains, we identified an excess of polymorphisms in the LRR domain compared with the TIR domain in two of the three TLRs studied (*TLR4*: 8 LRR vs. 0 TIR;

TLR7: 6 LRR vs. 0 TIR). Interestingly, *P. papua TLR5* contained a greater number of TIR domain polymorphisms than would be expected from other species (11 LRR vs. 7 TIR), particularly given the LRR is over three times the length of the TIR domain. Of the seven *TLR5* TIR domain polymorphisms, five were nonsynonymous substitutions, suggesting that the TIR domain of *TLR5* has been under selection to modulate signalling intensity.

Somewhat surprisingly, we also identified nonsynonymous polymorphic sites in the transmembrane domains of both *TLR4* (659 Ala/Thr) and *TLR5* (667 Ser/Phe). The transmembrane domain is an uncommon location for TLR polymorphisms, presumably because the region is highly constrained by chemical and functional requirements. As such, the effects of polymorphisms in this region are often large. For instance, the human TLR1 602S variant is associated with disrupted cell surface localization of the receptor but is protective against pathology associated with leprosy. It is also noteworthy that the Gentoo penguin TLR5 transmembrane polymorphic site (667) identified in this study is highly conserved elsewhere in avian phylogeny. Of the other birds with published TLR5 sequences, displayed in the alignment (Appendix Figure A.3), 57 (89%) have a serine (ancestral *P. papua* genotype) at this position in the transmembrane domain, and only one other bird—the Northern Fulmar (*F. glacialis*)—has a phenylalanine residue (derived *P. papua* genotype). The high conservation of serine at this position in the protein points to a widespread pressure for maintenance of function across avian phylogeny, and provides more evidence of a positively selected residue with functional consequences.

2.5.2 Functional polymorphisms in *TLR5* support positive selection

We identified a number of positively selected codons in both *TLR4* and *TLR5*, making both of these receptors candidates for further functional investigation. Polymorphisms in TLR LRR domains have the potential to yield preferences for subtly

different microbial ligands, such as LPS or flagellins from different bacterial species [169, 422]. However, very limited data are available regarding which pathogens are present in the environments of each of the Gentoo penguin clades, and elucidation of any differences in ligand preference will therefore require further study. We identified one positively selected site in TLR5 (285) that is adjacent to two important resides for flagellin binding in Interface B [551, 659]. This site would be a good candidate for functional investigation of changes in flagellin ligand preferences, but this would be difficult in the absence of known flagellin variants in candidate *P. papua* pathogens. Instead, we chose to investigate TIR and transmembrane domain polymorphisms for functional consequences because these can yield signaling intensity differences in response to the same agonist [169]. Given that the TIR domain of TLR4 did not show any nonsynonymous polymorphisms, we focused on the TLR5 TIR/transmembrane domain, and in particular the two residues with the strongest signature of positive selection (667 and 845). Site 667 was likely to be of significant functional consequences because of its transmembrane location, nonconservative amino acid change (serine to phenylalanine) and both SIFT and PolyPhen-2 predicting the change to be of high importance.

The Gentoo penguin is reported to have undergone a circumpolar expansion, with ancestral populations in the Indian Ocean seeding northern populations that expanded into the Atlantic, and further expansions south of the Polar Front and to the West Antarctic Peninsula—the southernmost extreme of the range [137, 464]. It is interesting to note that one of the ancestral Indian Ocean clades (Marion and Crozet archipelagos) is completely dominated by birds of the 667S/845I genotype, whereas the most derived clade of Gentoo penguins south of the Polar Front are almost entirely dominated by the 667F/845V genotype. These data may reflect an incipient selective sweep of the 667F/845V genotype in Southern Gentoo penguins.

An alternative explanation for the reduction in diversity at residues 667 and 845 could be genetic bottlenecking during the expansion of *P. papua* south of the Polar Front. However, evidence from neutral markers in this study and a previous study

[357] reveals that neutral variation is maintained in the Southern Gentoo penguin colonies at levels comparable with the northern Atlantic Gentoo penguin clade up to the southernmost extreme of the range. In *TLR5*, we also saw no correlation between census population size and haplotype diversity. These findings support the hypothesis that a selective sweep, rather than a bottlenecking event, is responsible for the near-fixation of the TLR5 667F/845V haplotype in the Southern Gentoo penguin. Perhaps more importantly, the finding that the 667F/845V haplotype has enhanced signalling capability provides a functional basis for selection of this TLR5 haplotype.

2.5.3 Potential drivers of selection

Toll-like receptors, like other genes of the immune system, are subject to competing types of selection. Balancing selection works to maintain diversity at a population level in response to the diversity of pathogens in the environment, as was recently proposed in TLRs of the bank vole (*Myodes glareolus*) [313]. In contrast, purifying selection may predominate (to retain key functionality), which has been described in large-scale studies of human TLRs in different ethnic backgrounds [416]. Finally, positive selection may promote the fixation of novel variants that confer a fitness advantage in the response to pathogens. In the present study, we found evidence of positive selection in *TLR4* and *TLR5*, which likely indicates that the pathogen composition differs substantially between distinct locations in the Gentoo penguin's range.

Spatial heterogeneity in the profile of pathogens that afflict Gentoo penguins would be a key driver for the patterns of selection identified in the TLR variants. Latitudinal species diversity gradients have been described for pathogens (and their hosts) [148, 224, 499], which might suggest fewer pathogens in Antarctic species. However, a diverse range of pathogens is found in these environments (discussed below). Moreover, within the Gentoo penguin's range, there are diverse biotic and abiotic characters that exhibit spatial variation [45, 44, 87, 103, 587] and these factors will

affect the transmission of pathogens. Indeed, the regionalized selection of TLR alleles in different sectors of the Gentoo penguin's range supports the premise that different challenges are more prevalent or pathogenic in different populations.

The dense colonial conditions and ubiquitous guano (faeces) that characterize Gentoo penguin habitats provide ideal conditions for the transmission of a wide range of pathogens transmitted by direct contact or faeces. Furthermore, penguins as a group are known to be highly susceptible to a variety of infectious diseases including avian cholera [269], avian pox [286], avian malaria [177, 217], and aspergillosis [178]. A number of infection associated mass mortality events have been documented in both wild and captive penguin populations [218]. However, little is known about the pathogens that exist in sub-Antarctic and Antarctic regions, their prevalence, or their fitness costs on penguin populations. Limited data are available on the prevalence of diseases in penguin populations [45, 110, 218, 640], and most studies rely upon short notes, observations, and case reports closely tied to obvious signs of disease and mass mortality in well-studied and highly visited penguin colonies. Studies that survey the environmental and host microbiomes to characterize pathogen presence in polar regions remain limited to sites near major polar research stations, have small sample sizes, and/or do not cover large spatial and temporal ranges [144, 145, 172, 389, 665].

The presence of Gram-negative bacteria exhibiting both lipopolysaccharides and flagella, including *Campylobacter*, *Escherichia*, *Salmonella*, and others, has been demonstrated in Gentoo penguin colonies [44, 66, 147, 197, 210]. However, it is not known whether any of these (or other bacterial pathogens) vary across the Gentoo penguin's range, or may have played a role in the selection of Gentoo penguin *TLR4* or *TLR5* variants.

Studies of single-stranded RNA viruses (which would typically be recognized by TLR7) are similarly lacking in Gentoo penguins. Though single-stranded RNA viruses, including the causative agents of Newcastle disease virus and avian influenza, have occasionally been detected in *Pygoscelis* penguins through immunological assays and direct isolation [415, 427, 450, 614, 634], the fitness consequences of viral infection

on penguin populations are unknown. We know of only one case report from Signy Island, where evidence of a puffinosis outbreak (normally caused by a Coronavirus) was described in Gentoo penguin chicks [391]. The viral drivers behind the strong purifying selection we observed in *TLR7* are unknown, but it could be that the ssRNA viruses that affect Gentoo penguins are less diverse across the species range than flagellated Gram-negative bacteria.

Sympatric interactions with a diversity of migratory flying birds and their parasites may be important contributors to pathogen diversity in Gentoo penguin colonies. Birds such as albatrosses, petrels, shearwaters, sheathbills, shags, gulls, terns, and skuas are often observed in close proximity to Gentoo penguin colonies and there are 46 species recognized in Antarctica alone [352]. There is some evidence that ectoparasites and blood parasites are transmitted between co-occurring bird species [43, 356]. It is plausible that cross-species transmission events are important in cross-colony transmission and structuring the profile of pathogens afflicting particular Gentoo penguins.

It remains unclear why the two functionally tested TMD/TIR residues in TLR5 would confer increased responsiveness to flagellin in Gentoo penguins south of the Polar Front. TLR signaling must be tightly controlled and aberrant TLR-induced inflammation can lead to immune pathology, toxic shock syndrome, and death. It is unsurprising, therefore, that TLR polymorphisms have been described that confer reduced sensitivity to their agonist and a state of tolerance. For instance, the replacement of a highly conserved proline residue by a histidine at position 712 of TLR4 confers endotoxin resistance in certain strains of mice [486]. It could be that the exposure to (or diversity of) pathogens is decreased for the Southern Gentoo penguin clade compared with other clades of Gentoo penguins, and thus individuals can tolerate enhanced signaling to a prevailing infection. Alternatively, the enhanced signaling could be a manifestation of the Southern Gentoo penguin adapting to a particular pathogen that is present in the West Antarctic Peninsula and absent elsewhere. The finding of adaptive changes in the Gentoo penguin immune system necessitates

a much better understanding of the pathogen threats faced by Gentoo penguins in order for their significance to be realized.

2.5.4 Concluding remarks

This wide-ranging immunogenetic study of TLRs in wild Gentoo penguins reveals differential selection and adaptation to local pathogen pressure. Although the drivers behind the observed patterns of diversity and selection remain unclear in the context of currently available data, it is clear that the Gentoo penguin has undergone adaptation to local pathogen assemblages across its range.

Infectious disease threats to penguins are likely to become ever more severe in the coming decades given the rapidly changing polar climate [388, 397]. There is also evidence of reverse zoonosis of enteric bacteria being transmitted from humans to sea bird species in Antarctica [93], which could further increase transmission, especially in light of the increasing presence of tourists and scientific research programs in Antarctica. Although the Gentoo penguin is not currently one of the 13 out of 18 penguin species with a conservation state of threatened or near-threatened, certain sub-Antarctic populations have experienced sharp declines [122, 123, 124, 353]. Consequently, the vulnerability of pathogen-naïve populations of penguins should not be underestimated, nor should the importance of the Gentoo penguin as a sentinel species in the Southern Ocean [87].

Our findings have important implications for the conservation of not just Gentoo penguins but also many other vertebrate species, both in the wild and in captivity. Until now, most efforts to genetically delineate conservation units have relied mostly on neutral markers. The ever-increasing availability of genomic data allows targeted analysis of pathogen-recognition and other immune genes to assess whether different populations possess specific functional adaptations to their environments and should therefore be conserved separately. The approach used here, together with pathogen discovery and surveillance systems, could better define conservation units in species

that occupy varied habitats and ecological niches in order to focus resources on potentially susceptible populations.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

Title of Paper	Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird
Publication Status	Published
Publication Details	Hila Levy*, Steven R Fiddaman*, Juliana A Vianna, Daly Noll, Gemma V Clucas, Jasmine K H Sidhu, Michael J Polito, Charles A Bost, Richard A Phillips, Sarah Crofts, Gary D Miller, Pierre Pistorius, Francesco Bonnadonna, Celine Le Bohec, Andres Barbosa, Phil Trathan, Andrea Raya Rey, Laurent A F Frantz, Tom Hart, Adrian L Smith, Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird, Molecular Biology and Evolution, Volume 37, Issue 6, June 2020, Pages 1708–1726, doi: 10.1093/molbev/msaa040

Student Confirmation

Student Name:	Steven R Fiddaman		
Contribution to the Paper	I conducted PCRs and sequencing of *TLR5* from all samples in conjunction with JKHS—an undergraduate student—and assisted in sequencing of *TLR7*. HL carried out all remaining PCRs and sequencing. HL carried out all population genetic analyses. I performed all positive selection analyses and identification of functional sites. I performed all cloning and functional work, including generating a CRISPR-edited *TLR5*-null reporter cell line. I generated all figures and wrote the paper with HL. ALS, LAFF and TH supervised the project and provided editorial assistance. All other authors provided samples and/or sequence data.		
Signature: Steven R Fiddaman		Date	13/07/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Adrian L Smith		
Supervisor comments		
Signature 	Date	13/07/2021

Chapter 3

Adaptation and cryptic pseudogenization in penguin Toll-like receptors

3.1 Abstract

Penguins (Sphenisciformes) are an iconic order of flightless, diving seabirds, distributed across a large latitudinal range in the Southern Hemisphere. The extensive area over which penguins are endemic is likely to have fostered variation in pathogen pressure, which in turn will have imposed differential selective pressures on the penguin immune system. At the front line of pathogen detection and response, the Toll-like receptors (TLRs) provide insight into host evolution in the face of microbial challenge. TLRs respond to conserved pathogen-associated molecular patterns and are frequently found to be under positive selection, despite retaining specificity for defined agonist classes. I undertook a comparative immunogenetics analysis of TLRs for all penguin species, and found evidence of adaptive evolution that was largely restricted to the cell surface expressed TLRs, with evidence of positive selection at, or near, key ligand-binding sites in *TLR1B*, *TLR4* and *TLR5*. Intriguingly, *TLR15*, which is activated by fungal products, appeared to have been pseudogenized multiple

times in the *Eudyptes* spp., but a full-length form was present as a rare haplotype at the population level. However, *in vitro* analysis revealed that even the full-length form of *Eudyptes TLR15* was non-functional, indicating an ancestral cryptic pseudogenization prior to its eventual disruption multiple times in the *Eudyptes* lineage. This unusual pseudogenization event could provide an insight into immune adaptation to fungal pathogens such as *Aspergillus*, which is responsible for significant mortality in wild and captive bird populations.

3.2 Introduction

Penguins (Sphenisciformes) are a unique order of flightless, wing-propelled diving seabirds widely distributed across diverse coastal and island environments across the Southern Hemisphere. Between 18-20 extant species of penguin have been documented, represented within six well-defined genera (*Aptenodytes*, *Pygoscelis*, *Eudyptula*, *Spheniscus*, *Eudyptes* and *Megadyptes*), with some debate surrounding the taxonomy (species/sub-species status) of certain recently diverged lineages [112, 117, 118, 189, 190, 457, 604].

The recent release of high-coverage genomes from all extant penguin species [457] as part of the wider 'Bird 10k' genomes project [667] provides a unique opportunity to address comparative evolutionary questions across an entire avian order. One such comparison is to assess how the penguin immune system has adapted to the many, varied pathogen environments in which penguins are endemic. There is evidence that pathogen species richness changes as a product of biotic and abiotic factors, which are known to vary significantly across the ranges of penguins [125, 285, 479, 591]. It is likely, therefore, that a key part of penguin species radiation has been adaptation to the novel pathogens they encounter.

One body of evidence for pathogen-driven immunogenetic selection in wild animals comes from the major histocompatability complex (MHC; [148, 443, 516]), which is important for presenting antigens to the adaptive immune system. How-

ever, investigating functional genetic diversity in the MHC is challenging in non-model organisms because of the complexity associated with extensive gene duplication and gene loss events [126]. Instead, other immune genes are increasingly being recognized as targets for pathogen-mediated selection, particularly those encoding pattern-recognition receptors (PRRs) of the innate immune system. One such family is the Toll-like receptors (TLRs), which are cell surface and endosomal receptors responsible for recognizing pathogen-associated molecular patterns (PAMPs) and initiating intracellular signalling cascades [568]. TLRs are functionally conserved across vertebrates [493, 71, 81], but are frequently found to be under positive selection, making them attractive loci to study adaptation to different pathogen pressures [25, 220, 325, 358, 565, 603].

Birds typically have 10 Toll-like receptors, of which *TLR3*, *TLR4*, *TLR5* and *TLR7* are clear one-to-one orthologues of mammalian TLRs [71, 81]. In birds, *TLR1* and *TLR2* are both present as two-copy tandem duplications. Recent evidence suggests that the *TLR1* duplication pre-dates the reptile-mammalian divergence, since *TLR1A* clusters with mammalian *TLR10* in phylogenetic analysis [603]. The *TLR2* duplication is less well understood; some mammals also have a pseudogenized *TLR2* as a second copy [493], but it remains unclear whether this duplication occurred prior to the reptile-mammalian divergence or twice independently [264, 579, 603]. *TLR15* is generally considered to be an avian- and reptilian-specific TLR [72]. However *TLR15* has recently been identified in the Australian ghost shark (*Callorhinchus milii*; [612]), suggesting that the origin of the receptor predates the divergence of the Chondrichthyes fish and tetrapods. Finally, *TLR21* is found in fish, amphibians, reptiles and birds [301, 657], but is absent in mammals.

Having found evidence of functional adaptation in Gentoo penguin (*Pygoscelis papua* ssp.) *TLR4* and *TLR5* [358], we undertook a comprehensive comparative analysis of penguin TLRs to consider processes that may have shaped the evolution of TLR-mediated immunity across this entire order of vertebrates. Using sequences derived from 21 high-coverage genomes encompassing all penguin species and sub-

species [359, 457], we identified patterns of adaptive evolution showing increased positive selection in cell surface TLRs compared to endosomal TLRs. Surprisingly, multiple pseudogenized *TLR15* haplotypes were identified in *Eudyptes* penguin reference genomes. To investigate the pseudogenization of *TLR15* further, we analyzed *TLR15* sequences from >100 individuals representing fives species/subspecies of the *Eudyptes* genus. There was evidence of at least eight independent gene disruptions, but interestingly, the full-length gene was present at the population level as a rare haplotype. Functionally testing the full-length haplotype with known and novel agonists, we found that the gene is non-functional, despite not showing any signs of pseudogenization. This sugests that the 'cryptic', full-length, pseudogene represents the ancestral state onto which additional mutations accumulated as a result of relaxed selection. These additional truncating mutations have become near-fixed in the population through drift. Since TLR15 responds to fungal pathogens, in particular *Aspergillus fumigatus*—an important pathogen of penguins and other birds—these findings may have important implications for the immune response to mycoses in penguins. In addition, our findings have broader significance to the dynamics of molecular trait erosion in the genomes of vertebrates.

3.3 Materials and methods

3.3.1 Sequence retrieval

Twenty-one assembled penguin genome sequences were downloaded from the GigaScience Database (doi: http://gigadb.org/dataset/100649) from two studies exploring penguin evolution (Li, et al. 2014 [359]; *Aptenodytes forsteri* and *Pygoscelis adeliae*); (Pan, et al. 2019 [457]; all other penguin species). TLR sequences from the annotated Emperor penguin (*Aptenodytes forsteri*) genome were used as query sequences (*TLR1A*: XM_009280175.1; *TLR1B*: XM_009280152.2; *TLR2A*: XM_009283352.2; *TLR2B*: XM_009283317.1; *TLR3*: XM_009277378.2;

TLR4: XM_009282256.2; *TLR5*: XM_009275754.1; *TLR7*: XM_009278529.2; *TLR15*: XM_009288440.1; *TLR21*: XM_009282498.1). Datasets of predicted coding genes were queried using default sensitivity parameters in a local BLASTn search. Significant hits of E $\approx$ 0 were retrieved. On occasions where significant hits were not retrieved from the predicted coding gene database, assembled genomes were queried using the genomic records from which the above *A. forsteri* mRNA sequences were derived. Finally, the collection of TLR sequences was supplemented by the available representatives from the Adélie penguin (*Pygoscelis adeliae*; *TLR1A*: XM_009321398.1; *TLR1B*: XM_009332470.1; *TLR2A*: XM_009329794.1; *TLR2B*: XM_009329795.1; *TLR3*: XM_009330657.1; *TLR4*: XM_009319316.1; *TLR5*: XM_009333665.1; *TLR7*: XM_009318873.1; *TLR15*: XM_009319611.1) and with subclades of the Gentoo penguin (*Pygoscelis papua* ssp.) from the Indian ocean and West Antarctic Peninsula [358]; accessions: MN394307, MN394343, MN313097, MN313136, MN312948, MN312985). Population-level data for *TLR15* in the Eudyptes penguins (*E. chrysolophus chrysolophus*, N = 40; *E. chrysolophus schlegeli,* N = 7; *E. chrysocome*, N = 22; *E. filholi*, N = 27; *E. moseleyi*, N = 13) were obtained from whole genome data generated in [189, 604] and unpublished data.

3.3.2 Alignments and polymorphism identification

Following the removal of introns (where appropriate) by reference to the annotated *Aptenodytes forsteri* sequence, sequences were aligned using MUSCLE (v. 3.8.425; [157]) in Geneious Prime (2019.0.3). Sequences were truncated to the length of the coding sequence, with the start codon confirmed by the presence of a signal peptide in the amino acid sequence immediately downstream as determined by the Phobius web server [281, 282]. Polymorphisms in TLR sequences were identified using DnaSP v. 6.12 [506].

3.3.3 Positive selection analysis

To detect signatures of positive selection, multiple alignments were analyzed with the codeml program in PAML v. 4.9 [652, 654] with the F3X4 codon frequency model, using the phylogenetic tree from [457]. Various models were fitted to the multiple alignments: M1a (neutral model; two site classes: $0 < \omega_0 < 1$ and $\omega_1 = 1$); M2a (positive selection; three site classes: $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 > 1$); M7 (neutral model; values of ω fit to a beta distribution where $\omega > 1$ is disallowed); M8 (positive selection; similar to M7 but with an additional codon class of $\omega > 1$). Likelihood ratio tests were performed on pairs of models to assess whether models allowing positively selected codons gave a significantly better fit to the data than neutral null hypothesis of neutral codon evolution could be rejected ($P < 0.05$). The posterior probabilities of codons under selection in M2a and M8 were inferred using the Bayes Empirical Bayes algorithm [655]. To further analyze sequences for positive selection, the Mixed Effects Model of Evolution (MEME; [420]) and Fast Unconstrained Bayesian Approximation (FUBAR; [419]) tools were implemented on the Datamonkey server (http://www.datamonkey.org; accessed April 2020; [140, 475, 625]) using the same alignments (excluding pseudogenes) that were used for the PAML analysis. Default significance thresholds were used for both tools (P-value of 0.1 for MEME and posterior probability of 0.9 for FUBAR).

3.3.4 Homology modelling

We used I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/; [505] to generate 3D models of the *Aptenodytes forsteri* TLR exodomains. As homology templates for AfoTLR1B and AfoTLR2B, we selected human TLR1/TLR2 (PDBID: 2z7x), for AfoTLR4 we used mouse TLR4 (PDBID: 5ijb), for AfoTLR5 zebrafish TLR5b (PDBID: 3v44), and for AfoTLR7 mouse TLR7 (PDBID: 5gmh). For further analysis, only I-TASSER models with the highest C values were used (in all cases this confidence score was $C > 0.5$, estimated RMSD max. 5.7 ± 3.6Å). To visualize

the location of positively selected sites and functional residues and to measure their molecular distances we used the PyMOL Molecular Graphics System (version 2.3.3, Schrödinger, LLC).

3.3.5 Phasing of TLR15 haplotypes

TLR15 haplotypes were inferred for individual birds using Beagle [80, 79]. Data were prepared in accordance with GATK Best Practice recommendations [142, 601]. First, HaplotypeCaller in GATK was used to reassemble reads in putative indel sites and assign per-sample genotypes, as well as infer physical phasing of variants on contiguous reads [477]. Per-sample genotypes were aggregated using CombineGVCFs in GATK and then joint-genotyping was performed using GenotypeGVCFs. Variants were filtered using VariantFiltration with parameters of read depth $\geqslant 6$ and base quality $\geqslant 30$. Following phasing using Beagle, bcftools [362] was used to extract the two haplotypes per individual for further analysis.

3.3.6 Fungal lysate preparation

Saccharomyces cerevisiae lysates were prepared as previously described [72]. For *Aspergillus fumigatus*, lysates were prepared by streaking a fresh Sabouraud dextrose agar plate (with chloramphenicol) with a swab from the lung of a Humboldt penguin with aspergillosis at necropsy, growth at 37 °C for 48 hours, then homogenization in a bead beater (Biospec) in ice-cold sterile PBS, followed by centrifugation at 13 000 × G at 4 °C for 15 minutes to clear the lysate.

3.3.7 Cloning of TLR15 constructs

Gblocks comprising the condon-optimised full-length sequences of North fulmar (*Fulmar glacialis*), Emperor penguin (*Aptenodytes forsteri*), consensus *Eudyptes* intact *TLR15*, sequences were designed (IDT), and cloned into the p3×FLAG-CMV-14 vector (Sigma) using a Gibson assembly approach [204]. Chicken TLR15 in pTarget

(Promega) was used previously by [72]. Mouse CD4-TIR15 and Eudyptes-chicken TIR15 chimaeras were designed as previously described [72, 402] and cloned into pTarget (CD4 chimaeras) or p3×FLAG-CMV-14 (chicken chimaera).

3.3.8 Transient transfection and stimulation

All constructs were transiently transfected into HEK-Blue Null1 cells (which express the SEAP reporter gene under the control of the IFN-β minimal promoter fused to five NF-κB and AP-1 binding sites; Invivogen) using TransIT-2020 (Mirus) in 24-well plates with three technical replicates per condition. Cells were maintained in high glucose Dulbecco's Modified Eagle's Medium (Gibco). Forty-eight hours after transfection, cell culture medium was replaced with medium containing the appropriate agonist (*Aspergillus fumigatus* or *Saccharomyces cerevisiae* lysate at 1 μg/ml, or medium control). After a further 24 h, cell supernatants were aspirated to measure the enzymatic activity of the secreted alkaline phosphatase (SEAP), which is a proxy for NF-κB activation. Supernatants were mixed with p-Nitrophenyl Phosphate (Sigma) according to the manufacturer's instructions, and absorbance was measured at 405 nm on a FLUOstar Omega microplate reader (BMG Labtech) at 37 °C. Differences between means were tested statistically using an analysis of variance model with Dunnett's test [153] for multiple comparisons to a control in R (v. 4.0.2).

3.3.9 Immunoblotting

Immunoblotting was performed on the same samples as the SEAP assay to confirm protein expression. Briefly, cells were lysed directly in NuPAGE LDS Sample Buffer (4×; Novex), supplemented with 0.1 M DTT, boiled for 10 minutes at 95 °C, resolved on a 4-20% gradient gel (Bio-Rad), transferred using a semi-dry system (Bio-Rad) at constant current of 200 mA for 1 h, blocked using StartingBlock PBS blocking buffer (Thermo Scientific), probed using mouse monoclonal anti-FLAG (M2; 1:1000; Sigma) overnight at 4 °C, then after washing, probed with a goat anti-mouse HRP-conjugated

secondary antibody (Cusabio; 1:10 000), and visualized on an ImageQuant LAS 4000 (Cytiva) using SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Scientific).

3.4 Results

3.4.1 Toll-like receptor and pseudogene identification

BLAST [13] was used to identify TLR sequences in assembled penguin genomes generated using Illumina short-read sequences [359, 457]. The majority of penguin species possessed representatives of each of the 10 avian Toll-like receptors (*TLR1A*, *TLR1B*, *TLR2A*, *TLR2B*, *TLR3*, *TLR4*, *TLR5*, *TLR7, TLR15*, and *TLR21*). Representatives of the TLR1/2 family were missing in several species, likely due to assembly issues resulting from almost identical regions of gene conversion that are difficult to resolve with short-read sequencing. *TLR3*, *TLR4*, *TLR5* and *TLR7*, and *TLR15* were identified in all penguin species. *TLR21* was absent or found only as a fragment in several penguin species, similar to other studies of avian TLRs [10, 603], suggesting there may be technical difficulties assembling the genomic region comprising *TLR21* rather than true absence in these species [542].

We identified a pseudogene for *TLR5* in the Snares Crested penguin (*Eudyptes robustus*), caused by a cytosine to thymine substitution at position 1180 which led to a premature stop codon. In contrast to a previous report on the African penguin (*Spheniscus demersus*; [129]), we found no evidence of pseudogenization of *TLR7,* and the gene was present and intact in all other penguin species. As described in detail below, we found a pseudogenized copy of *TLR15* in all of the eight *Eudyptes* spp., and an independent heterozygous pseudogenization in the Chinstrap penguin (*Pygoscelis antarcticus*) *TLR15*. Pseudogenes were removed from subsequent diversity and selection analyzes due to the propensity to accumulate mutations through drift following pseudogenization [365].

3.4.2 Patterns of diversity in penguin Toll-like receptors

Alignments of penguin Toll-like receptors were analyzed for polymorphic sites. The TLR1/2 family in birds is known to have undergone multiple gene-conversion events [264, 603], which results in regions with almost exact sequence identity between the *TLR1A* and *TLR1B*, and *TLR2A* and *TLR2B* paralogs. Since all currently available high-coverage penguin genomes used short-read sequencing technology [359, 457] it remains difficult to confidently assign a read in a gene-converted region to the correct TLR paralog, which potentially inflates diversity statistics. We therefore excluded known gene-converted regions in the TLR1/2 family from diversity analysis. A total of 970 SNPs were identified in penguin TLRs, of which 550 (56.7%) were non-synonymous changes (Figure 3.1). In non-gene-converted TLRs, total polymorphism number was broadly similar, ranging from 114 (*TLR7*) to 163 (*TLR15*; Figure 3.1). *TLR15* yielded the highest number of polymorphisms despite having comparatively few sequence representatives due to excluded pseudogenes in the *Eudyptes* spp. and the Chinstrap penguin (*Pygoscelis antarcticus*).

3.4.3 Positive selection in penguin Toll-like receptors

To investigate whether penguin Toll-like receptors have experienced diversifying selection, which could be an indicator of adaptation to different pathogen pressures, we fitted different site models (M1a/M2a and M7/M8) to multi-species sequence alignments using the codeml package in PAML v. 4.9 [652, 654]. All pseudogenes were removed prior to analysis, and only non-gene-converted regions of TLR1/2 family members were analyzed. In all cases, the M1a/M2a and M7/M8 results were comparable, so only the more conservative M1a/M2a p-values are reported in the following section (see Appendix Table B.1 for all results). The cell surface TLRs *TLR1B*, *TLR4*, *TLR5* and *TLR15* were found to evolve under positive selection in penguins (likelihood-ratio test; $P = 9.2 \times 10^{-4}$, $P = 9.1 \times 10^{-4}$, $P = 4.8 \times 10^{-6}$, $P = 0.023$ respectively; Appendix Table B.1). For the intracellular TLRs (*TLR3*, *TLR7* and

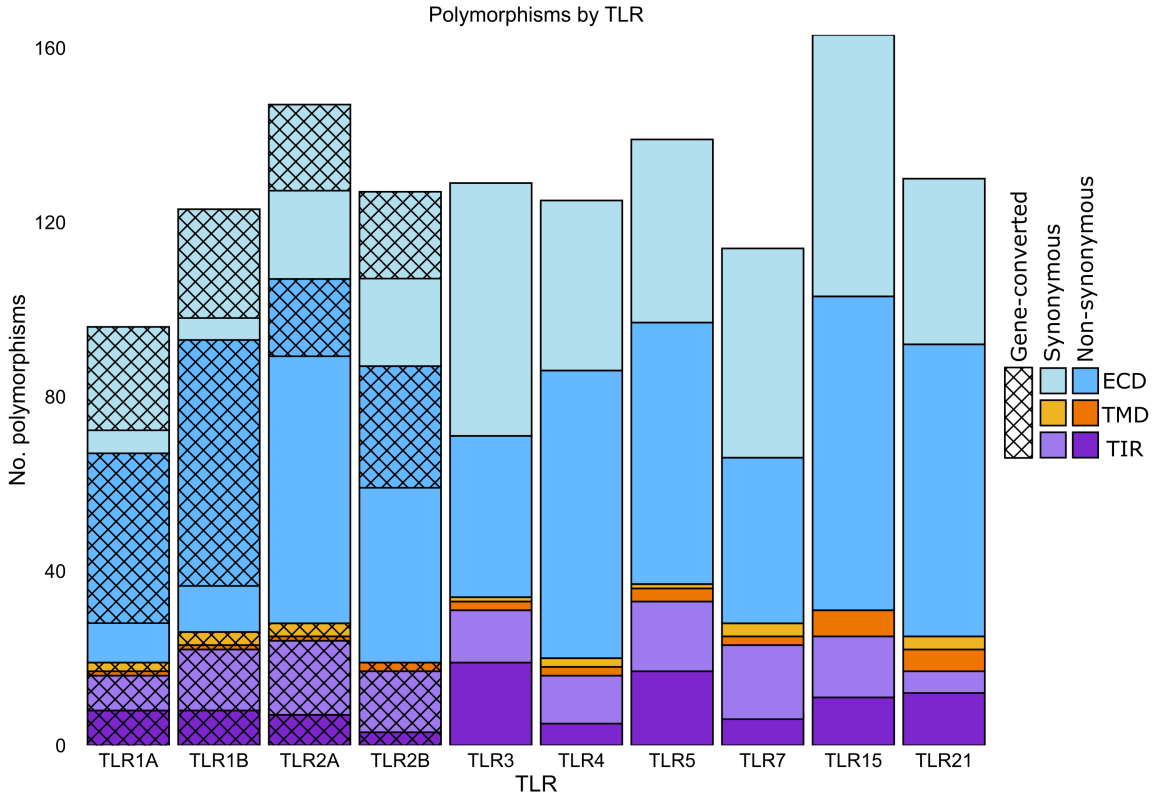


Figure 3.1: Number of polymorphisms per TLR (excluding pseudogenes), stratified according to domain and polymorphism type. Abbreviations: ECD, extracellular domain; TMD, transmembrane domain; TIR, Toll/interleukin-1 receptor domain. Apparent polymorphisms in gene-converted regions of the TLR1/2 family are shown in hatched boxes as sequence differences between orthologs cannot be distinguished from differences between the gene-converted paralogs in genomes generated using exclusively short-read technology.

TLR21), positive selection was only supported for *TLR7* ($P = 0.13$, $P = 0.028$, $P = 0.65$, respectively; Appendix Table B.1).

Using the Bayes Empirical Bayes procedure for inferring codon sites under selection implemented in model 2a of PAML [655], 33 positively selected sites were identified (posterior probability > 0.90) across the 10 TLRs examined (Figure 3.2, Figure 3.3, Figure 3.4, Appendix Table B.2). Of these, 28 were located in the extracellular (ligand-binding) domain, two were located in the transmembrane domain, and three were in the Toll/interleukin-1 receptor (TIR) homology domain (Figure 3.2).

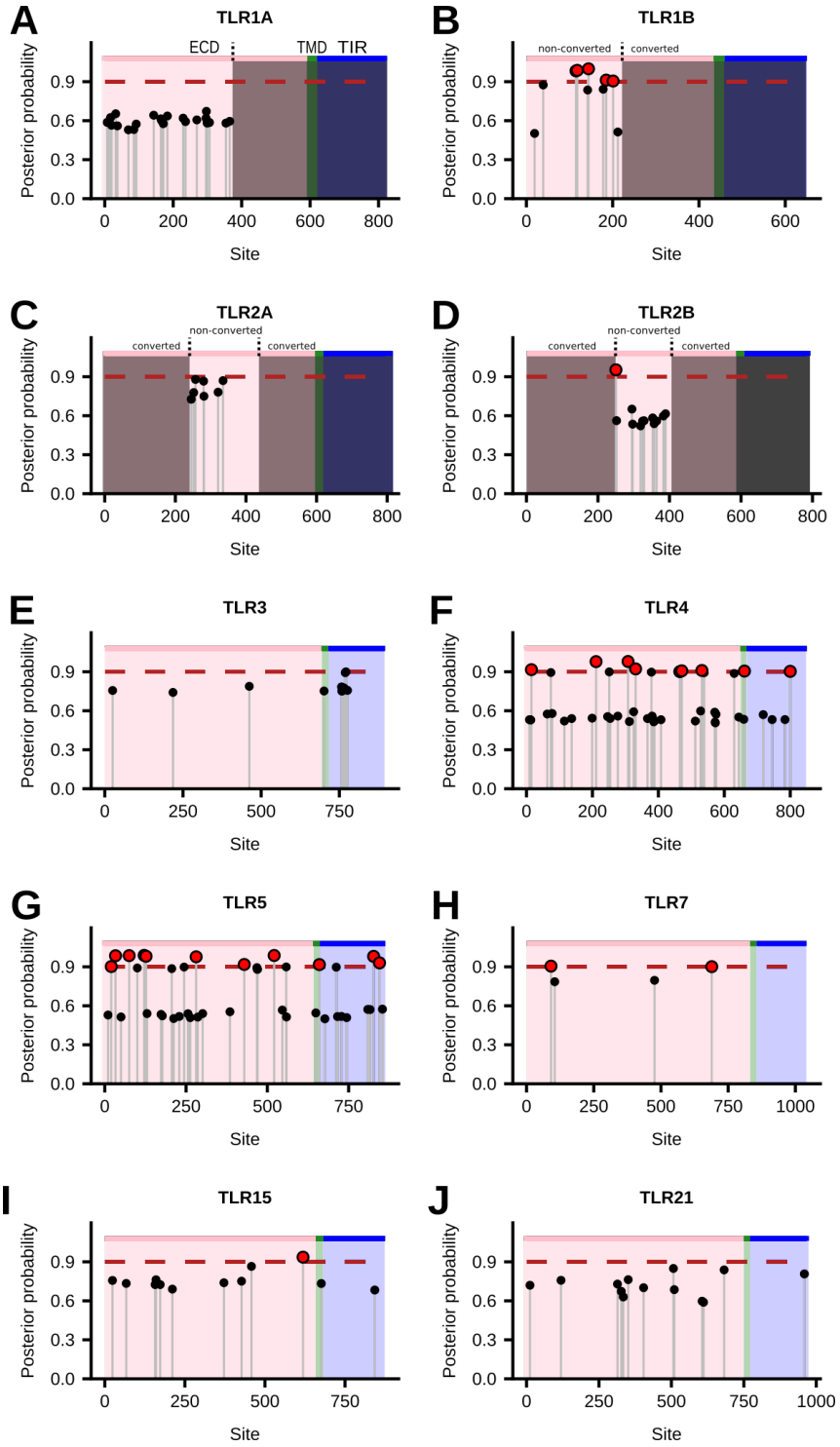


Figure 3.2: Locations of positively selected sites in penguin Toll-like receptors. Posterior probability of positively-selected sites was determined using the Bayes Empirical Bayes procedure described in [655]. Red points indicate sites with posterior probability of selection of $\geqslant 0.90$, while pink, green and blue boxes denote the extracellular domain, transmembrane domain and TIR (Toll/Interleukin-1 receptor) domains of the TLR, respectively. Greyed-out regions in the TLR1/2 family denote excluded gene-converted regions.

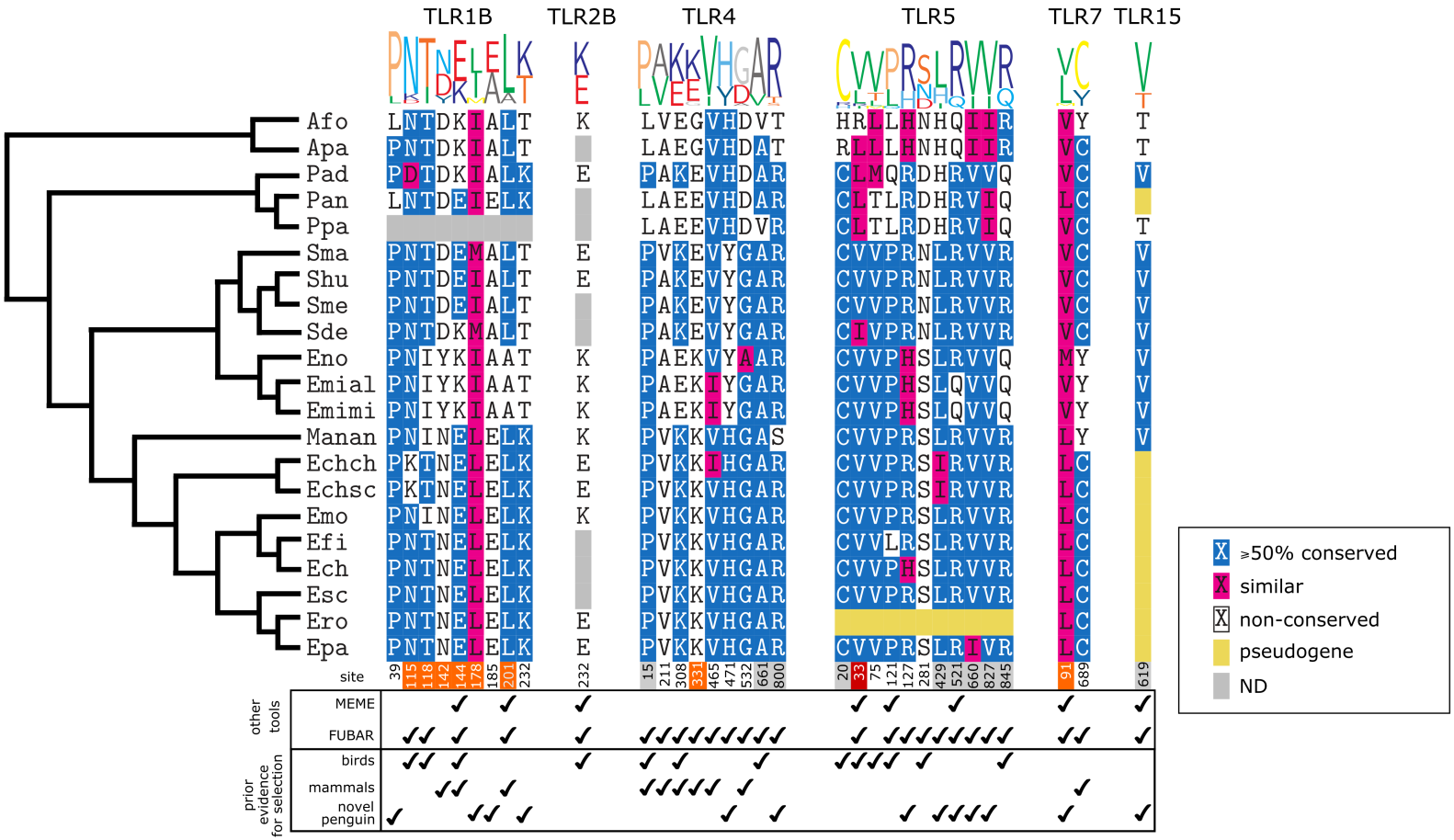


Figure 3.3: (Continued on the following page.)

Figure 3.3: Alignments of positively selected sites. Amino acid sequence alignments of positively selected sites in all TLRs (excluding gene-converted regions of the TLR1/2 family). Sequence logo is coloured according to the RasMol colouring convention, where amino acids of similar properties are grouped. Amino acids in the alignments are coloured according to physicochemical conservation. Pseudogenes were excluded from the analysis, and are indicated by yellow boxes. Grey boxes in alignments indicate missing data. Site position (relative to the Afo coding sequence) is given below the alignment, and is coloured according to whether the selected site is proximal to a site of known function (orange: the site is within 5 Å of a site of known function; red: the site has a known function; grey: distance from sites of known function could not be measured). For each site, 'other tools' indicates whether a site was also found to be under positive selection by using the alternative methods of MEME and FUBAR, implemented in Datamonkey. Additionally, an indication is given under the alignment ('prior evidence for selection') as to whether the site has previously been described as being positively selected in birds and/or mammals, or whether the site is a novel selected site in penguins (see Appendix B.2 for details). Sequence alignments and images were generated in the msa (v. 1.18.0) package in R [62] which implements the MUSCLE algorithm [157]. Abbreviations: Afo, *Aptenodytes forsteri* (Emperor penguin); Apa, *Aptenodytes patagonicus* (King penguin); Pad, *Pygoscelis adeliae* (Adélie penguin); Pan, *Pygoscelis antarcticus* (Chinstrap penguin); Ppa, *Pygoscelis papua* (Gentoo penguin); Sma, *Spheniscus magellanicus* (Magellanic penguin); Shu, *Spheniscus humboldti* (Humboldt penguin); Sme, *Spheniscus mendiculus* (Galapagos penguin); Sde, *Spheniscus demersus* (African penguin); Eno, *Eudyptula novaehollandiae* (Fairy penguin); Emial, *Eudyptula minor albosignata* (White-flippered penguin); Emimi, *Eudyptula minor minor* (Little penguin); Manan, *Megadyptes antipodes antipodes* (Yellow-eyed penguin); Echch, *Eudyptes chrysolophus chrysolophus* (Macaroni penguin); Echsh, *Eudyptes chrysolophus schlegeli* (Royal penguin); Emo, *Eudyptes moseleyi* (Northern Rockhopper penguin); Efi, *Eudyptes filholi* (Eastern Rockhopper penguin); Ech, *Eudyptes chrysocome* (Southern Rockhopper penguin); Esc, *Eudyptes sclateri* (Erect-crested penguin); Ero, *Eudyptes robustus* (Snares Crested penguin); Epa *Eudyptes pachyrhynchus* (Fiordland penguin).

To further validate the sites identified as being positively selected in PAML, we employed the Mixed Effects Model of Evolution (MEME; [420]) and Fast Unconstrained Bayesian Approximation (FUBAR; [419]) tests of positive selection. Sites identified by FUBAR largely agreed with those identified by PAML, and 26/33 sites were also significant for positive selection (posterior probability > 0.90; Figure 3.3, Appendix Table B.3). MEME analysis was not as consistent, and only 8/33 sites identified by PAML were corroborated in MEME (P-value $\leqslant 0.10$; Figure 3.3, Appendix Table B.3), perhaps indicating that MEME is less powerful or more conservative than FUBAR.

Nine positively selected sites were identified in non-gene-converted regions of *TLR1B* (codons 39, 115, 118, 142, 144, 178, 185, 201 and 232; Figure 3.2, Figure 3.3, Figure 3.4) and a further codon was identified in *TLR2B* (232; Figure 3.2, Figure 3.3, Figure 3.4). Three of the sites in *TLR1B* have previously been identified as being positively selected in birds (115, 118 and 144; Figure 3.3; [603]), suggesting further diversification in the penguin lineage, and three of the sites have been identified in mammals (142, 144 and 201; Figure 3.3; [637, 264]). Interestingly, some of the sites identified in *TLR1B* are in close proximity to residues critical for TLR function. Residues 115, 118, 142, 144 and 178 of TLR1B are proximal (<5Å) to sites involved in lipopeptide (Pam3CSK4) binding [273, 603], suggesting there may have been changes in lipopeptide agonist recognition or specificity in penguins (Figure 3.4). Sites 142 and 201 are proximal (<5Å) to known receptor dimerization sites (site 201 is adjacent to a known dimerization site) which could have functional implications for the formation of TLR heterodimers [273]. Codons 43, 182, 189 and 237 (*TLR1B*) have not been previously identified as positively selected in birds or other taxa, so are diversified sites that are novel to penguins (Figure 3.3). Codon 232 (*TLR2B*) has previously been found to be positively selected in birds, but is not proximal to a site of known function (Figure 3.3).

Almost all of the positively selected codons identified in penguin *TLR4* have been reported as positively selected in other taxa including mammals (codons 15, 211,

308, 331, 465 and 532; Figure 3.3; [25, 423, 535, 608, 637]) and elsewhere in birds (codons 4, 308 and 661; Figure 3.3; [603]). Codons 471 and 800 have not been described as positively selected in other taxa, and therefore are uniquely diversified in the penguin lineage (Figure 3.3). In addition, we identified sites that could be functionally important due to their proximity to sites of known function. Site 308 is homologous to human site 298, adjacent to a site that has been implicated in LPS hyporesponsiveness [24, 446], and may be protective against malaria [174]. Site 331 is <5Å from the LPS-binding site 314 [196, 458, 460, 530] so could be functionally important in the recognition of bacterial LPS. Taken together, these observations suggest that *TLR4* has undergone ectodomain diversification in different lineages, which may be indicative of adaptation to changes in the profile or structure of agonists.

Several codons identified as positively selected in penguin *TLR5* have been identified elsewhere in birds (Figure 3.3; codons 20, 33, 75, 121, 281 and 845; [220, 603, 606]). Furthermore, site 33 has previously been identified as a flagellin binding site [659] which could imply a change of ligand-binding function. Several selected sites in TLR5 could not be modelled because no suitable crystal structure template covering the sites was available (sites 20, 429, 521, 660, 827 and 845) and distance from sites of known function could therefore not be determined. However, it is interesting to note that site 845 is one of two co-segregating sites that has recently been implicated in TLR5 signaling intensity differences in the Gentoo penguin species complex (*Pygoscelis papua* ssp.; [358]), suggesting this functional change is not restricted to the Gentoo penguin.

Two positively selected codons were identified in *TLR7*, one of which (689) is also selected in mammalian *TLR7* [25, 637]. Importantly, the other selected site (91) is adjacent to a known dimerization site of *TLR7* [672], which could have implications for TLR7 homodimer formation. We found one positively selected codon (619) in *TLR15* which has not been identified elsewhere in birds (Figure 3.2, Figure 3.3). Currently, insufficient data exist to determine whether this site is likely to be of functional significance, and the site could not be plotted on a homology model because

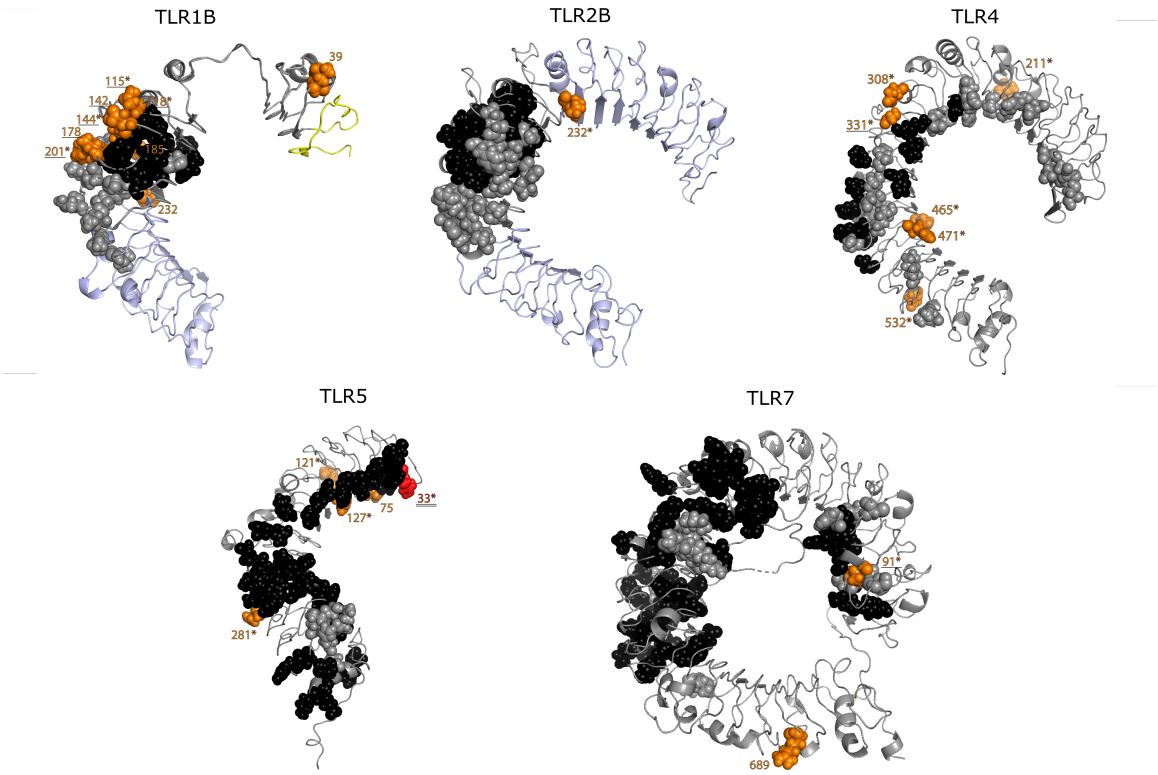


Figure 3.4: Locations of positively selected sites on penguin Toll-like receptor ectodomains, relative to sites of known function. Sites that were determined to be under positive selection in PAML analysis were plotted on homology models of penguin TLRs (orange residues). Amino acids that are known to be ligand binding sites are shaded black, while those involved in receptor dimerization are shaded grey. An underlined site indicates that it is proximal (<5Å) to a site of known function, and a site that is shaded red highlights a site of known function. An asterisk after the selected site position indicates that the site was determined to be positively selected by PAML and at least one other method (MEME or FUBAR). TLR backbones shaded blue indicate gene-converted regions in which positive selection could not be tested, and a yellow TLR backbone indicates a signal peptide. Certain sites are not displayed because they were omitted from the homology model template crystal structure (TLR4: sites 15, 661 and 800; TLR5: sites 20, 429, 521, 660, 827 and 845; TLR15: site 619). Distances to sites of known function could not be measured for these sites. Higher resolution versions of the homology models can be found in Appendix Figures B.1-B.5.

a suitable crystal structure template is lacking.

3.4.4 Pseudogenization of *TLR15*

Analysis of *TLR15* sequences indicated that this gene has been disrupted (at least as a heterozygote) in seven of the eight reference genomes for the *Eudyptes* genus. The disrupting mutations (premature stop codon or indel introducing a frame shift) were often distinct in different species, suggesting several independent pseudogenization events. To explore this in more detail, the dataset was supplemented with resequencing data from a larger number of individuals from five *Eudyptes* species/subspecies obtained from various locations around their natural ranges (Figure 3.5; *E. chrysolophus chrysolophus*, n = 40; *E. chrysolophus schlegeli*, n = 6; *E. moseleyi*, n = 12; *E. filholi*, n = 28; and *E. chrysocome*, n = 21; total n = 107; Appendix Table B.4). *TLR15* sequences were extracted from these genomes, phased, and examined for the presence of pseudogene-causing mutations.

Most individuals had more than one mutation that, in isolation, could represent a pseudogene-causing event (i.e., more than one indel/premature stop codon). In order to identify independent pseudogenizations, we focused on individuals with only one disrupting mutation to exclude mutations arising subsequently by drift. Eight such events were identified in the *Eudyptes* spp., comprising two nonsense SNPs (C143G and C185A), four single nucleotide insertions (681+A, 1006+A, 1826+T and 1996+T), and two larger insertions (2273+31bp, and a large insertion at position 1391 of unknown length, denoted "1391+?"; Figure 3.6A, B, C). While we were unable to determine whether the large insertion would have resulted in a frameshift, the first codon of this insertion is a stop codon, and thus yields a truncated sequence. In the case of all other indels, the reading frame of the gene was disrupted and the ORF was terminated by a stop codon a short distance from the indel site. A further independent pseudogenization in *TLR15* was noted in the Chinstrap penguin (*Pygoscelis antarcticus*)—a heterozygous nonsense mutation (A2530T; Figure 3.6A).

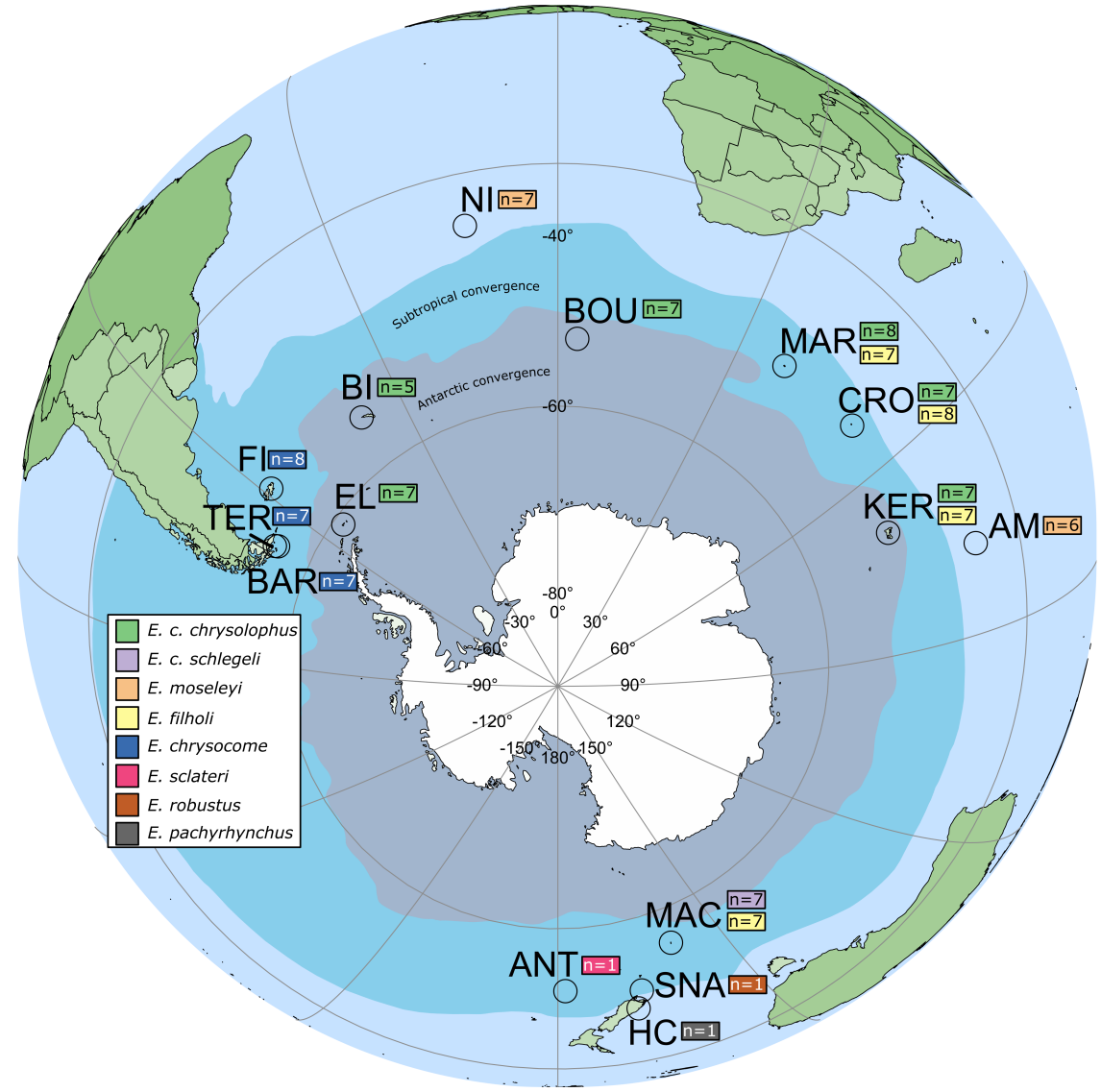


Figure 3.5: *TLR15* sequences used in pseudogene analysis by location and species. Number of individuals included in the analysis is given in boxes beside location labels, with colour indicating the species. Location abbreviations are: ANT, Antipodes Islands; AM, Amsterdam Island; BAR, Barnevelt Islands; BI, Bird Island, South Georgia; BOU, Bouvet Island; CRO, Crozet Islands; EL, Elephant Island, South Shetland Islands; FI, Falkland/Malvinas Islands; HC, Harrison Cove, New Zealand; KER, Kerguelen Islands; MAC, Macquarie Island; MAR, Marion Island; NI, Nightingale Island; SNA, The Snares, New Zealand; TER, Terhalten Island, Tierra del Fuego.

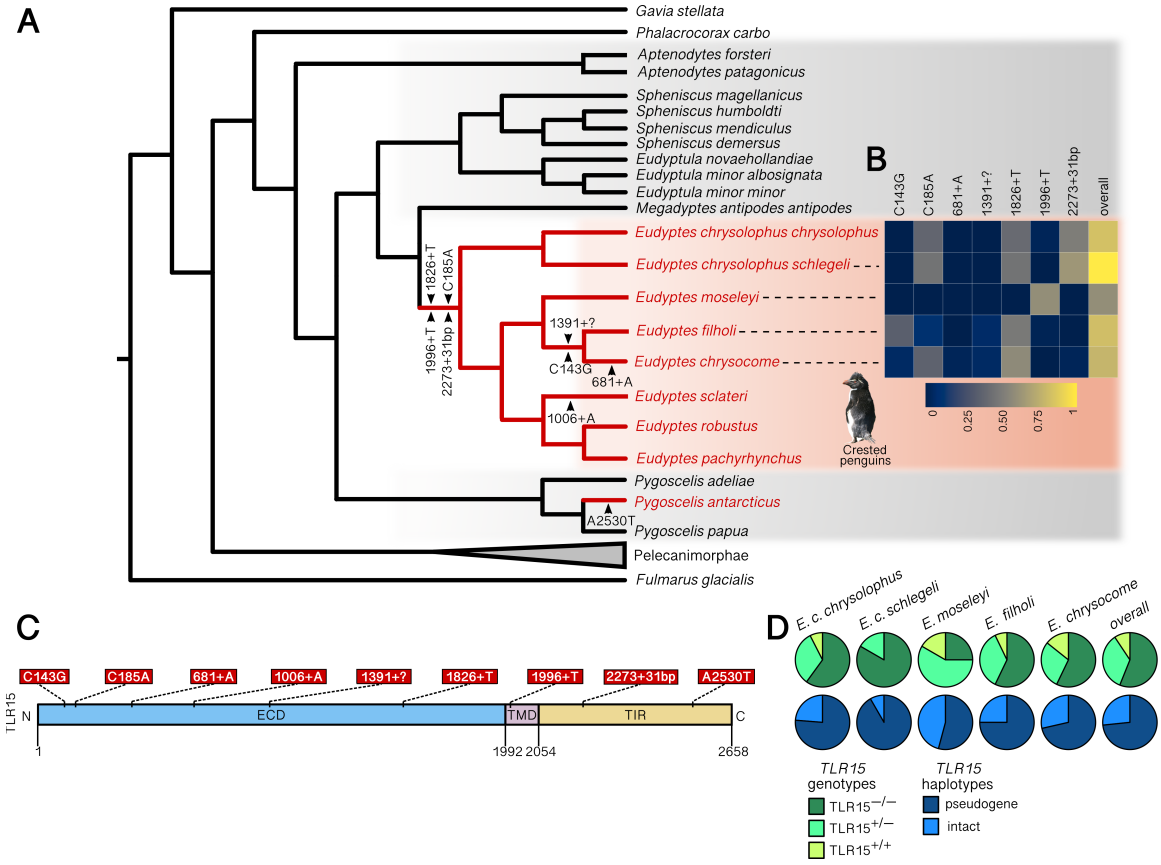


Figure 3.6: *TLR15* appears to have been pseudogenized repeatedly in *Eudyptes* penguins. (A) Cladogram of penguins and related Aequorlitornithes indicating species where at least one individual was found to be at least heterozygous for a pseudogene-causing mutation in *TLR15* (red labels). Independent pseudogene events (as determined by at least one bird having only the specified pseudogene mutation) are indicated by black arrows at the latest branch at which the mutation could have occurred, given the taxonomic distribution. (B) Heat map indicates the proportion of each of the sampled individuals that were at least heterozygous for the specified pseudogene-causing mutation in species for which population-level *TLR15* sequence data were available. (C) Schematic diagram of *TLR15* shows the position of each independent pseudogene event relative to the gross architecture of the receptor (numbers are nucleotide positions). Six mutations were found in the extracellular domain (ECD), one in the transmembrane domain (TMD), and two in the TIR domain. "1391?" indicates an insertion that is too large to be resolved using short-read sequencing technology, but is likely to be in excess of ∼150bp (and contains a stop codon). (D) *TLR15* haplotypes and genotypes of species for which population-level data were available. $TLR15^{-/-}$ birds were homozygous for a pseudogene-causing mutation (or had one independent pseudogene in each copy); $TLR15^{+/+}$ birds had no evidence of a pseudogene-causing mutation; $TLR15^{+/-}$ birds had one heterozygous pseudogene event.

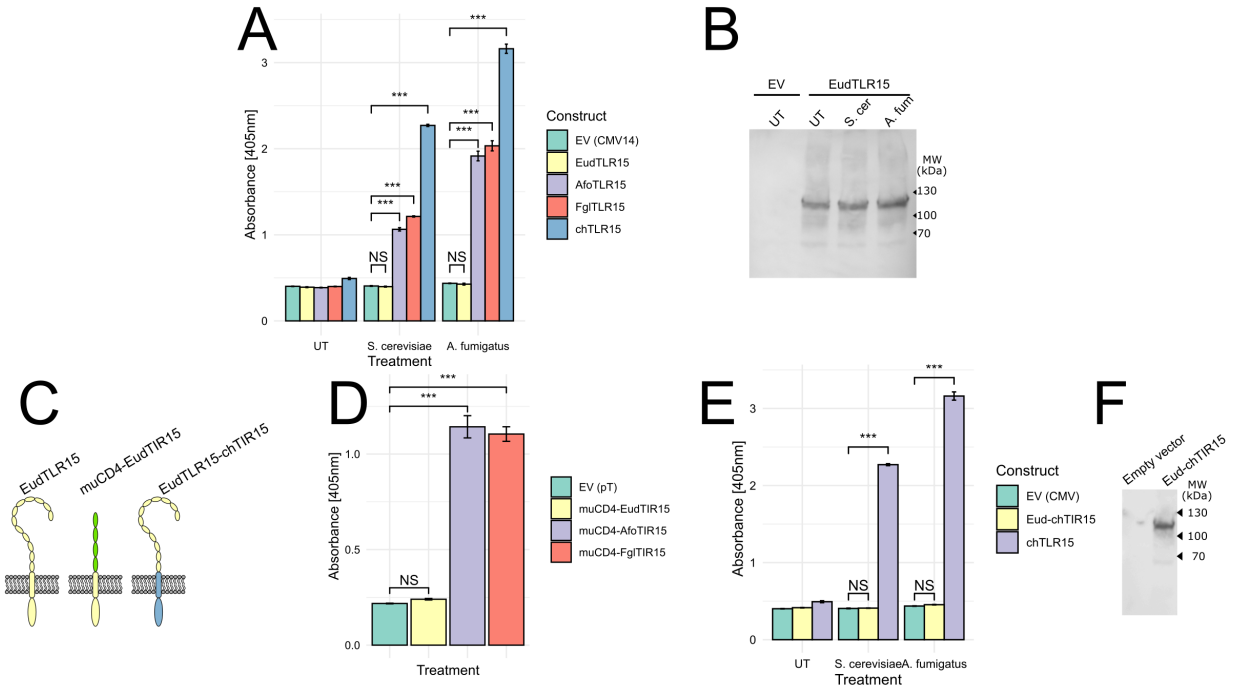


Figure 3.7: Full-length *Eudyptes* consensus *TLR15* is non-functional, and is therefore a cryptic pseudogene. (A) Expression and stimulation of TLR15 from chicken (ch), Northern fulmar (Fgl), Emperor penguin (Afo), and a consensus intact *Eudyptes* TLR15 (Eud) in HEK-Blue Null1 NF-κB reporter cells. Cells expressing TLR15 constructs, or empty vector (EV) control, were challenged with brewer's yeast (*Saccharomyces cerevisiae*) or *Aspergillus fumigatus* lysates, or untreated (UT; medium control). (B) Immunoblot expression analysis of *Eudyptes* TLR15 in HEK-Blue Null1 NF-κB reporter cells with the same treatments as (A). (C) Schematic diagram of *Eudyptes* TLR15 chimaeras generated to functionally test the extracellular domain and intracellular domain independently. *Eudyptes* TLR15 TIR domain was attached to the extracellular domain from murine CD4 (muCD4-EudTIR15), and *Eudyptes* TLR15 extracellular domain was attached to the TIR domain from chicken TLR15 (EudTLR15-chTIR15). Equivalent chimaeric constructs were made for muCD4-AfoTIR15 and muCD4-FglTIR15. (D) NF-κB response to expression of muCD4-TIR domain chimaeras (and no additional stimulation). (E) NF-κB response to expression of Eud-chTIR15 or chTLR15 and stimulation with *S. cerevisiae* or *A. fumigatus* lysates. (F) Immunoblot expression analysis of Eud-chTIR15 in HEK-Blue Null1 cells. CMV14 and pT (pTarget) refer to the expression vectors used for each construct. Statistical significance was considered to be $P < 0.05$ and '***' denotes $P < 1{\times}10^{-7}$ following Dunnett's procedure for multiple comparisons [153]; NS, not significant.

Since the mutation is towards the end of the TIR domain, it is unclear whether the protein would still retain functionality in this species, and it is not known how prevalent such mutations are in the absence of population-level data.

Based on their phylogenetic distribution, the *Eudyptes TLR15* pseudogenes were inferred to have different times of origin (Figure 3.6A, B). 681+A was only observed in *E. chrysocome* (1/42 haplotypes) and C143G and 1391+? were only observed in *E. filholi* and *E. chrysocome* (C143G: 18/56 haplotypes and 2/42 haplotypes respectively; 1391+?: 3/56 haplotypes and 2/42 haplotypes respectively). The 681+A, C143G and 1391+? mutations were therefore inferred to have arisen within the Rockhopper penguin lineage. C185A, 1826+T, 1996+T and 2273+31bp were found in multiple species and were therefore considered to be more ancient in origin. For instance, the 1826+T mutation was found in *E. chrysolophus chrysolophus* (30/80 haplotypes), *E. chrysolophus schlegeli* (5/12 haplotypes), *E. filholi* (25/56 haplotypes) and *E. chrysocome* (22/42 haplotypes).

Analysis of phased haplotypes indicated that a majority of *Eudyptes* spp. individuals had evidence of pseudogenization in at least one copy of the gene. Frequency of pseudogenization ranged from 83.3% (*E. moseleyi*) to 100% of individuals (*E. c. schlegeli*; Figure 3.6B) with a mean of 90.8% across all species. We found that 157/214 (73.36%) of haplotypes were pseudogenes, 60/107 (56.07%) of birds had pseudogenes for both copies ($TLR15^{-/-}$), 37/107 (34.58%) were heterozygous for one pseudogene and one intact haplotype ($TLR15^{+/-}$), and 10/107 (9.35%) were homozygous for an intact haplotype ($TLR15^{+/+}$). At the species level, the prevalence of the $TLR15^{-/-}$ genotype was broadly similar (50-60%; Figure 3.6D) in the best-sampled species (*E. c. chrysolophus*, *E. filholi E. chrysocome*) but varied more in *E. c. schlegeli* (83.33%; N = 6) and *E. moseleyi* (25.00%; N = 12), which likely indicates variation due to under-sampling. In all species, $TLR15^{+/+}$ prevalence was low (0-16.67%; Figure 3.6D), indicating that individuals with intact genotypes are rare. Phasing also revealed an overwhelming preponderance of unique haplotypes (209 unique from a possible 214 derived from the 107 individuals included in the analysis), which may

indicate a relaxation in selection pressure following pseudogenization and the acquisition of mutations by drift.

3.4.5 Functional analysis of full-length TLR15

Having multiple independent, but non-fixed, pseudogenizations in a single genus would be highly unusual. We therefore sought to test the alternative, more parsimonious hypothesis that the apparently intact *TLR15* in the *Eudyptes* was in fact a cryptic pseudogene that has subsequently acquired mutations that truncated the coding sequence. FLAG-tagged full-length TLR15 from *Aptenodytes forsteri* (Emperor penguin), *Fulmarus glacialis* (Northern fulmar), *Gallus gallus* (chicken), and a consensus *Eudyptes* intact TLR15 (which was itself a haplotype present in wild birds) were expressed in HEK-Blue Null1 NF-κB reporter cells. Cells expressing these constructs were then challenged with a known TLR15 agonist—*Saccharomyces cerevisiae* lysate [72]—alongside the lysate of an important fungal pathogen in penguins, *Aspergillus fumigatus*. Interestingly, the Emperor penguin, Northern fulmar and chicken TLR15 all responded to both *S. cerevisiae* and *A. fumigatus* lysates, but neither could stimulate the activation of the *Eudyptes* TLR15 (Figure 3.7A), despite full-length protein clearly being expressed in the system (Figure 3.7B). This supports the hypothesis that even the seemingly intact *Eudyptes TLR15* is non-functional, in contrast to other bird species. We then sought to identify the location of the lesion that gives rise to the non-functional phenotype by means of chimaeric TLR constructs (Figure 3.7C). To test for TIR domain functionality, a constitutively active form of the *Eudyptes* TLR15 transmembrane and TIR domain was constructed with the extracellular domain of murine CD4 [402](Figure 3.7C). Similar constitutively active constructs were made for TLR15 from the Emperor penguin and the Northern fulmar as controls. The muCD4-EudTIR15 chimaera did not signal (whereas the control constructs did), suggesting a critical function-affecting lesion in the *Eudyptes* TLR15 TIR domain (Figure 3.7D). Finally, to test the extracellular domain

of *Eudyptes* TLR15, a further chimaera was constructed by attaching the chicken TLR15 TIR domain (as a known-functional TIR domain) to the *Eudyptes* TLR15 extracellular domain. This construct was also non-responsive to both *S. cerevisiae* and *A. fumigatus* lysates (Figure 3.7E), despite clearly being expressed (Figure 3.7F). Taken together, these results indicate that the seemingly intact *Eudyptes TLR15* is a cryptic pseudogene—one or more disabling (but not truncating) mutations rendered the receptor non-functional prior to its eventual overt pseudogenization by drift in subsequent lineages.

3.4.6 Loss-of-function mutation

Since the mechanism of action of TLR15 signaling has not been elucidated, we could not identify candidate sites for cryptic pseudogenization based on their proximity to sites of known function. Instead, we used a homology-based method to identify sites that are distinct in Eudyptes compared to other birds (Appendix Table B.4). Ten sites were identified that were distinct in *Eudyptes* compared to the rest of birds, and of these, three sites were highly conserved across a wider analysis of vertebrates (161, 736 and 787; Appendix Table B.5 and Appendix Figure B.7). Two polymorphisms were also found to be likely deleterious (L161P, extracellular domain, and L683S, Transmembrane/TIR domain; Appendix Tables B.7 and B.6) when applied to the chicken TLR15 sequence.

3.5 Discussion

Immune system genes, particularly the Toll-like receptors, are frequently found to be hotspots of diversity and positive selection in birds [10, 220, 325, 542, 565, 603]. This is unsurprising given the strong bilateral selection pressures at the host-pathogen interface. Penguins radiated from Australia/New Zealand in the early Miocene into highly diverse niches [604], from the tropical Galápagos Islands, through the temperate forests and rocky shores of southern hemisphere coastlines and islands, to the

sea ice of Antarctica [502, 457, 604]. These diverse biotic and abiotic conditions are likely to foster distinct pathogen assemblages [125, 285, 479, 591], which in turn has exerted a selection pressure on penguins to adapt to the different threats they pose. In the present study, we found evidence of adaptive evolution at key functional sites in penguin Toll-like receptors, which strongly implies change of pathogen recognition function. In addition, we described a highly unusual cryptic pseudogenization event in the relatively understudied *TLR15* in the crested (*Eudyptes* spp.) penguins, which could have significant implications for how penguins respond to important pathogens such as *Aspergillus* spp.

3.5.1 Penguin Toll-like receptors have undergone diversifying selection at key functional sites

Toll-like receptor orthologs retain specificity for particular agonists over large evolutionary distances [493]. Despite this, TLRs in vertebrates frequently undergo diversifying selection through coevolution with pathogens [25, 374, 603]. Even closely related taxa may show substantial variability in molecular phenotypes of individual TLRs [325, 358, 582]. Our findings are similar to those reported elsewhere in birds. Positive selection is more common in lipid-, glycan-, and protein-recognizing TLRs (*TLR1B*, *TLR4*, *TLR5*, *TLR15*), compared to nucleic acid-recognizing TLRs as reported by others [603]. Furthermore, the extracellular (LRR) domains of TLRs (which interact directly with PAMPs) were found to have many more positively selected sites than the intracellular (TIR) domains. An excess of positively selected sites in the extracellular domain of TLRs has been widely documented [25, 220, 565], likely due to the interaction with components of PAMPs that are themselves under selection to evade recognition [14]. The TIR domain, on the other hand, interacts with host adaptor proteins and is typically under purifying selection [407]. The finding of an excess of positively selected sites in the extracellular domains of penguin TLRs could indicate adaptation to different pathogen pressures in different penguin

species.

Further evidence of changing receptor function comes from several positively selected sites being at, or near, positions of known functional importance. For instance, residues 115, 118, 142, 144 and 178 of *TLR1B*, residues 308 and 331 of *TLR4*, and residue 33 of *TLR5* are all at, or within 5 Å of, sites of known ligand binding function [24, 196, 273, 446, 458, 460, 530]. All but one (*TLR1B*, site 178) of these sites have been found to be positively selected in mammals and/or birds in other studies [25, 185, 220, 264, 423, 535, 603, 608, 637], implying that these sites undergo recurrent positive selection to change function in different vertebrate lineages.

3.5.2 Pseudogenization of *TLR5* in the Snares Crested penguin

While gathering sequences, we discovered a *TLR5* pseudogene in the Snares Crested penguin. *TLR5* pseudogenization has occurred multiple independent times in different bird species across several orders (Passeriformes, Psittaciformes, Cariamiformes, Trogoniformes, Phaethontiformes, Eurypygiformes, and Apodiformes; [10, 36, 603]). The discovery of a *TLR5* pseudogene in the Snares Crested penguin represents the first description of a *TLR5* pseudogene in the Sphenisciformes. *TLR5* has also been lost independently from several other vertebrate lineages, including pigs, Yangtze river dolphin, pinnipeds, pangolins, tuatara and clownfish [374, 533]. The conditions under which *TLR5* pseudogenization occurs remain unclear, although it is interesting to note that *TLR5* pseudogenization invariably occurs at or near terminal branches in vertebrate phylogeny, suggesting a recent change in pathogen exposure. It is also interesting to note that the loss of *TLR5* is not without consequence in other vertebrates. For instance, the common TLR5-392STOP variant in humans is associated with increased susceptibility to Legionnaires' disease [236]. However, it is also plausible that the loss of *TLR5* is adaptive, akin to *TLR4*-null mice being resistant to endotoxic shock [486].

3.5.3 Cryptic pseudogenization in *Eudyptes* spp. *TLR15*

TLR15 is a comparatively understudied receptor, first described as being upregulated in the caecum of chickens infected with *Salmonella enterica* serovar Typhimurium [252]. Unlike most other TLRs, a canonical agonist has not been described for TLR15. There is evidence that TLR15 responds to the N-terminal diacylated lipopeptide of the haemagglutinin protein of *Mycoplasma synoviae* in a TLR1- and TLR2-independent manner [455]. Another study identified a heat-labile and PMSF-abrogable yeast-derived agonist as being capable of activating TLR15 [72], while a further report postulates that TLR15 is activated not by a conventional agonist, but rather by receptor cleavage induced by microbe-derived proteases [138] TLR15 faces the extracellular space [138] and is unusual for TLRs of this type for having a low rate of adaptive evolution (despite high levels of variation) in birds that is more akin to the intracellular nucleic-acid recognizing TLRs (*TLR3*, *TLR7* and *TLR21*; [603]). This could be indicative of an unusual mechanism of stimulation, such as through cleavage by microbial proteases.

Initial examination of penguin reference genomes suggested that *TLR15* had been pseudogenized multiple times in the *Eudyptes* spp., with a further independent pseudogenization in the Chinstrap penguin (*Pygoscelis antarcticus*). Since several of the mutations were widespread in the *Eudyptes* penguins in our population-level analysis of >100 penguins, we concluded that at least some of the disruptions had taken place in the common ancestor to all *Eudyptes* penguins. It could be argued that this pattern could have arisen through modern introgression between lineages. Indeed, between specific pairings of *Eudyptes* penguins, introgression is high; for instance Royal and Erect-crested (*E. chrysolophus schlegeli* and *Eudyptes sclateri*, respectively) introgression is evident in ∼25% of the genome [604]. However, in other *Eudyptes* spp. pairings, introgression is absent or minimal; for example between Royal and Eastern rockhopper penguins (*E. chrysolophus schlegeli* and *E. chrysocome filholi*, respectively) and between Macaroni and Eastern rockhopper penguins (*E. chrysolo-*

phus chrysolophus and *E. chrysocome filholi*, respectively), rates of introgression are <1% [604]. Despite this, the 'basal' disrupted *TLR15* haplotypes were shared widely across species. For instance, the 1826+T frameshift mutation was found at high frequencies in pairs of species with little evidence of introgression; 30/80 haplotypes in *E. chrysolophus chrysolophus* and 25/56 haplotypes in *E. chrysocome filholi* (introgression <1%; [604]). This implies that the 1826+ mutation was acquired by descent, rather than introgression. Since *E. chrysolophus chrysolophus* and *E. chrysocome filholi* occupy the two sister clades in the genus, the only likely explanation for the mutation appearing in both lineages is that it occurred in the stem branch to extant *Eudyptes* spp. The fact that four independent pseudogenes were found to be basal to the genus is likely to have been a result of incomplete lineage sorting in the early evolution of the crested penguins. Furthermore, intact haplotypes were present in the populations at low levels as part of heterozygote ($TLR15^{+/-}$; 35.51%) or as homozygote ($TLR15^{+/+}$; 9.35%) genotypes. Given that it would be highly unusual to have so many independent pseudogenizations of a gene in a single genus, we elected to functionally test the consensus intact genotype to ascertain whether it was likely functional prior to the independent accumulation of multiple, unique inactivating mutations.

Although a canonical agonist has not been defined for TLR15, it is clear that this TLR is activated by yeast and fungal-derived products [72, 138]. *Aspergillus* spp. are common fungal pathogens of penguins and other birds, both captive and wild [54, 178, 213, 258, 304, 441]. Given the importance of Aspergillus spp., we hypothesized that TLR15 may play a role in responding to Aspergillus spp. infection. It is clear that TLR15 from chicken, Northern fulmar and Emperor penguin responded to *Aspergillus fumigatus* lysate, as well as a previously defined agonist, brewer's yeast (*Saccharomyces cerevisiae*) lysate [72]. By contrast, the consensus intact *Eudyptes* spp. TLR15 was incapable of signaling in response to either of these agonists, despite being expressed in its full-length form. The failure of constitutively active forms of the *Eudyptes* TLR15 TIR domain or chimaeras (extracellular domain of *Eudyptes*

TLR15 fused with the strongly signaling chicken TIR domain) to rescue activity suggests that there are multiple lesions across the TLR sequence (at least one in each domain). These findings suggest that receptor functionality was lost prior to the gene disruptions across *Eudyptes*, which were likely a result of relaxed selection pressure and genetic drift.

It is therefore likely that the intact *Eudyptes* spp. *TLR15* haplotype represents a cryptic pseudogene, in the sense that the open reading frame is complete in the genome and is capable of yielding a full-length protein, but it lacks functionality. Cryptic pseudogenes are difficult to detect without functional analysis or an overt phenotype, and have only rarely been reported [39, 638]. More commonly, a gene that undergoes duplication may experience functional redundancy and subsequent loss of one copy through pseudogenization [669]. However, a duplication/pseudogenization scenario does not fit with *Eudyptes TLR15*, since many of the truncating mutations were homozygous, we did not identify any triallelic sites, and the observed read depth of *TLR15* in reference genomes [457] was comparable to coding sequences across the rest of the genome (Appendix B.3). It is also unlikely that *TLR15* experienced any functional redundancy with other TLRs, since *TLR15* is the sole family member and recognizes agonists that are distinct from other TLR families [72]. It is important to note that the radiation of *Eudyptes* penguins coincided with, and was perhaps driven by, a period of great flux with the emergence of sub-Antarctic and temperate islands in the Plio-Pleistocene [118, 604]. Small population dynamics associated with colonization of new islands may have fostered the survival and local fixation of a loss-of-function *TLR15* variant that subsequently drifted to fixation across the genus.

Although the mechanism of action of TLR15 has not been fully elucidated, it was clear that functionally disruptive changes exist in both the extracellular (leucine-rich repeat) domain and in the transmembrane/TIR domain. We identified a number of sites that differ in *Eudyptes* penguins that are otherwise well conserved across birds and other vertebrates (Appendix Table B.4). Using PROVEAN [101] and SIFT [432], we identified two sites that were predicted to be deleterious when the chicken

TLR15 sequence was altered (L161P and L647S; site positions refer to the Emperor penguin sequence; Appendix Tables B.6 and B.7). Amino acid site 161 is located in the extracellular domain in an extended loop between LRR3 and LRR4, while site 647 is within the transmembrane domain. Additional work to functionally test the identified sites could give a better insight into the mechanism of TLR15 signaling, and into the origin of the cryptic pseudogene in *Eudyptes* spp.

The consequence of *TLR15* loss in penguins remains unclear. A simple prediction would be that birds with intact *TLR15* (e.g. the Emperor penguin) are more resistant to *Aspergillus fumigatus* infection compared to *Eudyptes* species. However, comparing incidence of aspergillosis in Emperor penguins and *Eudyptes* spp. is problematic because they are not sympatric in the wild, and are maintained at different climate conditions in captive settings [527]. There is evidence that several other penguin species are susceptible to aspergillosis, such as Gentoo penguins (genus: *Pygoscelis*; [178]), Magellanic penguins (genus: *Spheniscus*; [88, 329, 642]), and Little penguins (genus: *Eudyptula*; [258, 441]). However, multiple other factors, such as environmental conditions and bird health, are known to increase risk of mycoses [54], so elucidating the contribution of genetics to aspergillosis resistance is difficult. Future work to evaluate the functionality of *TLR15* in other penguin and bird species, followed by a prospective cohort study of aspergillosis infection in comparable penguin populations will greatly enhance our understanding of the genetic contribution to disease resistance. Penguins are popular attractions at zoos around the world, and the ability to identify genetically susceptible populations will help safeguard vulnerable animals. Similarly, climate change is predicted to introduce novel pathogens to wild populations, and so identifying genetically at-risk populations is valuable from a conservation perspective [116].

3.6 Conclusions

Toll-like receptors are under constant evolutionary pressure to adapt to a changing pathogen landscape. Penguins are an ideal taxon to study TLR adaptation because their extensive geographical range undoubtedly means than pathogen exposure is distinct between species and genera. In this study, we provided evidence of widespread adaptive evolution in TLRs across penguin phylogeny. Further, we recapitulated patterns of adaptive evolution of particular TLRs and functional sites that have been reported in other birds and vertebrate taxa. We also reported a highly unusual cryptic pseudogenization event in *Eudyptes TLR15*, which recognizes fungal products and may be involved in the recognition of pathogens such as *Aspergillus* spp. This gene evidently lost function in the common ancestor to extant *Eudyptes* and has undergone subsequent disruption and overt pseudogenization among different lineages. Aside from potentially contributing to the understanding of penguin susceptibility to aspergillosis, the cryptic pseudogenization of *TLR15* provides an insight into the process of trait erosion in wild animals and the mechanism of gene disruption.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

Title of Paper	Adaptation and cryptic pseudogenization in penguin Toll-like receptors
Publication Status	In review
Publication Details	Steven R Fiddaman, Michal Vinkler, Simon G Spiro, Hila Levy, Christopher A Emerling, Amy C Boyd, Evangelos A. Dimopoulos, Juliana A Vianna, Theresa L Cole, Hailin Pan, Miaoquan Fang, Guojie Zhang, Tom Hart, Laurent AF Frantz, Adrian L Smith, Adaptation and cryptic pseudogenization in penguin Toll-like receptors, Molecular Biology and Evolution, in review.

Student Confirmation

Student Name:	Steven R Fiddaman		
Contribution to the Paper	I conceived the project and conducted all analyses except for homology modelling, which was done by MV. I wrote the paper and made all figures. LAFF, ALS and TH supervised the project and provided editorial assistance. All other authors provided samples or sequence data.		
Signature: Steven R Fiddaman		Date	13/07/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Adrian L Smith		
Supervisor comments		
Signature 	Date	13/07/2021

Chapter 4

A genome-wide analysis reveals extensive positive selection at immune loci in penguins

4.1 Abstract

Having explored the effect of pathogen-driven selection on penguin Toll-like receptors in the previous two chapters, I now turn to a more global assessment of immunogenetic changes in penguins. Here, I present a genome-wide analysis of selection in all extant species of penguin. Over 50 immune genes were positively selected, a significant proportion of which were pathogen-facing immune components, which is suggestive of host-pathogen co-evolution. Analysis of selected sites revealed considerable overlap with sites of known function. Of particular importance, sites lining a viral RNA-binding cleft in interferon-induced protein with tetratricopeptide repeats (*IFIT5*) and sites forming part of a region known to be indispensable for viral binding and cell entry in *CD81* were found to be positively selected. Taken together, these data provide evidence for extensive pathogen-driven immunogenetic selection in addition to that which we previously reported in the Toll-like receptors in penguins.

4.2 Introduction

Penguins occupy a significant proportion of the southern hemisphere, with ranges spanning from the equatorial Galápagos Islands (Galápagos penguin; *Spheniscus mendiculus*) to 77.74 °S in Antarctica (Emperor penguin; *Aptenodytes forsteri* [67]). Penguins spend most of their lives at sea, but are dependent on land for breeding and nesting [67]. The physiological adaptations required for a transition towards a diving, flightless, semi-aquatic life history are numerous and have been well studied [604]. Many other aspects of penguin physiology, however, are currently unknown. For instance, penguins nest in densely colonial conditions, characterised by ubiquitous guano and interactions with other bird species. The opportunity for pathogens to spread in penguin colonies is therefore substantial, evidenced by outbreaks of infectious disease causing widespread mortality in wild and captive penguin populations [45, 110, 218]. Precisely how pathogens have shaped the penguin immune response is an important avenue for investigation.

It has been argued that penguins are 'immunologically naïve' because of their presumed origin in the inhospitable Antarctic climate leading to a reduced pathogen pressure [218, 37]. However, fossil and molecular evidence now indicate that *Aptenodytes* and *Pygoscelis* penguins only secondarily colonised Antarctica [200, 330], and the origin and radiation of penguins was in the milder Miocene and may have been driven by island formation [118, 200, 562, 604]. This is inconsistent with the crown penguins having evolved in a low pathogen environment.

Despite this, penguins do appear to be highly susceptible to novel pathogens, as indicated by the high impact of disease outbreaks in ectopic captive settings. For instance, outbreaks of aspergillosis [304], mycobacteriosis [425], Newcastle disease [230], malaria [217] and other diseases [218] occur in captive penguins and cause considerable mortality. Furthermore, the introduction of novel diseases into wild penguin colonies can cause significant disease burden. For example, an outbreak of highly pathogenic avian influenza subtype H5N8 caused substantial mortality in a

population of African penguins (*Spheniscus demersus*) in Namibia [413], and a disease resembling puffinosis was responsible for a mass mortality event of Gentoo penguin (*Pygoscelis papua*) chicks on Signy Island [391].

Tourist numbers to Antarctica have risen from a few hundred per annum at the beginning of the tourism industry in the 1950s to over 55,000 in 2019-20 [268]. However, despite the measures implemented by tour operators, current biosecurity practice may by insufficient to prevent spread of disease to wild penguins [128, 265]. There is already some evidence of humans acting as vectors of penguin pathogens [93]. Furthermore, in terms of climate change, penguins occupy some of the most at-risk environments [388, 557]) and are therefore likely to suffer from the spread of novel pathogens by non-native bird species.

The vulnerability of penguin species is further compounded because 12 out of 18 penguin species listed on the IUCN Red List are considered threatened or near-threatened. Penguins therefore do not have the resilience of species with large populations to cope with losses due to infectious disease. An understanding of the penguin immune system is a crucial first step to understanding penguin susceptibility to pathogens. Coupled with a greater understanding of pathogen diversity in penguin-endemic regions, identifying at-risk species and populations could inform conservation efforts and mitigate the effects of a changing climate. In this chapter, I present a genome-wide analysis of positive selection in penguins as a first step in understanding how pathogens have shaped the penguin immune system. The results reveal that immune genes are enriched in positive selection compared to genes of other functions. Moreover, positive selection was concentrated in pathogen-facing genes, and selection in several pathogen-binding sites was noted, indicating that co-evolution with pathogens has played a significant role in the evolution of immunity in penguins.

4.3 Methods

4.3.1 Protein-coding gene annotation

Genomes sequenced in Li, et al (2014 [359]; Adélie penguin [*Pygoscelis adeliae*] and emperor penguin [*Aptenodytes forsteri*]), together with genomes from all other extant species of penguin [457] were included in the analysis. The methods for protein-coding gene identification are given in [457]. Briefly, an avian reference gene set of 20,181 protein-coding genes was curated using non-redundant sequences from the zebra finch (*Taeniopygia guttata*) and chicken (*Gallus gallus*) published genomes. Next, TBLASTN was used to identify a candidate genomic region for each of the protein queries with a e-value cutoff of 1×10^{-5}. Each candidate genomic region was then used as an input for GeneWise (v. 2.4.1 [58]) and the gene model with the highest identity with the corresponding protein homologue was identified. Avian transcriptome data from 71 different species were also assembled and the protein products mapped to the penguin genomes. In addition, human protein data were mapped to the penguin genomes. Together, the three different gene models (avian reference gene set, human protein data set, transcriptome data set, used in this order of priority) were used to annotate approximately 16,000 genes in each penguin genome, which is comparable to the zebra finch and chicken genomes.

4.3.2 Gene function annotation and orthologue alignment

The methods for gene function annotation are given in Pan *et al.* (2019) [457]. Briefly, each gene was queried against three functional databases using BLASTP: Swiss-Prot release-2019 03 [63] InterPro v. 68.0 [277], and KEGG v89.1 [287]. Gene Ontology (GO) terms were assigned to each gene based on its corresponding InterPro entry. Orthologues from each genome were identified using Cactus [27]. Orthologue sets were aligned using a codon alignment in PRANK [383] and the final dataset only included genes which included orthologues from all species (n = 8716).

4.3.3 Selection analysis

In order to detect positive selection within the penguin lineage, without a priori predictions about the branches along which selection occurred, the site models in PAML (v 4.9; [652, 654]) were employed. The PAML site models M1a (Nearly Neutral) and M2a (Positive Selection) were fitted to the alignments. Likelihood ratio tests were performed on pairs of log-likelihood values and P-values obtained from the χ^2 distribution with df = 2. To control for multiple hypothesis testing, q-values were computed based on the distribution of P-values using the `qvalue` package in R (v. 4.0.2), and a q-value threshold of $q < 0.01$ was considered significant. Alignments for all genes identified as positively selected were manually checked to confirm the absence of alignment error, assembly error or non-codon-based variation that could falsely be interpreted as positive selection. Identities of sites under selection were obtained using the Bayes Empirical Bayes (BEB) procedure and were considered significant at posterior probability > 0.9.

4.3.4 Immune gene identification

The R package `GO.db` was used to identify all offspring gene ontology (GO) terms from the highest level immune system process GO term (GO:0002376). The resultant list of GO terms were cross referenced with the positively selected genes using a custom bash script and positively selected immune genes were identified. In total, 51 immune genes were positively selected with $q < 0.01$.

4.3.5 Homology modelling

In order to relate positively selected sites in penguin genes to regions of known function, homology models were built using the I-TASSER program [505, 649, 670]. The protein sequences used to generate the homology models were derived from the Eastern rockhopper penguin (*Eudyptes filholi*; CD81) and Emperor penguin (*Aptenodytes forsteri*; IFIT5), and are given in Appendix Section C.1. The best model in each

case was then taken forward for annotation in Pymol (v. 2.5.0). Both CD81 and IFIT5 yielded highly accurate models (CD81: C-score = 1.02; TM-score = 0.85 ± 0.08; RMSD = 3.7 ± 2.5 Å; IFIT5: C-score = 0.94; TM-score = 0.84 ± 0.08; RMSD = 5.2 ± 3.3 Å).

4.4 Results and Discussion

The genome-wide positive selection analysis identified 616 positively selected genes with q < 0.1 (275 genes with q < 0.01). Of these, metabolic genes formed the biggest functional class (n = 307), followed by developmental genes (n = 183), signalling components (n = 170) and immune genes (n = 109; Figure 1). Genes are likely to fall into multiple categories, and it is unsurprising that the most populous categories are also the most generic (metabolic genes, developmental genes and signalling genes). Controlling for the different numbers of genes in each functional category, immune genes had the highest rate of positive selection (9.14%; average 6.40%; Figure 1), indicating that immune genes have undergone functional diversification in penguins.

4.4.1 Positively selected immune genes

There were 109 positively selected genes with function relevant to the immune system. The following section focuses on some of the 51 genes that were highly significant for positive selection (q < 0.01). A full list of these genes can be found in Appendix Table C.1.

4.4.1.1 Toll-like receptors and adaptors

Two Toll-like receptors (*TLR4* and *TLR5*) were positively selected (q = 2.13E-03 and q = 2.14E-05, respectively). Toll-like receptors are responsible for binding pathogen-derived molecules directly and initiating a downstream signalling cascade to activate the immune response [6]. Both TLR4 and TLR5 are responsible for recognising bacterial-derived molecules (lipopolysaccharide for TLR4 [102] and flagellin for TLR5

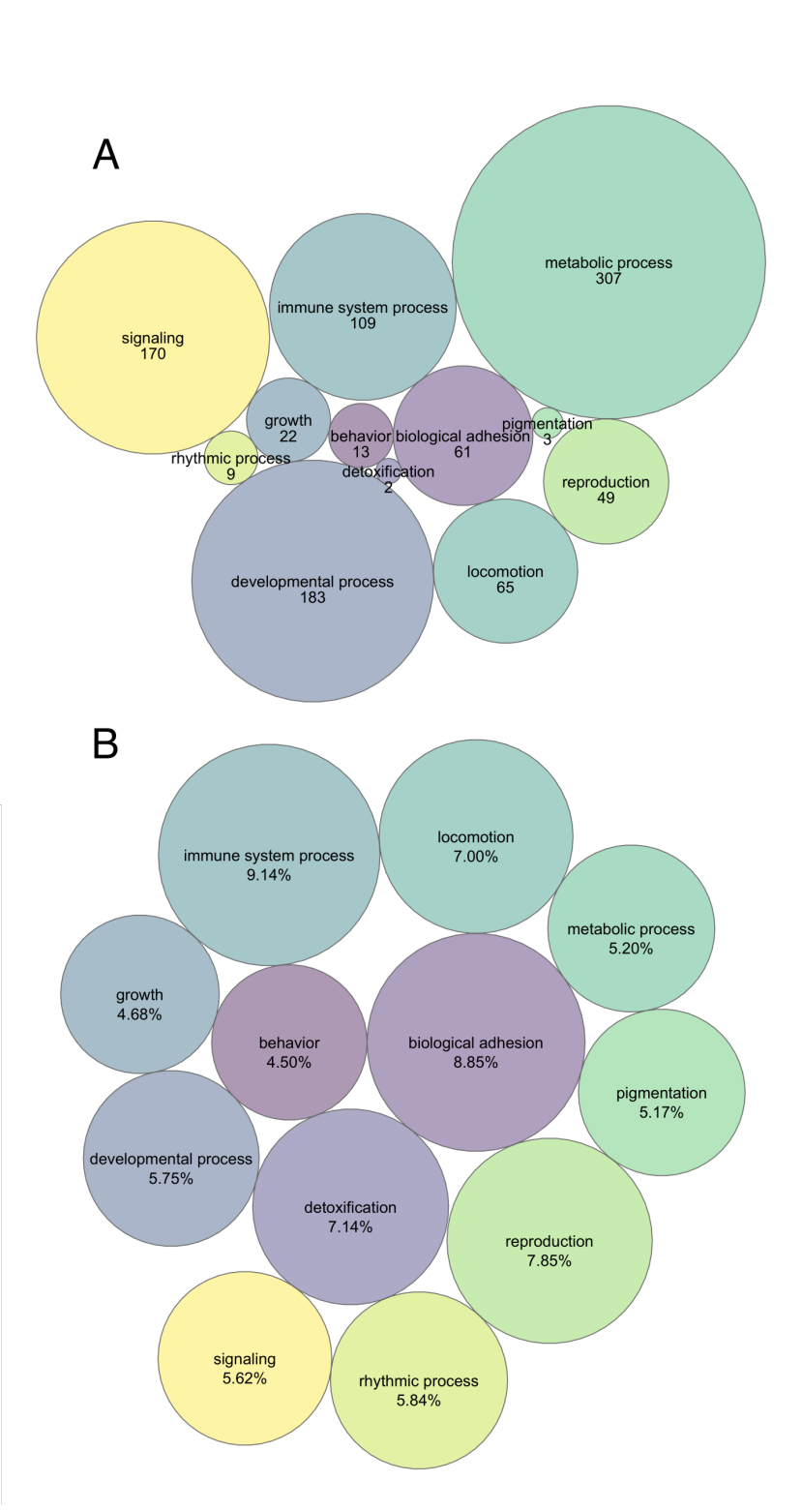


Figure 4.1: Numbers and proportions of positively selected genes by functional category. (A) Raw numbers of positively selected genes per functional category (categories are not mutually exclusive and genes can occupy more than one). (B) Proportion of input genes in each category that were positively selected. For instance, of all the immune genes included in the analysis, 9.14% of these were positively selected. In each case, bubble size reflects relative size of each category.

[239]). *TLR4* and *TLR5* were found to be under positive selection in the Gentoo penguin sub-lineages ([358]; Chapter 2), across all penguin lineages (Chapter 3), as well as elsewhere in birds [10, 220, 603, 606]. The positively selected sites in *TLR4* and *TLR5* are discussed in detail in Chapters 2 and 3 of this thesis, but key sites include a site proximal to an LPS-binding site in *TLR4* (codon 302; [603]) and at a flagellin-binding site in *TLR5* (codon 33; [659]). As discussed previously, positive selection of *TLR4* and *TLR5* implies that penguins, like other birds, have adapted to changes in bacterial pathogens which may be associated with their radiation into different climates. In addition to the Toll-like receptors themselves, a TLR adaptor protein—TIR Domain-Containing Adapter Molecule 1 (*TICAM1*, a.k.a. *TRIF*)—was also found to be under positive selection (q = 4.46E-06). TRIF forms the basis of the myeloid differentiation primary response protein 88 (MYD88)-independent signalling pathway used by TLR3 and TLR4 [485]. Interestingly, *TRIF* appears to be under strong purifying selection in mammals [423, 485], but is known to be under extensive diversifying selection in birds [542]. This implies a fundamental difference in MYD88-independent signalling between birds and mammals, and may suggests birds have fine-tuned the responsiveness of TLR3 and TLR4 activation.

4.4.1.2 *IFIT5*

In addition to the Toll-like receptors, other pattern recognition receptors were found to be positively selected. One such example is interferon-induced protein with tetratricopeptide repeats (*IFIT5*; q = 5.98E-11), which binds viral RNA bearing a 5′-triphosphate group [1], as opposed to host mRNA which is protected with a 5′ cap. Mechanistically, IFIT5 has an RNA-binding cleft that is lined with positively charged residues contributed by tetratricopeptide repeats and associated linking structures [291, 292]. Interestingly, penguin *IFIT5* has undergone diversification at sites known to form part of the RNA-binding cleft (codons 407, 409, 413 and 421; corresponding to human codons 412, 414, 418 and 426; [291, 292]; Figure 4.2). This strongly implies penguin *IFIT5* has undergone functional change to bind different viral RNA motifs

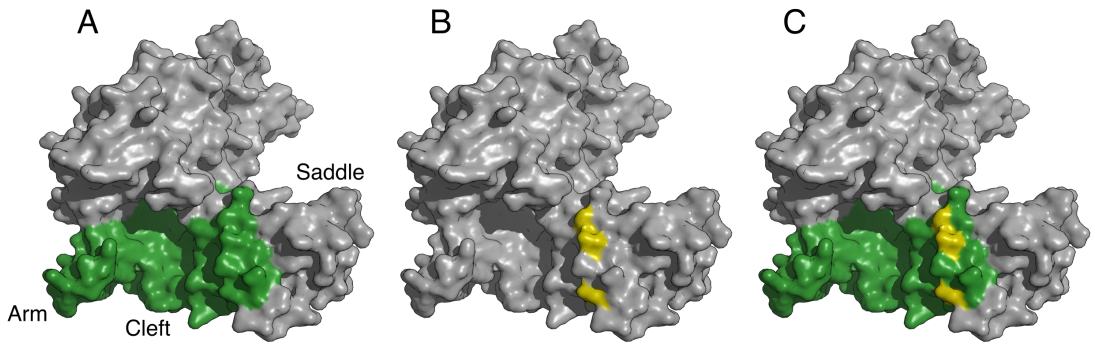


Figure 4.2: Homology model of penguin *IFIT5* showing positively selected sites and RNA-binding cleft. RNA-binding cleft is shown in green and positively selected sites are shown in yellow. (A) shows just the RNA-binding cleft, (B) shows just the positively selected sites and (C) shows the RNA-binding cleft with the positively selected sites superimposed.

in response to pathogen pressure.

4.4.1.3 *CD81*

There was also evidence of evidence of positive selection at a viral target of cell entry. CD81 is a cell surface protein that plays an important role in cell entry of hepatitis C by interaction with the HCV E2 glycoprotein [49, 668] and was positively selected in penguins (q = 2.09E-04). Positive selection has previously been reported at the glycoprotein interface in bat *CD81* [184], and interestingly, the same interface contained a cluster of positively selected sites in penguin *CD81* (sites 181, 182, 186, corresponding to human sites 180, 181 and 185; Figure 4.3), and penguin site 186 (human site 185) was positively selected in both bat and penguin analyses [184]. Although birds are not known to be natural hosts for mammalian hepatitis C virus, a related virus hepacivirus-like virus (Flaviviridae: Hepacivirus) is known to cause a lethal syndrome in bald eagles (*Haliaeetus leucocephalus*; [207]). It is therefore plausible that non-HCV hepacivirus-like viruses use the same mechanism of cell entry into avian cells, which would explain the concentration of positively selected sites at the host-virus interface.

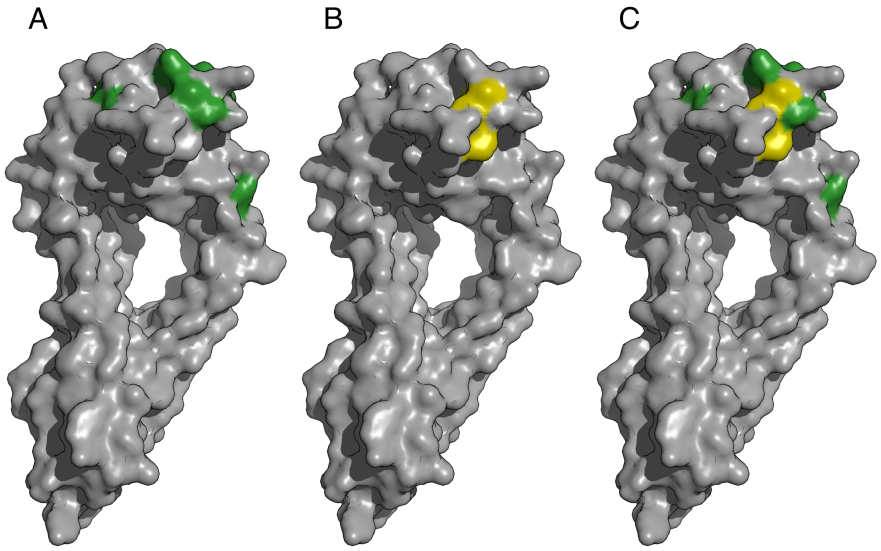


Figure 4.3: Homology model of penguin *CD81* showing positively selected sites and putative HCV E2 glycoprotein-binding site. RNA-binding cleft is shown in green and positively selected sites are shown in yellow. (A) shows just the putative HCV E2 glycoprotein-binding site, (B) shows just the positively selected sites and (C) shows the putative HCV E2 glycoprotein-binding site with the positively selected sites superimposed.

4.4.1.4 Transferrin

Additionally, positive selection was detected in penguin transferrin (*TF*; q = 5.66E-03), which is part of the "nutritional" immune system that sequesters iron from iron-scavenging pathogens [92]. Positive selection in transferrin as an indicator of the host-pathogen co-evolutionary arms race to sequester and scavenge iron has also been detected in mammals [41] and various fishes [183, 564, 650].

4.4.1.5 *PIGR*

The polymeric immunoglobulin receptor (*PIGR*) is responsible for transcytosis of IgA and IgM into the gut lumen, and was found to be positively selected in penguins (q = 3.69E-10). In particular, a cluster of selected sites was identified in the D2 domain (penguin sites 169, 182, 184, 191, which correspond to human residues 161, 173, 178, 185), which is known to be under recurrent positive selection in mammals [120]. The D2 domain is thought to be involved in transcytosis of IgM, but not IgA [436], but

its precise function is yet to be elucidated. Given the extensive positive selection in the D2 domain, it has been argued that it may be involved in a host-pathogen co-evolutionary arms race and thus may have as-yet undiscovered role in immunity [120].

4.4.1.6 Other positively selected genes

Complement components *C4*, C4-binding protein (*C4bp*), *C5* and *C8b* were positively selected across penguins (q-values: 2.13E-04, 2.17E-06, 7.58E-09 and 7.58E-04, respectively). C5 and C8b form part of the membrane-attack complex (MAC), which causes pore formation in the membranes of target cells. Certain pathogens have evolved ways to circumvent the MAC, and thus may exert a selective pressure on MAC components. For instance, herpesvirus saimiri encodes a mimetic of CD59, which is a host molecule which functions to inhibit complement-mediated lysis of host cells [9, 53]. It is feasible that functional change in MAC components enables penguin complement to undo evasion by pathogens.

RNaseL (q = 5.45E-10) is an endonuclease that forms part of the interferon antiviral response. Similar to mammalian *RNaseL*, positively selected sites in penguin *RNaseL* were concentrated in the C-terminal endoribonuclease domain which is responsible for cleaving viral RNA [274].

Several other types of immune component were positively selected, including regulatory factors (e.g. Interferon regulatory factor 1, *IRF1*; q = 2.64E-05), antigen presentation components (e.g. *CD4*; q = 1.97E-03, and cathepsins E (*CATE*) and S (*CATS*); q = 1.98E-03 and 2.10E-04, respectively), cytokines (e.g. interferon gamma (*IFNG*); q = 3.36E-03), and signalling components (e.g. *CD40L*; q = 3.63E-03).

4.4.2 Conclusions

Of the 51 positively selected genes in the penguin immune system, a large proportion were pathogen-facing components. This strongly implies that penguins have expe-

rienced co-evolution with pathogens as part of a host-pathogen arms race. Further evidence comes from the localisation of selected sites within these genes: sites were clustered at ligand-binding residues in Toll-like receptors, in the RNA-binding cleft of *IFIT5*, in the hepacivirus-binding interface of *CD81*, and in the endoribonuclease domain of *RNaseL*. Immune genes are known to be hotspots of selection in birds and mammals [542] but such studies typically focus on selection at the level of the gene, rather than predicting functional consequences based on the distribution of sites and cross-referencing with known function [34, 424]. By combining these approaches, the true significance of immune system adaptation can be realised.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

Title of Paper	A comprehensive genomic history of the evolution of penguins
Publication Status	Pre-submission
Publication Details	Theresa L. Cole†, Chengran Zhou†, Miaoquan Fang†, Hailin Pan*, Daniel T. Ksepka*, Daniel Thomas*, Steven R. Fiddaman*, Christopher Emerling*, Xupeng Bi, Qi Fang, Martin R. Ellegaard, Adrian L. Smith, Tracy Heath, Alan J. D. Tennyson, Pablo Garcia Borboroglu, Jamie R. Wood, Charles-Andr´e Bost, Peter Hadden, Misha Vorobyev, Jason Turuwhenua, Yves Cherel, Thomas Mattern, Tom Hart, Mikkel-Holger S. Sinding, Lara D. Shepherd, Richard A. Phillips, Petra Quillfeldt, Juan F. Masello, Stefanie Grosser, Juan L. Bouzat, Peter Ryan, David Thomas, Ursula Ellenberg, Peter Dann, Gary Miller, P. Dee Boersma, M. Thomas P. Gilbert, De-Xing Zhang, Guojie Zhang, A comprehensive genomic history of the evolution of penguins. Manuscript in preparation.

Student Confirmation

Student Name:	Steven R Fiddaman	
Contribution to the Paper	This chapter forms part of a larger paper that is about to be submitted. While the work contained within the chapter is entirely my own (with supervision from ALS), all other authors contributed to other parts of the final manuscript. My site-based selection analyses are being used as a basis for other parts of the paper concerning physiological adaptation in penguins.	
Signature: Steven R Fiddaman	Date	13/07/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Adrian L Smith		
Supervisor comments		
Signature 	Date	13/07/2021

Chapter 5

A comprehensive description of diversity in chicken Toll-like receptors

5.1 Abstract

Chickens are highly valued worldwide as a source of protein in the human food supply. Domestication and artificial selection for desired commercial traits has resulted in fast-growing birds with excellent yields of meat and eggs. Typically, however, such intensive selection brings about a depression of diversity in the genome. From an immune perspective this may be problematic: a lack of genetic variation in the population may be associated with susceptibility to novel and existing pathogens. To investigate immune diversity in the chicken further, I conducted the largest investigation into Toll-like receptor (TLR) diversity in a domestic animal, to my knowledge. Using TLR sequences derived from the genomes of >1600 domestic chickens and congeneric junglefowl species, I found that, contrary to expectations, chicken TLRs are rich in allelic diversity. Furthermore, I found evidence of positive selection, strongest at the *TLR2* locus, which was contemporaneous with chicken domestication from its red junglefowl ancestor (*Gallus gallus spadiceus*). This may indicate that the *TLR2*

locus was an important adaptive locus for domestication, perhaps linked to the increased pathogen pressure associated with higher population densities. I also catalogued a large number of high-impact variants (e.g., frameshift, nonsense and splice variant mutations) and discovered that several pseudogenes are prevalent in well-known modern chicken breeds. Taken together, these observations suggest that the chicken did not experience reduced immune diversity through domestication, which may have been a result of repeated introgression with wild junglefowl.

5.2 Introduction

Chickens are the most numerous livestock animal, contributing 68 billion individuals and 127 million tonnes of meat to the human food supply annually (source: UN Food and Agriculture Organization). Intensive selection for production traits has vastly increased the efficiency of commercial birds relative to their landrace counterparts. Modern broiler chickens have an exceptional feed conversion ratio and reach slaughter weight in 5-6 weeks of age [531, 544].

The increasing demand for chicken protein forces poultry farms to increase in size and house more birds at greater densities. Larger, denser flocks increase the risk of disease outbreaks which negatively affect production efficiency. Great effort has been invested in vaccination, biosecurity, and early detection and warning systems [29, 430] in an effort to combat infectious disease in chickens, arguably more so than other food animals. However, comparatively little attention has been paid to the immunological resilience of the chickens themselves, particularly gleaning information from the rich genetic resource of feral birds and wild red junglefowl from which chickens were domesticated. Balancing selection, resulting from heterozygotic advantage, is a driver for the maintenance of diversity in the innate immune system [95, 173, 313] which presumably increases resilience in wild populations challenged with novel pathogens. This evolutionary process is in contrast to positive selection, which acts to favour one allele and may drive it to fixation in a population. The potential for a reduction

in diversity at innate immune loci following domestication and artificial selection for production traits may have undesired consequences of eliminating the heterozygote advantage afforded to wild populations.

Chickens, like most birds, have 10 Toll-like receptors which respond to highly conserved microbe-associated molecular patterns (MAMPs; [81]). The TLR1/2 family comprises two pairs of genes found in tandem on chromosome 4 and are responsible for recognising di- and triacylated lipopeptides through heterodimerisation [253]. TLR3, TLR4, TLR5 and TLR7 are one-to-one orthologues of mammalian TLRs and recognise double-stranded RNA [529], lipopolysaccharide [474, 540], flagellin [202], and single-stranded RNA [385], respectively. TLR15 is an avian and reptilian-specific TLR which responds to fungal agonists and/or proteolytic cleavage [72, 138], while TLR21 is a CpG DNA-recognising TLR distinct from mammalian TLR9 [82]. A subset of TLRs are expressed at the cell surface (TLR1A, TLR1B, TLR2A, TLR2B, TLR5, TLR15), while the remainder are localised to endosomes (TLR3, TLR7, TLR21). TLR4 is unusual in that it can be both cell surface and endosomal in localisation [194].

The genomic sequencing era has led to the availability of highly granular whole-genome datasets for a variety of animals. Chickens are one such animal, and there are several hundred publicly availably genomes which can be analysed for genetic diversity at key loci such as the innate immune system. Compiling a large dataset of >1600 genomes from commercial birds, feral/backyard, fancy breeds, all five subspecies of red junglefowl, and all three congeneric *Gallus* spp. (green junglefowl, *Gallus varius*; grey junglefowl, *Gallus sonneratii*; and Ceylon junglefowl, *Gallus lafayettii*), this study aimed to curate the most comprehensive collection of chicken Toll-like receptor diversity of any study to date. These data can form the basis on which inbred lines are assessed for their immune health, and further investigation into association of variants with susceptibility to infectious disease.

5.3 Methods

5.3.1 Data gathering and assembly pipeline

To ensure consistency in the analysis pipeline, all reads were re-mapped to the latest version of the chicken genome (Galgal6). Publicly available sequencing reads of chicken breeds and junglefowl species were downloaded from NCBI SRA or were obtained from [620] (a full list of accession numbers can be found in Appendix Table D.1). Reads were mapped to the chicken genome using `bwa mem` [361] and potential PCR duplicates were removed using `samtools rmdup` [362]. Next, putative indel sites were identified and realigned using GATK `RealignerTargetCreator` and `IndelRealigner` [442].

5.3.2 Variant discovery pipeline

Using analysis-ready reads from the previous step, GATK `HaplotypeCaller` was used to call per-sample SNPs and indels to make GVCF files. Next, a GenomicsDB was created using `GenomicsDBImport` and samples were joint-called using `GenotypeGVCFs`, creating a multi-sample VCF [442].

For analyses involving ancient data, indels and transition polymorphisms were excluded in order to mitigate the effects of DNA damage on variant discovery. Detection of indels in ancient samples (which have very short read lengths) is unreliable, and transition polymorphisms are artificially elevated in samples that have experienced DNA damage. Metadata associated with ancient samples are presented in Table 5.1. DNA extractions for ancient samples were carried out as part of ongoing work by the Palaeogenomics & Bio-Archaeology Research Network (PalaeoBARN) at Oxford University following strict protocols for aDNA.

Table 5.1: Metadata associated with ancient samples

Sample ID	Site	Country	Date range
OL1011	Maresha	Israel	C3rd-1st BC
OL1014	Lyminge	England	C8th-12th AD
OL1231	Vienna Csokorgasse	Austria	C8th AD
OL1385	Buda Castle, Budapest, Telaki Palace	Hungary	C16th-17th AD
OL1391	Buda Castle, Budapest, Telaki Palace	Hungary	C16th-17th AD
OL1401	Hódmezővásárhely-Gorzsa	Hungary	C11th-14th AD
OL1588	Cheboksary City	Russia	C16th-19th AD
OL1677	Ostolopovskoe settlement	Russia	C10th-12th AD
OL1682	Cille Pheadair	Scotland	C10th-11th AD
OL1771	Lyminge	England	C8th-9th AD
OL2267	Naderi Tepe	Iran	Medieval
OL2272	Naderi Tepe	Iran	Medieval

5.3.3 Phasing TLR loci

To facilitate accurate phasing of each TLR locus, genomic regions 500 kbp either side of the locus were extracted from the multi-sample VCF using `tabix` [360]. Plink was then used to generate bed/bim/fam files for each locus [482]. Next, samples with entirely missing data across the specified region were removed using `vcftools --remove-indv` [131]. To phase TLR loci, SHAPEIT (v2.r904) was used with bed/bim/fam files generated in the previous step with a genetic map of 1 Mbp = 1 cM [139]. Output haps files were converted back to VCF with `shapeit --convert`.

5.3.4 Phylogenetic reconstruction

Genomic regions corresponding to the TLR locus and 10 kbp upstream/downstream were extracted from phased VCFs. VCFs were then normalised using `bcftools norm` and consensus fasta sequences for each haplotype were generated using `bcftools consensus` [132]. Consensus fasta sequences were then aligned using `mafft`. Unique sequences were used for phylogenetic reconstruction using RAxML with 100 rapid bootstrap searches and 20 maximum likelihood searches [554]. Trees were visualised and annotated using Interactive Tree of Life (iTOL; [354]).

5.3.5 Variant identification

To identify the types of variant discovered in TLR sequence, `VEP` was used [399]. Since the nature of a genetic variant can vary according to the splice isoform, `VEP` was run for each known splice isoform for each TLR.

5.3.6 Neutrality and diversity statistics

Neutrality (Tajima's D; [567]) and diversity statistics (Nei's Pi; [426]) were calculated for each TLR using PopGenome in R [468].

5.3.7 Genome-wide Tajima's D comparison

Coordinates for all chicken genes that are designated as 1:1 orthologs of human genes (i.e., are likely to have been correctly annotated) were obtained from BioMart (Ensembl). Excluding mitochondrial genes and genes on microchromosomes, the total number of genes included in the analysis was $N = 13373$. For each gene in the ortholog gene set, and the Toll-like receptors, Tajima's D was calculated using `vcftools` [131]. Loci with insufficient variation for Tajima's D to be calculated were excluded (leaving $N = 12781$ genes in the final dataset). A density function of Tajima's D values was plotted using `geom_density` in ggplot2 in R (v. 4.0.2). To compute data centiles, an empirical cumulative distribution function was calculated in R using the `ecdf(x)` function.

5.3.8 Positive selection analyses

The Hyphy suite of programs was used for positive selection analysis. Several site-based models were used: Single-Likelihood Ancestor Counting (SLAC; [322]; Fast, Unconstrained Bayesian AppRoximation (FUBAR; [419]); Fixed Effects Likelihood (FEL; [322]); Mixed Effects Model of Evolution (MEME; [420]). All analyses were conducted using the command line version of Hyphy (v. 2.5.31), using alignments containing only full-length sequences with no internal stop codons.

5.4 Results

5.4.1 Dataset description

To obtain a representative sample of chicken and junglefowl genetic diversity, an extensive genomic dataset was curated. In total, 1660 genomes were included in the analysis (Figure 5.1), comprising 13 green junglefowl (*Gallus varius*), three grey junglefowl (*G. sonneratii*), five Ceylon junglefowl (*G. lafayetii*), 253 red junglefowl with representatives from all five subspecies (*G. gallus bankiva*, *G. gallus gallus*, *G. gallus jabouillei*, *G. gallus murghi*, and *G. gallus spadiceus*), and 1382 domestic chickens (*G. gallus domesticus*). Of the 1382 domestic chickens, 195 were commercial chickens (broilers and layers), 209 were European fancy breeds, and 978 were feral/backyard/local breeds from Africa ($N = 11$), East Asia ($N = 540$), North America, ($N = 76$), Polynesia ($N = 24$), South America ($N = 5$), South Asia ($N = 135$), and West Asia ($N = 187$). Additional sample origin information and accession numbers are given in Appendix Table D.1. Raw reads from each sample were re-aligned to the most recent chicken reference genome (Galgal6) and processed for variant discovery. Mean depth of coverage was 11.3 ± 11.2 (Figure 5.2).

5.4.2 Diversity in chicken Toll-like receptors

Indices of diversity were calculated for each Toll-like receptor coding sequence, initially considering all chicken and junglefowl species as a single population. Nei's pi was highest in *TLR1A/1B* (0.00585 and 0.00633, respectively) and lowest in *TLR2A* (0.00092; Table 5.2). Although the present study considered a large number of domestic chickens and also included junglefowl, the estimates of pi were remarkably similar to a study that investigated *TLR3*, *TLR4*, *TLR5* and *TLR7* diversity in European fancy breeds [565] (Table 5.3). Estimates of haplotype diversity indicate that chicken Toll-like receptors are highly diverse, ranging from 0.84 (*TLR5*) to 0.99 (*TLR1A* and *TLR1B*; Table 5.2). Once again, estimates of haplotype diversity were similar to

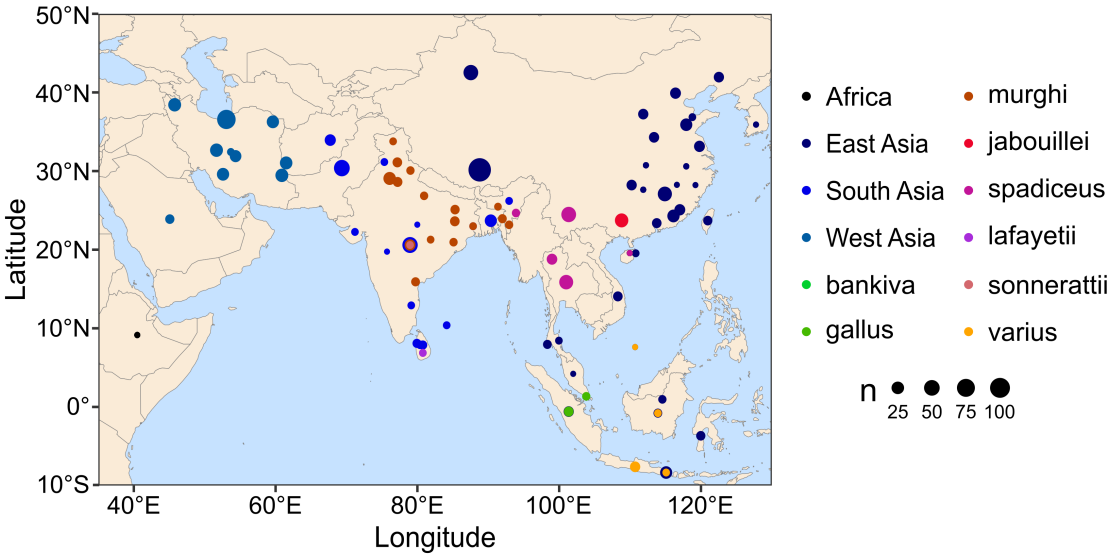


Figure 5.1: Map showing sampling locations of feral/backyard chickens and related wild junglefowl species across Asia and Africa. Where available, accurate GPS coordinates of sampling sites were plotted, but when only country- or province-level data were available, the central coordinates of the country or province were used. These samples are in addition to several hundred commercial birds, European fancy breeds, North and South American breeds etc. that were also analysed but for which geographic information has little meaning due to recent translocation (see Appendix Table D.1 for full information).

those in [565] (Table 5.4).

Table 5.2: Diversity and neutrality statistics for chicken TLRs. Tajima's D was calculated using chickens only, while diversity statistics were calculated using the entire dataset of junglefowl and chickens.

	Tajima's D	Seg. sites	Hap. diversity	π	CDS length	Scaled π
TLR1A	-0.758729	188	0.99425	14.38265	2457	0.00585
TLR1B	-0.509395	168	0.99392	12.40258	1959	0.00633
TLR2A	-1.9519	101	0.86278	2.64556	2868	0.00092
TLR2B	-1.92845	73	0.90016	3.54005	2346	0.00151
TLR3	-0.869335	132	0.96556	9.67782	2691	0.00360
TLR4	-0.90962	99	0.97614	7.40232	2532	0.00292
TLR5	-1.31981	103	0.84060	4.02985	2586	0.00156
TLR7	-1.24649	107	0.90036	3.96550	3195	0.00124
TLR15	-1.15919	107	0.94535	6.92067	2777	0.00249
TLR21	-1.457	187	0.94151	5.83257	3463	0.00168

Table 5.3: Comparison of diversity statistics between studies.

	This study	Świderská *et al.* (2018)	Downing *et al.* (2010)	Levy *et al.* (2020)
TLR1A	0.00585	N/A	0.00168	N/A
TLR1B	0.00633	N/A	0.00320	N/A
TLR2A	0.00092	N/A	0.00070	N/A
TLR2B	0.00151	N/A	0.00033	N/A
TLR3	0.00360	0.00302	0.00417	N/A
TLR4	0.00292	0.00258	0.00261	0.00061
TLR5	0.00156	0.00111	0.00262	0.00148
TLR7	0.00124	0.00106	0.00199	0.00010
TLR15	0.00249	N/A	0.00189	N/A
TLR21	0.00168	N/A	N/A	N/A

Świderská et al. (2018), [565]; Downing et al. (2010), [151]; Levy et al. (2020), [358]

Table 5.4: Comparison of haplotype diversity between studies.

	Świderská *et al.* (2018)	This study
TLR3	0.903	0.966
TLR4	0.947	0.976
TLR5	0.854	0.841
TLR7	0.849	0.900

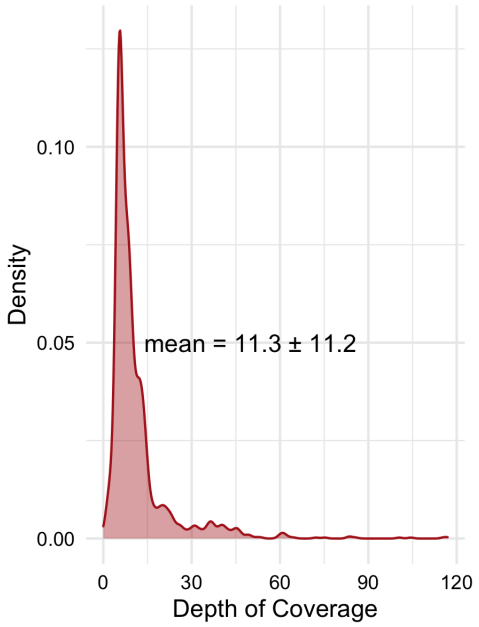


Figure 5.2: Depth of coverage for chicken and junglefowl genomes.

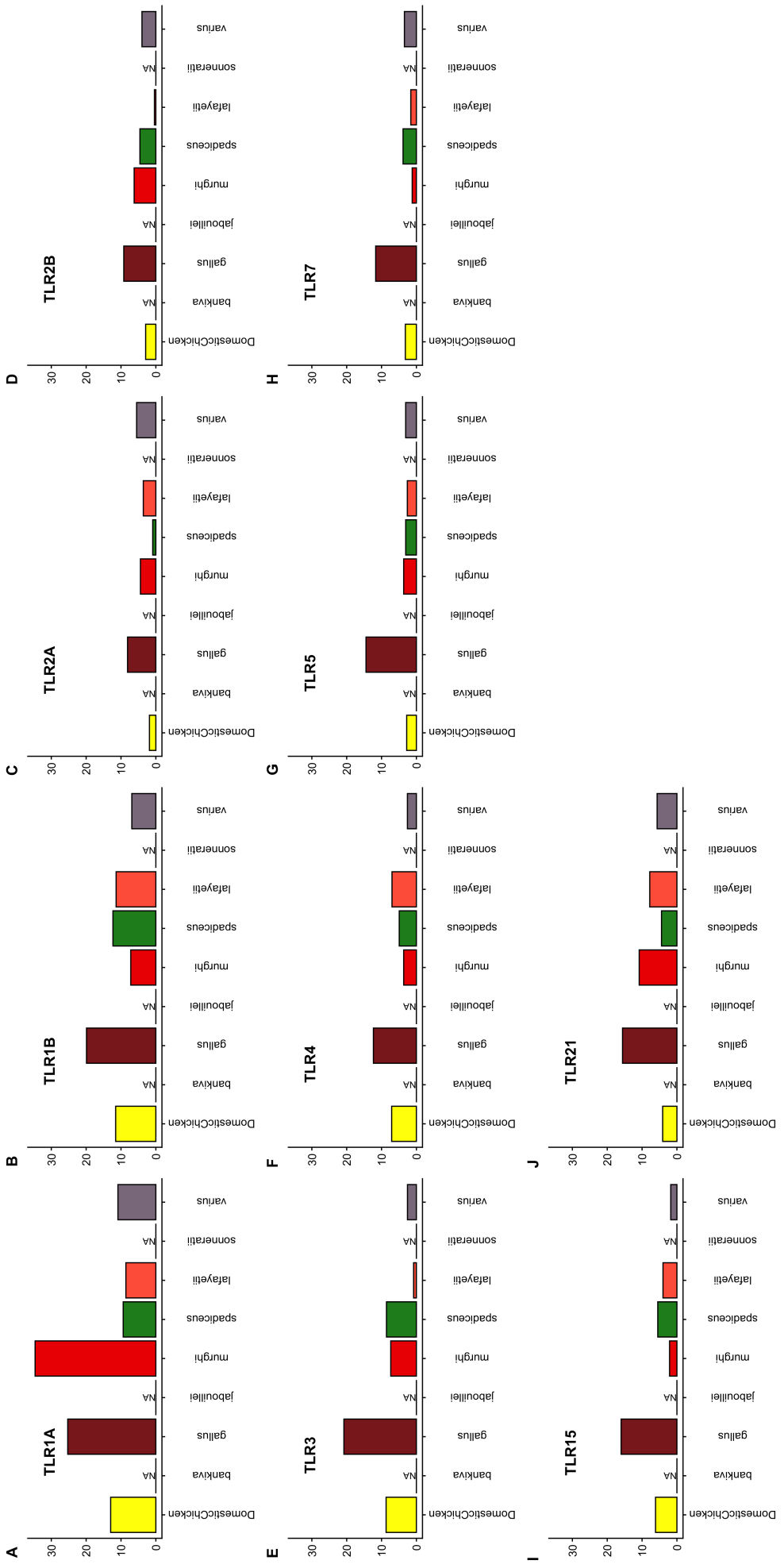


Figure 5.3: Within-population nucleotide diversity of chicken TLRs. Average number of nucleotide differences between pairs of TLR sequences (y-axis) in chickens and related junglefowl. NA: populations with insufficient numbers of individuals or sufficient diversity among sequences to be analysed.

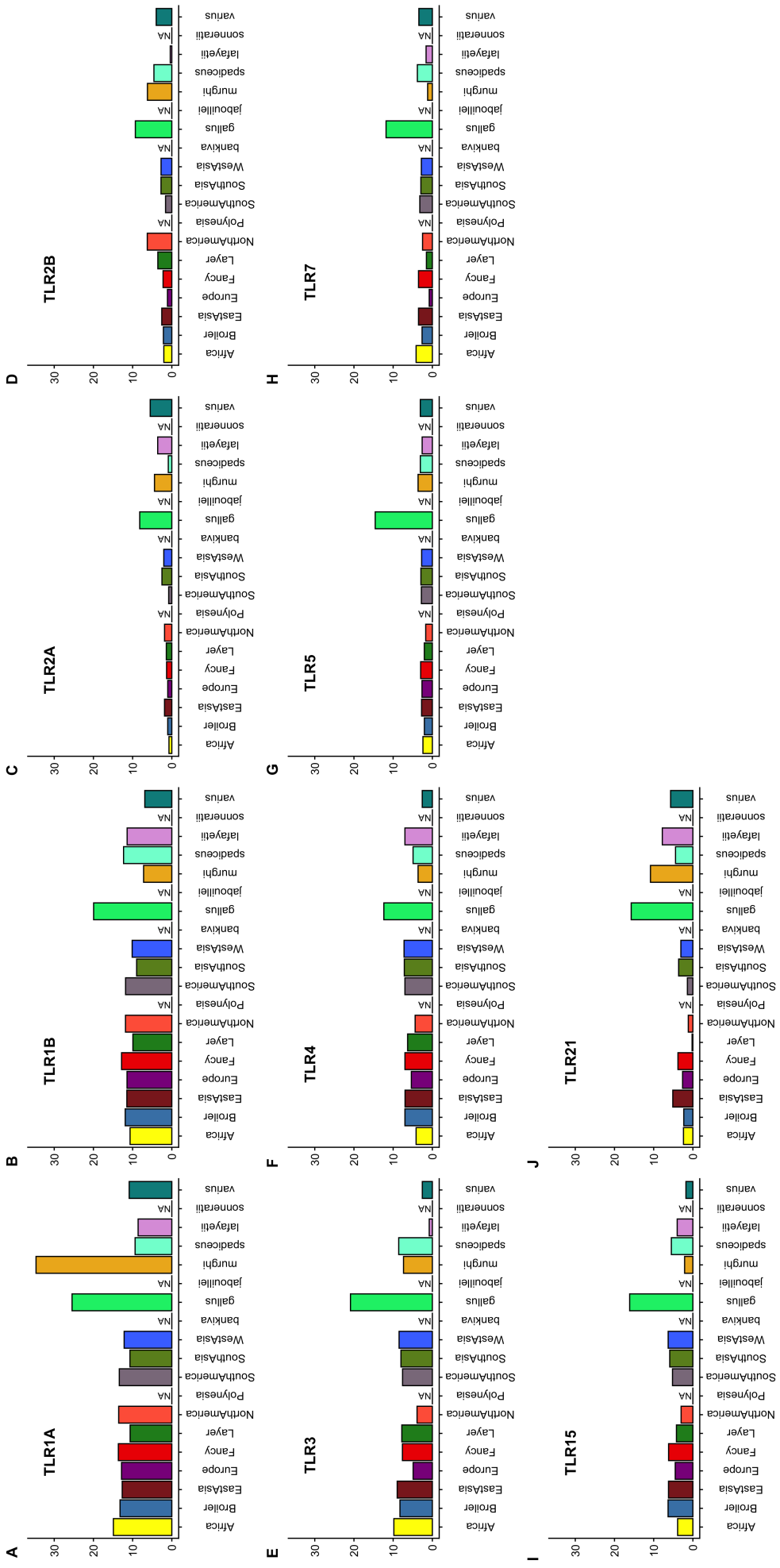


Figure 5.4: Within-population nucleotide diversity of chicken TLRs, stratified by chicken subpopulations. Average number of nucleotide differences between pairs of TLR sequences (y-axis) in domestic chicken subpopulations and related junglefowl. NA: populations with insufficient numbers of individuals or insufficient diversity among sequences to be analysed.

To investigate diversity within different species/subspecies, nucleotide diversity was calculated for domestic chickens and each junglefowl species/subspecies (Figure 5.3). Domestic chickens displayed lowest nucleotide diversity in *TLR2A* and *TLR2B*. The red junglefowl subspecies *Gallus gallus gallus* had the greatest within-population diversity at all but one locus *TLR1A* where *G. g. murghi* was greatest. For most loci, domestic chicken TLR diversity was comparable to *G. g. spadiceus* (the red junglefowl subspecies from which chickens were domesticated), indicating that domestic chickens did not experience a significant genetic bottleneck during domestication, or (more likely) chickens have repeatedly hybridised with their wild counterparts post-domestication which has increased diversity.

Further stratifying domestic chicken populations indicates that within-population TLR diversity is relatively consistent for different populations of chickens (Figure 5.4). This is even true for commercial broiler and layer birds, which may be expected to have undergone a diversity bottleneck during selection for production traits. However, it should be stressed that the broiler and layer populations comprise several independently derived populations of commercial birds from around the world, and further stratifying according to commercial origin may reveal a depression in diversity within each population. Domestic chicken diversity was lowest at the *TLR2A/2B* locus and greatest at the *TLR1A/1B* locus, although *G. g. gallus* diversity was still greater than any domestic chicken subpopulation at every locus.

As anticipated, between-population diversity measurements indicated that domestic chicken and red junglefowl populations are more self-similar than they are to other junglefowl species (Figures 5.5 and 5.6). As the outgroup to the rest of junglefowl, the green junglefowl (*G. varius*) was generally the most distinct population, with the grey (*G. sonneratii*) or Ceylon junglefowl (*G. lafayetii*) being more closely related to chickens and junglefowl. Of the red junglefowl subspecies, *G. g. gallus* was generally the most distinct from domestic chickens, but it was variable which of *G. gallus murghi* or *G. gallus spadiceus* was the most closely related to domestic chickens. This is presumably because chickens are known to have hybridised extensively with local

red junglefowl subspecies and TLRs will lie in different blocks of inheritance. Within domestic chickens, subpopulations were uniformly self-similar, indicating that there is limited genetic differentiation between different groups of domestic chickens.

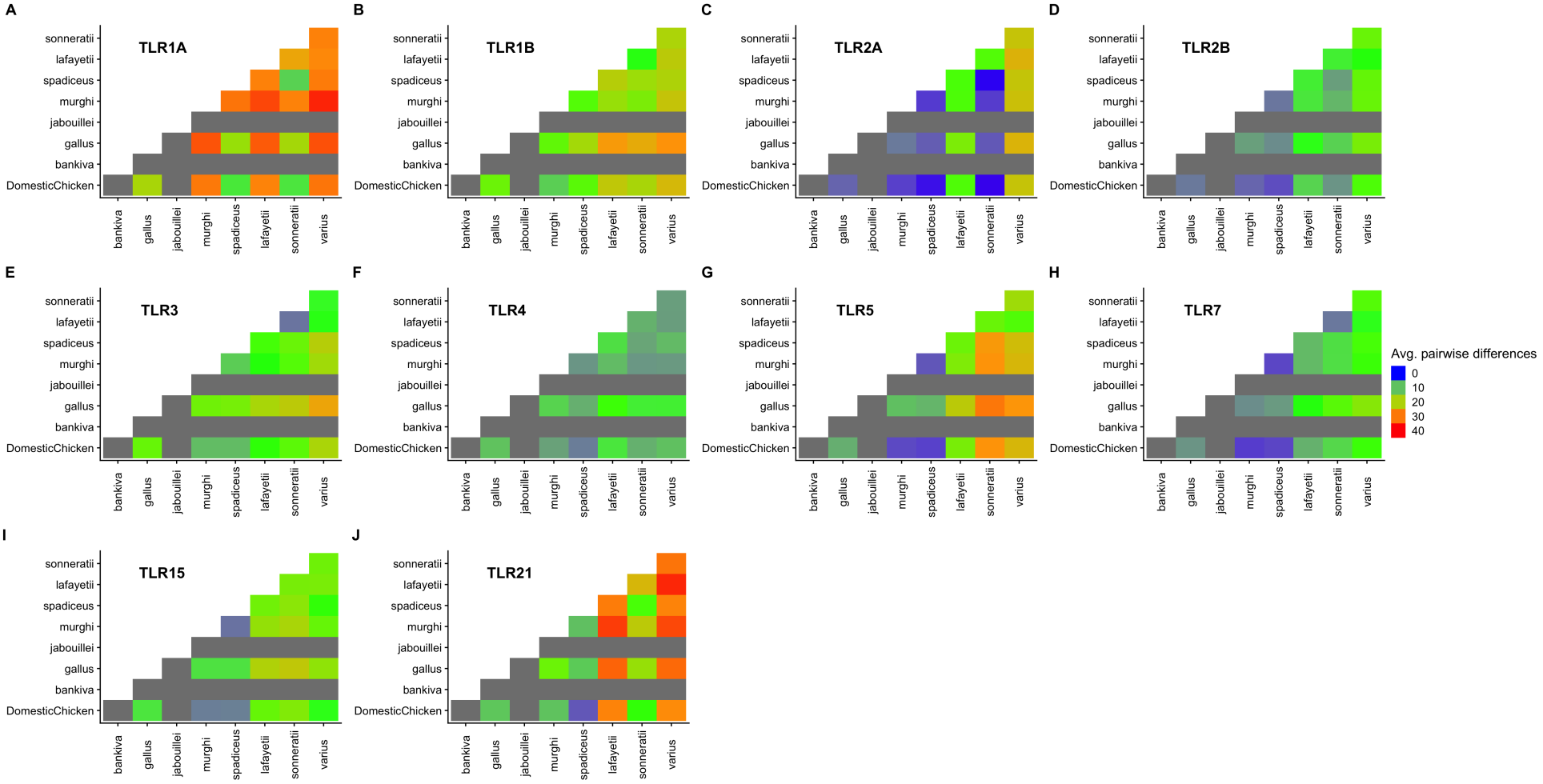


Figure 5.5: Between-population nucleotide diversity of chicken TLRs. Heatmaps of nucleotide diversity indicate the average number of nucleotide differences between pairs of individuals in different populations. Grey cells indicate populations with insufficient individuals for comparisons to be made.

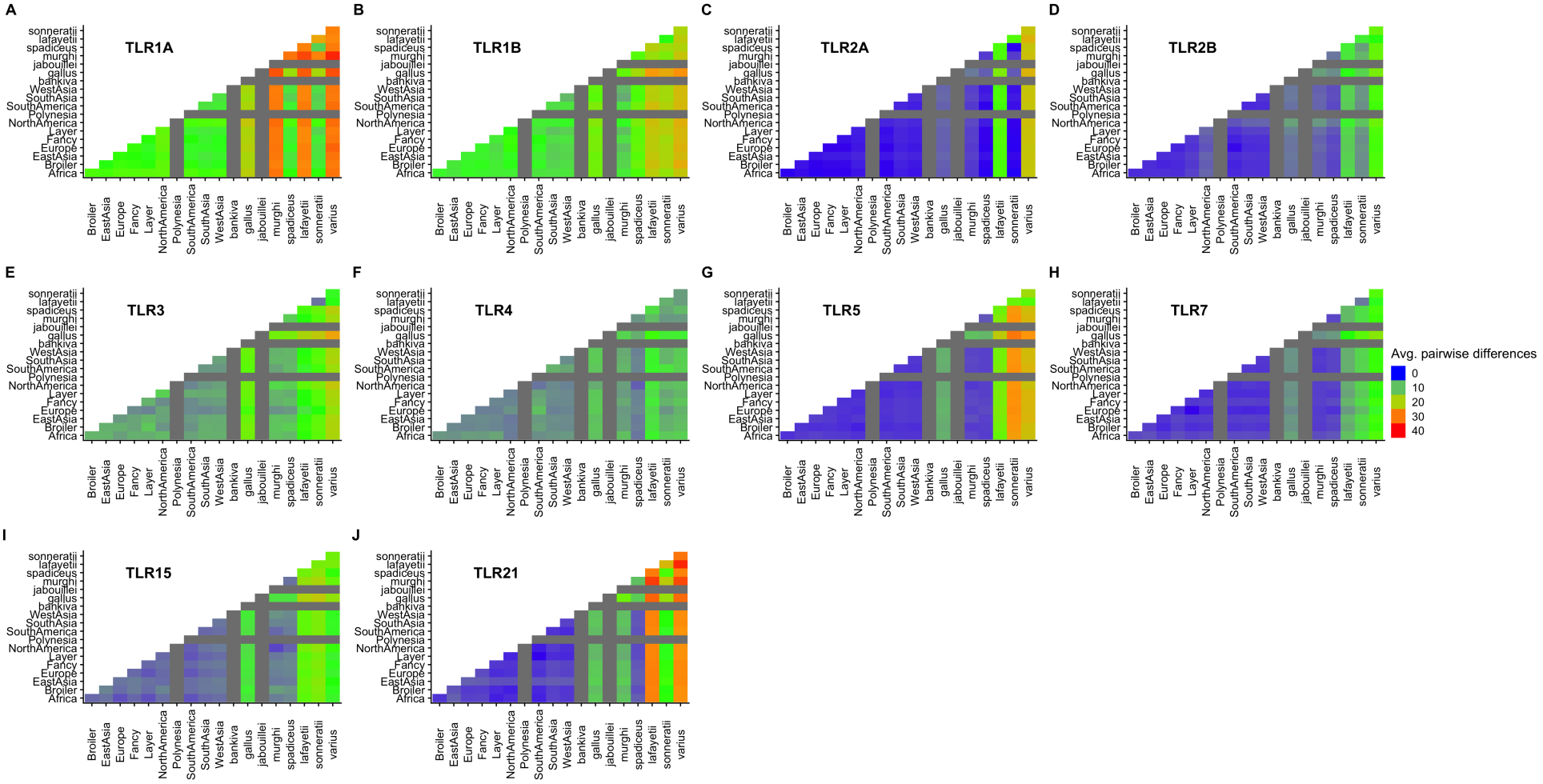


Figure 5.6: Between-population nucleotide diversity of chicken TLRs including chicken subpopulations. Heatmaps of nucleotide diversity indicate the average number of nucleotide differences between pairs of individuals in different populations. Grey cells indicate populations with insufficient individuals for comparisons to be made.

5.4.3 Neutrality and selective sweep analysis

As a measure of neutrality, Tajima's D was consistently negative in chicken Toll-like receptors (Table 5.2). In order to contextualise Tajima's D in the Toll-like receptors, Tajima's D was calculated for every gene in the chicken genome with predicted 1:1 orthology to human genes (i.e, genes which are likely to have been correctly annotated, Figure 5.7). Chicken TLRs were invariably skewed towards the lowest Tajima's D values, and all fell under the 25th percentile. *TLR3* and *TLR4* were under the 10th percentile; *TLR5*, *TLR7* and *TLR15* were under the 5th percentile; and *TLR2A*, *TLR2B* and *TLR21* were under the 1st percentile. These comparisons demonstrate that *TLR2A*, *TLR2B* and *TLR21* have a significantly more negative Tajima's D than would be expected for a coding gene in the chicken genome, possibly due to a recent selective sweep.

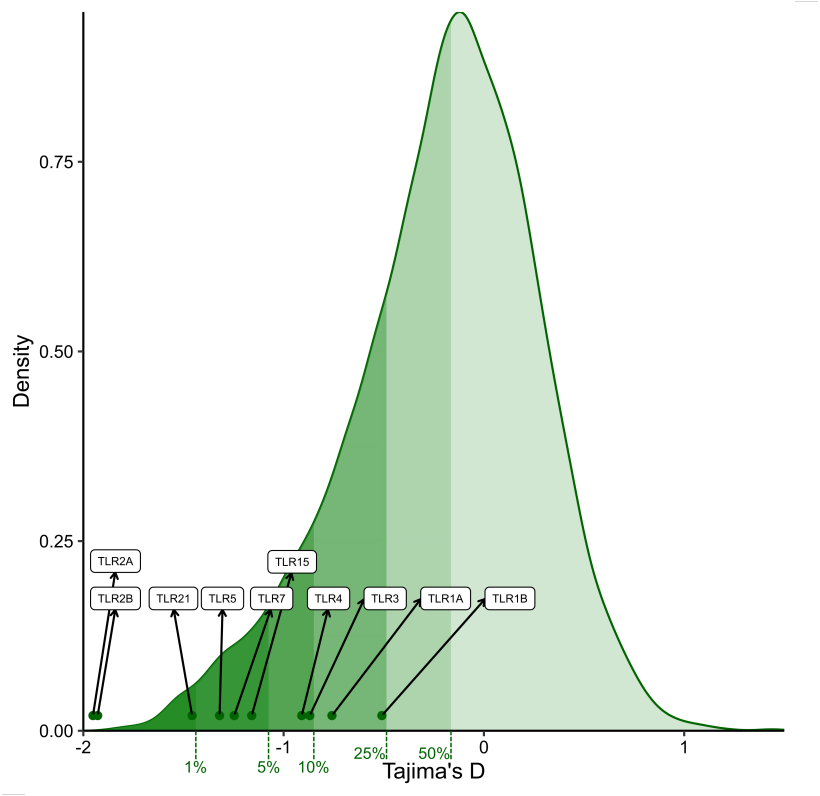


Figure 5.7: Values of Tajima's D for TLRs compared to the rest of the chicken genome. Tajima's D was calculated for each of the chicken TLR genes and are plotted on a cumulative distribution function (CDF) of all chicken genes that are 1:1 orthologs. Based on the CDF, the 50th, 25th, 10th, 5th and 1st centiles are indicated on the x-axis.

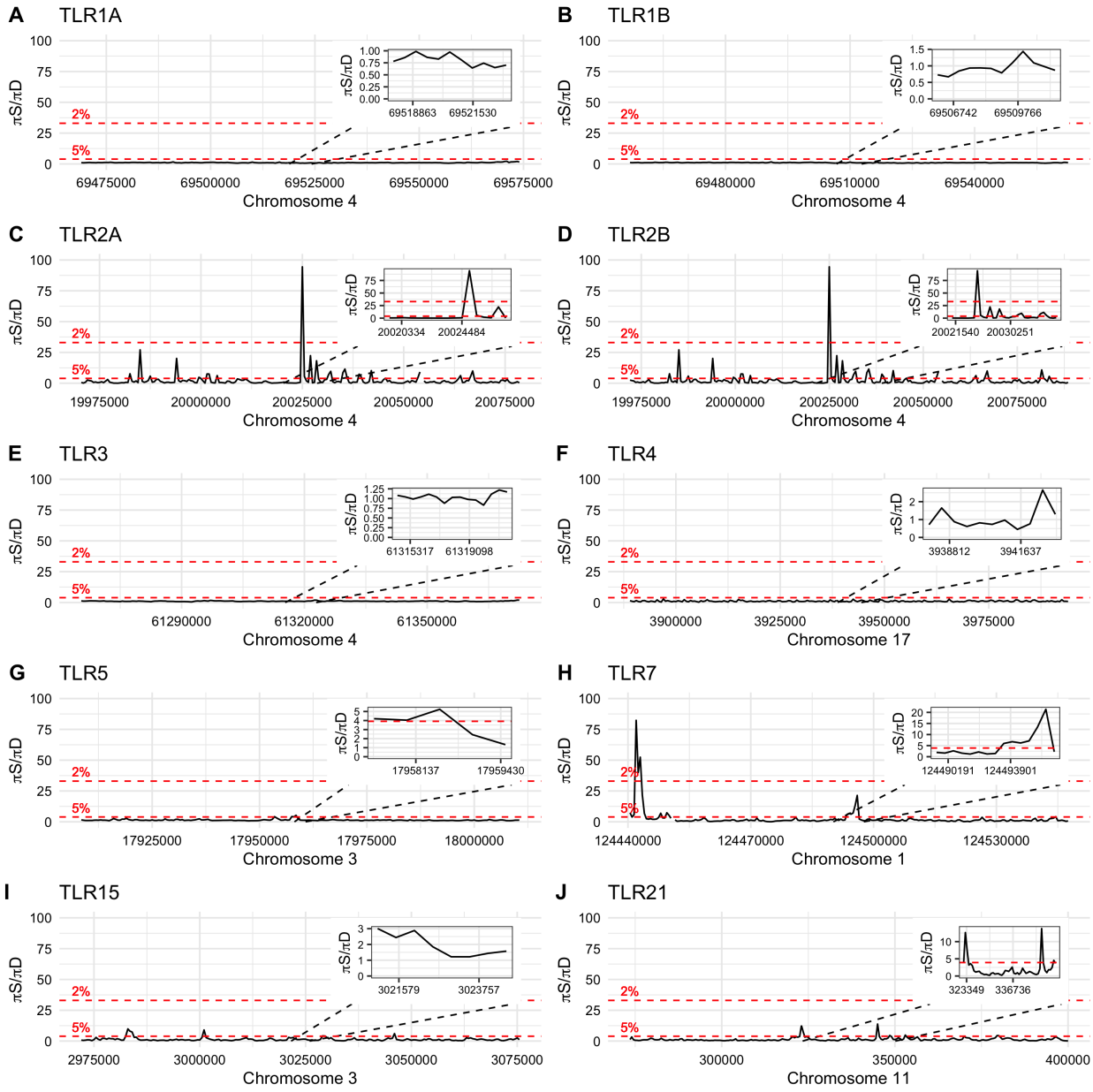


Figure 5.8: Selective sweep analysis of chicken TLRs. For 100 kb regions surrounding TLR loci, Nei's π was calculated with a sliding window analysis of 1000 bp windows and 500 bp step size. This analysis was done separately for *Gallus gallus spadiceus* (the red junglefowl progenitor to domestic chickens) and domestic chickens and the ratio of π (π_S/π_D) was taken. Higher π_S/π_D indicates reduction in diversity in domestic chickens relative to *Gallus gallus spadiceus*, which could be indicative of a selective sweep. Inset figure is the π_S/π_D over the TLR genomic region itself (note different scales in inset figures). Red dashed lines indicate π_S/π_D in the top 5% and 2%, as calculated using an empirical cumulative distribution of chromosomes 1-3.

The reduced nucleotide diversity (described in the previous section) and a negative Tajima's D for the *TLR2A/2B* locus is suggestive of a selective sweep. To more

formally test for evidence of a selective sweep in the *TLR2A/2B* locus and across all other TLRs, a sliding window pi analysis was conducted with 1000 bp windows and 500 bp step size for 100 kb surrounding each TLR locus (Figure 5.8). The sliding window analysis was conducted twice: once for the domestic chicken population (π_D), and once for the *Gallus gallus spadiceus* population (π_S; the junglefowl subspecies from which domestic chickens were derived). Then, a ratio was calculated between each window (π_S/π_D). Low ratio values indicate that ancestral diversity levels have been maintained or increased in chickens, while high ratio values indicate that the chicken is much less diverse than its ancestor, which could be evidence of a selective sweep. *TLR2A* and *TLR2B* showed significant π_S/π_D deviations compared to the rest of the genome, with a peak in the 1.5^{th} centile. There was also evidence of increased π_S/π_D in *TLR5*, *TLR7* and *TLR21* which could also be evidence of a selective sweep at these loci.

5.4.4 Positive selection analysis

To further investigate whether chicken/junglefowl Toll-like receptors have experienced positive selection, multiple sequence alignments were analysed using the Hyphy package of programs, implementing site-based models of selection analysis: FEL [322], FUBAR [419] MEME [420] and SLAC [322]. In total, 422 codons were found to have been positively selected across all TLRs, ranging from 18 (*TLR2A*) to 60 (*TLR1A*) sites per TLR (Figure 5.9). There was generally high concordance between the different methods, and 326/422 (77.3%) sites were supported by three or four of the methods. Further details of each model's results can be found in Appendix Tables D.2-D.5.

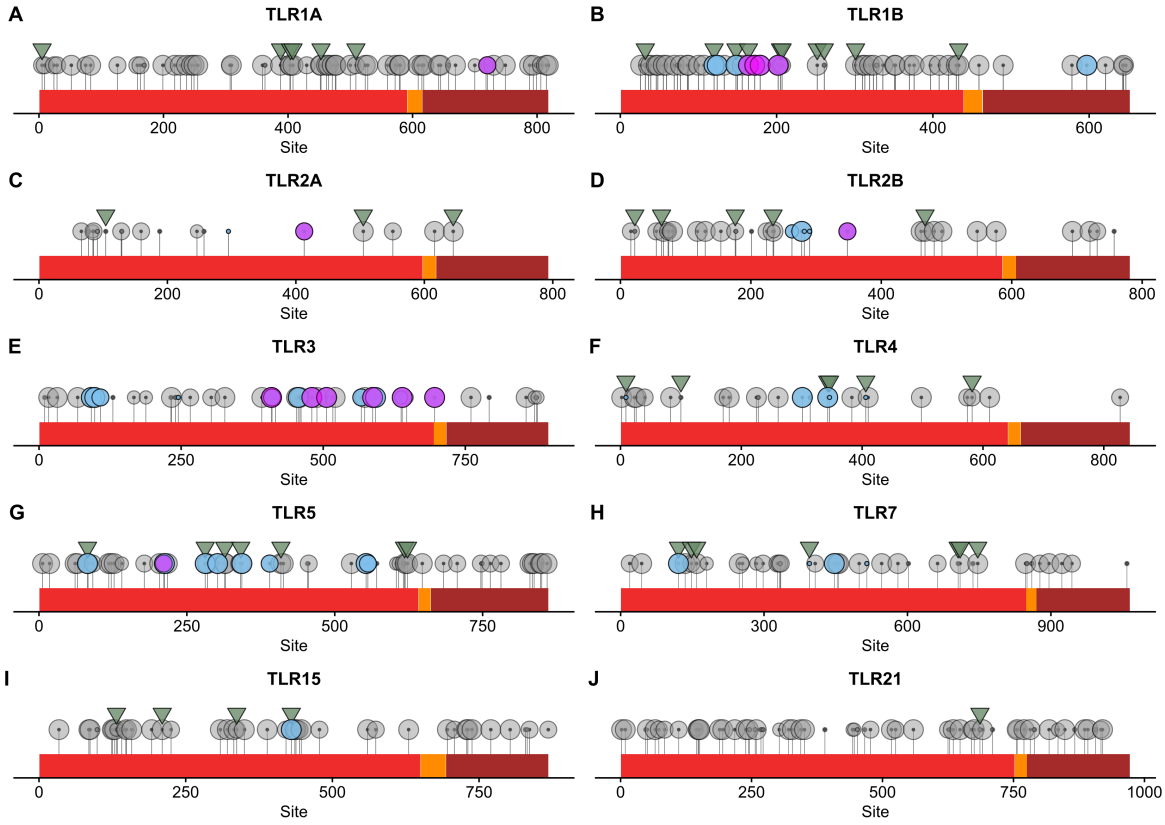


Figure 5.9: Positions of positively selected sites in chicken Toll-like receptors. TLR gene structure is shown in bright red (extracellular domain), orange (transmembrane domain) and dark red (TIR domain). Points show the positions of positively selected sites and the size of the point refers to how many independent methods identified the site as being positively selected (largest = 4; smallest = 1). Points coloured grey are not proximal to any sites of known function. Blue points are <2 amino acids from a site of known ligand binding or dimerization function, while pink points are known functional sites. Green triangles denote positively selected sites that have previously been identified in other organisms. Full results can be found in Appendix Tables D.2-D.5.

Of the 422 sites, 50 were proximal ($\leqslant$2 amino acids) to sites of known function (ligand binding or dimerization), and of these, 14 were located at exact locations of sites of known function. These sites comprised lipopeptide-binding sites (TLR1B: 164, 172, 179; TLR2B: 348), dsRNA-binding sites (TLR3: 409, 480, 506), a flagellin-binding site (TLR5: 211), and various dimerization sites (TLR1A: 720; TLR1B: 172, 179, 202, 413; TLR3: 506, 587, 639, 647) [100, 273, 288, 377, 386, 409, 440, 489, 659] One site, TLR3 site 409 (dsRNA-binding; [100]), had four amino acid variants (E, G, K and R), indicating extensive diversification and likely changes in dsRNA-binding

specificity. In addition, there were 49 sites that have been identified as positively selected in various taxa including birds [10, 220, 603, 606, 617, 636, 637] and mammals [25, 535, 548], suggesting there is recurrent positive selection at these sites across the vertebrates.

5.4.5 Phylogenetic reconstruction

To determine phylogenetic relationships between individuals, phased TLR sequences (plus 20 kb upstream and downstream) were aligned and trees were built using RAxML [554] (Figure 5.10). All TLRs displayed similarly complex patterns of phylogeny, so one representative TLR (*TLR4*) is shown in Figure 5.10) *Gallus varius*, *G. sonneratii* and *G. lafayetti* formed three monophyletic clades, indicating that gene flow is limited or absent between *Gallus gallus* and other *Gallus* spp. The picture for red junglefowl subspecies is, however, more complex. Each of the subspecies formed one 'main' clade that was monophyletic or almost monophyletic. However, with the exception of *G. g. bankiva* (of which only three individuals were included in the analysis), red junglefowl subspecies were also frequently found in mixed clades with domestic chickens. This indicates that there has been extensive and repeated introgression between red junglefowl subspecies and domestic chickens, which has been documented elsewhere [620]. Evidence of hybridisation was not limited to animals which are known to be sympatric (e.g. , *G. g. murghi* was not exclusively found in mixed clades with domestic chickens from India, and *G. g. spadiceus* was not found exclusively with chickens from Thailand or southern China). This could mean that chickens are frequently translocated, or that the introgression occurred when the ranges of red junglefowl were different compared to the present day.

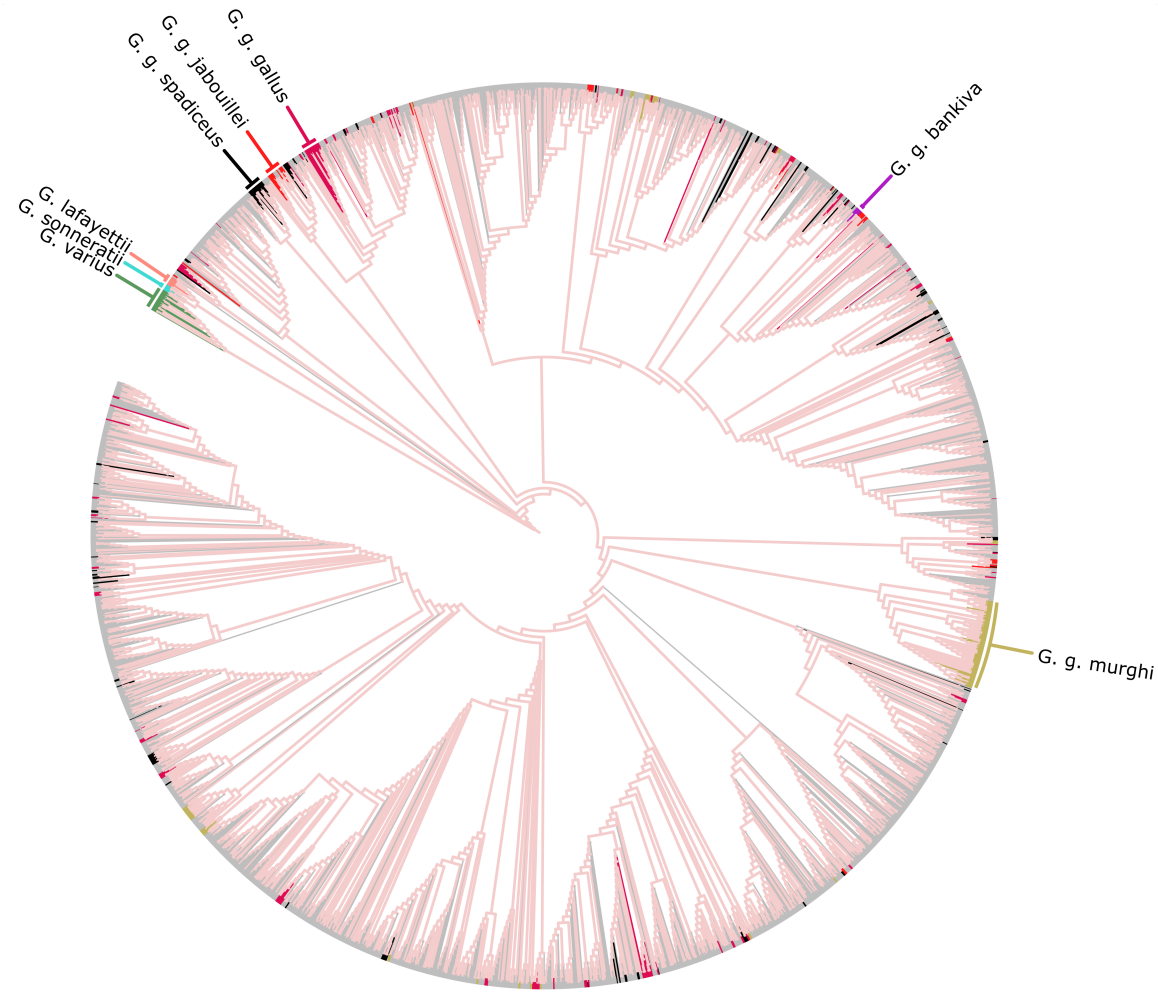


Figure 5.10: Phylogeny of a representative TLR (*TLR4*) in chickens and junglefowl. Internal branches are shaded light pink and domestic chicken tips are shaded light grey. Junglefowl tips are shaded according to species/subspecies and monophyletic clades are highlighted.

Further phylogenies were generated by including ancient samples in the alignments (excluding indels and transition polymorphisms). As with the previous analysis, all TLRs showed broadly the same patterns of phylogenetic relationships, so the phylogeny for *TLR4* is presented (Figure 5.11). As expected, ancient samples were widespread in the tree and tended not to cluster together .There were some exceptions however; two samples from the same site (Naderi Tepe, Medieval Iran) were closely related, and three samples from Hungary (C11th-14th AD), UK (C8th-12th AD) and Israel (C3rd-1st BC) clustered together. The fact that the two Iranian samples from the same location and period in history clustered together is reassuring and suggests

that the effect of DNA damage has been suitably controlled for. Three samples from different geographic locations and separated in time by ∼1000 years clustering together is intriguing, and may suggest that diversity in European chickens was low until at least the Medieval period (more probably until the C19th, when there was an influx of Asian breeds for hobby breeders). The three clustering samples were in a clade predominantly with red junglefowl, while the two ancient Iranian samples were in a clade with mostly Chinese breeds.

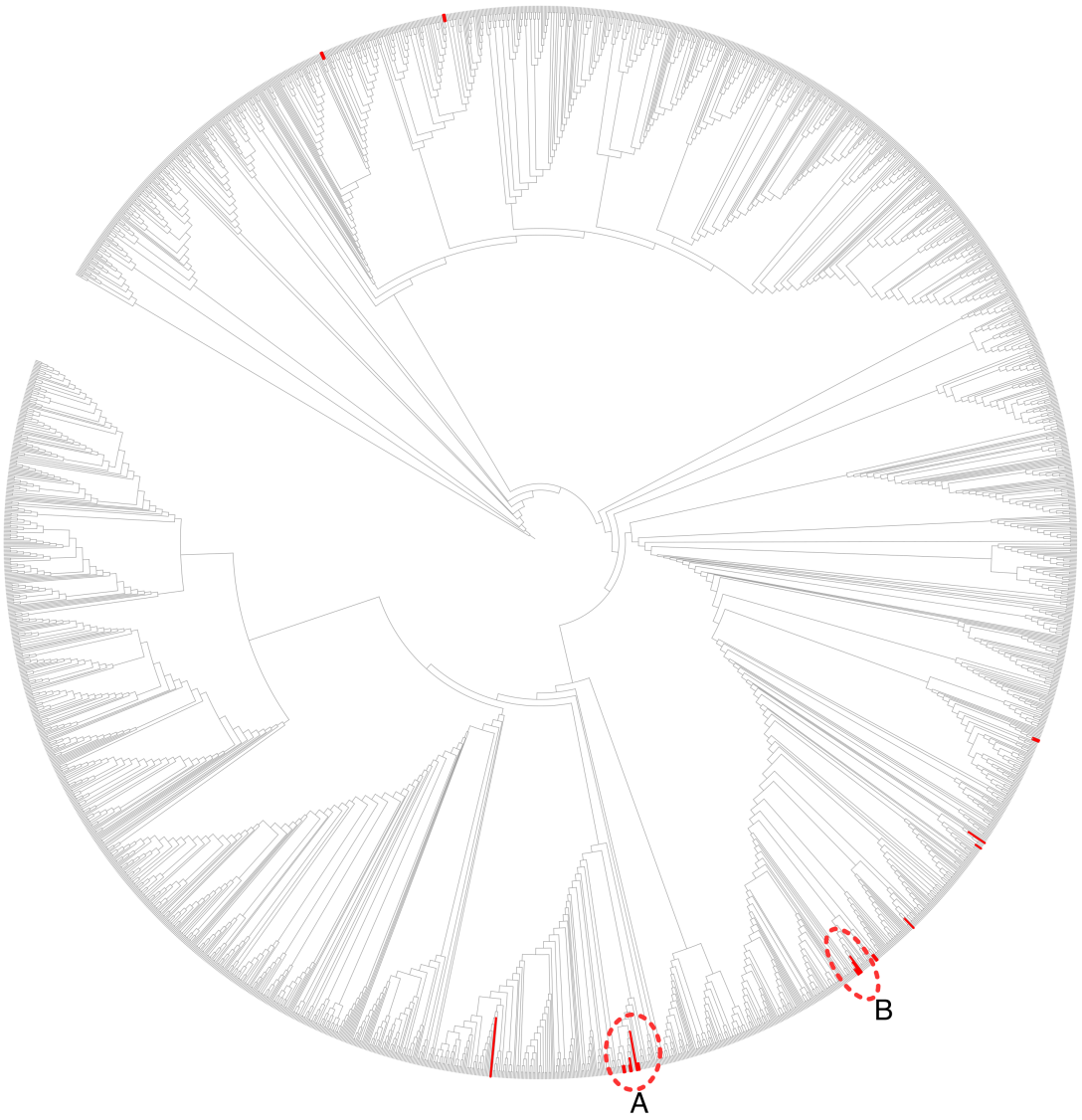


Figure 5.11: Phylogeny of a representative TLR (*TLR4*) in chickens and junglefowl including ancient sequences. Two clades are highlighted. Clade A comprises samples from Hungary (C11th-14th AD), UK (C8th-12th AD) and Israel (C3rd-1st BC) along with predominantly red junglefowl, while clade B comprises two samples from the same site in Medieval Iran along with mostly Chinese breeds of chicken.

5.4.6 Variant analysis

To identify sequence variants in chickens and junglefowl, Variant Call Format (VCF) files were analysed using the Variant Effect Predictor (VEP) program [399]. The analysis was repeated for each of the splice isoforms of the TLR, since, for example, an intronic variant in one isoform may be a coding variant in another isoform (Figure

5.12). In total, there were 1259 coding SNPs, of which 592 were synonymous changes and 657 were non-synonymous changes. There were 3293 intronic variants and 670 variants located in the 5′- and 3′-UTR regions. There were also several variants of high impact, including 13 internal stop codons, 34 frameshift mutations, 36 splice site variants and three lost start codons (Figure 5.12).

TLR1A, *TLR3*, *TLR4* and *TLR5* all exist as single splice forms, and therefore any high impact mutations cannot be escaped by the use of a different splice variant. Almost all of the high impact mutations in these TLRs were detected in single individuals, suggesting they are de novo variants and are likely to be of low significance at the population level (Table 5.5).

TLR2A, *TLR2B*, *TLR7*, *TLR15* and *TLR21* exist as multiple splice forms, but each was found to have high impact mutations in exons that are shared across all splice variants (Figure 5.12), and therefore cannot be escaped by the use of a different splice variant. A stop-gained variant at 4:20026321 affects both *TLR2A* splice forms and is found in almost half of White Plymouth Rock birds (29/62; 46.8%), suggesting that a null *TLR2A* allele is common in this breed of chicken. The variant was also found in all five Ceylon junglefowl (*Gallus lafayetii*) included in the analysis, suggesting that the origins of the stop-gained variant predate the emergence of domestic chickens. Interestingly, the same 29 White Plymouth Rock birds with the stop variant in *TLR2A* also had a stop variant in *TLR2B* (4:20033955). Since the stop variant was in the gene converted region shared by both TLRs, care must be taken in the interpretation of this result. However, since all of the birds with the *TLR2A* variant were determined to be homozygous, it is likely real and the birds are null for both *TLR2A* and *TLR2B*. The frameshift mutation at 4:20033867 affects both *TLR2B* splice forms and is very widespread in chickens and junglefowl, affecting 483 birds in total, including 41 White Leghorn and 46 White Plymouth Rock breeds, 95 Chinese birds and 81 European fancy breed birds. Hence, it appears that the *TLR2* family is a widespread pseudogene in chickens and junglefowl.

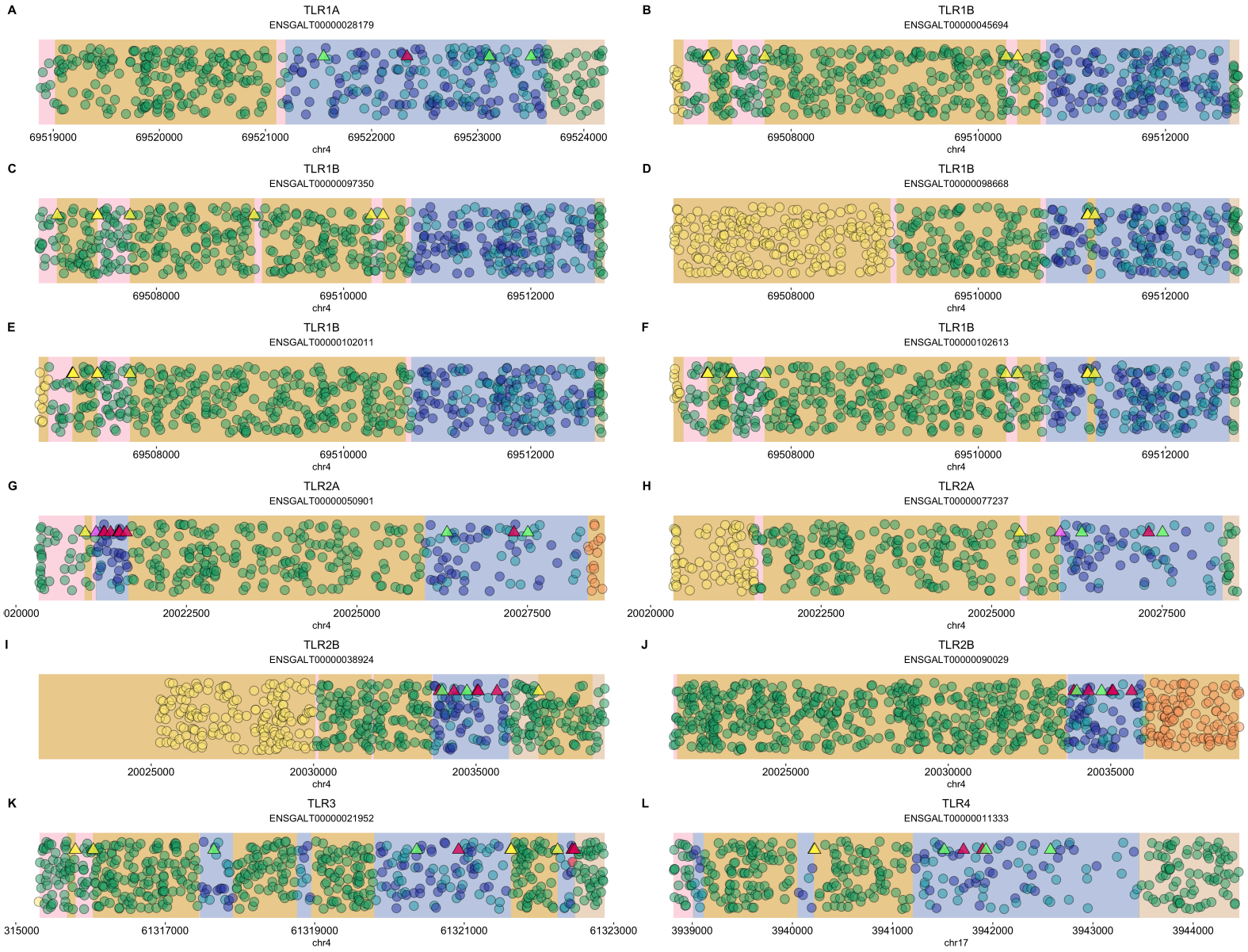


Figure 5.12: *Continued on next page.*

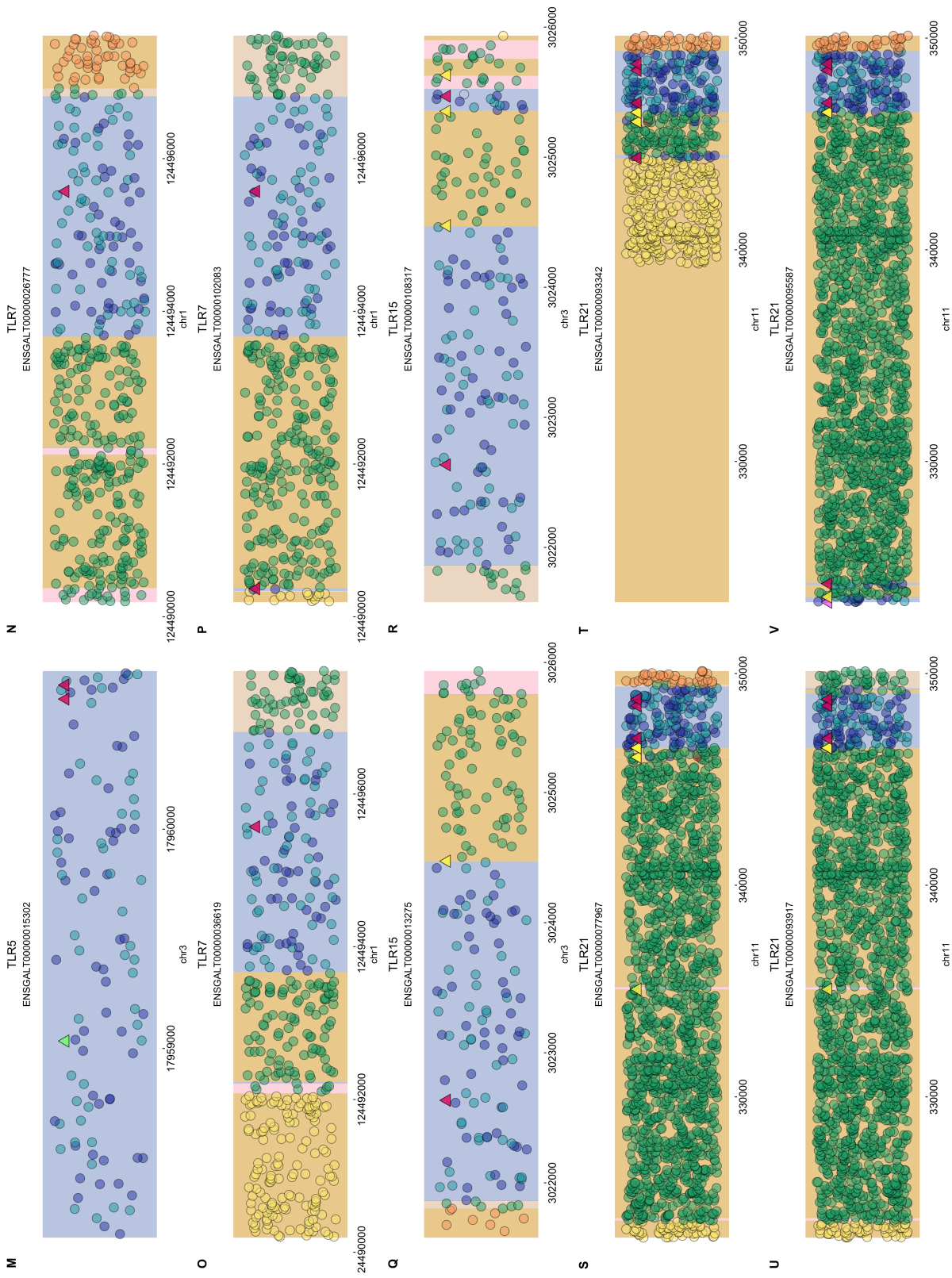


Figure 5.12: *Continued on next page.*

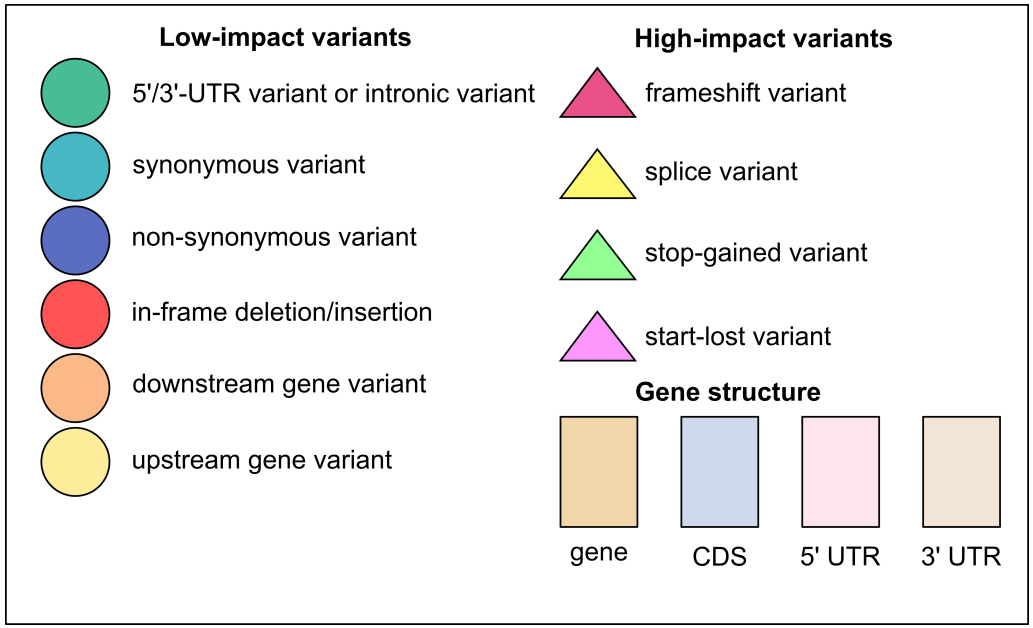


Figure 5.12: Positions of variants discovered in chicken and junglefowl TLR sequences. Gene structure is plotted in coloured boxes with variants overlaid. For each TLR, all known transcript isoforms are represented since variant impact may be different for each isoform. Low-impact variants are presented as filled circles while high-impact variants are filled triangles. Low-impact variants are plotted with jitter to reduce overplotting and aid visualisation, and therefore y-axis deviation has no meaning in the plots.

Gene structure has undergone significant variation across chickens and junglefowl. Almost all of the TLRs exhibited some degree of polymorphism in splice sites, which could impact the splice isoforms that are transcribed. Furthermore, there appears to be use of different splice isoforms between chickens and junglefowl. Almost all (*G. varius*, 10/13; *G. lafayetii* 5/5; *G. sonneratii* 3/3) of the congeneric junglefowl species exhibit a frameshift mutation at 3:3025468 which would preclude the use of the ENSGALT00000108317 transcript of *TLR15*, suggesting that this transcript has been secondarily co-opted by chickens and is not used in junglefowl.

Table 5.5: Distribution of high-impact TLR variants in chickens and related junglefowl

TLR	Variant coordinates	Transcript accession	Variant type	CDS position	Variant distribution
TLR1A	4:69521548	ENSGALT00000028179	stop_gained	358	China (1)[1]; murghi (1)
TLR1A	4:69522335-69522336	ENSGALT00000028179	frameshift	1145-1146	fancy (1)
TLR1A	4:69523111	ENSGALT00000028179	stop_gained	1921	China (1)
TLR1A	4:69523503	ENSGALT00000028179	stop_gained	2313	China (1)
TLR2A	4:20021175	ENSGALT00000050901	start_lost	2	Indonesia (1)
TLR2A	4:20021288	ENSGALT00000050901	frameshift	115	gallus (3)
TLR2A	4:20021298-20021307	ENSGALT00000050901	frameshift	125-134	varius (4)
TLR2A	4:20021388-20021395	ENSGALT00000050901	frameshift	215-222	sonneratii (2)
TLR2A	4:20021503-20021504	ENSGALT00000050901	frameshift	330-331	China (2); commercial (1); spadiceus (1); White Plymouth Rock (1)
TLR2A	4:20021505	ENSGALT00000050901	frameshift	332	bankiva (3); China (38); commercial (6); fancy (61); gallus (6); India (1); Iran (1); Japan (3); Java (1); lafayettii (4); Malaysia (1); North America (2); Saudi Arabia (1); sonneratii (2); Sri Lanka (4); varius (7); White Leghorn (32); White Plymouth Rock (26)
TLR2A	4:20021620	ENSGALT00000050901	frameshift	447	gallus (2)
TLR2A	4:20026006	ENSGALT00000077237	start_lost	2	gallus (1)

Continued on next page

[1]Presented in the form A (n) where A is the country/breed/species/subspecies of the bird and n is the number of birds with the variant. For junglefowl, the species or subspecies name is given; for wild/backyard domestic chickens, the country of origin is given; European fancy breeds are combined into the "fancy" category. As the most numerous categories of production birds, White Leghorn and White Plymouth Rock birds are enumerated individually.

TLR	Variant coordinates	Transcript accession	Variant type	CDS position	Variant distribution
TLR2A	4:20026321	ENSGALT00000050901	stop_gained	803	fancy (2); Iran (1); lafayettii (5); White Plymouth Rock (29)
TLR2A	4:20026321	ENSGALT00000077237	stop_gained	317	*as above*[2]
TLR2A	4:20027303	ENSGALT00000050901	frameshift	1785	gallus (1)
TLR2A	4:20027303	ENSGALT00000077237	frameshift	1299	*as above*
TLR2A	4:20027506	ENSGALT00000050901	stop_gained	1988	India (1)
TLR2A	4:20027506	ENSGALT00000077237	stop_gained	1502	*as above*
TLR2B	4:20033867	ENSGALT00000038924	frameshift	196	Afghanistan (2); Africa (3); bankiva (1); China (95); commercial (35); fancy (81); gallus (24); India (2); India (23); Indonesia (7); Iran (61); Japan (4); Java (2); Korea (1); lafayettii (4); Malaysia (1); murghi (3); North America (1); Pakistan (10); Saudi Arabia (4); sonneratii (1); South America (4); spadiceus (11); Sri Lanka (6); Thailand (1); varius (8); Vietnam (1); White Leghorn (41); White Plymouth Rock (46)
TLR2B	4:20033867	ENSGALT00000090029	frameshift	196	*as above*
TLR2B	4:20033869-20033870	ENSGALT00000038924	frameshift	198-199	China (6)
TLR2B	4:20033869-20033870	ENSGALT00000090029	frameshift	198-199	*as above*
TLR2B	4:20033955	ENSGALT00000038924	stop_gained	284	commercial (1); fancy (3); Indonesia (1); lafayettii (4); White Plymouth Rock (29)

Continued on next page

[2]This refers to instances where there is more than one predicted splice isoform and the variant occurs in coding sequences of each of the listed isoforms.

TLR	Variant coordinates	Transcript accession	Variant type	CDS position	Variant distribution
TLR2B	4:20033955	ENSGALT00000090029	stop_gained	284	*as above*
TLR2B	4:20034321	ENSGALT00000038924	frameshift	650	China (2)
TLR2B	4:20034321	ENSGALT00000090029	frameshift	650	*as above*
TLR2B	4:20034722	ENSGALT00000038924	stop_gained	1051	China (2); gallus (8); Indonesia (1); spadiceus (1)
TLR2B	4:20034722	ENSGALT00000090029	stop_gained	1051	*as above*
TLR2B	4:20035037-20035038	ENSGALT00000038924	frameshift	1366-1367	fancy (2)
TLR2B	4:20035037-20035038	ENSGALT00000090029	frameshift	1366-1367	*as above*
TLR2B	4:20035071-20035072	ENSGALT00000038924	frameshift	1400-1401	Malaysia (2)
TLR2B	4:20035071-20035072	ENSGALT00000090029	frameshift	1400-1401	*as above*
TLR2B	4:20035648-20035657	ENSGALT00000038924	frameshift	1977-1986	gallus (1)
TLR2B	4:20035648-20035657	ENSGALT00000090029	frameshift	1977-1986	*as above*
TLR3	4:61317650	ENSGALT00000021952	stop_gained	187	murghi (1)
TLR3	4:61320364	ENSGALT00000021952	stop_gained	1201	jabouillei (1)
TLR3	4:61320925	ENSGALT00000021952	frameshift	1762	murghi (1)
TLR3	4:61322443-61322444	ENSGALT00000021952	frameshift	2656-2657	China (1)
TLR3	4:61322445-61322446	ENSGALT00000021952	frameshift	2658-2659	*as above*
TLR3	4:61322448-61322449	ENSGALT00000021952	frameshift	2661-2662	*as above*
TLR3	4:61322478	ENSGALT00000021952	frameshift	2691	China (1)
TLR4	17:3941515	ENSGALT00000011333	stop_gained	586	murghi (1)
TLR4	17:3941707-3941708	ENSGALT00000011333	frameshift	778-779	China (1)
TLR4	17:3941897-3941898	ENSGALT00000011333	frameshift	968-969	fancy (1)
TLR4	17:3941932	ENSGALT00000011333	stop_gained	1003	India (1); Java (1)
TLR4	17:3942576	ENSGALT00000011333	stop_gained	1647	commercial (1)
TLR5	3:17958188-17958189	ENSGALT00000015302	frameshift	52-53	China (1)
TLR5	3:17959034	ENSGALT00000015302	stop_gained	898	murghi (1)
TLR5	3:17960595	ENSGALT00000015302	frameshift	2459	China (2)
TLR5	3:17960657-17960658	ENSGALT00000015302	frameshift	2521-2522	China (1)

Continued on next page

TLR	Variant coordinates	Transcript accession	Variant type	CDS position	Variant distribution
TLR7	1:124490368	ENSGALT00000102083	frameshift	30	Afghanistan (5); Africa (5); bankiva (3); China (96); commercial (9); fancy (27); gallus (37); India (20); Indonesia (16); Iran (38); jabouillei (9); Japan (1); lafayettii (5); murghi (22); North America (1); Pakistan (9) sonneratii (3); spadiceus (31); Sri Lanka (3); Thailand (5); varius (11); Vietnam (2); White Leghorn (1)
TLR7	1:124495572	ENSGALT00000026777	frameshift	1918	gallus (1)
TLR7	1:124495572	ENSGALT00000036619	frameshift	1918	*as above*
TLR7	1:124495572	ENSGALT00000102083	frameshift	1942	*as above*
TLR15	3:3022636-3022637	ENSGALT00000013275	frameshift	1835-1836	gallus (1)
TLR15	3:3022636-3022637	ENSGALT00000108317	frameshift	2000-2001	*as above*
TLR15	3:3025468-3025469	ENSGALT00000108317	frameshift	57-58	lafayetii (5); sonneratii (3); varius (10)
TLR21	11:323351	ENSGALT00000095587	start_lost	3	varius (2)
TLR21	11:324223-324224	ENSGALT00000095587	frameshift	313-314	bankiva (1); China (21); fancy (7); gallus (3); India (6); Indonesia (12); Iran (21); Pakistan (2); spadiceus (6); Sri Lanka (1); Vietnam (1)
TLR21	11:344336-344337	ENSGALT00000093342	frameshift	8-9	lafayettii (1)
TLR21	11:346936	ENSGALT00000077967	frameshift	465	murghi (2)
TLR21	11:346936	ENSGALT00000093342	frameshift	639	*as above*
TLR21	11:346936	ENSGALT00000093917	frameshift	450	*as above*
TLR21	11:346936	ENSGALT00000095587	frameshift	798	*as above*

Continued on next page

TLR	Variant coordinates	Transcript accession	Variant type	CDS position	Variant distribution
TLR21	11:348479-348480	ENSGALT00000077967	frameshift	2008-2009	lafayetii (5); murghi (1)
TLR21	11:348479-348480	ENSGALT00000093342	frameshift	2182-2183	*as above*
TLR21	11:348479-348480	ENSGALT00000093917	frameshift	1993-1994	*as above*
TLR21	11:348479-348480	ENSGALT00000095587	frameshift	2341-2342	*as above*
TLR21	11:348790	ENSGALT00000077967	frameshift	2319	fancy (1)
TLR21	11:348790	ENSGALT00000093342	frameshift	2493	*as above*
TLR21	11:348790	ENSGALT00000093917	frameshift	2304	*as above*
TLR21	11:348790	ENSGALT00000095587	frameshift	2652	*as above*

5.5 Discussion

Chickens are thought to be rich in genomic diversity, with an average of 5 SNPs/kb [639], six- or seven-fold higher than humans [511]. Moreover, immune genes are frequently found to be hotspots of diversity and selection [47, 323, 484, 542] in genome-wide analyses. Toll-like receptor agonists are also increasingly being used as vaccine adjuvants in the livestock industry [115, 228, 245, 613]. Despite their importance, comparatively little is known about genetic variability in TLRs that may underlie differences in adjuvant efficacy in different breeds, or render certain breeds susceptible or resistant to pathogens. Few studies of livestock animals have investigated TLR diversity, and often only consider either (i) a subset of TLRs, (ii) a portion of the coding sequence, (iii) relatively few breeds, or (iv) relatively few individuals [134, 151, 437, 565]. Large-scale, comparative studies are needed to fully appreciate the pool of genetic diversity available to maintain immunologically healthy livestock. Using 1660 chicken and junglefowl genomes, this study forms the most comprehensive description of Toll-like receptor diversity in a livestock animal.

A rich pool of genetic diversity in domestic chickens is contrary to the expectation that domesticated animals experience significant genetic bottlenecks, firstly during domestication and secondly through intensive artificial selection for desired traits [187, 394, 351, 525]. Chickens were originally thought to have been domesticated multiple times [379, 405]. However, such analyses were based on mitochondrial DNA, which is sensitive to spurious signals of repeated domestication due to the propensity of domesticated animals to hybridise with their wild counterparts [337, 338]. Using whole-genome data from >800 chickens and junglefowl, it is now thought that chickens were domesticated once from the red junglefowl subspecies *Gallus gallus spadiceus*, which has a modern range covering southwestern China, northern Thailand and Myanmar [620]. Initial domestication of the chicken has been followed by repeated admixture with sympatric red junglefowl lineages. For instance, *Gallus gallus murghi* has contributed significant portions of ancestry to chickens in India

(∼17.6%) and Bangladesh (∼22.4%; [620]). It was continued hybridisation with wild junglefowl that led to both the high diversity in domestic chickens and the spurious finding of multiple independent domestications.

The repeated hybridisation of chickens and red junglefowl also means the phylogenies of single loci are complex. At the genomic level, it is possible to resolve the five subspecies of red junglefowl into five clades [620], but TLR phylogeny reveals mixed patterns of inheritance. From the phylogenies constructed at TLR loci, it is possible to resolve the green junglefowl (*G. varius*), Ceylon junglefowl (*G. lafayetti*) and Grey junglefowl (*G. sonneratii*) to three monophyletic clades, separate from domestic chickens and red junglefowl. This is in contrast to a previous report which concluded there has been hybridisation between chickens and grey junglefowl (which has also been confirmed for the yellow skin locus on chromosome 24 [163]), that is evident in *TLR1A* and *TLR2A* because they share a clade in phylogenetic analysis [151]. Despite the fact that the *TLR1A/B* and *TLR2A/B* loci are physically proximal on chromosome 4 (meaning they are likely to be inherited as single blocks), the small sample size and the construction of phylogenies using only TLR sequence (rather than including a genomic region upstream and downstream) renders these conclusions unreliable because of a lack of diversity, particularly for *TLR2A*. Furthermore, using whole-genome analysis, chromosome 4 does not have any predicted gene flow from grey junglefowl to chickens [339, 620].

Indices of genetic diversity obtained in this study support previous observations that the chicken genome is rich in variants [565, 639]. Chicken TLRs were found to be, on average, nine times more diverse than corresponding human TLRs derived from diverse populations [565]. In general, the values of Nei's Pi in the present study are highly concordant with the values derived from the next largest study of chicken TLR diversity [565]. Diversity indices are less consistent with another study [151], but this is likely due to a small sample size and including roughly as many junglefowl as chickens in the dataset, so pairwise comparisons will tend to exaggerate diversity. In general, the present study found that levels of within-population diversity

in domestic chicken TLRs were comparable to those seen in *G. g. spadiceus* and other red junglefowl subspecies (with the exception of *G. g. gallus*, which was the most diverse population in all groups of chickens and junglefowl). This indicates that domestic chickens have maintained diversity at TLR loci and do not display typical reductions in diversity that would be associated with a domestication bottleneck. Furthermore, diversity is maintained within subpopulations, indicating that it is not a small number of diverse populations that exaggerate diversity across all chickens.

Comparing chicken TLR diversity to that of the Gentoo penguin (*Pygoscelis papua* as an example of an outbred, wild population), it is clear that chicken TLR diversity is significantly greater for *TLR4* and *TLR7* (4.8× and 12.4×, respectively), but roughly the same for *TLR5* (Chapter 2, [358]). Additional large-scale investigations into TLR variability in wild birds would be needed to determine whether the Gentoo penguin has typical or atypical levels of TLR diversity, but it appears that chickens are unusual amongst other birds for being so diverse.

Since Toll-like receptors are frequently found to be under positive selection in birds [220, 358, 603], it is important to investigate whether chicken TLRs have experienced non-neutral evolution. All chicken TLRs had negative values of Tajima's D, which at a superficial level is consistent with chickens having expanded greatly in population size after a domestication bottleneck and acquired novel variants from repeated hybridisation with junglefowl. However, such a process would be expected to affect the whole genome. Despite this, chicken Toll-like receptors were skewed towards the most negative values of the distribution of Tajima's D across coding genes. The most negative genes, *TLR2A*, *TLR2B* and *TLR21*, may therefore have undergone a recent selective sweep in chickens. Especially for *TLR2A* and *TLR2B*, this hypothesis is also supported by a low within-population diversity of chickens at these loci. Further evidence of selection at the *TLR2A*, *TLR2B* and *TLR21* loci comes from taking ratios of diversity in sliding windows across TLR loci (π_S/π_D). *TLR2A* and *TLR2B* showed significantly depressed (within the 2nd percentile) diversity in domestic chickens (π_D) relative to *G. g. spadiceus* (π_S), while *TLR5*, *TLR7* and *TLR21*

also had values between the 2nd-5th percentiles. *TLR5* and *TLR7* have been found to be under positive selection in chickens in a previous analysis (although sites in *TLR5* were marginally non-significant; *TLR2A*, *TLR2B* and *TLR21* were not included in their analysis [565]) This may provide evidence for adaptation to bacterial (*TLR2A* *TLR2B*, *TLR5*, *TLR21*) and viral (*TLR7*) Toll-like receptors during the evolution of the chicken.

Further evidence for positive selection comes from codon-level tests for positive selection. In four Hyphy analyses, all 10 chicken TLRs showed some evidence of positive selection with a high degree of concordance between the different models. The selected sites included ligand binding sites in *TLR1B*, *TLR2B*, *TLR3* and *TLR5*, as well as additional dimerization sites in *TLR1A*, *TLR1B*, *TLR2A* and *TLR3*. Birds typically show lower rates of adaptive evolution in the nucleic acid-recognising TLRs (*TLR3*, *TLR7* and *TLR21*; [603]). However, chickens appear to be unusual in that regard, and the three nucleic acid-recognising TLRs had on- or above-average numbers of positively selected sites (*TLR3*: 44; *TLR7*: 38; *TLR21*: 59; average: 42.2). Furthermore, almost half (6/14) of the selected sites of known function were found in *TLR3*, indicating there has been extensive co-evolution with viral pathogens. *TLR3* and/or *TLR7* are important for the chicken immune response to influenza [559, 226, 290] Newcastle disease virus [98] and infectious bursal disease virus [244], which are together responsible for vast economic losses in the poultry industry [11, 146, 223, 326]. It is interesting to postulate that these viral disease impose significant selection pressures on domesticated chicken populations, and that this may be linked to the enormous population expansion, increased population density and widespread trade of birds.

Variant discovery from the 1660 genomes yielded a rich dataset of SNPs and other variants at TLR loci. In total, there were 1259 coding SNPs, of which 592 were synonymous changes and 657 were non-synonymous changes. In addition, there were ∼4000 non-coding variants, and 86 high-impact variants (including stop-gained, start-lost, frameshift and splice site variants). For context, this is 3.7 times more variants than are currently listed on Ensembl for chicken TLRs (1429; Galgal6; accessed May

2021). Given their high phenotypic impact, the truncating variants are likely to have a significant effect on the functioning of the chicken immune system. Previous reports of chicken TLRs did not identify any truncating variants [151, 565, 300] despite TLR pseudogenisation appearing to be relatively common across birds [10, 36, 603]. Most high impact variants were rare, affecting one or very few birds. However, certain variants are significantly more common in certain breeds. For instance, the stop variants at 4:20026321(*TLR2A*) and 4:20033955 (*TLR2B*) had a high prevalence in White Plymouth Rock birds and thus could significantly impact the ability of this breed to combat bacterial pathogens. Given that both *TLR2A*/*TLR2B* variants are widespread, additionally affecting feral birds in Asia, European fancy breeds and the Ceylon junglefowl (*G. lafayettii*), it seems likely that the variant was already present when the White Plymouth Rock breed was isolated in the mid-19th century [227].

5.5.1 Future directions and conclusions

Chickens appear to be unusual in their high levels of genomic diversity; typically a domesticated animal may show reduced diversity as a result of bottlenecking. However, the unique ecology of the chicken has meant that its history is inextricably linked with the congeneric wild junglefowl, with which it has hybridised extensively. This has resulted in a pangeneric pool of alleles which has bolstered domestic chicken diversity. This study has found substantial levels of diversity at TLR loci, strong deviations from neutral evolution and widespread truncated alleles. Future work to characterise disease associations with variants will demonstrate their phenotypic consequences, and selective breeding to remove deleterious alleles (e.g. *TLR2* pseudogenes in the White Plymouth Rock breed) may improve the immune health of birds.

The rich catalogue of TLR polymorphisms will be of considerable interest to the wider community. As such, variants will be deposited in the European Variation Archive for use by the community. Having appreciated the value of the large dataset used in this study, my co-authors and collaborators have set up a Chicken Genome

Consortium, comprising groups at several institutions across Europe who have contributed genomic data. The aims of the Consortium are numerous and investigations of immune system evolution, production traits, feather colour (and other phenotypic traits), copy number variants, and endogenous retroviral elements are soon to be underway. The complete variant dataset is shared with all members of the consortium, and will be made publicly available following the first round of analyses by consortium members.

Chapter 6

Evolutionary repurposing of TLR21 as a broad-spectrum CpG-DNA recogniser in the birds and reptiles

6.1 Abstract

The ability to detect non-self nucleic acid as an indicator of pathogen presence is a major component of the pattern-recognition receptor repertoire. Specifically, the abundance of unmethylated CpG dinucleotides in microbial genomes, compared to their relative scarcity in vertebrate nuclear genomes, provides a point of difference that can be used to differentiate between self and non-self. To date, two Toll-like receptors that respond to CpG DNA motifs have been identified: *TLR9* and *TLR21*. *TLR9* is present in the genomes of fishes, amphibians and mammals, but is absent in birds and reptiles. *TLR21* is found in fishes, amphibians, birds and reptiles, but is absent in mammals. Therefore, fishes and amphibians have both *TLR9* and *TLR21*, mammals have only *TLR9*, and birds/reptiles have only *TLR21*. The ability of TLR9 or TLR21 to respond to CpG motifs in DNA is dependent on numerous factors,

including the number of CpGs and the surrounding sequence context. In fishes, TLR9 is a generalist CpG recogniser, responding to a wide range of different CpG motifs, while TLR21 is fastidious, responding to few CpGs motifs. In stark contrast, chicken TLR21 is a broad CpG recogniser. I hypothesised that the absence of *TLR9* in the birds and reptiles led *TLR21* to be repurposed to fill the role of a broad CpG recogniser. Consistent with this hypothesis, I found that amphibian TLR21 is fish-like, and reptilian TLR21 is bird-like in CpG-recognition profile using functional assays. Furthermore, I found evidence of strong and recurrent positive selection in the avian and reptilian *TLR21*, but not in the fish *TLR21*, which is consistent with functional change following the loss of *TLR9*. More generally, these results point to a necessity for vertebrates to maintain a broad CpG-recognition capability in their innate immune system.

6.2 Introduction

Detecting "non-self" is a crucial part of the innate immune system [271] and can be achieved through detection of pathogen-associated molecular patterns (PAMPs) which are present in microbial life, but not in the animal host. One such PAMP is the unmethylated cytosine-phosphate-guanine (CpG) DNA dinucleotide motif, found in hexamers with the general formula RRCGYY (where R represents a purine and Y a pyrimidine; [328]). Umethylated CpG motifs are frequent in bacterial and viral genomes, but are greatly underrepresented in vertebrate genomes [335]. The Toll-like receptors form a key part of the host detection system for unmethylated CpG motifs in the innate immune system.

To date, two Toll-like receptors have been identified which respond to unmethylated CpG motifs—TLR9 and TLR21. These two Toll-like receptors are phylogenetically unrelated, which implies they have independently acquired the ability to respond to unmethylated CpG-DNA and are functional analogs [82]. *TLR9* is found in fishes, amphibians and mammals [493], and has recently been described in chelonians (tur-

tles) [149], but is absent from the majority of avian and reptilian genomes [81]. *TLR21* is also found in fishes and amphibians, and is present in birds and reptiles, but has been lost from mammalian genomes [81, 82, 301]. The phylogenetic distribution of CpG-recognising TLRs means fishes and amphibians have two (*TLR9* and *TLR21*), but mammals and birds/reptiles typically only have one (*TLR9* in mammals, *TLR21* in birds/reptiles).

The ability of TLR9 and TLR21 to recognise CpG-DNA (or its short, synthetic counterpart: CpG-oligodeoxynucleotide, CpG-ODN) is not absolute, but rather dependent on the length of the DNA fragment, the number of CpG hexamers, and the spacing and position of the bases flanking the core CpG [107]. TLR9 is also known to exhibit species-specificity in its responsiveness to CpG motifs—murine TLR9 preferentially responds to CpG-ODNs with a GACGTT motif, while TLR9 from human, pig, dog, sheep, goat and horse is activated more by CpG-ODNs with a GTCGTT motif [327, 472, 490]. This implies that CpG-recognising TLRs can develop different CpG-recognition profiles in response to pathogen-induced selection pressure.

Furthermore, there is also evidence from animals with both *TLR9* and *TLR21* that the receptors can undergo specialisation within the same animal. For instance, fish TLR9 is known to recognise a broad array of CpG motifs from different classes [657], whereas fish TLR21 is a fastidious recogniser of a small number of CpG motifs from the Class B CpG-ODNs [107, 364, 657]. Moreover, the recognition profiles of TLR9 and TLR21 in fishes appear complementary, and the two receptors form part of a cooperative immune response against an important pathogen (*Edwardsiella tarda*) in zebrafish [657].

Even prior to the discovery and characterisation of *TLR21* in birds, it was established that chicken cells respond to a broad range of different CpG motifs [242, 643]. More recent studies have confirmed this observation, and attributed the CpG-responsiveness to TLR21, rather than other pattern-recognition receptors [107] It is therefore important to highlight the transformation of TLR21 in fishes (fastidious recogniser of a small number of CpG motifs) to birds (robust, generalist responder).

Furthermore, it is interesting to postulate that the repurposing of TLR21 as a generalist CpG recogniser was driven by (or was a driver of) the loss of *TLR9* in birds and reptiles.

To date, the CpG response profile in TLR21 in birds (apart from chicken), reptiles and amphibians is unknown. In this study, I investigated whether TLR21 has undergone functional repurposing in birds and reptiles following the loss of TLR9. To support this conclusion, extensive positive selection was noted in the ligand binding domain of TLR21 of birds and reptiles, which is indicative of a change in ligand binding preference. Furthermore, cloning of *TLR21* from a wide range of different animals and challenge with a CpG panel revealed that TLR21 in birds and reptiles has a highly diverse recognition profile. This is in contrast to fish and amphibian TLR21, which was a weak responder with a slight preference for Class B CpG-ODNs. These data support the hypothesis that TLR21 has undergone functional repurposing in birds and reptiles and highlights the plasticity of PRR recognition profiles in the innate immune system.

6.3 Materials and Methods

6.3.1 Sequence retrieval

TLR11, *12*, *13* and *21* sequences were retrieved from Ensembl using mouse *TLR11* (ENSMUSG00000051969), mouse *TLR12* (ENSMUSG00000062545), mouse *TLR13* (ENSMUSG00000033777) and chicken *TLR21* (ENSGALG00000033068) to query for orthologous sequences. Truncated or incorrectly annotated (e.g., *TLR13* sequences included with *TLR21* orthologs) were removed from the analysis. A full list of accession numbers for sequences used can be found in Appendix Table E.1.

6.3.2 Maximum-likelihood tree

In order to elucidate the phylogenetic relationship between *TLR11*, *12*, *13* and *21*, a maximum likelihood tree was built using RAxML [554] with 100 rapid bootstrap searches and 20 maximum likelihood searches, based on an alignment generated with MAFFT [294].

6.3.3 Positive selection analyses

Both the 5′- and 3′-ends of *TLR21* are variable across vertebrates, and are not conducive to codon-based tests of positive selection. Instead, a section of the leucine-rich repeat domain was aligned (corresponding to the start of LRR1 to the end of LRR-CT) and used for analysis. RAxML [554] was used to generate a tree based on the *TLR21* alignment. First, the adaptive Branch-Site Random Effects Likelihood (aBSREL) tool from Hyphy [547] was used with all branches specified as test branches. A P-value threshold of 0.05 was used (following correction for multiple comparisons with the Holm-Bonferroni method). The branch-site model in PAML [652, 654] (model = 2; NSsites = 2) was also used, specifying the bird and reptilian lineage as the foreground branch, and all others as the background branch. Two models were compared: (i) negative/neutral model, where omega is fixed at 1, and (ii) positive selection model, where omega is allowed to vary. The fit of the two models was compared using a likelihood ratio test, and P-value was calculated based on the χ^2 distribution with df = 1.

6.3.4 Homology modelling

A homology model of the chicken TLR21 ectodomain was generated using using the web server of iTASSER [505]. The template for homology modelling was mouse TLR13 (PDB: 4Z0C). Positively selected sites and a putative DNA-binding site were mapped onto the homology model using PyMOL v. 2.5.

6.3.5 TLR21 cloning strategy

TLR21 coding sequences from fugu (*Takifugu rubripes*), axolotl (*Ambystoma mexicanum*), newt (*Lissotriton vulgaris*), gecko (*Gekko japonicus*), crocodile (*Crocodylus porosus*), bearded dragon (*Pogona vitticeps*) and Northern rockhopper penguin (*Eudyptes moseleyi*) were obtained as GBlocks (IDT). Due to a region of low complexity towards the end of the TIR domain in avian TLR21, penguin TLR21 needed to be codon optimised to allow synthesis. All other constructs were made using the original reference sequence. Coding sequences for *TLR21* from zebrafish (*Danio rerio*), *Xenopus laevis* and chicken (*Gallus gallus*) were generated by PCR. In all cases, a consensus Kozak sequence (GCCGCCACC) was added immediately upstream of the start codon, and the native stop codon was removed to allow the reading frame to continue into the 3× FLAG tag of the p3×FLAG-CMV14 expression vector (Sigma).

6.3.5.1 RNA extraction

Carcasses of *Xenopus laevis*, zebrafish (*Danio rerio*), and chicken (*Gallus gallus*) euthanised for other purposes were provided by Roger Patient and Philip Pinheiro (MRC Weatherall Institute of Molecular Medicine, University of Oxford; *Xenopus* and zebrafish) and Tommaso Pizzari (Zoology Department, University of Oxford; red junglefowl). Animals were dissected to obtain a variety of tissues and stored in RNAlater. 20 mg tissue were mixed with 600 μl RLT buffer and homogenised for 1 minute using a bead-beater (Biospec). Lysates were cleared by centrifugation for 3 minutes at maximum speed at 4 °C. RNA extraction was achieved using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Briefly, the homogenised lysate was mixed with 1 volume of 70% ethanol, mixed by pipetting, and transferred to an RNeasy spin column for centrifugation. Genomic DNA was then digested with an on-column DNA digestion. The column membrane was washed twice using RPE buffer and centrifuged one extra time to limit the carryover of ethanol. RNA was eluted in 30 μl RNAse-free water and stored at -80 °C until further processing.

6.3.5.2 cDNA synthesis

RNA samples were reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, ThermoFisher) according to the manufacturer's instructions. The 2× RT mastermix was prepared as follows (per reaction): 10× RT buffer, 2 μl; 25× dNTP mix (100 mM), 0.8 μl; 10× RT random primers, 2 μl; Multiscribe reverse transcriptase, 1 μl; nuclease-free water, 4.2 μl (total = 10.0 μl). Ten microlitres of RNA were mixed with 10 μl of 2× RT mastermix and incubated using a thermocycler as follows: 25 °C, 10 minutes; 37 °C, 120 minutes; 85 °C, 5 minutes. cDNA purity was assessed using a NanoDrop-2000 spectrophotometer (ThermoFisher) and concentration was measured using a Qubit (ThermoFisher).

6.3.5.3 Vector and insert PCRs

All PCRs in the following section used the high fidelity Q5 polymerase (NEB). Chicken, *Xenopus* and zebrafish *TLR21* were amplified from cDNA using primers listed in Table 6.1. Each PCR was designed to include a ∼20 bp region homologous to the vector sequence on either end, a human consensus Kozak sequence (GCCGCCACC), and a 3′ spacer to ensure the FLAG sequence remained in frame. For chicken, *Xenopus* and zebrafish *TLR21*, inserts were ligated into the p3×FLAG-CMV14 vector which had been linearised using HindIII and XbaI restriction enzymes (both NEB) using the Gibson assembly protocol (see below). For the remaining constructs, vector stock for Gibson assembly was generated using a universal vector PCR (Table 6.1).

Table 6.1: Primer sequences and PCR conditions used for generating inserts, linearised vector and for sequencing.

PCR	Type	Primers (5′-3′)		T_a	T_{ex}
chTLR21	Insert	F	ttagtgaaccgtcagaattagccgccaccATGATGGAGACAGCGGAG	61	2.30
		R	tgtagtcagcccgggatcctggCATCTGTTTGTCTCCTTCC		
xlTLR21	Insert	F	ttagtgaaccgtcagaattagccgccaccATGGTGCAAAGTAGACAAGTC	63	2.30
		R	tgtagtcagcccgggatcctggGTACCGCTCAGCTGAACAATG		
zfTLR21	Insert	F	ttagtgaaccgtcagaattagccgccaccATGGCACACTCTGCATGC	63	2.30
		R	tgtagtcagcccgggatcctggGGGCATCAGGTAATATATTTCATCTTC		
CMV14	Vector	F	GACTACAAAGACCATGAC	57	5.00
		R	TCTAGAGTCGACTGGTAC		
amTLR21	Sequencing	F	CAACCTGTCCATTCTCTGC	N/A	N/A
		R	CCAAAGACATCGCTAGTGAG		
chTLR21	Sequencing	F	TGACATCCAACCTCGCAGCCTGG	N/A	N/A
		R	CATCTGTTTGTCTCCTTCC		
cpTLR21	Sequencing	F	GGCAGCAACATGTCAAGGC	N/A	N/A
		R	AAAGCTGCAATGTGGTTCGG		
emTLR21	Sequencing	F	TCTTTGGACTTCTCTGGCC	N/A	N/A
		R	CGCATTCTCCAGATTAAGG		
fuTLR21	Sequencing	F	GAAAGCTGAGCAGGATAACC	N/A	N/A
		R	AGGTCTGCATTACGTTGTGAT		
gjTLR21	Sequencing	F	TCGAAACTGTAAATGTAACTG	N/A	N/A
		R	TACTACCTCAGCTTCGATAAC		
lvTLR21	Sequencing	F	GCAACCTGTCCAGTCTCTG	N/A	N/A
		R	CATTCTTCATCAGATGCAGGAC		
pvTLR21	Sequencing	F	CAGTTAATCGAGCTGTGCCA	N/A	N/A
		R	CTAGAAAGAGGTTGCGCAATC		
xlTLR21	Sequencing	F	TCATACAGAATTCTACTGCTC	N/A	N/A
		R	AGGTCTAGAACATGAAGG		
zfTLR21	Sequencing	F	AAGCATAACAAACCTGACAAG	N/A	N/A

PCR	Type	Primers (5′-3′)		T_a	T_{ex}
CMV14	Sequencing	R	AAGCATAACAAACCTGACAAG	60	1.40
		F	GGCACCAAAATCAACGGGAC		
		R	ACTTGCCCCTTGCTCCATAC		

Bases in lower case are overhanging ends used to generate homologous sequence overlaps for Gibson assembly. Underlined bases in F primers are the consensus Kozak sequence; Underlined bases in R primers are spacers to keep the FLAG-tag in frame.

T_a, annealing temperature (°C).

T_{ex}, extension time (min.sec).

Primers with "N/A" for T_a and T_{ex} were only submitted as sequencing primers for PCR products generated by using the vector sequencing primers.

Abbreviations: ch, chicken; xl, *Xenopus laevis*; zf, zebrafish; am, *Ambystoma mexicanum* (axolotl); cp, *Crocodylus porosus* (saltwater crocodile); em, *Eudyptes moseleyi*, northern rockhopper penguin; fu, fugu (*Takifugu rubripes*); gj, *Gekko japonicus*, Schlegel's Japanese gecko; lv, *Lissotriton vulgaris* (common newt); pv, *Pogona vitticeps* (central bearded dragon).

6.3.5.4 Gibson assembly/enzyme-free assembly

All *TLR21* coding sequences were cloned into the p3×FLAG-CMV expression vector (Sigma) using Gibson assembly. For each assembly, 0.025 pmol of vector were used, and inserts were added in a 2:1 insert:vector ratio (i.e. 0.05 pmol insert). Ligations were assembled on ice and made up to 10 μl with nuclease-free water. Ligations were then mixed 1:1 with the 2× Gibson Assembly Master Mix (NEB) and incubated at 50 °C for 1 hour.

Over the course of the project, it became apparent that the addition of the 2× Gibson Assembly Master Mix (NEB) was superfluous for the correct ligation of insert into vector. Control reactions containing only water in addition to the fragments to be ligated worked with equal efficiency compared to reactions where the mastermix was included, indicating that the enzymes are not required for ligation. The same phenomenon was observed with purified vs. unpurified PCRs, with and without the 50 °C incubation stage normally required for Gibson assembly, and with freeze-thawed and fresh samples. The mechanism behind this is unclear, but it seems likely that transformed *E. coli* are capable of assembling simple constructs without a pre-ligation step. It is plausible that ligation enzymes are required for more complex assemblies, but for simple assemblies, the removal of the superfluous Gibson assembly step streamlined the process and saved unnecessary expense.

6.3.5.5 Transformation and colony screening

Using ligations from the previous step, 10-β *E. coli* were transformed according to the high-efficiency transformation protocol (NEB). Pre-aliquoted tubes of 50 μl competent cells were kept on ice until just thawed. Two microlitres of ligation were added to the cells and the tube was flicked five times to mix. The mixture was incubated on ice for 30 minutes before heat shock at exactly 42 °C for exactly 30 seconds. Tubes were incubated for 5 minutes on ice and mixed with 950 μl room-temperature SOC (Super Optimal broth with Catabolite repression) medium. Cells were incubated at 37 °C

for 60 minutes with vigorous shaking to facilitate outgrowth before being plated on warmed LB (lysogeny broth) plates with Ampicillin (100 μg/ml) as a selective agent. Plates were incubated overnight at 37 °C.

For colony screening, individual colonies were picked with a sterile pipette tip which was then used to inoculate 40 μl LB-ampicillin, then 100 μl nuclease-free water. Next, PCR reactions were assembled using MyFi 2× mastermix (12.5 μl), 1.25 μl forward and reverse primers (final concentration 500 nM), nuclease-free water (8 μl), and 2 μl inoculated water as template. PCR cycling conditions were as follows: 95 °C initial denaturation (3 minutes), 40 cycles of 95 °C (15 seconds), 60 °C (15 seconds), 72 °C (1 minutes 40 seconds), then a final extension of 72 °C (7 minutes). Colony screening primers were universal to all constructs in the p3×FLAG-CMV expression vector and spanned the multiple cloning site, yielding a ∼400 bp product for an empty vector and a ∼3 kbp product for a correct insertion (F: 5′-GGCACCAAAATCAACGGGAC-3′; R: 5′-ACTTGCCCCTTGCTCCATAC-3′). Using the same MCS-spanning primers and additional species-specific internal primers (Table 6.1), PCR products with suitably sized inserts were submitted for full sequencing to ensure fidelity of the insert.

6.3.5.6 Endo-free maxipreps

Constructs with fully sequenced inserts were grown up from the colony screening inoculate. Twenty microlitres of the inoculate were used to inoculate a loosely capped tube containing 5 ml LB-Ampicillin, which was then incubated for 8 h at 37 °C with vigorous rotation. The starter culture was then diluted 1/500 in 110 ml LB-Ampicillin in a conical flask, which was further incubated for 12-16 h at 37 °C with vigorous rotation. Plasmid DNA was then extracted and purified using the EndoFree Plasmid Maxi Kit (Qiagen). Bacterial cultures were harvested at 4700 ° G for 30 minutes at 4 °C and the supernatant was discarded. The bacterial pellet was fully resuspended in 10 ml buffer P1 (which was supplemented with LyseBlue). Lysis of bacterial cells was achieved by the addition of 10 ml buffer P2 and incubation at room temperature for

5 minutes, followed by neutralisation by addition of 10 ml buffer P3. The lysate was cleared using the Qiafilter Maxi Cartridge into a 50 ml tube. To remove contaminating endotoxins, the filtered lysate was treated with 2.5 ml buffer ER, followed by mixing and incubation on ice for 30 minutes. The filtered lysate was then added to an equilibrated Qiagen-tip and allowed to pass through the resin by gravity flow. The Qiagen-tip was then washed twice using 30 ml buffer QC and then DNA was eluted in 15 ml buffer QN. DNA was precipitated by adding 10.5 ml room-temperature isopropanol, followed by mixing and centrifugation at 4700 $\times$ G for 1 hour at 4 °C. The DNA pellet was washed with 5 ml endotoxin-free room-temperature 70% ethanol and centrifuged at 4700 $\times$ G for 1 hour at 4 °C. The supernatant was decanted and the DNA pellet was air-dried for 5 minutes, with care being taken not to over-dry the pellet. Finally, the pellet was redissolved in 500 μl endotoxin-free TE buffer.

Plasmid purity was assessed using a NanoDrop-2000 spectrophotometer (ThermoFisher) and concentration was measured using a Qubit (ThermoFisher). Before being used for transfection, all construct inserts were direct sequenced using the same primers as for colony screening.

6.3.6 Transfection and stimulation

For all transfection experiments, HEK-Blue Null1 cells (Invivogen) were used. HEK-Blue cells express a secreted alkaline phosphatase (SEAP) reporter gene under the control of the IFN-β minimal promoter fused to five NF-κB and AP-1 binding sites, and therefore secrete SEAP into the cell culture medium in response to TLR stimulation. HEK-Blue cells were maintained in high glucose Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% foetal bovine serum (FBS; Gibco), 4 mM L-glutamate, 50 U/ml penicillin and streptomycin (ThermoFisher), and 50 μg/ml Zeocin (InvivoGen). Cells were cultured in a humidified incubator at 37 °C with 5% CO2.

Twenty-four hours prior to transfection, cells were seeded into 24-well plates at

a concentration of 3x10^5 cells/ml in order to be 60-80% confluent at the time of transfection. One the day of transfection, plasmid DNA was diluted in OptiMEM reduced serum medium (Gibco) to achieve 500 ng DNA per well. The transfection reagent TransIT-2020 was diluted in OptiMEM to achieve a final concentration of 1.5 μl per well. The diluted transfection reagent was mixed with the plasmid DNA and incubated at room temperature for 15 minutes. Complexes were then added to wells dropwise, followed by gentle rocking to distribute the complexes throughout the well. Transfected cells were incubated for 48 h prior to stimulation.

Prior to stimulation, CpG-ODNs (Invivogen) were reconstituted in 200 μl ultrapure nuclease-free water, yielding stock concentrations of between 107.0-157.1 μM, depending on the length of the ODN. CpG-ODN sequences used in this study are given in Table 6.2. The working concentration for all CpG-ODNs was 2 μM. CpG-ODNs were diluted to the appropriate concentration in fresh complete medium, which was then used to replace the medium on transfected cells. Control agonists (dsDNA from *E.coli*; ssDNA from *E. coli* with Lyovec; ORN Sa19, with and without Lyovec; all Invivogen) were made up according to the manufacturer's instructions and used at a final concentration of 1 μg/ml. Ultrapure nuclease-free water was added to negative control wells. Cells were incubated with agonists for 24 h before harvest of supernatants for SEAP analysis. Supernatants were mixed with p-Nitrophenyl phosphate (pNPP; Sigma) according to the manufacturer's instructions, and absorbance was measured at 405 nm on a FLUOstart Omega microplate reader (BMG Labtech) at 37 °C. Differences between means were tested statistically using ANOVA with Dunnett's correction [153] for multiple comparisons to a control in R (v. 4.0.2).

6.3.7 Immunoblotting

Cells were washed in ice-cold PBS and lysates were harvested in NuPAGE LDS Sample Buffer (4×; Novex) supplemented with 0.1 M DTT and boiled for 10 minutes at 95 °C. Samples were resolved on a 4-20% gradient gel (Bio-Rad), transferred using a semi-dry

Table 6.2: CpG-ODN sequences used for stimulation of TLR21.

Class	CpG-ODN	Sequence	Length
A	ODN 1585	5'- ggGGTCAACGTTGAgggggg	20
A	ODN 2216	5'-ggGGGACGATCGTCgggggg-3'	20
A	ODN 2336	5'- gggGACGACGTCGTGgggggg -3'	21
B	ODN 1668	5'-tccatgacgttcctgatgct-3'	20
B	ODN 1826	5'-tccatgacgttcctgacgtt-3'	20
B	ODN 2006	5'-tcgtcgttttgtcgttttgtcgtt-3'	24
B	ODN 2007	5'- tcgtcgttgtcgttttgtcgtt -3'	22
B	ODN BW006	5'-tcgacgttcgtcgttcgtcgttc-3'	23
B	ODN D-SL01	5'-tcgcgacgttcgcccgacgttcggta-3'	26
C	ODN 2395	5'-tcgtcgttttcggcgcgcgccg-3'	22
C	ODN M362	5'-tcgtcgtcgttcgaacgacgttgat-3'	25
C	ODN D-SL03	5'-tcgcgaacgttcgccgcgttcgaacgcgg-3'	29

Underlined sequences are palindromes.
Bases in capital letters are phosphodiester and those in lower case are phosphorothioate.

apparatus (Bio-Rad), blocking using StartingBlock PBS blocking buffer (Thermo Scientific), probed using mouse monoclonal anti-FLAG (M2; 1:1000; Sigma) overnight at 4 °C, then, after washing, probed with a goat anti-mouse HRP-conjugated secondary antibody (Cusabio; 1:10 000), and visualised on an ImageQuant LAS 4000 (Cytiva) using SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Scientific). Sizes of immunoreactive products were compared to the PageRuler Prestained Protein Ladder (Thermo Scientific).

6.4 Results

6.4.1 *TLR21* is distinct from *TLR13*

Designations of *TLR21* and *TLR13* are often used interchangeably by genome annotation pipelines, presumably depending on which species is used as a reference for gene identification. As such, the evolutionary history between *TLR21* and *TLR13* is unclear. To investigate the phylogenetic relationship between TLR11/12/13/21 family members, TLRs from various representative species were aligned and a phy-

logenetic tree was built in RAxML (Figure 6.1). As expected, *TLR11* and *TLR12* were sister clades with only mammalian representatives. *TLR13* was an outgroup to *TLR11* and *TLR12*, again only containing mammalian sequences. The *TLR21* clade, which was a sister clade to the *TLR11/12/13* clade, contained fish, amphibians, reptiles and birds, but no mammals. If *TLR13* were simply 'mammalian *TLR21*', it would be expected that the *TLR13* clade would be nested within the *TLR21* clade. Instead, *TLR13* forms a distinct monophyletic clade, which implies that *TLR13* and *TLR21* do not share a recent origin. *TLR21* is therefore distinct from *TLR13* and should be corrected in genomes where is has been mis-annotated as *TLR13* (e.g. *Coturnix japonica*, XM_032446984.1; *Parus major*, XM_033511680.1; *Anolis carolinensis*, XM_008124928.2; *Alligator sinensis*, XM_006038111.2), or mis-annotated as a 1:1 ortholog with a mammalian *TLR13* in Ensembl.

6.4.2 Features of *TLR21*

In the chicken, *TLR21* comprises a 2852-3252 bp open reading frame (depending on splice isoform) located on chromosome 11. The intraluminal (ligand-binding) domain comprises a signal peptide, 27 leucine-rich repeats spaced at regular intervals, and a putative DNA-binding site D456, L457, Y458 based on comparison with human TLR9 sequence (Figure 6.2; [301]). TLR21 does not, however, have the z-loop between LRR14 and 15 which is required for cleavage and dimerization [641, 333]. Following the single transmembrane domain, there are three conserved boxes of IL-1R family motifs (typical of all TLRs and other IL-1R family members), along with a conserved proline residue that is implicated in MyD88-activation [474, 592].

Uniquely for a Toll-like receptor, chicken TLR21 also has a poly-glutamic acid motif towards the end of the TIR domain—a feature associated with intrinsically disordered proteins (Figure 6.3; [595]). The poly-glutamic acid motif is present in TLR21 across fish, birds and reptiles, but interestingly, it has expanded in some passerines to include an extra one or two poly-glutamic acid motifs. This has resulted

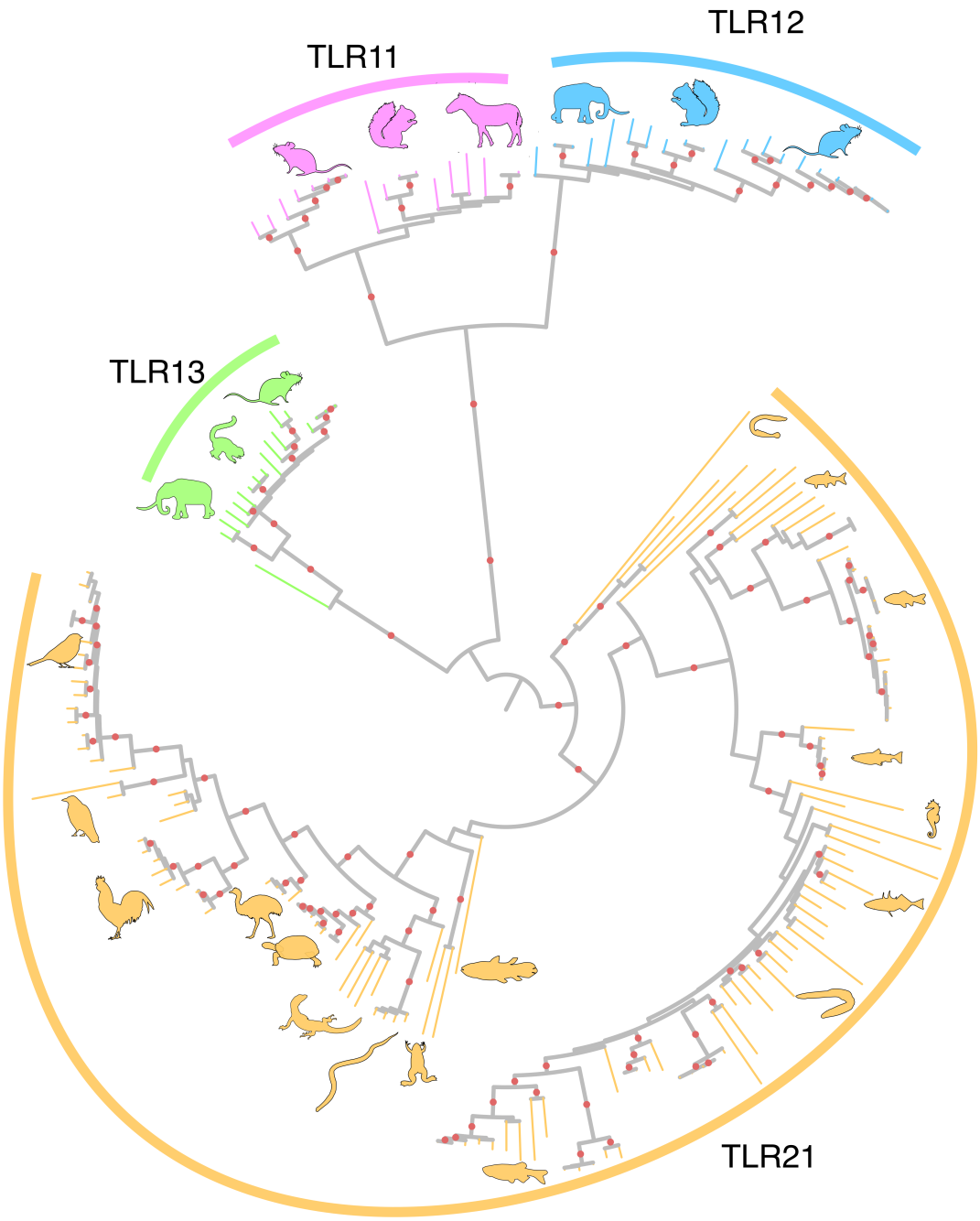


Figure 6.1: Phylogenetic relationship between *TLR11*, *TLR12*, *TLR13* and *TLR21*. Maximum likelihood tree was made using RAxML with 100 rapid bootstrap searches and 20 maximum likelihood searches. Red circles indicate that the node has a bootstrap support of 100.

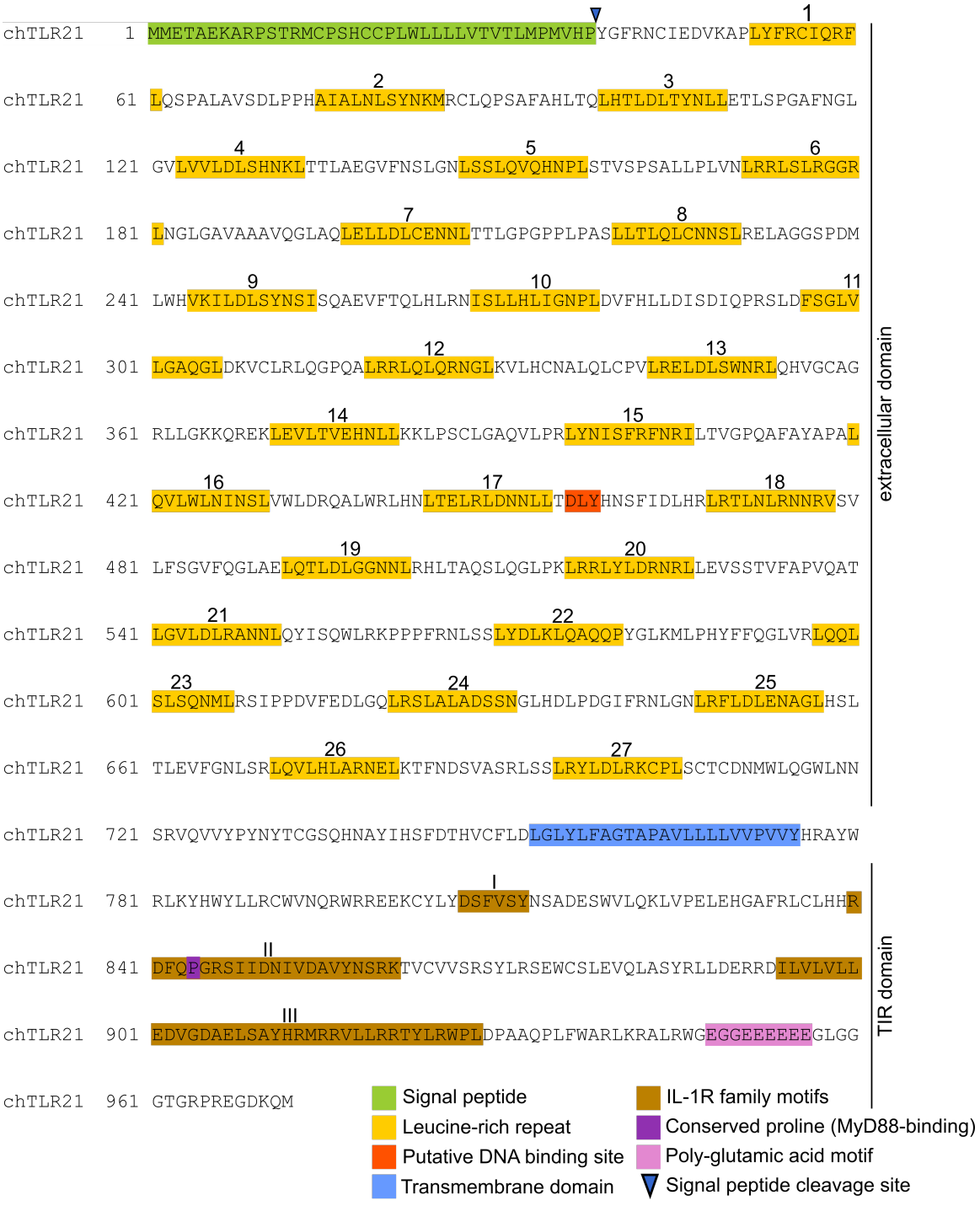


Figure 6.2: Conserved features of the TLR21 protein. The chicken TLR21 sequence is presented with the key conserved features of TLR21 highlighted. There is a signal peptide ending in a likely cleavage site at residue P38 and a single transmembrane domain spanning residues L753-Y775. TLR21 has 27 leucine-rich repeats in regular intervals within the extracellular/intra-endosomal domain. There is a putative DNA-binding site (D456, L457, Y458) reported in [301] based on comparison with the human TLR9 sequence. Three conserved boxes of IL-1R family motifs (typical of all TLRs and other IL-1R family members) are found in the TIR domain, along with a conserved proline residue that is implicated in MyD88 activation [474, 592].

in an elevation of the proportion of charged residues in the C-terminus of passerine TLR21. For instance, the first 956 amino acids (to just before the start of the poly-glutamic acid motif) of crow TLR21 is 20.5% charged amino acids, while the last 149 amino acids is 48.3% charged amino acids. It is unclear why TLR21, and in particular passerine TLR21, should benefit from an intrinsically disordered region, but such regions are known to have diverse functions including as a flexible linker region [152], for metal binding [171], and for thermostability [578].

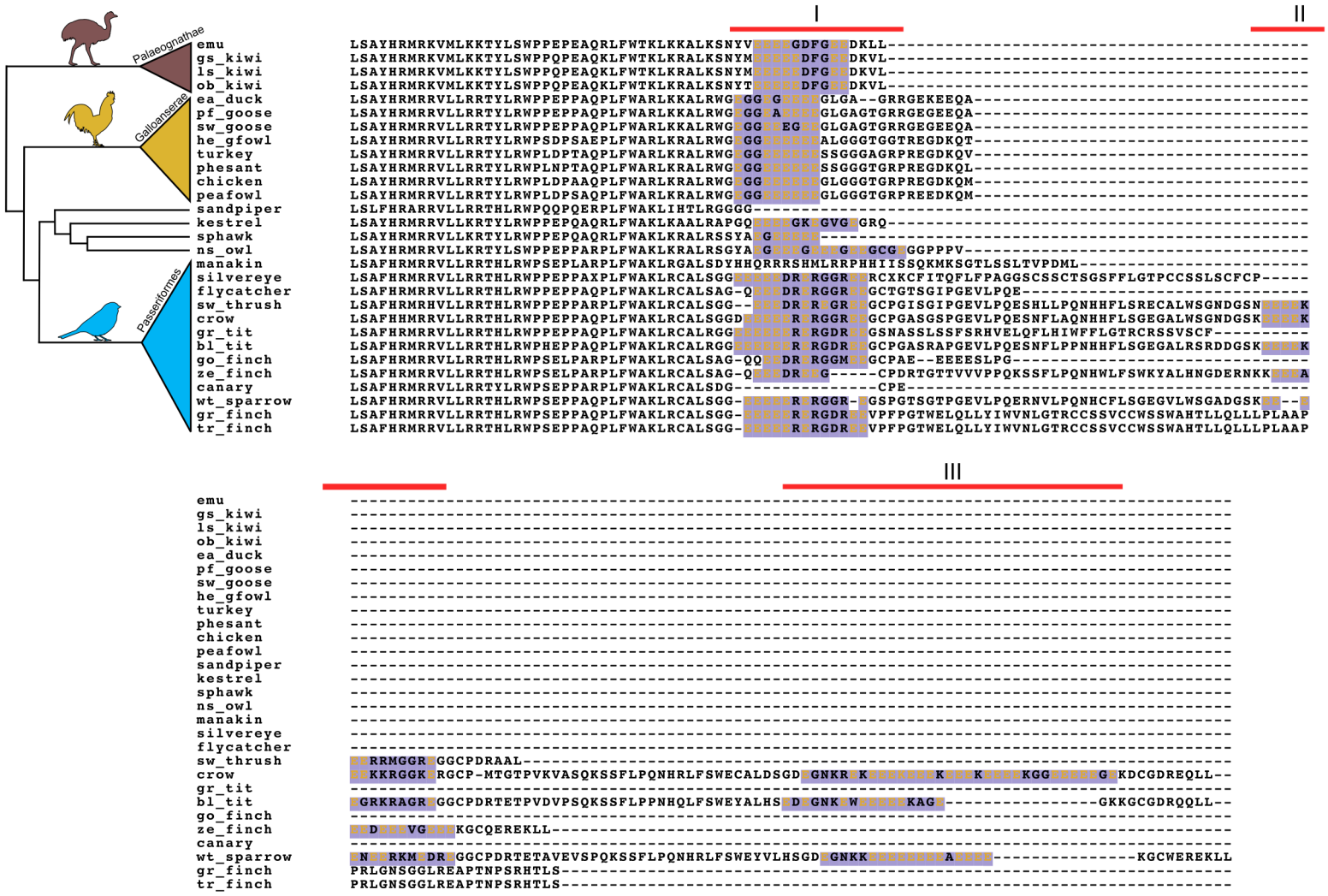


Figure 6.3: Identification of poly-glutamic acid repeats in the TIR domain of avian TLR21. A section of alignment of the TIR domain of avian TLR21 is shown (spanning chicken residue 908 to the end of each sequence). Poly-glutamic acid stretches are highlighted in coloured boxes and numbered, and have increased in number in the passerines, but not elsewhere in birds.



Figure 6.4: Direct sequencing of zebrafish *TLR21*-containing plasmid. All inserts were fully sequenced prior to cloning, and all final constructs were direct sequenced to confirm successful integration of the coding sequence and ensure there was no disruption of the reading frame. To be considered successfully sequenced, all bases needed to have been covered by at least one (preferably two or three) high-quality Sanger reads. This figure represents an example of the high-quality sequencing reads obtained for all *TLR21* constructs.

6.4.3 Molecular cloning and expression of TLR21

To functionally investigate the agonist specificity profiles of TLR21, the coding sequence of *TLR21* from fish (fugu and zebrafish), amphibians (axolotl, newt, and *Xenopus*), reptiles (bearded dragon, crocodile, and gecko), and birds (chicken and penguin) were cloned into the p3×FLAG-CMV-14 expression vector, incorporating a 3×FLAG-tag on the C-terminus of the protein. Following cloning, all constructs were direct sequenced to confirm the correct integration of the intact coding sequence. The results of sequencing for all constructs are given in Table 6.3 and an example of the sequencing traces obtained is given in Figure 6.4. All inserts generated by synthesis were faithful copies of their parental sequence, with the minor exception of bearded dragon *TLR21*, which had a synonymous change in the coding sequence (C792T). Two of the three constructs with PCR-derived inserts had deviations from the reference sequence. The chicken *TLR21* PCR was derived from an outbred population of red junglefowl, and thus has several deviations from the domestic chicken reference sequence. The majority of these variants have been described previously (Table 6.4). The *Xenopus TLR21* PCR was derived from a laboratory animal, but it is not known whether the variants listed are common in *Xenopus* populations.

Table 6.3: Cloning of *TLR21* from fish, amphibians, reptiles and birds. All plasmids were direct sequenced and any deviations from the reference sequence are listed in the final column.

Species	**Species**	**Taxon**	**Strategy**	**Optimised**	**Reference sequence**	**Sequence variants**
Fugu	*Takifugu rubripes*	Fish	Gblock	NO	ENSTRUG00000004103	None
Zebrafish	*Danio rerio*	Fish	PCR	NO	NM_001199335.1	None
Axolotl	*Ambystoma mexicanum*	Amphib.	Gblock	NO	[1]	None
Common newt	*Lissotriton vulgaris*	Amphib.	Gblock	NO	[2]	None
Xenopus	*Xenopus laevis*	Amphib.	PCR	NO	XM_018226707.1	[4]
Schlegel's Japanese gecko	*Gekko japonicus*	Reptile	Gblock	NO	XM_015416616.1	None
Saltwater crocodile	*Crocodylus porosus*	Reptile	Gblock	NO	XM_019534916.1	None
Central bearded dragon	*Pogona vitticeps*	Reptile	Gblock	NO	ENSPVIG00000003139	C792T (S)[6]
Chicken	*Gallus gallus*	Bird	PCR	NO	ENSGALT00000077967	[5]
N. rockhopper penguin	*Eudyptes moseleyi*	Bird	Gblock	YES	[3]	None

[1] https://ambystoma.uky.edu/index.php; SuperContig_180230

[2] http://newtbase.eko.uj.edu.pl/download; c146344_g1_i1

[3] Chapter 2; Pan et al. 2019; MW793864

[4] T609A (H236Q); G1478A (S526N); G2061A (S)

[5] C25T (W9R); C498A (S); A566T (A189V); G632A (G211E); T717G (D239E); G1788A (S); G1989A (S); C2364T (S); G2403A (S); C2418T (S); T2830C (S)

[6] (S), synonymous polymorphism.

Table 6.4: Variants in cloned chicken *TLR21* compared to known SNPs.

Nucleotide pos.	Variant	AA pos.	Variant	Known?	Ensembl ID
25	C→T	9	R→W	YES	rs314633105
498	C→A	166	N/A	YES	rs15598749
566	C→T	189	A→V	YES	rs14017085
632	G→A	211	G→E	NO	N/A
717	T→G	239	D→E	YES	rs14017086
1788	G→A	596	N/A	NO	N/A
1989	G→A	663	N/A	YES	rs14017087
2364	C→T	788	N/A	YES	rs13689175
2403	G→A	801	N/A	YES	rs14017088
2418	C→T	806	N/A	YES	rs14017089
2830	T→C	944	N/A	YES	rs316232358

Having successfully cloned and sequenced each of the *TLR21* constructs, plasmids were then transfected into the human HEK-Blue Null1 reporter cell line using TransIT-2020 transfection reagent. Expression levels were assessed by western blotting using an anti-FLAG antibody. Most constructs expressed well and appropriately sized (100-130 kDa) products were detectable following western blotting (Figure 6.5. Surprisingly, and despite no detectable issues with the insert sequences, the zebrafish, axolotl and *Xenopus* TLR21 constructs failed to express at detectable levels in HEK-Blue cells.

6.4.4 Functional testing of cloned TLR21 constructs

To investigate the breadth of CpG reactivity of TLR21 in different animals across evolutionary time, all constructs with confirmed expression were challenged with a panel of CpG oligonucleotides. CpG-ODNs from all three classes (A, B and C) were included, as well as double- and single-stranded DNA from *E. coli*, and ORN Sa19 (a mammalian TLR13 agonist).

Firstly, TLR21 in fish (fugu) and amphibians (newt) were tested, as representatives of species which also have the broad-recognising TLR9 (Figure 6.6). Neither fugu nor newt TLR21 responded robustly to any of the panel of CpG agonists. The response with the greatest magnitude was ODN 2006 (a class B CpG) in fugu, al-

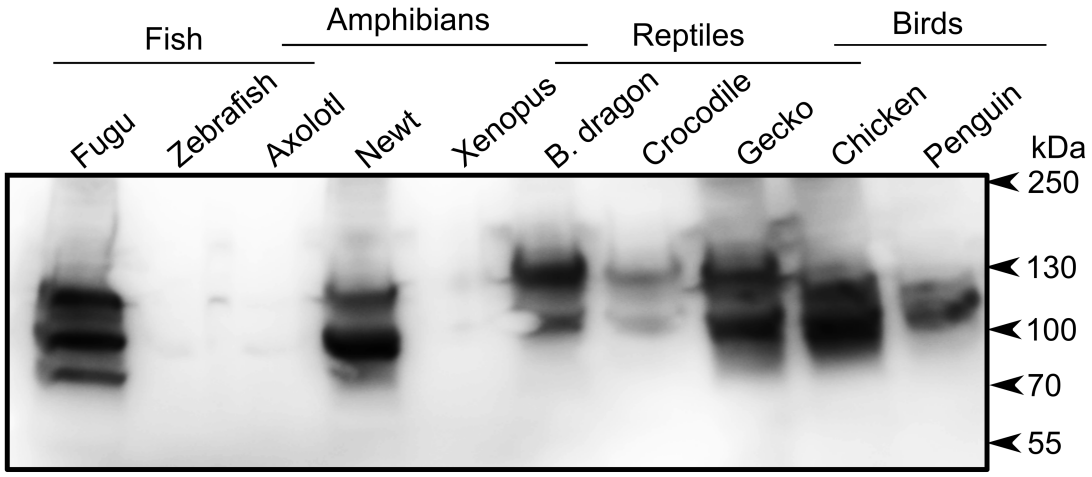


Figure 6.5: Western blot analysis of cloned TLR21 construct expression. Plasmids containing FLAG-tagged full-length *TLR21* from different species were expressed in HEK-Blue Null1 cells and analysed with an anti-FLAG western blot.

though after correction for multiple comparisons, the difference was not statistically significant (post-correction P-value, 0.12). ODN 2006 is one of the two CpG-ODNs that was found to stimulate zebrafish TLR21 in a previous study [657]. In addition, the lack of responsiveness to *E. coli* DNA in either of the constructs may also imply a generalised insensitivity (or high specificity) to CpG motifs. Despite their lack of signalling, both of the constructs were expressed at strong levels. Neither of the constructs responded to the TLR13 agonist ORN Sa19, suggesting these TLRs are distinct, as predicted from phylogeny.

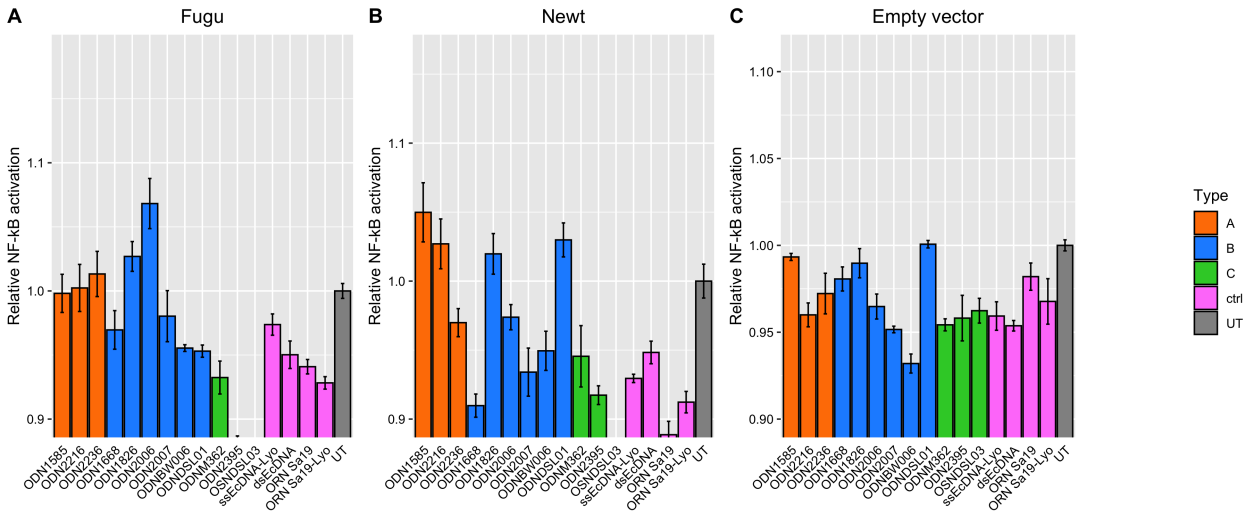


Figure 6.6: Specificity of TLR21 from a fish and an amphibian to a panel of CpG-ODN agonists. TLR21 from fugu and newt were challenged with a panel of CpG-ODNs from classes A, B and C, in addition to single- and double-stranded DNA from *E. coli* and ORN Sa19 (a mammalian TLR13 agonist), with and without Lyovec transfection reagent. Bars are presented as fold change relative to untreated ("UT") wells. One-sided (treatment greater than control) statistical differences were calculated using ANOVA with Dunnett's multiple comparison correction [153]. No comparisons were statistically significant following multiple comparison correction. Error bars represent standard error of the mean. NF-κB activation was normalised to untreated wells and is presented as relative numbers.

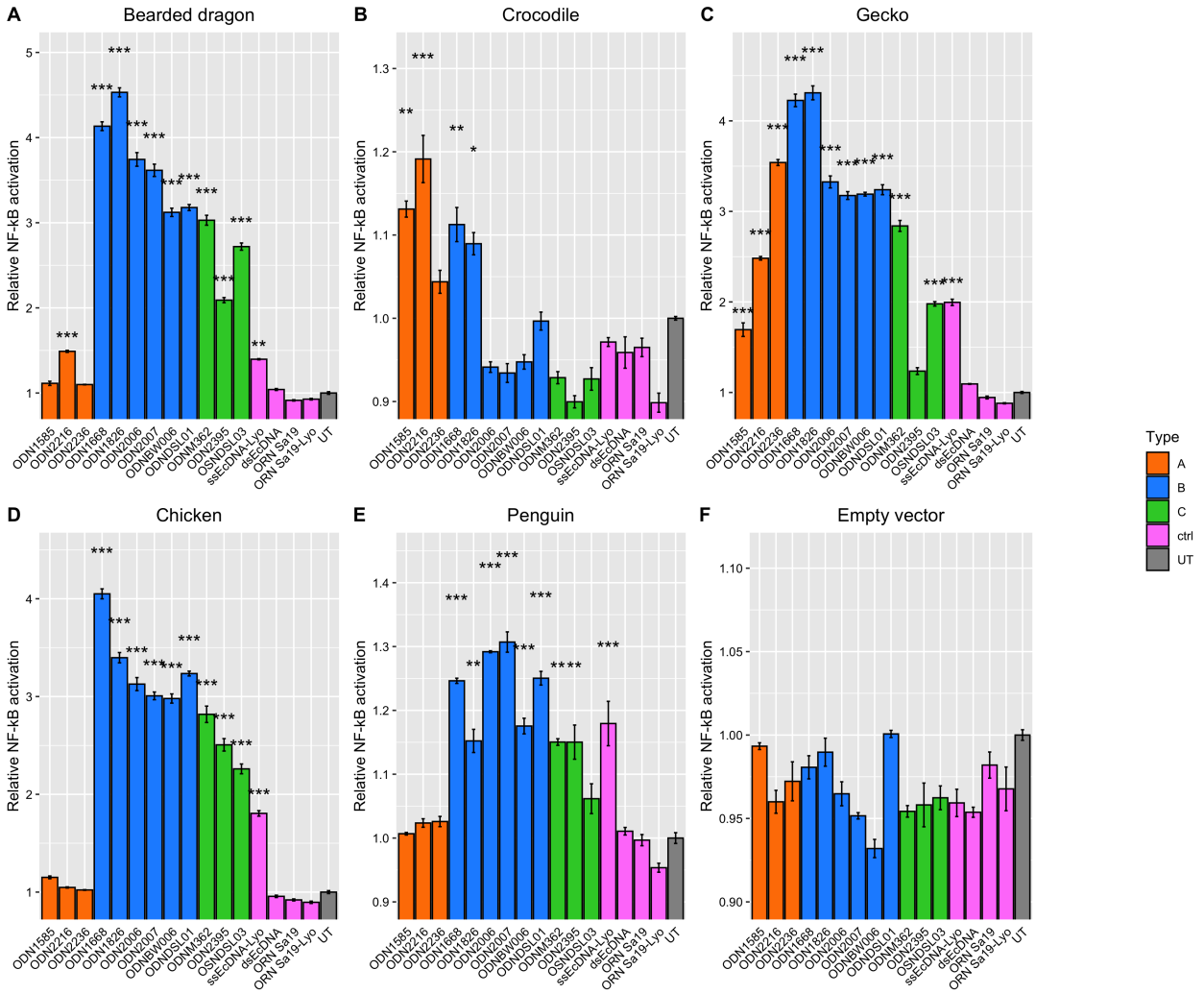


Figure 6.7: Specificity of TLR21 from birds and reptiles to a panel of CpG-ODN agonists. TLR21 from bearded dragon, crocodile, gecko, chicken and penguin were challenged with a panel of CpG-ODNs from classes A, B and C, in addition to single- and double-stranded DNA from *E. coli* and ORN Sa19 (a mammalian TLR13 agonist), with and without Lyovec. Bars are presented as fold change relative to untreated ("UT") wells. One-sided (treatment greater than control) statistical differences were calculated using ANOVA with Dunnett's multiple comparison correction [153]. Asterisks denote levels of statistical significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Error bars represent standard error of the mean.

Next, TLR21 in bird and reptiles species (which have lost TLR9) were tested (Figure 6.6). The magnitudes of response to agonists were substantially higher in the birds and reptiles compared to the fish and amphibians, with bearded dragon, gecko and chicken all exceeding four-fold enhancement of NF-κB activity with B-class CpG agonists. Reptiles appear to have little preference for particular classes of CpG, and responded to A-, B- and C-class CpGs (with the exception of crocodile TLR21 not responding to C-class CpGs). In contrast, birds showed a preference for B- and C-class CpGs, and neither of the birds tested responded to A-class CpGs. The increased promiscuity and heightened responsiveness of avian/reptilian TLR21 to CpG motifs is borne out by the observation that birds and reptiles, except crocodile (which had the lowest protein expression), all responded to single-stranded *E. coli* DNA, which is effectively a collection of different CpG motifs at low concentration. Taken together, these data suggest that TLR21 has undergone repurposing in the avian/reptilian lineage to be a CpG-recognising TLR with broad specificity. Similar to fugu and newt, none of the bird/reptile TLR21 responded to the TLR13 agonist ORN Sa19, further confirming the different origins of TLR13 and TLR21.

6.4.5 Signatures of positive selection in avian/reptilian *TLR21*

In order to trace the broadening of TLR21 CpG-recognition profiles in the avian and reptilian lineage to the nucleotide level, branch and branch-site models of positive selection were applied to multiple sequence alignments of the leucine-rich repeat (LRR) domain of *TLR21*. First, the adaptive Branch-Site Random Effects Likelihood (aBSREL) tool from Hyphy [547] was used. aBSREL detected positive selection in several lineages depicted in Figure 6.8. Most strikingly, positive selection appeared particularly prevalent in the birds, with the Galliformes and Neoaves displaying multiple rounds of selection. Positive selection in the Lepidosauria (lizards and snakes) was marginally non-significant after correction for multiple comparisons (uncorrected P-

value = 0.00039; corrected P-value = 0.089). Within the fishes, positive selection was detected along 5 branches, although typically selection was detected along terminal branches, suggesting that it is phylogenetically more recent than in the birds.

Next, a branch-site comparison was run in PAML (model = 2 NSsites = 2), specifying the birds and reptiles as the 'foreground' lineage, in which selection is explicitly tested (Figure 6.9). This analysis used the same LRR alignment as was used for aBSREL. Two models, one allowing omega (dN/dS) to vary (selection model), and another where omega is fixed (neutral model), were compared. The selection model was a significantly better fit than the neutral model (2ΔLnL = 10.065; P = 7.56×10^{-4}), indicating that bird and reptile *TLR21* has experienced positive selection. Most sites in the alignment were under either neutral or purifying selection in both background and foreground branches (87.3%), but the remaining 12.7% of sites were predicted to be under positive selection in the birds and reptiles (dN/dS, omega = 1.55), while being under neutral or purifying selection in the background branches. The majority of selected sites were found in helices between leucine-rich repeats (161S, LRR5/6; 236G, LRR8/9; 264H, LRR9/10; 285D, LRR10/11; 361R, LRR13/14; 689A, LRR26/27; 693, LRR26/27), while the remainder were located within the LRRs themselves (224Q, in LRR8; 599Q, in LRR23) Posterior probabilities for selected sites are given in Appendix Table E.2.

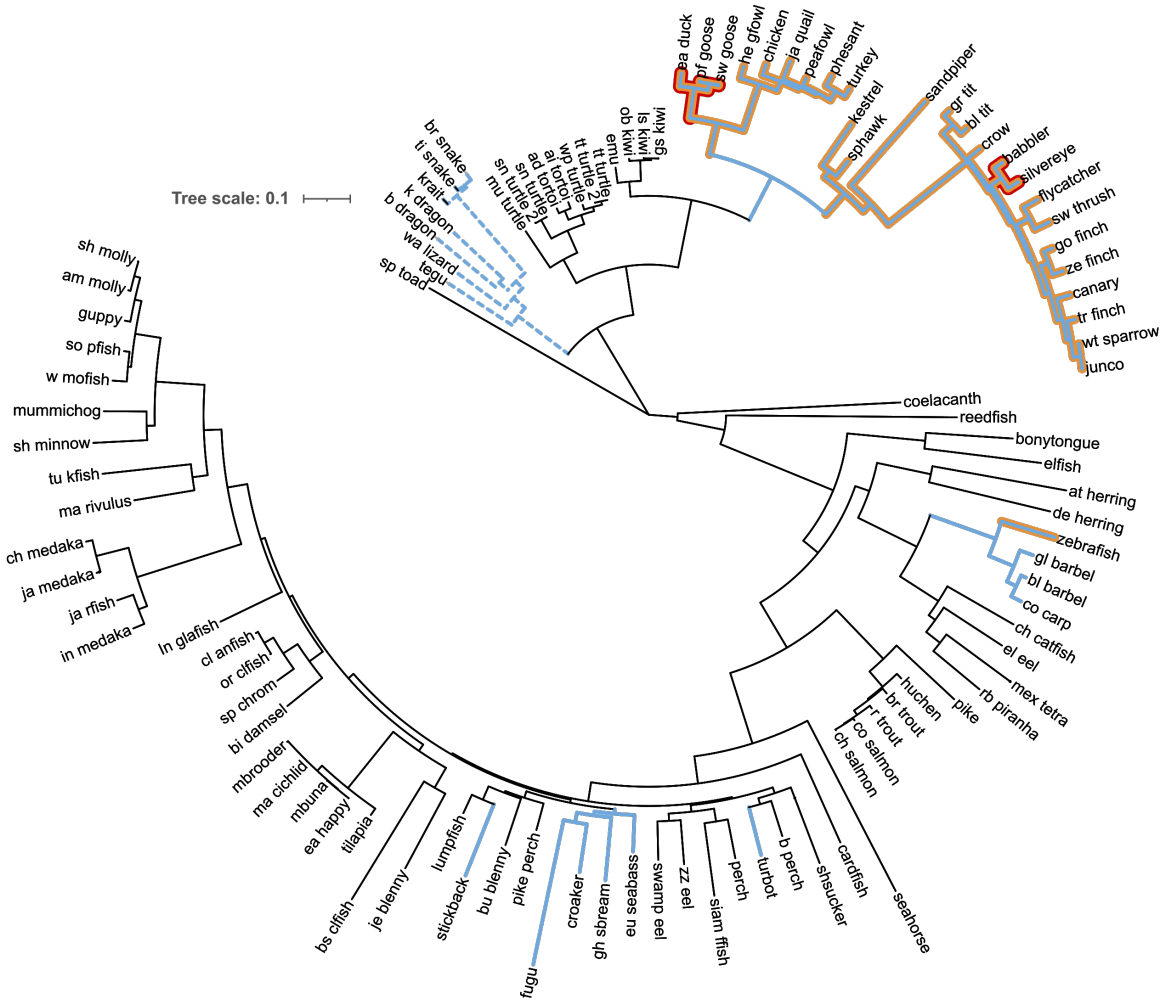


Figure 6.8: Strong positive selection in avian/reptilian *TLR21*. ABSREL [547] was used to test all branches in the *TLR21* phylogeny for evidence of positive selection. Lineages with evidence of positive selection P-value < 0.05 (following multiple comparison correction with the Holm-Bonferroni method) are highlighted on the tree (blue, orange and red). Evidence of recurrent selection is indicated by superimposed colours on branches. For instance, there was evidence of positive selection in the ancestor to the Neoaves and Galloanserae, with further selection at the base of the Galloanserae, and finally in the Anatidae. The marginally non-significant selection (uncorrected P-value = 0.00039; corrected P-value = 0.089) in the Lepidosauria is indicated by a dashed lined. Full metadata associated with shortened names in the figure are given in Appendix Table E.1

.

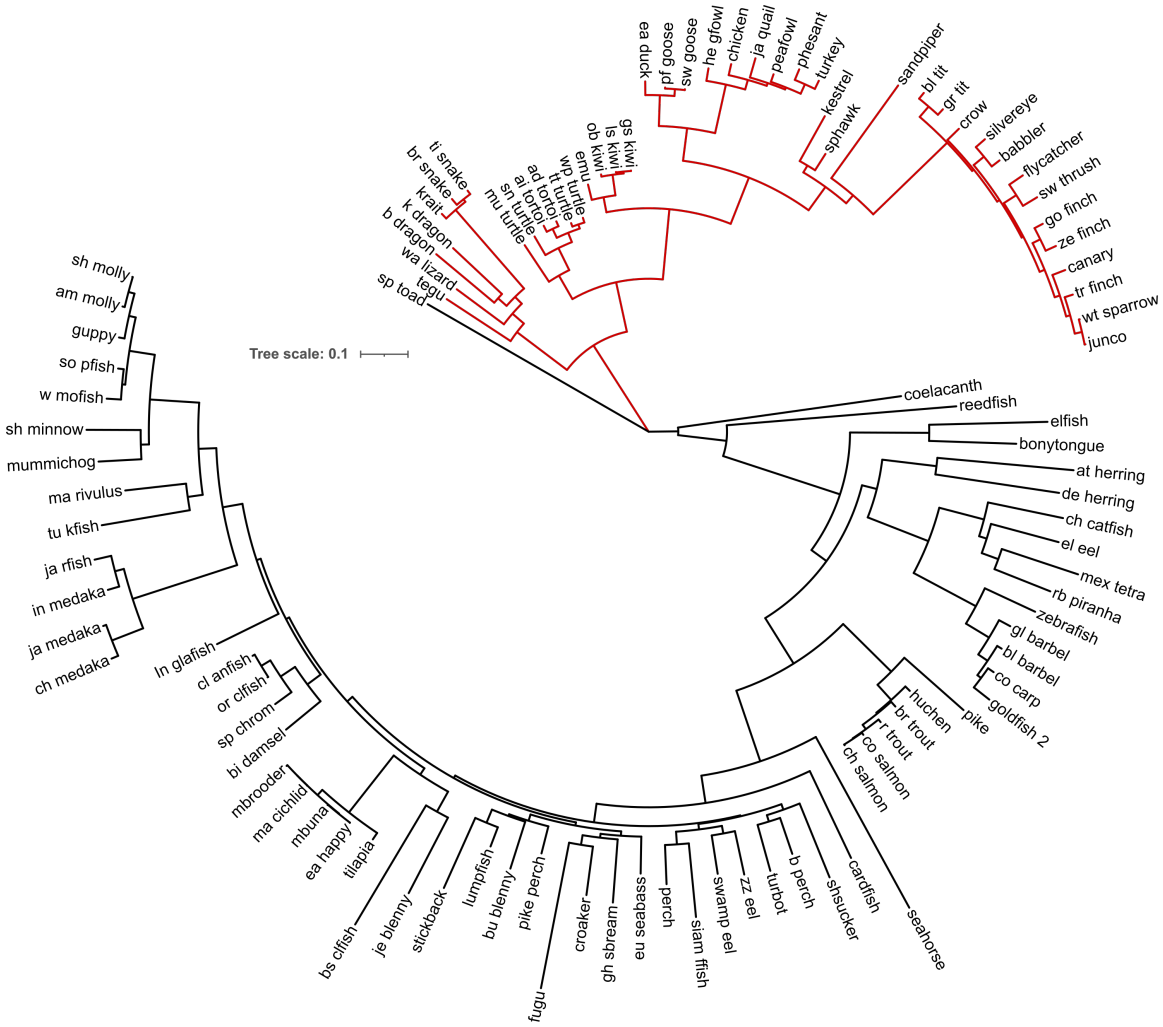


Figure 6.9: Test branches in *TLR21* phylogeny for branch-site model in PAML. The 'foreground lineage' (in which selection is formally tested by PAML) is highlighted in red. Full metadata associated with shortened names in the figure are given in Appendix Table E.1

.

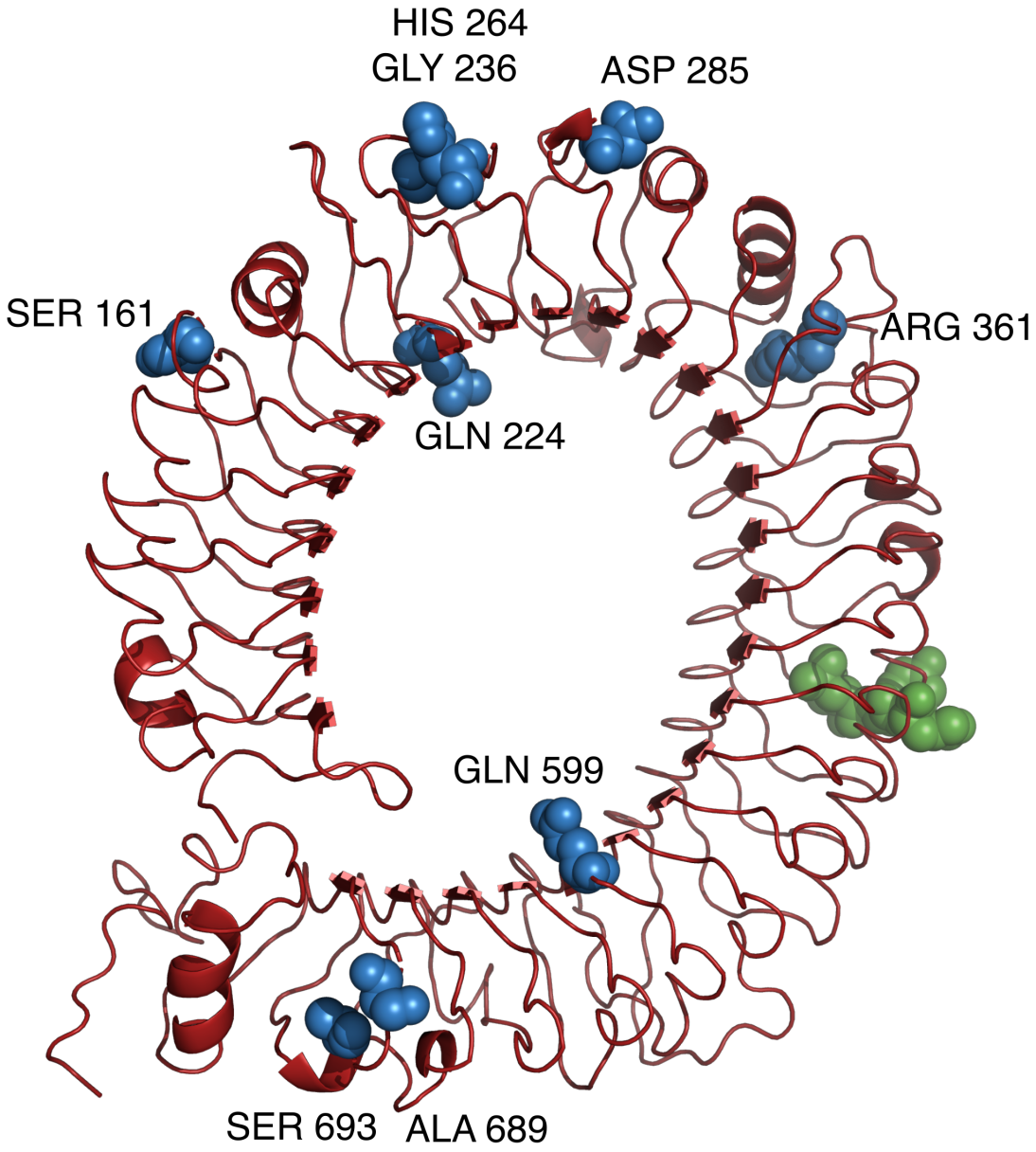


Figure 6.10: Homology model of the TLR21 ectodomain showing positively selected sites. The chicken TLR21 ectodomain is shown in red, positively selected sites in blue, and a putative DNA-binding site (D456, L457, Y458; [301]) in green.

6.5 Discussion

Animals can sense potential pathogens through a diverse repertoire of pattern-recognition receptors that recognise conserved PAMPs. The immunostimulatory properties of bacterial DNA were first noted using the nucleic acid fraction of *Mycobacterium bovis* (strain Bacillus Calmette-Guérin) on tumours [539, 585]. Further work noted that the immunostimulatory activity is only present when DNA contains unmethylated CpG dinucleotides [328, 500]. Unmethylated CpG dinucleotides are widely distributed in microbial genomes [375], but are comparatively scarce in vertebrate genomes [335], thus explaining why these motifs can be used to positively identify microbial DNA and avoid self-recognition. The two CpG-DNA-recognising Toll-like receptors—*TLR9* and *TLR21*—are phylogenetically unrelated, suggesting they each acquired the ability to detect these motifs independently [82, 493, 579].

6.5.1 *TLR21* structure and phylogeny

TLR21 was first discovered and characterised in the chicken [82, 301, 493, 579], and was found to be potently stimulated by CpG oligodeoxynucleotides (ODNs) ODN2006 and ODN2007. Using inhibitors of endosomal acidification, chloroquine and bafilomycin A, combined with confocal microscopy, TLR21 was found to localise to the endoplasmic reticulum and endolysosomal system and was dependent on endosomal acidification for its activity [82, 301], as are other nucleic acid-recognising TLRs [48]. Structurally, TLR21 has 27 leucine-rich repeats (LRRs) in its intraluminal domain, a single-span transmembrane domain, and a highly conserved Toll/interleukin-1 receptor (TIR) homology domain [301]. Although it shares no substantive homology to TLR9, chicken TLR21 also has two conserved amino acids in TLR9 implicated in DNA binding (Asp534 and Tyr536; [301, 508, 628]) but the functional relevance of these residues to TLR21 remains unknown. TLR21 lacks the z-loop between LRR14 and 15 which has been implicated in LRR domain cleavage and receptor activation [166, 333, 641], suggesting it has a different mode of activation to TLR9.

Phylogenetically, *TLR21* lies within the TLR11/12/13 clade of nucleic acid-recognising TLRs [81, 493, 641]. There has been some confusion regarding the relationship between *TLR13* and *TLR21*. *TLR13* is largely restricted to mammals, while *TLR21* is found in fishes, amphibians, reptiles and birds, but is absent in mammals. Some authors working with limited numbers of species have claimed that *TLR13* is an ortholog of *TLR21* [641]. The problem is further compounded by automatic genome annotation pipelines, which in many cases have mis-annotated *TLR21* in birds/reptiles/fishes/amphibians as *TLR13* (the closest TLR relative in mammalian species). More comprehensive phylogenetic analysis of the TLR11/12/13/21 family (this study) reveals that *TLR13* is a sister clade to *TLR11/12*, and the whole *TLR11/12/13* clade is a sister to the *TLR21* clade. The absence of the *TLR13* clade nested within the *TLR21* clade confirms that *TLR13* and *TLR21* are not orthologs, but rather ancient paralogs. This is further supported by their specificity for different ligands (typically Toll-like receptors retain specificity for a defined agonist class)—TLR21 responds to CpG DNA motifs, while TLR13 responds to bacterial RNA [251, 366]. The present study confirms that none of the TLR21 from the animals tested responds to the synthetic TLR13 ligand, ORN Sa19. Therefore, genomes in which *TLR21* has been mis-annotated as *TLR13* should be corrected.

6.5.2 CpG preference

It is well documented that mammalian TLR9 recognises a broad array of CpG motifs, although there is some species specificity in the CpG profile [327, 472, 490]. It follows that TLR21 may exhibit the same plasticity in ligand binding to adapt to different pathogens. Synthetic CpG ODNs have been widely used to investigate TLR9 responses in mammalian species. They are short, synthetic, single-stranded DNA molecules with a phosphorothioated backbone to prevent degradation. Four classes (A, B, C, and P) of CpG ODN have been described, based on their structure and biological effects on peripheral blood mononuclear cells (PBMCs) in mammals. Class

A CpG ODNs induce large amounts of IFN-α production and induce plasmacytoid dendritic cell (pDC) maturation, but have little effect on B cell activation. Class B CpG ODNs strongly induce B cell proliferation, as well as pDC and monocyte maturation, NK cell activation and cytokine production. Class C CpG ODNs can induce responses similar to both Class A and Class B, but more weakly. Class P CPG ODNs are characterised by their high capacity to induce IFN-α production and NF-κB activation [327, 333].

Chickens have previously been found to respond to a wide range of CpG-ODNs [107, 242, 643], but appear to have a preference for those in Class B and C [82]. The chicken TLR21 used herein confirmed a broad response to all CpG-ODNs from Class B and C. A similar pattern of preference for Class B and C was also demonstrated in the Northern rockhopper penguin (*Eudyptes moseleyi*), indicating that divergent avian TLR21 proteins display a Class B/C CpG preference. Reptile TLR21 CpG preference has not previously been documented. Interestingly, reptilian TLR21 may have an even broader CpG response profile than avian TLR21, with the central bearded dragon (*Pogona vitticeps*), saltwater crocodile (*Crocodylus porosus*), and Schlegel's Japanese gecko (*Gekko japonicus*) responding to Class A, as well as Class B and C CpG-ODNs (apart from saltwater crocodile, which did not respond to Class C CpGs). However, it is worth noting that the protein levels of crocodile TLR21 were relatively low compared with other reptilian TLRs, which may have underestimated the responsiveness of the receptor.

Several studies have documented the narrow CpG preference of fish TLR21, which primarily respond to Class B CPG-ODNs [107, 364, 657] and generally signals more weakly than TLR9 [657]. So far, no study has documented amphibian TLR21 CpG preferences. In the present study, there was a slight (but not statistically significant) preference of fugu (*Takifugu rubripes*) TLR21 for ODN2006, a B-class CpG ODN that has been previously shown to stimulate zebrafish TLR21 [657]. Newt TLR21 showed no strong preference for any CpG ODN tested, and the magnitude of response was very low. This suggests that, similar to fish TLR21, amphibian TLR21 is a weak

responder to CpG ODNs and may only display a slight preference to a certain CpG motif.

More insight into fish/amphibian TLR21 preference could be gained by successful testing of the remaining TLR21s that were cloned but failed to express at appreciable levels *in vitro*. TLR21 from zebrafish (*Danio rerio*), axolotl (*Ambystoma mexicanum*) and *Xenopus laevis* were not detected at the protein level following transfection. Coding sequences from all constructs were direct sequenced to ensure there were no aberrations that would prevent successful expression. Despite sequence confirmation, there are several reasons why the corresponding protein may not have been expressed. Firstly, at the nucleotide level, rare codons may prevent the coding sequence from being transcribed effectively. Coding sequences were generally not codon-optimised (except for the Northern rockhopper penguin, which needed to be codon-optimised to allow for synthesis) to make comparisons between PCR-derived and synthetic sequences fair. Secondly, the accessory protein UNC93 *Caenorhabditis elegans* homolog of B1 (UNC93B1) is known to regulate the cellular localization and targeting of TLR9 [5, 342, 343] and TLR21 [657], and the activity of zebrafish TLR21 in response to CpG-ODNs is enhanced by simultaneous expression of UNC93B1 [657]. Given that TLR21 from other species signalled effectively, it seems likely that human UNC93B1 was expressed and functional in the HEK-Blue reporter cells, but it could be that it does not interact efficiently with the constructs tested. Interestingly, it has recently been shown that UNC93B1 is required for maintenance of expression of intracellular TLRs [463], which could explain why TLR21 protein was not detected in zebrafish, axolotl and *Xenopus*. Finally, although parental vector stocks were used in the preparation of TLR21 constructs, it is feasible that some error was introduced into the plasmid during the PCR of the vector. Given the very high fidelity Q5 polymerase (NEB) was used, this seems unlikely but it is possible. Future work to fully sequence the entire plasmid and using q-RTPCR to determine mRNA expression levels of each coding sequence would help troubleshoot the lack of expression.

6.5.3 Signatures of selection

The different functionality of the successfully tested TLR21 in fishes/amphibians compared to birds and reptiles is stark. The fish/amphibian TLR21 (which presumably represents the ancestral state) responds only weakly to very few CpG ODN motifs, while the avian/reptilian TLR21 responds robustly to the majority of CpG ODN motifs tested. An obvious hypothesis is that the apparent repurposing of TLR21 as a broad CpG recogniser was coincidental with the loss of TLR9 (which is known to be a broad CpG recogniser in fishes [657]) in the lineage of bird and reptiles.

One way of predicting functional change in a gene is to analyse multiple sequence alignments for signatures of positive selection. Branch and branch-site analyses employed in this study unequivocally point to diversification of the leucine rich repeat (ligand-binding) domain of TLR21 in the bird/reptilian lineage. One branch in the avian/reptilian lineage where selection was conspicuously absent in *TLR21* was the chelonians (turtles; uncorrected and corrected P-value = 1). Interestingly, a recent study has identified *TLR9* in the genomes of at least some chelonians [149]. As an exception that may prove the rule, if chelonian *TLR9* is intact, they may possess a fish-like fastidious *TLR21*. Functional testing of chelonian TLR21 will provide additional insight into the co-evolutionary sharing of the CpG recognition space between TLR9 and TLR21.

Nine codon sites were found to be under positive selection in the birds/reptiles relative to fishes and amphibians, and the majority were located in helices linking leucine-rich repeats. Since the mechanism of activation of TLR21 has not been elucidated, it is not currently possible to determine which of these changes are most likely to have contributed to a change in function. The sites identified in the branch-site analysis do not correspond to the putative DNA-binding site identified by conservation with a TLR9 DNA-binding site Asp534 and Tyr536 [301, 508, 628], but given that TLR21 lacks the z-loop between LRR14 and LRR15 [333, 641] that is responsible for TLR9 dimerisation [166], it seems likely that TLR21 has a different mechanism of

action to TLR9. This is further suggested by the fact that TLR9 and TLR21 appear to be functional analogs of one another, rather than true homologs [82]. The sites identified as positively selected in birds and reptiles in the present study may give an indication of the putative DNA-binding sites in TLR21 and warrant further testing.

6.5.4 Conclusions

Toll-like receptors typically maintain specificity for a defined agonist class across large evolutionary distances [493]. This is despite the continual co-evolution with pathogens to counter subtle changes in ligand conformation [611]. Nucleic acids are unusual among the patterns recognised by the innate immune system because of their combinatorial diversity. CpG motifs can be flanked by many combinations of nucleotides, and thus different patterns. It is perhaps this wealth of possible diversity that drove the evolution of two CpG-recognising TLRs in the first instance; one TLR (TLR21) could afford to be a fastidious CpG recogniser because its counterpart (TLR9) is a CpG generalist. The loss of TLR9 in the avian/reptilian lineage may have imposed a selection pressure on the fastidious TLR21 to occupy more of the CpG recognition space. The result of this repurposing is the archetypal chicken TLR21, which is known to be a robust responder to a variety of different CpG motifs. This study aimed to track the journey of TLR21 from fastidious to generalist across evolutionary time. The ligand recognition profiles of the previously untested amphibian and reptile TLR21 were as predicted; a fish-like weak responder for amphibian TLR21 and a robust generalist response for reptile TLR21. Furthermore, this study documents the selective events that coincide with the repurposing of TLR21. Extensive positive selection was found, particularly in the birds, which supports the hypothesis that the function of TLR21 changed following the loss of TLR9. Taken together, these data support the hypothesis that TLR21 underwent functional reassignment in the birds and reptiles, and more generally, that there is a need to retain broad CpG recognition as part of the vertebrate PRR system.

Chapter 7

Characterisation of key anti-influenza effectors in the highly susceptible mute swan (*Cygnus olor*)

7.1 Abstract

Waterfowl are major reservoirs of influenza and migrating birds may represent a route of global influenza transmission. Among them, swans are particularly susceptible to infection with highly pathogenic avian influenza (HPAI). In recent years, a number of HPAI outbreaks have caused significant mortality at the managed population of mute swans (*Cygnus olor*) at the Abbotsbury Swannery (Dorset, UK). Mechanistically, it is not known why swans are more susceptible to influenza than other waterfowl. In this chapter, I sought to characterise various innate immune components in the swan that are known to form part of the anti-influenza immune response. For all genes (with the exception of IFN-β, which was not found in the current mute swan genome assembly), an intact coding sequence of appropriate size was found. This indicates that the susceptibility of the mute swan to influenza infection is not a result of a

missing innate immune component, which is true in the analogous case of the highly susceptible chicken, which lacks a critical anti-influenza effector.

7.2 Introduction

Influenza A and B viruses cause seasonal epidemics in humans, resulting in ∼500,000 deaths annually worldwide [384]. Additionally threatening the food chain, highly pathogenic avian influenza (HPAI) viruses are a major disease burden in poultry, with > 400 million birds killed and economic losses totalling US$20 bn in the period between 2000-2010 [223]. Wild birds, and particularly waterfowl, are major zoonotic reservoirs of influenza [627], and migrating birds may be at least partly responsible for its global spread [452]. The vast majority of the 16 haemagglutinin (HA) and 9 neuraminidase (NA) subtypes circulating in birds are low pathogenic (LPAI) and do not cause mortality. However, H5Nx and H7Nx subtypes can evolve into HPAI strains by acquisition of a novel cleavage site in HA [326].

Among the waterfowl, mute swans (*Cygnus olor*) appear to be highly susceptible to HPAI infection. Almost invariably during screening of wild bird populations in Europe, the highest rates of HPAIV-positivity and virus-positive deaths have been reported in swans [18, 21, 22, 77, 208, 577]. It has been suggested that this may be in part due to the conspicuousness of the swan, resulting in the species being overrepresented in sampling efforts and easily found carcasses [431, 581]. However, experimental evidence corroborates the increased susceptibility of the swan compared to other species. Experimental infection of waterfowl with HPAI indicates that swans are highly susceptible to infection [283], and that swans are at greater risk of mortality compared to geese [78]. Ducks, on the other hand, show minimal or no clinical evidence of infection [299, 318, 600]. They do, however, shed significant amounts of virus and are thought to act as long-distance vectors of HPAI [299]. There are therefore clear species-specific differences in host vulnerability and response to influenza among waterfowl.

There are also within-species differences in HPAI susceptibility, and this is likely to be as a direct result of the acquisition of adaptive immunological memory following previous exposure. The breadth of seropositivity to different AIV strains is known to increase with age in mute swans [255], and older birds in good health with pre-existing virus-specific antibodies are more resistant to experimental infection with HPAI [283]. Furthermore, the mortality age profile of swans in the Abbotsbury study population during HPAI influenza outbreaks have invariably been skewed towards younger birds [254, 255, 483]. This suggests that immunologically naïve young birds are at a greater risk of HPAI infection, and that the retention of adaptive immune memory likely forms a crucial part an effective response to influenza in swans.

Although adaptive immunological memory can explain the age profile of susceptibility in swans, it cannot explain the interspecific differences in influenza susceptibility in younger waterfowl and poultry species. In contrast to ducks, chickens are known to be extremely susceptible to infection with HPAI, with near-complete mortality within four days of experimental infection [272, 513]. The difference in susceptibility has been attributed to the retinoic acid-inducible gene I (*RIG-I*), a critical sensor of viral RNA that is present in ducks but absent in chickens [42]. Expression of duck RIG-I in chicken cells is sufficient to rescue IFN-β promoter activity and restrict growth of influenza virus [42]. Further evidence that innate immune function can directly influence susceptibility to influenza comes from genetic polymorphism associations with disease severity. For instance children with the rs5743313/CT polymorphism in *TLR3* are at increased risk of pneumonia following infection with A/H1N1/2009 influenza virus [164, 344], and the rs12252 T > C polymorphism in *IFITM3* was found to be associated with risk of severe influenza [480]. Additionally, polymorphisms in *RIG-I* [279], *IRF-7* [108], *IRF-9* [75, 249], *IL1-A* [378], *TNF-*α [17, 198] and *CCR5* [303], among others, are associated with increased influenza susceptibility in humans. Taken together, these observations indicate that the innate immune system can influence host susceptibility to influenza on different levels, (i) by the complete absence of critical innate immune components, e.g. *RIG-I* in chickens, and (ii) by changes in

function due to polymorphism.

The well-studied population of wild mute swans at the Abbotsbury Swannery (Dorset, UK) has experienced a number of severe HPAI outbreaks in the last 10-15 years. An HPAI H5N1 outbreak in Abbotsbury in December 2007–January 2008 killed 11 swans [483] An outbreak of HPAI H5N8 in December 2016–January 2017 killed approximately 180 swans, including almost half of the < 1 year-old birds [254]. HPAI H5N6 caused a further outbreak at Abbotsbury in January 2018 [19, 20] causing $\sim$60 swan deaths with a similar over-representation of year-old birds. Most recently, an outbreak of HPAI H5N8 was detected in > 80 mute swans in January 2021 [23] The frequency and severity of HPAI outbreaks at the Abbotsbury site are concerning. Swans co-habit the lagoon with large numbers of migratory waterbirds from several orders which begin to arrive from Europe and Asia in September and increase in number until November [483]. Migratory birds are highly likely to be the primary vectors into sites like Abbotsbury, and may have caused the numerous severe outbreaks of HPAI in the last decade.

Considering both the experimental data and epidemiological data, swans are highly susceptible to HPAI, far in excess of what could be considered observational biases due to their large size and conspicuousness. This chapter is, to my knowledge, the first attempt to characterise the mute swan innate immune system with respect to searching for anti-influenza effectors. I take a candidate gene approach to identify and characterise the key mediators of the innate immune response to influenza in swans by (i) MinION sequencing, (ii) whole genome sequencing[1], and (iii) making use of a recent publicly available mute swan genome from the Vertebrate Genomes Project.

[1]Originally, this chapter was also meant include a genome-wide association study, comparing swans known to have died in the 2018 H5N6 outbreak at Abbotsbury Swannery to those known to have been H5N6-positive and recovered. Despite having collected many samples in each of these categories which underwent H5N6-screening in a suitable containment facility at the Animal and Plant Health Agency (APHA), the samples later became unavailable. Having already arranged and purchased a portion of the sequencing, we elected to continue with resequencing 20 individuals from previously collected blood samples (by this time, the mute swan genome had been released by the Vertebrate Genomes Project). While these genomes were not of great use to the present study, apart from to confirm regions of poor assembly in the published genome, the genomes will be valuable to other projects and will be made available to the wider community.

The results indicate that mute swan has all of the critical components required to initiate an immune response to influenza. The observed high susceptibility of mute swans to influenza is therefore likely to be more subtle than a lack of an obvious immune component.

7.3 Materials and methods

7.3.1 Study population

Abbotsbury Swannery, on the Fleet Lagoon in Dorset, United Kingdom (50.6537 °N, 2.6028 °W) is an important study site of a large population of wild mute swans (*Cygnus olor*). Population sizes vary seasonally in accordance with immigration of swans to the site from Europe during the summer moult, but is typically between 500 and 1000 birds. Hatched cygnets are sexed and tagged within 24 hours of hatching; tags are replaced with adult leg rings at around 5 months of age. Additionally, swans are involved in biennial weighings where birds are caught, weighed, and sexed (where possible). Detailed information about hatch date, sex and pedigree are thus known for most birds. During influenza outbreaks at the lagoon, blood samples and cloacal swabs are taken by veterinarians in order to monitor infections. The blood samples used in the present study were collected for other purposes (e.g. serological assays in [255]).

7.3.2 PCR of *RIG-I*, *TLR3* and *TLR7*

As the mute swan genome was not available at the initiation of this project, primers to amplify mute swan *RIG-I*, *TLR3* and *TLR7* were designed on the swan goose (*Anser cygnoides*) sequence (Table 7.1; accessions: NW_013185741.1, NW_013185689.1, NW_013185780.1). PCR products were designed to be ∼2 kb in length such that they could be easily pooled in equimolar amounts for sequencing. All PCRs were carried out using the Phusion High-Fidelity polymerase (NEB) with

the following cycling conditions: 95 °C for 3 minutes; then 40 cycles of 95 °C for 30 seconds, annealing temperature (see Table 7.1) for 30 seconds, 72 °C for 2 minutes; then a final extension step of 72 °C for 7 minutes. PCR products were resolved on a 1% agarose gel to confirm the presence of bands of the correct size.

Table 7.1: PCR primer sequences and conditions for amplifying *TLR3*, *TLR7* and *RIG-I* genes

Gene	Primers	Length (kb)	Primer sequences (5' → 3')	PCR Ta (°C)
TLR3	E1_start_F + E2_end_R	1.63	GATTTCTTGTGCCTGGCTTG	63
			TGGGATAAGTCAGGATAGTTCC	
	E3_start_F + E3_end_R	1.96	CCAAATTTAGGCTTCACGGGT	58
			TTTCCCCTTACATACAGTGT	
	E3_mid_F + E4_end_R	2.43	GACTGTCCTGGACGTCAGCA	68
			TCAGAGCACATTTCCCTCATTC	
TLR7	194_F + 2425_R	2.20	TGTCAGGAGCTTGGTTTCCT	65
			GGGAAGCTAGATTTCTTAATTATTTGG	
	1096_F + full_R	2.16	AAGGAGATTGGAGATGCTCAGT	65
			AGGTTGTTGTAAGCCAGGGT	
RIG-I	E1_F + E2-3_R	1.60	TGTCAGTCACAAGCCAAAACAA	69
			AGGAGAAATGCAGTCTGATGGT	
	E4_F + E6_R	2.20	TCATTCACCTTGCCATCCCA	67
			CCCTGTCACCAGCTAAACCA	
	E7_F + E9_R	2.11	ACTTTCCTCTAGGGCTTGCTG	66
			TGTTGTAGCAGCCATGAGGA	
	E10-11_F + E12_R	1.86	TTCCAGCCTCAGCAGTTCAG	66
			TCATTTCCCATTGTAGGAGCTT	
	E13_F + E15_R	2.82	ACTCCCTGGTTTGCTTTCCA	65
			AGTAGATGAAGGAAAATGCACGT	
	E16_F + E17_R	2.33	GTCCCCAGATGCTACCTACC	65
			TGCACTAAAATGGTGGGTACAA	

7.3.3 MinION sequencing of PCR products

7.3.3.1 PCR cleanup using SPRI beads

SPRI beads were warmed to room temperature and vortexed. For a 25 μl PCR reaction, 12.5 μl beads were added for a 0.5× cleanup. The mixture was flicked and left to incubate at room temperature with occasional flicking to allow binding of the DNA. Following a brief centrifugation, the tubes were placed on a magnet for 5 minutes until the supernatant was completely clear. The supernatant was aspirated and discarded. The pellet was washed with 200 μl fresh 80% ethanol while still on the magnet, leaving for 2 minutes in case any beads were dislodged. The supernatant was removed and the wash step was repeated for a total of two washes. After removing the second wash, the tubes were spun briefly to allow for removal of additional ethanol by aspiration. Then, the bead pellet was left until just dry (without letting the pellet crack). As soon as the beads were dry, 25 μl nuclease-free water were added to elute the DNA. The tubes were flicked to resuspend the beads and were then left for 5 minutes at room temperature. Finally, beads were once again pelleted on the magnet and the DNA-containing supernatant was removed and retained.

7.3.3.2 PCR pooling

The DNA concentration of each PCR was measured by Qubit. PCRs were pooled according to the concentration, aiming for a total flow cell input of 800 ng DNA (6.06 ng target for the 11 PCRs for 12 birds). Since the PCRs were relatively similarly sized (∼2 kb), we assumed that mass was approximately equal to molar amount. PCRs were pooled by bird (12 pools) such that bird-specific barcodes could be ligated.

7.3.3.3 End-prep and barcode ligation

For the end-prep reactions, the following reactions were assembled for each of the 12 pools:

Table 7.2: Recipe for end-prep of PCRs.

Component	Volume (μl)
DNA (200-400 fmol)	25
Ultra II End Prep Reaction Buffer	3.5
Ultra II End Prep Enzyme Mix	1.5
Total	30

The reactions were incubated at room temperature for 10 minutes, and then at 65 °C for 10 minutes. Tubes were placed on ice for 30 seconds. Next, the following reactions were assembled, using the end-prepped reactions described above:

Table 7.3: Recipe for barcoding end-prepped PCRs.

Component	Volume (μl)
End-prepped DNA	30
A barcode (NBXX)	2.5
Ultra II Ligation Master Mix	40
Ligation Enhancer	1
Total	73.5

The barcoding reactions were incubates at room temperature for 20 minutes, and then at 65 °C for 5 minutes. A 0.5× SPRI bead cleanup was then performed on all samples as described in detail above. DNA from each pool was eluted in 2.5 μl Barcodes and associated sample metadata were as follows:

Table 7.4: Barcodes and associated metadata for birds included in the pilot experiment.

Barcode	Sample ID	Bird ID	Darvic ring	Sex	Hatch Date
BC01	A033	AS00895	WVXC	F	11/10/2016
BC02	A035	AS00472	WUPX	M	05/10/2014
BC03	A036	AS00830	WVSV	M	11/10/2016
BC04	A073	AS00595	WVBF	F	08/10/2015
BC05	A075	AS00499	WUTH	F	20/10/2014
BC06	A083	AS00870	WVUX	M	11/10/2016
BC07	A194	AS00805	WVPO	M	21/09/2016
BC08	A195	AS00488	WVJC	M	20/10/2014
BC09	A196	AS00941	WVZK	M	11/10/2016
BC10	A197	AS00931	WVYY	M	11/10/2016
BC11	A242	AS00915	WVYC	F	11/10/2016
BC12	A244	AS00691	WVIC	F	12/10/2015

7.3.3.4 Barcode adaptor ligation

Each 2.5 μl eluate from the previous step was combined and made up to 30 μl by adding nuclease-free water. Next, barcode adaptors were ligated by mixing 30 μl pooled barcoded sample, 20 μl Barcode Adaptor Mix (BAM) and 50 μl blunt TA ligase (2× MM). The mixture was gently mixed by flicking and incubated at room temperature for 10 minutes. DNA cleanup was performed using a similar protocol as the previous 0.5× cleanup, except ABB buffer was used in place of EtOH during washes. Barcoded and adaptor-ligated DNA was eluted in 15.5 μl elution buffer.

7.3.3.5 Flow cell loading, sequencing and data processing

Using the purified eluate, a MinION flow cell was loaded according to the manufacturer's instructions. Sequencing was performed without basecalling for 48 h using MinKnow 1.7.7. Raw files were basecalled using Albacore v. 1.2.5 (Oxford Nanopore Technologies), demultiplexed and trimmed using Porechop, and then mapped using BWA to a reference sequencing comprising concatenated references of the three genes, interspersed with 1 kb of Ns.

7.3.4 BUSCO analysis of swan genome

In order to assess the completeness of the publicly available mute swan genome, the Benchmarking Universal Single-Copy Orthologs (BUSCO; v. 5.1.2) analysis pipeline was used [532]. The curated mute swan genome is available from the Vertebrate Genome Project (https://vgp.github.io/genomeark/Cygnus_olor/; last accessed June 2021) or from GenBank (GCA_009769625.1). For BUSCO analysis, two datasets—Eukaryota and Aves—were used to compare against the swan genome.

7.3.5 Genomic DNA extractions

Three commercially available DNA extraction kits were tested to extract genomic DNA from mute swan blood samples. The Invitrogen/ThermoFisher PureLink gDNA

mini kit, Bioline Isolate II gDNA kit and Zymo Quick-DNA Miniprep Plus kit were compared. Protocols were followed according to the manufacturer's instructions except for the initial proteinase K digestion, which was increased to use 40 μl proteinase K for 1 μl whole blood, and the digestion was left to incubate overnight. This is a common and necessary alteration to DNA extraction protocols when using nucleated avian erythrocytes (e.g. [660, 661]). DNA was eluted twice from columns in all cases. gDNA yields were measured using a Qubit fluorometer (ThermoFisher), purity was measured using a Nanodrop 2000 (ThermoFisher), and DNA integrity was assessed on an agarose gel by comparing against the Quick-Load 1 kb Extend DNA Ladder (NEB). $A_{260}/_{280}$ values between ∼1.8 and ∼2.0 were considered pure DNA.

Giving the best yields and acceptable purity and integrity, the Invitrogen/ThermoFisher PureLink gDNA mini kit was taken forward for processing of the remaining samples. Metadata for samples included in the final set are given in Table 7.5.

7.3.6 Whole-genome sequencing

Genomic DNA samples underwent TruSeq DNA PCR-free gel-free library preparation (350 bp insert). Libraries were sequenced using NovaSeq 150 bp paired-end sequencing (yielding >750M + 750M reads; Edinburgh Genomics). Basecalling and quality control was performed by Edinburgh Genomics. Total output was ∼225 Gbp (∼11 Gbp per sample, equating to 10× genomes). Reads also underwent in-house quality control using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed June 2021).

7.3.7 Read mapping

Reads were mapped to the publicly available mute swan genome described above using BWA `aln`/BWA `sampe`. Duplicate reads were removed using samtools rmdup, and mapped reads with mapping quality ⩾30 were taken forward for further analysis.

Table 7.5: Metadata for swan whole-genome samples.

Sample	Conc. [1]	Darvic ring	Bird ID	Sex	Hatched	Died	Influenza notes
A036	36.8	WVSV	AS00830	M	2016	16/01/18	
A166	34.0	WVNF	AS00758	M	2016	N/A	H5 titres, spinning [2]
A375	18.7	WXSI	AS01154	F	2017	15/01/18	HPAIV+ [3]
A493	67.8	WXIO	AS01020	F	2017	25/01/18	HPAIV+
A498	54.2	WXON	AS01118	F	2017	19/01/18	HPAIV+
A519	35.2	WXJV	AS01044	M	2017	N/A	
A523	19.9	WXIB	AS01010	M	2017	N/A	
A525	28.4	WXNU	AS01103	F	2017	N/A	
A545	32.4	WXIF	AS01013	M	2017	N/A	
A549	21.0	WXHC	AS00991	F	2017	N/A	
A562	24.8	WUUJ	AS00520	F	2014	N/A	HPAIV+, survived
A564	24.8	WXLF	AS01073	M	2017	N/A	
A580	25.8	WXHO	AS01000	F	2017	N/A	
A632	48.4	WXND	AS01091	M	2017	N/A	
W404	84.8	W404	Z24999	M	1980 [4]	ND	
W408	63.2	W408	Z24974	M	1980	ND	
W414	53.4	W414	Z24980	F	1980	ND	
W426	66.6	W426	Z45512	ND	1980	ND	
W441	56.6	W441	Z45527	ND	1980	ND	
W447	41.8	W447	Z45533	F	1980	ND	

[1] Conc. = ng/μl

[2] Spinning in swans is a pathognomonic indicator of HPAI infection, presumably due to neurological effects of the virus.

[3] Although samples taken at the time of the 2018 outbreak became unavailable, previously collected samples for four birds that were known-HPAIV+ at the time of the outbreak were used. Only one of these birds, A562 (an older bird), survived the subsequent HPAI outbreak.

[4] In addition to modern birds, six birds with a 1980 hatch-year were included in the samples for whole-genome sequencing.

N/A, not applicable; ND, not determined.

Read depth was calculated using `bedtools genomecov` and is given in Table 7.6:

Table 7.6: Calculated read depth of swan whole genomes

Bird	Read depth
A036	11.15
A166	11.91
A375	12.64
A493	14.60
A498	12.08
A519	13.31
A523	11.58
A525	9.38
A545	10.55
A549	6.48
A562	8.51
A564	8.78
A580	9.95
A632	8.50
W404	10.46
W408	7.82
W414	10.76
W426	6.66
W441	6.98
W447	10.10

7.3.8 Candidate gene identification

Thirty-three key anti-influenza genes were identified (Table 7.7). Using genomic query sequences from the swan goose (*Anser cygnoides*; Table 7.7), BLASTn searches were performed for each of the 33 putative anti-influenza genes, using the publicly available mute swan genome as a reference. For each gene, hits of $E \approx 0$ were considered to be almost certain orthologs, and genomic regions 10 kb either side of the hit were extracted from the swan genome. Swan and goose orthologs were aligned along with the known mRNA sequence of the goose gene to identify exon structure using MAFFT [294]. Using the likely exon structure of the mute swan gene, coding sequences were extracted and examined for length and presence of any putative disruptive mutations (frameshifts or premature stop codons). On occasions where the mute swan reference genome had poor quality assemblies in the region of the gene of interest, the equivalent regions were extracted from the 20 resequenced mute swan genomes to resolve the

true sequence. In all instances where a putative pseudogene was identified in the reference genome, comparison with the resequenced genomes revealed the true, intact sequence.

Table 7.7: Identification of candidate anti-influenza genes.

Gene product	Function	Evidence for influenza
RIG-I	PRR; Binds dsRNA; signals via MAVS to activate the type I IFN response	[42, 279]
MDA5	PRR; Binds dsRNA; signals via MAVS to activate the type I IFN response	[371, 382]
LGP2	PRR/Regulatory; Binds dsRNA; downregulates RIG-I responses	[543]
ZAP	PRR; Degrades viral mRNAs via recruitment of ribonucleases	[238, 671]
TRIM25	PRR/Regulatory; Regulates RIG-I via K63-linked ubiquitination; Targets viral RNA synthesis	[319, 404]
Riplet	Regulatory; Regulates RIG-I via K63-linked ubiquitination	[241, 453, 487]
OASL	PRR; Degrades viral RNA in combination with RNase L	[674]
MAVS	Signalling; Activates transcription factors and upregulates pro-inflammatory genes	[621]
IRF7	Regulatory; Transcription factor responsible for upregulation of type I IFN	[108, 584]
STING	PRR/Signalling; Cytosolic nucleic acid sensor; Stimulates IFN genes	[259]
NALP3 (NLRP3)	PRR; Detects signals of cell damage, e.g. extracellular ATP	[575]
TLR3	PRR; Binds dsRNA and signals via TRIF to produce type I IFN	[340, 368]
TRIF	Signalling; Adaptor protein for TLR3 (exclusively) and TLR4	[541]
TLR7	PRR; Binds ssRNA and signals via MyD88 to produce type I IFN	[284]
HDAC11	Regulatory; down-regulates expression of IL-10	[439]
LSm14A	PRR; Binds nucleic acids and mediates IRF3 activation and type I IFN induction	[165, 367]
IL-1B	Cytokine; Important mediator of the pro-inflammatory response	[198, 307, 378]
IL-18	Cytokine; Induces type II IFN responses	[141, 372]
RSAD2 (Viperin)	Effector; Produces chain-terminating nucleotides; Disrupts lipid rafts and inhibits viral budding	[154, 622]
IFIT5	PRR; Binds and initiates degradation of uncapped (viral) RNA	[498]
CCL4	Chemokine; chemoattractant which functions to attract NK cells, monocytes and other immune cells	[510, 514]
EIF2AK2 (PKR)	PRR; Binds dsRNA and induces cellular apoptosis	[38, 363]
IL4I1	Regulatory; negatively regulates T cell activation and proliferation	[32]
USP18	Regulatory; negatively regulates type I IFN	[302]
EPSTI1	Regulatory; involved in M1 macrophage polarization	[632, 664]
CMPK2	Regulatory; involved in macrophage differentiation	[263]
CCL19	Chemokine; binds CCR7 and functions as a chemoattractant	[180]
CCL21	Chemokine; binds CCR7 and functions as a chemoattractant	[180]

Continued on next page

Table 7.7: Identification of candidate anti-influenza genes.

Gene product	Function	Evidence for influenza
MX	Effector; Decreases influenza NP synthesis and blocks endocytosis of viral particles	[165, 316, 623]
IFITM3	Effector; Inhibits viral entry to host cell by disrupting intracellular cholesterol homeostasis	[552, 631]
IFN-α	Effector; Type I interferon, important anti-viral cytokine	[305]
IFN-β	Effector; Type I interferon, important anti-viral cytokine	[305]
IFN-γ	Effector; Type II interferon, important anti-viral cytokine	[85]

7.4 Results

7.4.1 Pilot experiment: sequencing of *TLR3*, *TLR7* and *RIG-I*

As a pilot experiment, some of the most likely genes involved in the immune response to influenza were sequenced in 12 swans using MinION technology. Three genes (*TLR3*, *TLR7* and *RIG-I*) were fully sequenced by generating PCR products covering the entire coding sequence. TLR3 is a sensor of double-stranded RNA, and polymorphisms are known influence susceptibility to influenza [164, 344]. TLR7 is a single-stranded RNA recogniser, and is presumed to be involved in the recognition of GU-rich regions of the ssRNA influenza genome [246, 284]. RIG-I is a crucial sensor of intracellular RNA, and is known to be a critical determinant of influenza susceptibility in waterfowl [42].

MinION sequencing of mute swan *TLR3*, *TLR7* and *RIG-I* was effective, and all amplified regions had a sequence depth of ∼200-fold. Analysis of open reading frames indicated that the three genes are present as full-length nucleotide sequences in the swan and are therefore likely to yield full-length proteins. Within the 12 swans, polymorphisms were sparse in the coding regions of each of the genes: three polymorphisms were identified in *RIG-I*, three in *TLR3*, and one in *TLR7* (Table 7.8). Of these, only one polymorphism in each of *TLR3* and *TLR7* were non-synonymous (T3783G and T2013C). Both of these polymorphisms were present in only one bird, and T2013C was a heterozygous change. These data point to a high level of conservation within swan *TLR3*, *TLR7* and *RIG-I* and thus these genes are highly unlikely to contribute to the variation in innate immune response to HPAI.

Table 7.8: Single-nucleotide polymorphisms identified in pilot experiment candidate genes within the swan population.

Gene	Exon	Position	Change	S/NS	AA Change	Birds Affected
RIG-I	5	3818	T→C	S	-	A075, A197
RIG-I	10	9820	A→C	S	-	A073 A197
RIG-I	11	10171	T→C	S	-	A073
TLR3	1	323	T→C	S	-	A194
TLR3	3	3701	A→G	S	-	A194
TLR3	3	3783	T→G	NS	S→A	A033
TLR7	2	2013	T→C[1]	NS	Y→H	A244

[1] Heterozygote.

Polymorphisms in the swan sequences with respect to the swan goose (*Anser cygnoides*) *TLR3*, *TLR7* and *RIG-I* sequences were also analysed (Table F.1, Appendix F). This could provide an insight into a change in immune function in the HPAI-susceptible swan with respect to the comparatively HPAI-resistant goose. Swan goose and mute swan sequences exhibited very high nucleotide identity (typically >98%). In total, there were 26 polymorphisms in *RIG-I*, 46 in *TLR3*, and 20 in *TLR7*. Of these, 11 polymorphisms in *RIG-I* were non-synonymous, 17 in *TLR3*, and 6 in *TLR7*. Two sites in *TLR3* were proximal to sites of known function: G355A causes an alanine to threonine substitution at residue 119 which is adjacent to a residue involved in dsRNA binding [100], while T497G causes a leucine to tryptophan substitution which is within 5 Å of a dsRNA binding site [100, 603]. No polymorphisms in *TLR7* or *RIG-I* were proximal to sites of known function.

7.4.2 Completeness of published mute swan genome

During the early stages of this chapter, a genome for the mute swan was published by the Vertebrate Genomes Project [492]. To determine whether the newly released mute swan genome assembly could be used as a reference for the chapter, the Benchmarking Universal Single-Copy Orthologs (BUSCO) method [532] was used. BUSCO is a measure of genome completeness that, unlike the existing N50 measure, is not dependent on contig length. Instead, BUSCO scans assembled genomes for the pres-

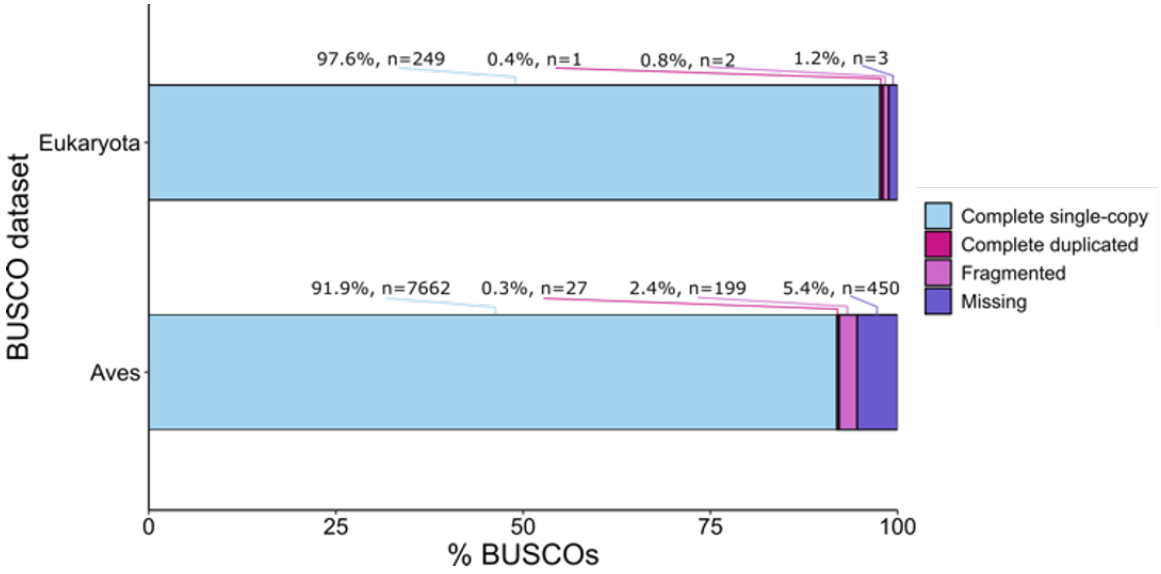


Figure 7.1: BUSCO analysis of completeness of the publicly available mute swan genome. The assembled mute swan sequence was analysed against two datasets of BUSCO genes (Eukaryota and Aves).

ence of a set of genes that are single-copy orthologues in >90% of other genomes. The mute swan genome was compared to two BUSCO datasets, firstly the Eukaryota and secondly the Aves (Figure 7.1). Both datasets indicated that the mute swan genome was of high completeness: for the Eukaryota dataset, 97.6% of genes were complete single-copy and for the Aves dataset, 91.9% were complete single copy. For reference, the human genome in 2015 was 89% complete with respect to the Eukaryota dataset [532]. These data indicate that the publicly available mute swan genome is of good quality and can be used to identify swan immune genes.

7.4.3 Preparation of genomic DNA for resequencing genomes

To prepare genomic DNA for sequencing, three commercially available kits were compared for DNA yield, purity, and integrity. Since avian erythrocytes are nucleated, only 1 μl of whole blood was used for DNA extraction for each bird. Additionally, in each protocol tested, the amount of proteinase K was increased (from 20 μl to 40 μl) and the length of time for digestion was also increased from 10 minutes to

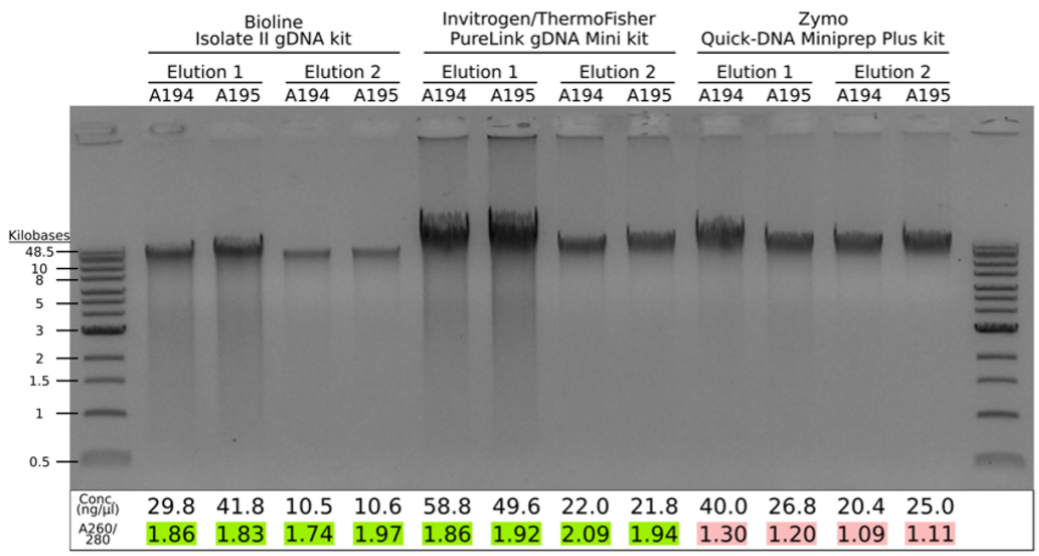


Figure 7.2: Testing three commercially available genomic DNA extraction kits. Kits were tested on two swan blood samples (1 μl of whole blood used as the input for each kit). Protocols were carried out according to the manufacturer's instructions, except where indicated in the text. DNA integrity was assessed on a 1% agarose gel. DNA concentration was measured using a Qubit fluorometer and DNA purity was assessed using a Nanodrop 2000. Measurements were made for two sequential elutions from the extraction column. The gel used 5 μl of input material.

overnight. Of the three kits tested, the Invitrogen/ThermoFisher PureLink gDNA mini kit was the superior, giving the greatest yields, high purity, and high integrity on a gel (Figure 7.2). Purity for the Bioline Isolate II gDNA kit was acceptable, but yields were the lowest of the tested kits. Yields were better for the Zymo Quick-DNA Miniprep Plus kit, but purity of the samples was unacceptable. Using these conditions DNA was also isolated from an additional 25 birds and presence of high-quality DNA was confirmed on a gel (Figure 7.3). Of these, 20 samples were taken forward for whole-genome sequencing.

7.4.4 Candidate gene finding in the mute swan genome

Since the number of available DNA samples was lower than originally planned, and were not delineated by survivor/non-survivor in one of the HPAI outbreaks, we elected

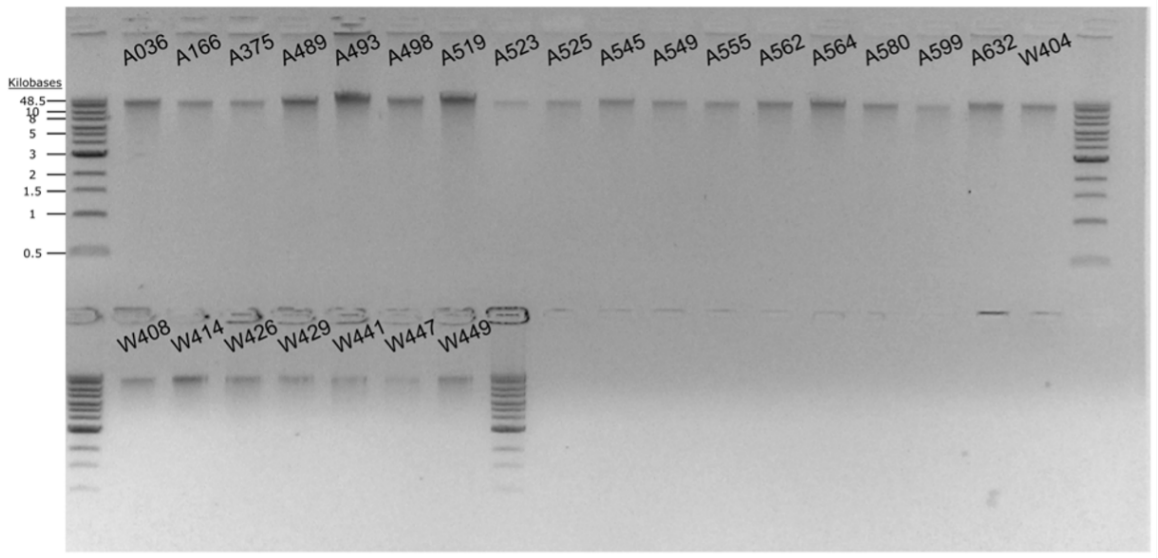


Figure 7.3: Extraction of genomic DNA from additional birds for whole-genome sequencing. The Invitrogen/ThermoFisher PureLink gDNA Mini kit was used according to the above-described protocol. The gel used 1.5 μl of input material.

to perform a candidate gene search for key genes known to be involved in anti-influenza immunity. The resequenced genomes were used to resolve potential errors in the publicly available swan genome. The swan genome was searched for 33 candidate genomes that are known to form the major axis of the anti-influenza response (Table 7.9). The genes include critical pattern-recognition receptors, cytokines, chemokines, anti-viral effectors, signalling molecules, and regulatory components. For each gene, the genomic sequence was identified in the goose (either the swan goose, *Anser cygnoides*, or the Pink-footed Goose, *Anser brachyrhynchos*) and this sequence was used as a query to BLAST the mute swan genome (https://vgp.github.io/genomeark/Cygnus_olor/). Due to poor assembly in the goose genomes, one gene—*ZAP* (*ZC3HAV1*)—used the duck sequence. The mute swan coding sequence was identified for each gene based on homology with the known coding sequence for the goose (or duck). The coding sequence was then analysed for length and the presence of any putative disabling mutations that may be disruptive to the anti-influenza immune response.

Table 7.9: Candidate gene analysis results.

Gene	Goose	Intact	Swan CDS length	Goose CDS length	ID Goose	Swan scaffold	Swan coordinates
RIG-I	ENSACDG00005011094	Y	2805	2805	98.895	SUPER_scf_60	80621272-80648877
MDA5	ENSACDG00005010923	Y	3015	3015	99.171	SUPER_scf_6	22520606-22548413
LGP2	ENSACDG00005008326	Y	2028	2028	98.028	SUPER_scf_28	2152057-2157365
ZAP[1]	KR025551.1	Y	2145	2145	96.177	SUPER_scf_1	72018808-72043611
TRIM25	ENSACDG00005001115	Y	1989	1989	98.013	scf_219	17014-17790
Riplet	ENSACDG00005000967	Y	1131	1131	99.469	SUPER_scf_18	8845664-8850849
OASL	ENSACDG00005000043	Y	1386	1386	96.609	SUPER_scf_10	9211547-9214274
MAVS	ENSACDG00005003614	Y	1914	1911	93.646	SUPER_scf_3	74724941-74733318
IRF7	ENSACDG00005015487	Y	1524	1524	97.9	SUPER_scf_5	46399106-46404000
STING	ENSACDG00005008888	Y	1149	1149	98.52	SUPER_scf_14	11943829-11949366
NALP3 a.k.a NLRP3	ENSACDG00005006748	Y	2880	2880	97.396	scf_118	41609-42305
TLR3	ENSACDG00005012587	Y	2688	2688	98.289	SUPER_scf_3	20069903-20078481
TRIF	ENSACDG00005013928	Y	2211	2211	96.161	SUPER_scf_27	3032621-3035978
TLR7	ENSACDG00005001824	Y	3144	3144	99.205	SUPER_scf_1	130457848-130467648
HDAC11	ENSACDG00005011025	Y	984	984	98.882	SUPER_scf_10	3405699-3435531
LSm14A	ENSACDG00005003043	Y	1386	1386	99.495	SUPER_scf_12	10151794-10172188
IL-1B	ENSACDG00005018484	Y	822	822	97.689	SUPER_scf_31	921961-924393
IL-18	ENSACDG00005012860	Y	594	594	98.99	SUPER_scf_24	2175749-2179433
RSAD2 (Viperin)	ENSACDG00005014756	Y	1158	1158	98.446	SUPER_scf_2	17664393-17671333
IFIT5	ENSACDG00005006844	Y	1440	1440	98.403	SUPER_scf_7	20206591-20211315
CCL4	ENSACDG00005010397	Y	282	282	98.936	SUPER_scf_20	533428-535683
EIF2AK2 (PKR)	ENSABRG00000018144	Y	1671	1668	96.709	scf_106	44092-45482
IL4I1	ENSACDG00005011260	Y	1536	1536	96.615	SUPER_scf_45	621642-627331
USP18	ENSACDG00005000018	Y	1131	1131	97.19	SUPER_scf_1	66667640-66684226
EPSTI1	ENSACDG00005002475	Y	855	855	96.956	SUPER_scf_1	177998012-178043530
CMPK2	ENSACDG00005014749	Y	958	1281	97.747	SUPER_scf_2	17669674-17682718
CCL19	ENSACDG00005018832	Y	300	300	96	SUPER_scf_60	9915486-9918479
CCL21	ENSACDG00005018833	Y	336	336	97.024	SUPER_scf_60	9922949-9925289

Continued on next page

Table 7.9: Candidate gene analysis results.

Gene	Goose	Intact	Swan CDS length	Goose CDS length	ID Goose	Swan scaffold	Swan coordinates
MX	ENSAPLG00000015976	Y	2274	2274	97.845	SUPER_scf_1	115384300-115403734
IFITM3	ENSABRG00000010112	Y	426	426	98.122	SUPER_scf_5	31966326-31968390
IFN-α	ENSACDG00005018759	Y	581	581	98.264	SUPER_scf_78	2117095-2117679
IFN-β	ENSABRG00000016262	N/A	-	-	-	-	-
IFN-γ	ENSABRG00000022157	Y	495	495	98.384	SUPER_scf_1	37770947-37775494

[1] Duck sequence.

For the majority of genes, one clear swan ortholog to each goose sequence was identified (Table 7.9). Furthermore, the nucleotide identity between the swan and the goose coding sequences was very high, often in excess of 97%. It initially appeared that the swan had disrupted copies of three genes (*IFN-α*, *TRIF* and *TRIM25*) based on the sequence in the publicly available genome. There were numerous disruptions to the coding sequences of these genes, which, if correct, would have yielded a truncated protein product. However, when the coding sequences for these genes were retrieved from the 20 genomes generated in the present study, intact sequences were found. This likely reflects poor quality regions of assembly in the publicly available genome that require correction. To investigate whether the Abbotsbury swans had disrupted copies of the immune genes, each coding sequence identified in the publicly available genome were also corroborated with the 20 Abbotsbury genomes. No disrupted genes were found in the Abbotsbury swans, likely indicating that the main components of the anti-influenza immune axis in swans are intact.

One gap in the results was interferon beta (*IFN-β*). *IFN-β* could not be identified in either the publicly available genome or the Abbotsbury swan genomes. Instead, the relatively closely related *IFN-α* genes were returned from BLAST searches. The complete absence of the *IFN-β* from searches likely indicates that there are difficulties in assembling this genomic region, rather than true loss (where a truncated or otherwise disrupted copy may have been isolated). Given the high identity between the swan and goose genomes (and the demonstration in this study that primers designed in the goose can effectively be used in the swan), sequencing of a locus-spanning PCR could reveal the presence or absence of the *IFN-β* locus.

7.4.5 Read depth of candidate genes

To investigate the possibility that genes involved in the swan innate immune response to influenza vary in copy number, read depth statistics were calculated for each of the 32 candidate genes for which swan orthologs were found. To provide an indication of

the reliability of the coverage statistics, calculations were performed for each of the 20 Abbotsbury swans (Figure 7.4).

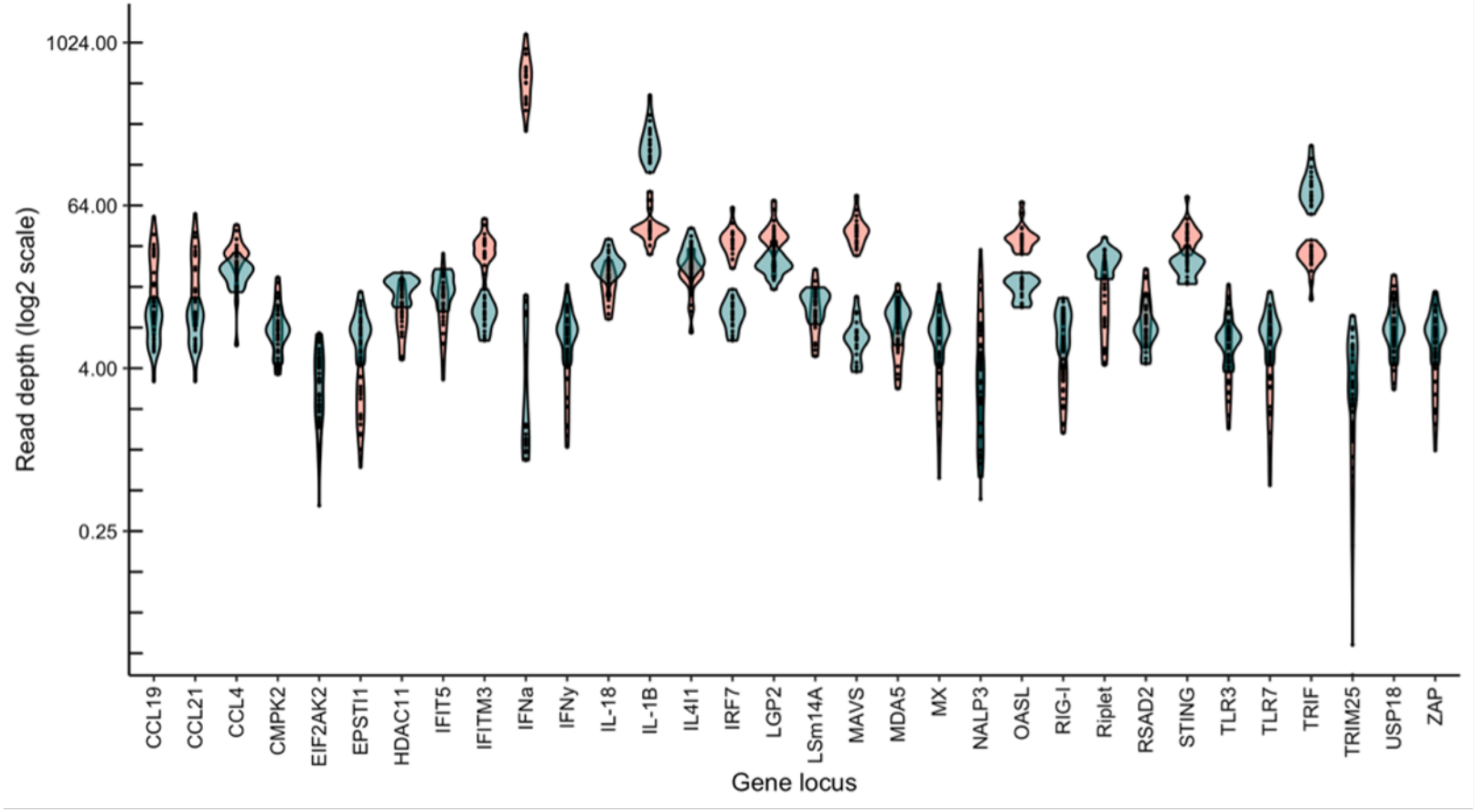


Figure 7.4: Read depth over candidate genes and their parental contig. Read depths for each locus were calculated for each of the 20 resequenced genomes, as well as for the parental contig on which the locus was found. Gene read depth is in red, while contig depth is in blue.

Most genes had average read depths that were comparable to their parent scaffold. However, some genes had clearly elevated read depth compared to the parent scaffold (*IFN-α*, *IRF7*, *MAVS*, and *OASL*; *P*-values from Welch's two-sample t-test 1.66×10^{-9}, 1.64×10^{-10}, 4.91×10^{-12}, 8.42×10^{-10}, respectively). The most pronounced increase in read depth was observed with *IFN-α*, which is known to have a number of paralogs in vertebrates (humans have 13 *IFN-α* subtypes [393]. Interestingly, the variation in read depth for *IFN-α*, both for the gene and the whole scaffold (designated SUPER_scf_78), was much larger than the rest of the genes measured. In chickens, the *IFN-α* locus is located on the Z chromosome [518], so, making use of the detailed metadata available for the Abbotsbury swans, the read depth was compared to the sex of the individual. This revealed significantly higher coverage over the whole contig in the homogametic (ZZ) male birds compared to the heterogametic (ZW) females (Welch's two-sample t-test; *P*-value = 6.16×10^{-5}; Figure 7.5). This implies that the SUPER_scf_78 is in fact the Z chromosome (or at least part of it) in the mute swan, and similar to other animals, *IFN-α* has undergone multiple duplications.

The other genes for which read depth appears to be elevated with respect to the parental contig (*IRF7*, *MAVS* and *OASL*) are autosomal in the chicken. Humans have three *OASL* paralogs—*OAS1*, *OAS2* and *OAS3*—but their identity to *OASL* is between 30-50% at the nucleotide level. It is not likely, therefore, that there has been widespread contamination of *OASL* reads with other OAS paralogs. Similarly, *IRF7* is a member of a family of interferon regulatory factor genes, but identity to other family members is low (around 20% at the nucleotide level). Finally, *MAVS* has no known paralogs in the vertebrate genome. Taken together, these observations suggest that, unlike *IFN-α*, there has not been extensive read sharing between these three genes and their paralogs. Instead, continued investigation will determine whether the observed increase in read depth at these loci is artefactual or related to duplication.

Conversely, there were two genes whose depth of coverage was markedly less than the parental scaffold (*IL-1β* and *TRIF*; *P*-values from Welch's two-sample t-test 4.66×10^{-10} and 3.27×10^{-9}, respectively). This may indicate that the parental contigs

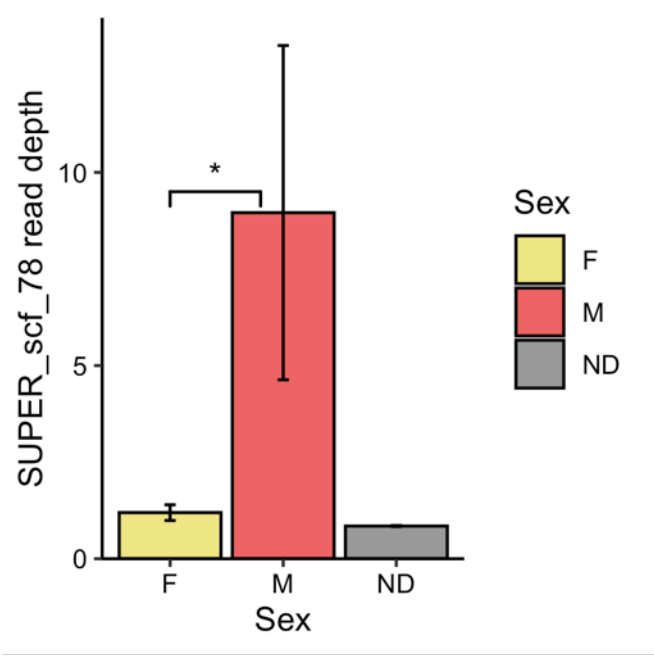


Figure 7.5: Read depth over the putative Z chromosome in the mute swan. Clear sex-related differences in read depth indicate that the contig designated 'SUPER_scf_78' in the publicly available mute swan genome is the Z chromosome (or at least part of it). *P = 6.16×10^{-5} (Welch's two-sample t-test).

have repeated regions that are accumulating reads. If true, this phenomenon would be more pronounced in shorter scaffolds. As expected, both *IL-1β* and *TRIF* lie on scaffolds of modest size in the mute swan genome (SUPER_scf_31, 5.8Mbp and SUPER_scf_2, 6.0Mbp, respectively). Additionally, the read depths of the gene loci themselves lie close to the genomic average, while the parental contig is markedly elevated. This indicates that the observed lower coverage of *IL-1β* and *TRIF* is likely due to the genes forming part of a repetitive region of the genome.

7.5 Discussion

The mute swan (*Cygnus olor*) is disproportionately susceptible to mortality from highly pathogenic avian influenza (HPAI) virus infection. The increased mortality is in excess of what would be anticipated from observation or sampling biases, which have been suggested as a contributing factor to the perceived susceptibility [483, 231]. Evidence from experimental studies [78, 283] suggests that swans are more susceptible

than other waterfowl such as geese, which themselves are quite conspicuous and live alongside swans in the wild [30]. The aim of this study was to characterise the innate immune components in the mute swan that are known to form the major pathways anti-influenza immunity in other animals.

Although bird migration is thought to be one of the primary modes of influenza spread [266, 306, 452], the relative contribution of the mute swan to the global spread of influenza is unclear [421, 456, 626]. On one hand, the high susceptibility of swans to HPAI means birds are likely to be unfit to fly any appreciable distance, and mute swans do not migrate over long distances like other waterfowl. However, evidence from experimental infection indicates that swans shed high titres of virus for several days post-infection [78, 283], and they are known to move during the winter months when waterways are frozen [421]. In birds, influenza is primarily an enteric infection [627], meaning swans contaminating waterways with infected faeces provides an efficient mechanism of faecal-oral transmission to other waterbirds in the same locale [452, 627]. Thus, although unlikely to contribute significantly to global spread, swans may be involved in local spread and maintenance of HPAI in waterways, which act as reservoirs for other migratory species like ducks which show minimal clinical evidence of infection and migrate long distances [299, 318, 599]. The importance of the swan to global influenza dynamics may therefore be in local influenza surveillance, rather than long-range transmission: their susceptibility to infection and conspicuousness means they are an excellent indicator species for influenza [231, 431]. Understanding what makes the swan unique among waterfowl in terms of its influenza responses helps to contextualise its role as a sentinel species.

Although it was speculative to suppose that there might be an obviously disrupted innate immune component in the mute swan, there is a clear precedent for this in the chicken, which has an absent *RIG-I* [42] and is known to be highly susceptible to influenza infection [272, 513]. The results from the present study, however, demonstrate that the swan has a full complement of genes that are typically involved in the anti-influenza response, at least in terms of having intact coding sequences of

suitable length. Of the 32 influenza response-related genes identified in the mute swan genome, all were found as complete coding sequences. Furthermore, nucleotide identity of swan genes to the goose ortholog was consistently in excess of 96%. A gene with a lost or redundant function may have undergone genetic drift and accumulated additional mutations. Only one gene—*IFN-β*—was not identified in the mute swan genome. Since the publicly available swan genome was used as a reference to map the resequenced genomes from 20 individuals, this could not be resolved using the resequencing data. In all likelihood, the apparent loss is due to an assembly error, but warrants further investigation, perhaps relying on the close identity of the goose and the swan to design a PCR to amplify the gene to determine its sequence.

None of the polymorphisms identified in the population-level analysis of *TLR3*, *TLR7* and *RIG-I* were predicted to cause loss of function. Two polymorphisms (G355A and T497G) yield non-synonymous polymorphisms that are proximal to dsRNA-binding sites in *TLR3* [100, 603]. However, while such polymorphisms are known to alter ligand-binding specificity [169], they would not typically lose function. A notable exception to this is the case of *Eudyptes* spp. penguin *TLR15*, which appears to have lost function in full-length haplotypes (Chapter 3). However, in the case of *Eudyptes TLR15*, multiple pseudogenised haplotypes were also present at the population level (presumably because of the relaxed selection pressure due to loss of function), and there was no evidence of this in the mute swan.

This study is the first attempt, to my knowledge, to attempt to characterise mute swan innate immunity as a potential explanation for susceptibility to HPAI. An understanding of what makes the mute swan susceptible to influenza is an important step in studying transmission dynamics of influenza in birds. Chickens appear to be equally as susceptible to HPAI, but this is explained by the lack of *RIG-I*, a crucial component of the anti-influenza immune response [42]. From the candidate gene approach taken in the present study, it is clear that there is no such obvious gap in the innate immune armoury in the mute swan: all of the well-documented components of the anti-influenza immune response are present with appropriately

sized coding sequences and no obvious truncating mutations. It follows that the susceptibility of the mute swan to HPAI is more subtle than that of the chicken. Possible explanations include the following: (i) swans lack a regulatory component, rather than effector component of the innate immune system; (ii) there is an as-yet unidentified effector component of the innate immune system that is non-functional in the swan; (iii) there is a non-truncating polymorphism in a swan innate immune component that alters function. Continued investigation into the functioning of the swan immune system could give an important insight into why swans experience heightened susceptibility to HPAI, and thus why they are such important sentinels of disease.

Chapter 8

General Discussion

In this discussion, I identify key results in each chapter of this thesis and discuss links between chapters. I also highlight the weaknesses or limitations of the work, particularly functional assays, and suggest avenues where future research could extend the conclusions drawn from this body of work. I conclude the discussion by considering possible future directions for the field as a whole.

8.1 Key results for each chapter

8.1.1 Chapters 2, 3, and 4

The work presented in Chapters 2, 3, and 4 represents an extensive characterisation of Toll-like receptor diversity and function in penguins (order: Sphenisciformes). Chapter 2 focused on the Gentoo penguin (*Pygoscelis papua*) as a model to investigate TLR diversity and adaptation. The choice of study species was a careful one: Gentoo penguins are now known to fall into at least four well-defined clades [466, 590], through which gene flow is absent or uncommon. This facet of Gentoo penguin ecology meant we were able to analyse Toll-like receptor sequences for evidence of adaptation to local pathogen assemblages. We found that variation in the three TLRs examined differed considerably, with the greatest diversity being found in *TLR5* and the lowest diversity in *TLR7*. The diversity also contributed to differential population struc-

ture, with most populations being genetically delineated in *TLR5*, and almost no population structure in the least diverse *TLR7*. *TLR4* and *TLR5* also exhibited a number of positively selected sites, most of which were located in the agonist-binding ectodomain, which could be an indicator of a change in function. Two sites in the *TLR5* transmembrane/TIR domain were strongly significant for positive selection, and appeared to co-segregate at the extremes of the Gentoo penguin's range: in populations in the Indian ocean vs. populations on the Antarctic peninsula, with intermediate frequencies in the Falkland/Malvinas Islands. Since the two selected polymorphisms were unusual in that they were located in the signalling domain of the TLR, rather than the more usual agonist-binding LRR domain, I hypothesised that they may be involved in signalling intensity differences. Functionally testing the two variants on the same TLR background revealed that the derived (Southern Gentoo) haplotype had a greater intensity of signalling in response to flagellin compared to the ancestral (Indian Ocean) haplotype, which indicates adaptation in the Southern Gentoo *TLR5*.

Having found evidence of adaptation within a single penguin species, I elected to expand my assessment to the entire order of penguins. Chapter 3 documents the patterns of selection across the TLRs of all penguin. Consistent with other studies, I noted that positive selection was more intense in (i) the agonist-binding LRR domain and (ii) TLRs expressed at the cell surface, consistent with several other studies. Furthermore, a comparative approach revealed that many of the selected sites were at positions known to be functional (typically using data from well-studied mouse and human TLRs), which strongly implies changes in TLR function. I also noted several sites that were not only selected in penguins, but also in mammals and elsewhere in birds (comparing my data to those from other studies). These sites are of interest because they imply a degree of recurrent positive selection, which, even in the absence of functional data, suggests they may be important. I also noted that the reference genome for several crested penguins (*Eudyptes* spp.) contained a pseudogene for *TLR15*. Moreover, the disabling mutation was distinct in different species. To

resolve this complex pattern, I obtained additional sequencing data covering *TLR15* from >100 crested penguins from five species. Based on the phylogenetic distribution of the different pseudogenic *TLR15* haplotypes, I reasoned that at least four of these had occurred no later than the stem branch leading to the crested penguins. Although there is known introgression between certain species in the *Eudyptes* genus [604], I argued that, since the same pseudogenic haplotypes were often found in species pairings for which there is no evidence of introgression, the disruption must have occurred in the common ancestor of the species pair (which was, in several cases, the common ancestor of the genus). I also observed that, along with several independently pseudogenised haplotypes, there were also seemingly intact (full-length) haplotypes present at low frequencies at the population level. At this point, there was a paradox: how could there have been strong enough selection pressure to independently pseudogenise *TLR15* as many as eight times in one genus, but not strong enough to drive any of these pseudogenes to fixation? The most parsimonious explanation was that the full-length haplotypes were also non-functional, and were cryptic pseudogenes. Despite generating a full-length protein product, the crested penguin *TLR15* did not respond to either of its agonists (*Saccharomyces cerevisiae* lysate, elucidated in [72], or *Aspergillus fumigatus* lysate, elucidated in this study), while the closely related Emperor penguin and Northern fulmar did respond. Since TLR15 under normal circumstances responded to *Aspergillus* lysate, I speculated that TLR15 is part of the normal immune response to *Aspergillus*, which is a pathogen of considerable concern in penguins. Therefore, a lack of *TLR15* may play a role in contributing to *Aspergillus* susceptibility in penguins.

Since Chapter 3 described extensive positive selection in penguin TLRs, there is evidence that penguins have undergone considerable co-evolution with their pathogens. To explore this further, I performed an unbiased whole-genome analysis of all penguin species to investigate other examples of host-pathogen co-evolution (Chapter 4). I first noted that the proportion of positively selected sites was higher in immune genes than any other functional class of genes. Over-represented

within this set of genes were pathogen-facing innate immune components such as pattern-recognition receptors. Pathogen-facing immune components may have increased rates of adaptive evolution to counteract pathogen-derived molecules that change to avoid detection. Genes encoding non-pathogen-facing molecules interact with host factors, and are therefore typically under purifying selection. Two examples of pathogen-facing genes in which positive selection was detected are *IFIT5* and *CD81*. In *IFIT5* (an RNA-binding pattern-recognition receptor), a cluster of positively selected sites was located within the RNA-binding cleft, which suggests penguin IFIT5 may have altered an RNA-binding capability. In contrast, in *CD81*, positively selected sites were found at an interface involved in glycoprotein-mediated cell entry by hepatitis C virus in mammals. Evidence of selection at these sites may indicate co-evolution of penguins with a virus that, like hepatitis C, uses CD81 as an entry receptor.

8.1.2 Chapter 5

Having elucidated patterns of diversity and adaptation in the Toll-like receptors of wild birds, my focus then moved to domesticated birds. The best candidate for a widespread domesticated bird was the domestic chicken, which diverged from its nearest wild relative—the red junglefowl subspecies *Gallus gallus spadiceus*—around 10 Kya [620]. Since its domestication, the chicken has undergone selection for production traits (egg and meat production), as well as for other phenotypic traits such as feather colour and structure, skin colouration, and other cosmetic traits which were important in the emergence of fancy breeds in the 19$^{\text{th}}$ century. Today, the chicken is our most numerous food animal, and commercial lines may have experienced a genetic bottleneck. Chapter 5 reports a comprehensive description of diversity of TLR genes in the domestic chicken (*Gallus gallus domesticus*). While a limited number of studies have examined TLR diversity in the chicken, especially in fancy breeds, none has compared diversity in fancy breeds, feral birds, commercial birds and all

congeneric junglefowl species and subspecies. Making use of >1600 genomes, I found that chicken TLRs are, generally, highly diverse, much more so than would be expected from a domesticated species. This is likely due, at least in part, to repeated hybridisation with sympatric red junglefowl species across Asia, as indicated by the phylogenetic analysis in the chapter. Similar to the penguin work, I identified positively selected sites that were at, or near, sites of known function, indicating that the observed variation may have phenotypic consequences. I also found evidence of selective sweeps affecting *TLR2A*, *TLR2B*, *TLR5*, *TLR7* and *TLR21*, which may indicate adaptation to changing pathogen pressures in higher-density commercialised settings compared to backyard flocks. Finally, I attempted to characterise TLR dysfunction in the chicken by identifying breeds and populations with high frequency pseudogenes in one or more TLRs. For instance, White Plymouth Rock birds were noted to have high rates of *TLR2A/B* pseudogenisation, which may impact their ability to handle bacterial pathogens. This dataset represents one of the most (or the most) extensive characterisations of TLR diversity outside of humans.

8.1.3 Chapter 6

In Chapter 6, I sought to test the hypothesis that the altered CpG-DNA recognition profile of TLR21 between fishes and birds was related to the loss of the other CpG-recognising TLR—TLR9—in the avian/reptilian lineage. Although some previous studies have characterised TLR21 CpG-recognition in several fish species and galliform birds, CpG recognition profiles of amphibian, reptilian and non-galliform TLR21 are largely unknown. I predicted that the amphibian TLR21 would be fish-like, since amphibians also have TLR9, and that the reptilian TLR21 would be bird-like, since the majority of reptiles have lost TLR9. This prediction was largely borne out: amphibian and fish TLR21 appear to be weak, specialist recognisers of Class B CpG-ODNs, while reptilian and bird TLR21 are robustly responding generalists. It was also interesting to note that reptile TLR21 appears to be even more of a

generalist than avian TLR21, responding to Class A, as well as Class B and C CpG-ODNs. While I had originally planned to test more TLR21 constructs, some of the fish and amphibian TLR21 did not express at detectable levels. The possible reasons for this are discussed in detail in the following sections. As well as documenting the functional change in TLR21, I also analysed a large number of *TLR21* sequences for evidence of positive selection that may coincide with the repurposing of *TLR21*. Interestingly, positive selection was extensive in the avian/reptilian lineage, and was much less prevalent in the fishes. Furthermore, one of the reptilian lineages that was conspicuously lacking positive selection in *TLR21* was the chelonians (turtles), which have recently been found to possess *TLR9* [149]. The complementary patterns of functional change and positive selection suggest that the repurposing of *TLR21* occurred in response to the loss of *TLR9* from the birds and reptiles.

8.1.4 Chapter 7

My final data chapter, Chapter 7, was a candidate gene-driven investigation of the innate immune system of the mute swan (*Cygnus olor*), which is particularly susceptible to highly pathogenic avian influenza (HPAI). A number of devastating HPAI outbreaks have affected the well-studied population of mute swans at Abbotsbury Swannery (Dorset, UK) over the last decade. Swans at the Swannery appear to experience disproportionately high mortality in response to HPAI, compared to other waterfowl. Supporting this observation, experimental infection of swans and other waterfowl has identified swans as being much more susceptible to HPAI, compared to geese and ducks. Given that the chicken (which is also highly susceptible to HPAI) has a demonstrable lack of *RIG-I*—a crucial viral-sensing pattern-recognition receptor—I hypothesised that swans have an analogous defect in their innate immune response to HPAI. Using MinION and whole-genome (Illumina) sequencing, I identified >30 genes that are known to be key mediators of the anti-influenza response in birds and other animals. All of the genes (with the exception of *IFN-β*) were identified in the

mute swan genome, and all were predicted to yield intact coding sequences of appropriate length. I also found no evidence of function-changing polymorphism in three genes—*TLR3*, *TLR7* and *RIG-I*—sequenced using MinION technology. I concluded that, if there is a defect in the swan immune system, it is more subtle than the analogous case in the chicken, and could be down to function-changing polymorphism in a key gene, or change of a regulatory component.

8.2 Integration of bioinformatic and functional methods

Immune genes are often identified as targets of positive selection in large unbiased genome-wide analyses (e.g. [323, 542]). Furthermore, Toll-like receptors are often chosen specifically to investigate pathogen-driven selection on immune genes (e.g. [25, 221, 220, 313, 358, 603, 637]). However, while positive selection and inferred functional changes are common in these scenarios, very few studies have made the direct steps to test bioinformatic predictions with functional assays. I consider this to be a gap in our understanding, and a break in the chain of logic to demonstrate phenotypic change arising from genetic polymorphism.

Toll-like receptors lend themselves to functional analysis for several reasons: (i) intracellular adapter proteins can interact with a wide range of different species' TIR domains and thus different TLRs can be tested in the same cell background (although see discussion below); (ii) orthologous TLRs across large evolutionary distances respond to well-defined classes of agonists so can be directly compared; (iii) diversity is germline-encoded, rather than the somatic diversity generated in the adaptive immune system; and (iv) reporter cell lines and agonists are commercially available and of high quality. Despite this, predictions of changes in TLR function are almost never tested *in vitro*. Chapters 2 and 3 represent, to my knowledge, the only studies to have tested bioinformatic predictions of functional change in Toll-like receptors in a

non-model species.

Without testing, it is difficult to predict the functional effects of positive selection. For instance, the two positively selected and co-segregating sites in Gentoo penguin *TLR5* transmembrane/TIR domain were not anticipated to lead to increased signalling (Chapter 2). It may have been expected that penguins living in the colder West Antarctic Peninsula would have, if anything, a more muted response to pathogens, since the pathogen pressure may be lower compared to the warmer and more populated Crozet/Marion Islands. Furthermore, without functional analysis, the cryptic pseudogene haplotype of *TLR15* in *Eudyptes* spp. would have been incorrectly identified as a functional haplotype (Chapter 3). In Chapter 6, functional analyses were required to demonstrate a trajectory whose start and end points were already known. TLR21 in fish was already known to be a fastidious recogniser of few CpG-DNA motifs [657], while TLR21 in chickens was known to be a broad, robust recogniser of several CpG motifs [107]. It seemed likely that this transition coincided with the loss of the broad-recognising TLR9 in the bird/reptile lineage, but this had not been demonstrated. I showed that the missing links in the transition—amphibians and reptiles—responded as predicted to a CpG panel (amphibians with a fish-like TLR21 and reptiles with a bird-like TLR21). In an unexpected but interesting result, I demonstrated that the CpG-recognition profile of reptiles was even broader than in birds, responding to Class A, as well as Class B and C CpGs.

Of course, functional testing is not always feasible (and is expensive and time-consuming), and I also relied solely on bioinformatic methods elsewhere in this thesis (Chapters 4, 5 and 7). However, wherever possible, I attempted to contextualise the results from my positive selection analyses. Functional information, as well as a crystal structure, is widely available for many immune components, and I always cross-referenced selected sites with sites of known function. By doing this, I identified sites that likely contribute to change in function. For instance, positively selected sites were found at TLR ligand-binding sites in Chapters 3, 4 and 5, and in the RNA-binding cleft of *IFIT5* and viral-binding surface of *CD81* in penguins (Chapter

4). Although, as described above, testing *in vitro* provides a more reliable indication of functional change, cross-referencing selected sites with those of known function provides good-quality indirect evidence.

The immunogenetics literature (outside of model organisms) is flooded with descriptive studies of diversity and selection in Toll-like receptors without accompanying functional evidence. While such studies are often cited as evidence of functional change in certain lineages, these conclusions are not reliable. In this thesis, I have attempted to relate genetic polymorphism to functional change, either directly or indirectly, and firmly believe this should be the benchmark for future immunogenetic studies, especially in non-model organisms.

8.3 Limitations of functional assays

> All models are wrong, but some are useful
>
> *George E.P. Box*

Ethical considerations and sampling challenges in wild animals preclude the sorts of experiments historically conducted in the mouse and other laboratory animals to unpick the fine details of innate immunity *in vivo*. As such, we are mostly limited to modelling TLR biology *in vitro* using reporter systems. Above, I argue that this approach should be the benchmark for inferring functional change in TLRs, and, while I believe this to be true, there also at least four caveats.

Firstly, TLR response to agonists is known to be heavily dependent on concentration [215]. Moreover, it has been argued that the artificially high and non-physiological agonist concentrations used during *in vitro* experiments yield unexpected results. For instance, it has been reported that TLR2 responds to various lipopeptide analogs in the absence of TLR1 and TLR6 (which normally mediate the TLR2 response to lipopeptides through heterodimerisation [84]). However, the degree to which this response is relevant at physiological agonist concentrations present dur-

ing infection is debated, and TLR1/6 may still be required *in vivo* [451]. During my experimental work in Chapters 2, 3 and 6, I generally performed agonist titrations to determine the lowest agonist dose that gave maximal responsiveness (with reference to the manufacturer's range of working concentrations). Full experiments were then conducted with the chosen agonist concentration, rather than using a range (which, in the case of the CpG-ODN panel in Chapter 6 would have been technically unfeasible). While TLR agonists have been used *in vitro* for hundreds of studies, there is a chance that some conclusions in my work may be unreliable because physiological concentrations of agonists are significantly lower. In Chapter 6, I mitigated the impact of agonist concentration differentially affecting TLRs from different species by only making relative statements within each CpG panel tested, rather than explicitly comparing TLRs. Where I did make explicit comparisons between Gentoo penguin TLR5 haplotypes in Chapter 2, I ensured that protein levels were normalised such that the effect of agonist concentration would be at least comparable between the TLRs tested.

Secondly, TLRs operate at the front of a complex signalling network to activate NF-κB and other signalling pathways. To effectively transduce the signal of a detected agonist, TLRs need to be able to interact with the intracellular adaptor proteins. While the signalling regions of TIR domains and adaptor molecules such as MyD88 are highly conserved [398], there is no guarantee that a particular TIR-adaptor combination will be compatible. Throughout my functional work, TLRs from various species were tested in a human cell background (HEK-Blue cells). While previous studies have successfully demonstrated TLR reactivity in human cells from a wide range of species, including fishes [657], birds [107], reptiles [611], and even amphioxus [662], there remains a possibility that TIR-adaptor interactions have undergone species specialisation. The testing of heterospecific TLRs in a human cell background may therefore give spurious results. Indeed, in a study comparing full-length (native) TLR5 from different species to chimaeras including the human TLR5 TIR domain in human cells, pig and chicken TLR5 showed reduced functionality to

their chimaeric counterparts [169], suggesting that the human TIR domain was more compatible with the human adaptor proteins. While making chimaeric TLRs may also have unintended consequences (such as altering the stability or structure of the protein), it is important for these to be included as controls. Clearly, for Chapter 2, the focus of the functional work was to elucidate the role of a transmembrane and a TIR domain polymorphism in Gentoo penguin TLR5, which precluded the use of a TIR domain chimaera. However, for Chapter 6, it could be that standardising the TIR domains of the various TLR21 tested would increase the signalling capability of the weaker responders and make for more reliable comparisons between species. While this is a considerable investment of resources, the conclusions drawn from the chapter may be more reliable.

Thirdly, TLRs are only functional when targeted to the correct cellular compartment. Intracellular TLRs are dependent on endosomal acidification and/or cleavage to become functional, which adds additional stages of interaction with the host human cell. The targeting of endosomal nucleic acid-recognising TLRs is dependent on UNC93B1 [566], which, in my expression system would also require interaction between human UNC93B1 and a heterospecific TLR. This is likely to only affect Chapter 6 (TLR21), as this was the only functional work to consider an endosomal TLR. It has recently been shown that UNC93B1 is also required for maintenance of intracellular expression [463], which could also explain why TLR21 from some species was not detectable by immunoblotting in my experiments, despite transfecting an intact plasmid. Furthermore, signalling of zebrafish TLR21 in human cells was found to be enhanced by co-expression with zebrafish UNC93B1 [657], suggesting there are species-specific differences in the interaction between TLR21 and UNC93B1. A logical extension to Chapter 6 would be to co-express each TLR21 with a homotypic UNC93B1 to promote expression and enhance signalling capability in human cells. This may raise activity of the problematic TLR21 constructs to detectable levels and allow signalling to be measured.

Finally, different species have well-documented codon biases that may be linked

to, among other things, tRNA availability in the cell [193]. Attempted expression of a gene from a species with a different codon usage profile in human cells may lead to attenuated or absent protein expression. TLR-specific codon biases have also been noted. For instance, *TLR15* in some reptilian species shows significant codon bias, and this directly abrogates expression in human cells [612]. Other, more evolutionarily conserved, TLRs (such as *TLR3*) do not display the same levels of interspecific codon variation [612]. In Chapter 6, cloned sequences were not codon optimised (with the exception of *Eudyptes moseleyi TLR21*, which required codon optimisation in order to be successfully synthesised), because the constructs were cloned with a combination of PCR- and synthesis-based approaches. To investigate whether codon bias be contributing to the lack of expression in certain fish and amphibian TLR21, these constructs could be re-made using codon-optimised sequences. To ensure optimal expression, the *TLR15* constructs in Chapter 2 were all codon-optimised prior to cloning.

8.4 Toll-like receptor evolution at different scales

In Chapters 2-6, I demonstrated that TLR adaptation can take place over multiple nested timescales. Over the greatest evolutionary distances (hundreds of Ma), TLR21 was found to completely alter its ligand recognition profile from a narrow to a broad recogniser of CpG-DNA motifs between fishes and birds/reptiles. Although TLRs are known to retain specificity for defined agonists over evolutionary time [82, 493], the functional change in TLR21 is arguably the best evidence for extensive functional repurposing of a Toll-like receptor.

In the tens of Ma, I demonstrated adaptation of Toll-like receptors within an order of vertebrates—the Sphenisciformes (penguins; time to MRCA 61-62 Ma; [545]). Radiation of penguins is thought to have been driven, at least in part, by island formation, since the evolution of several penguin clades is contemporaneous with the origin of their host islands [118]. Adaptation to these new environments, and

dispersal of penguins to different climates (north to the tropical Galápagos Islands and south to the Antarctic peninsula [457]), has likely meant changes to in pathogen profiles, which exert selection pressure on the innate immune system to adapt. These changes have evidently manifested in extensive positive selection in penguin TLRs, particularly in ligand-binding domains.

Within the last 5 Ma, there is additional evidence of further specialisation of TLRs in penguin sub-lineages. Chapter 2 documents the adaptation of *TLR4* and *TLR5* during the relatively recent radiation of the Gentoo penguin (*Pygoscelis papua*) to form four clades (Crozet/Marion Islands, Kerguelen Islands, Northern and Southern) in different climates [358]. The finding that genetic change was accompanied by functional change in TLR5 provides further evidence that TLRs can undergo relatively rapid specialisation to novel pathogens.

However, even at the shortest timescales considered in this thesis, there is still evidence of innate immune system adaptation. Domestic chickens diverged from the closely related red junglefowl subspecies *Gallus gallus spadiceus* around 10 kya [620], and the domestication process may have started somewhat later [467]. Despite this, I found evidence of a recent selective sweeps involving TLRs in the domestic chicken relative to *Gallus gallus spadiceus*. While this could be interpreted as a bystander effect of TLR loci being proximal to production loci, the concentration of selected sites in the ligand-binding domain, and specifically at ligand-binding sites, suggests that diversity of TLRs may be attributed to selectable immune phenotypes. This implies that, during a period of intense, anthropogenic selection for production traits, there was also considerable selection for immune traits, possibly linked to increased movement of food animals and their increased density.

Taken together, these observations indicate that selection in the innate immune system can operate over timescales in several orders of magnitude (from a few thousand to hundreds of millions of years). In TLRs, the more rapid selection over short timescales appears to fine-tune receptor specificity or responsiveness (as indicated by heightened TLR5 responsiveness in Chapter 2), while the longest timescales may

allow complete remodelling of a TLR ligand recognition profile (such as TLR21 in Chapter 6). The pattern-recognition system is therefore in a constant state of flux, likely reflecting the ancient co-evolutionary arms race with pathogens that will never ease.

8.5 Wildlife disease in a changing climate

Throughout this thesis, I have considered numerous examples of how animals adapt to the pathogens surrounding them. However, emerging infectious diseases—to which animals are maladapted—can pose significant threats. Some animals appear to be somewhat resilient to novel pathogens and can quickly adapt. For instance, eastern USA house finch (*Carpodacus mexicanus*) populations with a 12-year history of natural exposure to *Mycoplasma gallisepticum* (MG) displayed significantly upregulated adaptive immune responses compared to western USA, MG-naïve, populations during experimental infection. Given that none of the birds had any serological evidence of prior infection with MG, this implies a degree of immune adaptation to a novel pathogen in a short time [65]. However, other species appear to be more sensitive. For example, ongoing chytridiomycosis panzootic (caused by two fungal species, *Batrachochytrium dendrobatidis* and *B. salamandrivorans*) has led to population declines in >500 amphibian species [524]. Emerging pathogens are anticipated to be of particular concern as climate change causes rapid shifts in the ranges of hosts, pathogens and vectors. The ability to identify vulnerable species is a valuable way to mitigate the devastating consequences of emerging infectious diseases.

Environmental change and pathogen range expansion appear to be inextricably linked. Increased temperature is associated with more rapid development of protozoan parasites [434], increased rate of viral replication [26] (particularly important for ectothermic animals), increased susceptibility to viral diseases [73], increased growth rates of marine bacteria and fungi [260, 538], and increased biting rates of arthropod vectors [74]. These variables translate to differences in disease transmis-

sion. For instance, avian malaria appears to be undergoing increases in intensity and latitudinal/altitudinal range in response to climate warming [195, 663]. Other arthropod-vectored diseases, such as Lyme disease, yellow fever, dengue fever, plague, and tick-borne encephalitis have expanded their geographic range or increased in incidence in recent decades, which has been attributed to expansion of vector ranges [216, 233, 370]. Furthermore, an increased average temperature is likely to bring 'tropicalisation' of temperate regions, with the poleward expansion of cold-sensitive tropical species [454, 583]. While this acts to increase the ranges of disease-transmitting arthropod vectors, vertebrate hosts (which can also vector diseases) may also relocate to novel environments as conditions become less favourable in their native ranges. Many bird species appear to be expanding their ranges to higher latitudes in response to climate change [507], for example. Rapid range shifts have the potential to introduce novel pathogens to vulnerable populations and increase the intensity of transmission among sympatric species. The evidence for direct effects of climate change on disease transmission dynamics is increasing. For instance, ranavirus epidemics in frogs are known to occur with greater frequency and severity at higher temperatures, which is corroborated by infection in controlled experimental scenarios [481]. As well as being devastating for the many species affected (e.g. [61, 179, 495]) emerging infectious diseases have direct and significant consequences for humans, as exemplified by the ongoing SARS-Coronavirus 2 pandemic [616].

Animals in captivity are useful study systems to study how novel pathogens may affect naïve populations, in an analogy to emerging infectious diseases following species range shifts. Penguins in captive settings at more northerly latitudes are a case in point. *Aspergillus* spp. are ubiquitous in the environment, but infectious spores are present at higher densities in warmer, wetter conditions [31, 119]. Aspergillosis in captive penguin species is common and causes a significant burden of disease [3, 304, 642]. Although other bird species are also susceptible to aspergillosis [403], Chapter 3 identifies a *TLR15* pseudogene as a plausible explanation (or contributing factor) for susceptibility in the *Eudyptes* penguins. Since, within the penguins, a

functional *TLR15* was only identified in the basal Emperor penguin, it is plausible that other more derived penguin species also lack functional *TLR15*. Supporting this, a heterozygote pseudogene was also identified in the *Pygoscelis* genus (Chinstrap penguin, *Pygoscelis antarcticus*), which in *Eudyptes* spp. was indicative of the relaxed selection pressure due to cryptic pseudogenisation. Continued study to characterise the *Pygoscelis* spp. *TLR15* at the population level and in functional assays will reveal whether *TLR15* is functional, and may provide insights into differential *Aspergillus* susceptibility among penguins. Continuing my work in this chapter, I am engaged in collaborations with veterinarians in the UK and USA with extensive epidemiological records of aspergillosis deaths. I hope to test the hypothesis that an intact *TLR15* among penguin species is a positive predictor of aspergillosis resistance in penguins. Projects such as this not only help to characterise the immune response to problematic pathogens, but also may serve to identify species and populations most at risk from emerging infectious diseases due to a changing climate.

8.6 Future directions for the field

This thesis sits at an interface between several fields, including evolutionary biology, population genetics and immunology. The integration of a wide range of methods from these disciplines in this body of work has expanded current knowledge of Toll-like receptors, particularly those in wild animals.

8.6.1 Sequencing costs and technology

The cost of sequencing has fallen dramatically in the last decade with the advent of 'next-generation' or 'second-generation' sequencing technologies. The coveted '$1000 genome' target [602] has been surpassed, with many suppliers now competing for a '$100 genome' [509]. The low costs associated with sequencing have effectively democratised this branch of life sciences, meaning any project with a modest budget can afford to generate significant amounts of sequencing data. This is good

news for the immunogenetics field. Now that the repertoires of TLRs have been established in many species, the focus may now move to measuring TLR variation within populations, which, of course, relies on large numbers of samples. In addition, the most recent long-read sequencing technologies (e.g., Oxford Nanopore's MinION and PacBIO single molecule, real-time (SMRT) sequencing) can allow us to resolve highly repetitive and/or difficult-to-assemble regions of the genome such as the major histocompatibility complex (MHC), and large gene expansions, such as the various TLR expansions in the invertebrates. The ability to generate high-quality reference genomes for non-model organisms at low cost will revolutionise innate immune biology, and many other fields besides.

8.6.2 Large-scale patterns of host and pathogen diversity

While sequencing itself is becoming a fairly trivial exercise, sample collection for large-scale studies is as difficult now as ever before. Sample collection for the Gentoo TLR project (Chapter 2) took several seasons, travelling to colonies separated by thousands of miles of ocean at great expense. The result, however, was the largest spatial scale of an immunogenetics study in a wild animal (to our knowledge) and a sample set that can be used for a variety of purposes. Currently almost completely lacking, however, are any large scale investigations into how pathogen profiles change across large distances. Without these data, the variations in innate immune diversity and function are lacking context. While we may feel empirically justified in assuming that pathogen composition must change over large distances (due to the differences in biotic and abiotic features of the environment), there is a surprising lack of data supporting or refuting this hypothesis. The limited data that do exist are usually localised to easy-to-reach locations (e.g., polar research stations) and well-studied populations, and generally offer lists of species at one location, rather than changes in abundance over a spatial scale. Understanding pathogen compositions is also important from the perspective of a changing climate. Only by understanding

the 'baseline' abundances and distributions for putative pathogens can we determine how these patterns are changing in response to global climate change, and only then can we be proactive about risks affecting vulnerable populations. Large scale distribution and abundance data for pathogens would add another layer of richness to TLR immunogenetics, and it is worth remembering that immune diversity and pathogen diversity are inextricably linked.

8.6.3 Taking the comparative approach

We can learn much about the immune system by comparing and contrasting the different approaches animals have taken to counter pathogen threats. The comparative approach is one I have found useful throughout my thesis. The repurposing of a fastidious TLR21 to become a generalist CpG-recogniser would not have been appreciated without the comparisons being drawn between species. Furthermore, by comparing *TLR15* sequences from large numbers of crested penguins, I was able to appreciate the number of independent pseudogenisations there had been, leading me to hypothesise that the seemingly intact haplotype was a cryptic pseudogene. By comparing positively selected sites between my study populations and previously reported work, I was able to identify sites under recurrent positive selection in different taxa (e.g., Chapter 3) which are likely to be functionally important. Assisted by inexpensive sequencing, we should take every opportunity to compare the different immune strategies employed by different animals.

8.7 Data availability

Several of the chapters in this thesis generated large amounts of sequence data that will likely be useful to other projects. Being committed to reproducibility and data sharing, I wanted to ensure that all sequence data generated as part of this project is freely available. All sequence data generated in Chapter 2 (Gentoo penguins, [358]) has been deposited in GenBank with accession numbers MN394222–MN394376

(*TLR4*), MN313018–MN313169 (*TLR5*), MN312870–MN313017 (*TLR7*), and MN566362–MN566421 (mitochondrial control region, HVR1). All sequence data generated in Chapter 3 has been deposited in GenBank with accessions numbers MW793727–MW793735 (*TLR1A*); MW793736–MW793744 (*TLR1B*); MW793745–MW793756 (*TLR2A*); MW793757–MW793767 (*TLR2B*); MW793768–MW793786 (*TLR3*); MW793787–MW793805 (*TLR4*); MW793806–MW793824 (*TLR5*); MW793825–MW793843 (*TLR7*); MW793844–MW793854 (*TLR15*); MW793855–MW793869 (*TLR21*). Population-level *TLR15* data are available with accessions MW805883–MW806032. Penguin genomic data (Chapter 4) have already been made available as part of a previous publication [457]. All of the chicken data processed in Chapter 5 is being made publicly available through an international consortium of groups working on chicken adaptation and domestication. Finally, swan genomic data (Chapter 7) will soon be deposited both as short read entries and alignment map files.

8.8 Conclusions

Motivated by the aspiration to better understand how hosts and pathogens interact, my thesis contributes to existing knowledge of innate immune system diversity, adaptation, and function. I made use of extensive population-level and whole-genome datasets to describe variation between individuals and species, and identified likely targets for functional change. Moreover, in a combination of bioinformatic and laboratory methods that has scarcely been employed before, I tested bioinformatic predictions of functional change to more explicitly demonstrate the link between polymorphism and phenotype. In an era when sequencing data are ubiquitous and inexpensive to generate, we must not lose sight of the basic biological techniques that add richness and context to purely computational analyses.

Appendix A

Appendix to Chapter 2

A.1 CRISPR-Cas9 editing of endogenous *TLR5* in HEK-Blue Null1 cells

HEK-Blue Null1 cells express TLR5 protein at appreciable levels, and are therefore capable of responding to flagellin without transfection of additional *TLR5*. In order to test the function of Gentoo penguin TLR5 without the signal being masked by endogenous TLR5, a strategy was designed to generate a *TLR5*$^{-/-}$ cell line using the CRISPR-Cas9 system. A custom crRNA sequence with predicted complementarity only to human *TLR5* (5'-AUG AGC UCG AGC CCC UAC AAG UUU UAG AGC UAU GCU-3') was obtained from IDT. The CRISPR protocol was carried out according to the manufacturer's instructions. Briefly, resuspended crRNA and tracrRNA were mixed in equimolar concentrations to create a 1 μM working solution of guide RNA. Guide RNA was then heated at 95 °C for 5 minutes and allowed to cool to room temperature. To form the ribonucleoprotein (RNP) complex, the following mixture was assembled:

Component	Volume (μl)
Guide RNA oligos (1 μM)	22.5
Diluted Cas9 enzyme (1 μM)	22.5
Opti-MEM	330
TOTAL	**375**

Guide RNA and nuclease-negative controls were also assembled at this point. The mixtures were incubated for 5 minutes are room temperature to allow the formation of the RNP complexes. Next, RNP complexes were reverse-transfected into low-passage HEK-Blue Null1 cells using Lipofectamine RNAiMAX reagent (1.2 μl per well), without antibiotics. Cells were at a concentration of 4×10^5 cells/ml and were incubated in a 96-well plate for 48 hours.

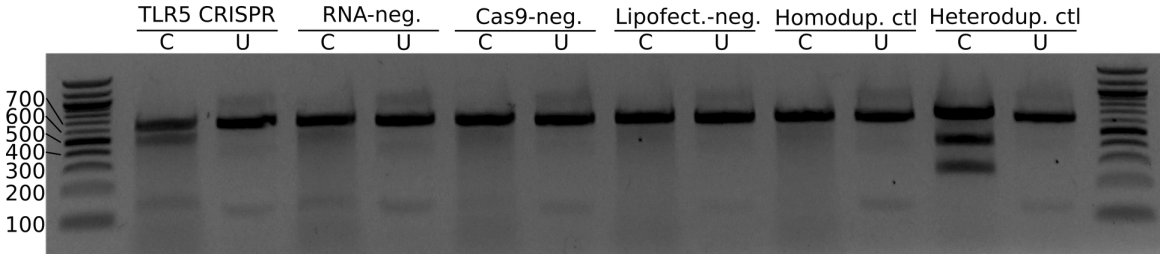


Figure A.1: T7EI mismatch cleavage assay for detection of edited *TLR5*. An asymmetric PCR was performed around the editing site (653 bp), which would yield ∼500 bp and ∼150 bp products when cleaved by the structure-sensitive T7 endonuclease I. RNA-, Cas9- and Lipofectamine RNAiMAX-negative controls were also included, as well as the technical homoduplex (should not be cleaved) and heteroduplex (should be cleaved) controls to optimise assay conditions. Abbreviations: C, cut (with T7EI); U, uncut.

In order to assess whether disruption of *TLR5* had taken place on the bulk-transfected cells, the T7EI mismatch cleavage assay (IDT) was used. DNA from transfected cells was used as a template for a PCR that amplifies the region surrounding the editing site (F primer: TGCTCCTTTGATGGCCGAAT; R primer: TCCACTGTCCAGCCATTTCC), generating 653 bp amplicon that is asymmetric around the editing site (∼500 bp from the F primer and ∼150 bp from the R primer). The T7EI assay relies on the activity of the structure-sensitive T7 endonuclease I, which cleaves dsDNA at the sites of DNA mismatches. Therefore, heat denaturation followed by formation of heteroduplexes comprising the two independently edited

copies of *TLR5* generates a substrate for T7EI if editing has taken place. T7EI-mediated cleavage then generates two fragments which can be resolved by agarose gel electrophoresis. The T7EI mismatch cleavage assay yielded a 500 bp product that was visible with gel electrophoresis in the *TLR5* CRISPR-treated cells, but not in the RNA-, Cas9, or Lipofectamine RNAiMAX-negative controls (Figure A.1)

Having established that editing had taken place on the bulk-transfectants, cells were trypsinised and passed through a sterile cell filter to obtain a single-cell suspension. Next, cells were diluted to concentration of 0.8 cells per well of a 96-well plate and seeded into 96-well plates. Wells were monitored for the formation of single colonies. Once at 50% confluence, wells with single colonies were isolated and grown on.

To functionally test whether clones were phenotypically null for *TLR5*, cells from different clones were seeded into 96-well plate, challenged with flagellin (1 μg/ml FLA-ST; Invivogen), and incubated for 24 h. Next, supernatant was harvested and analysed using the SEAP reporter assay according to the manufacturer's instructions (Sigma) on a microplate reader (FLUOStar Omega, BMG Labtech). Since the SEAP assay is an enzymatic process, multiple readings were taken over a 50 minute period (Figure A.2). Some clones, e.g. clone 13, responded robustly to flagellin, indicating the CRISPR disruption of the gene had not been successful. Others, e.g. clone 7, had a reduced response to flagellin, possibly indicating that only one copy of the gene had been disrupted, or that the clone was mixed. Three clones—3, 9, and 24—did not respond to flagellin, and were therefore good candidates to take forward for sequencing.

To confirm successful disruption of endogenous *TLR5*, the amplicon surrounding the editing site described above was subject to PCR cloning (using the NEB PCR cloning kit) in clones 3, 9, and 24. Twelve colonies were picked and Sanger sequenced for each clone, and the presence of two independent frameshift mutations was confirmed. Clone 9 was taken forward for use in the Gentoo penguin experimental work.

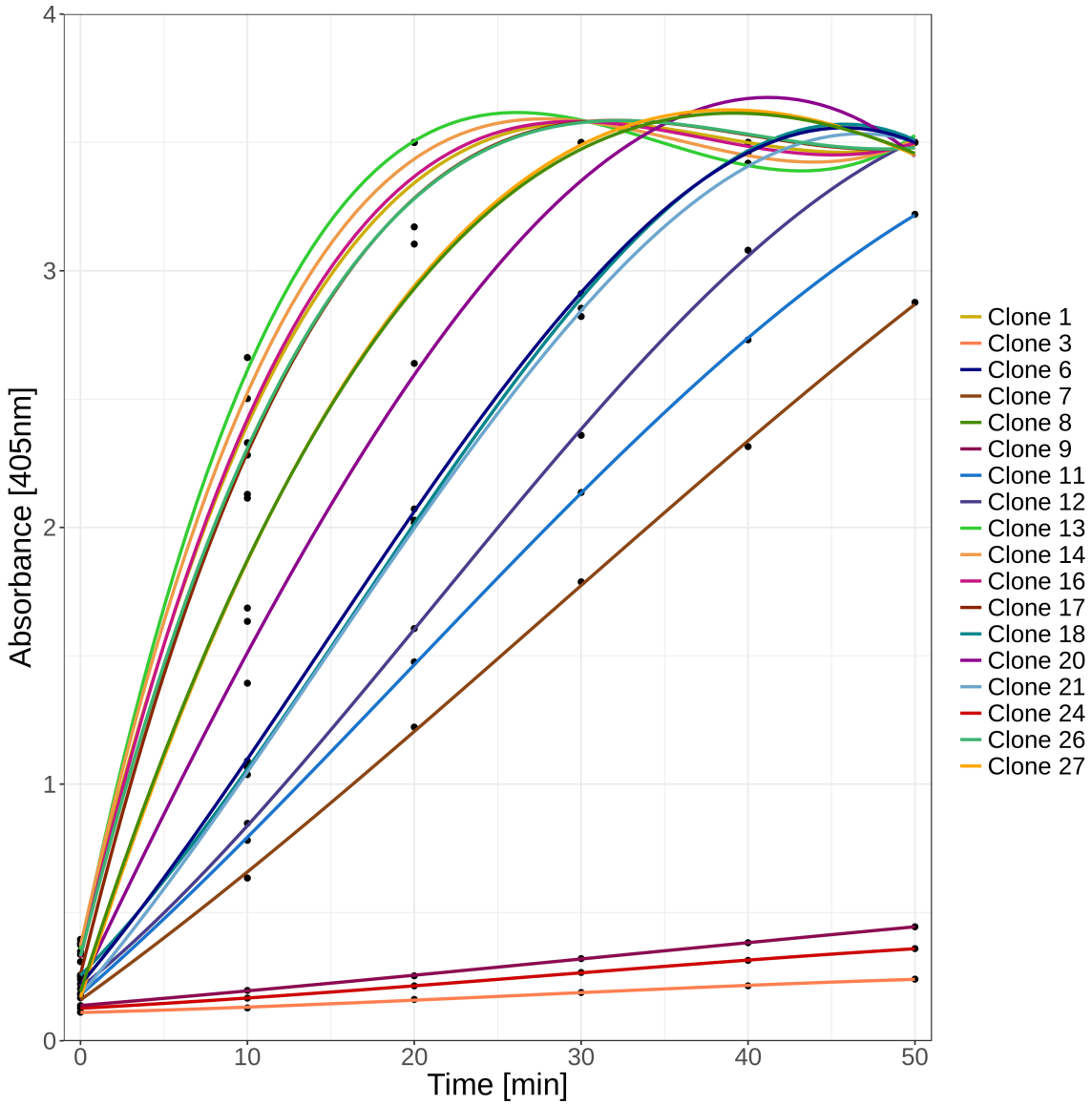


Figure A.2: Stimulation of putative *TLR5*$^{-/-}$ CRISPR-Cas9-edited cells with flagellin. Single-colony clones from bulk transfection of the CRISPR-Cas9 ribonucleoprotein complex were treated with flagellin (1 μg/ml) and supernatants were assessed for SEAP activity over a period of 50 minutes.

A.2 Tables

Table A.1: Polymorphisms and diversity indices for fourteen colonies of Gentoo penguin at three Toll-like receptor loci and the mitochondrial hypervariable control region.

Locus	Population (N)	S	H	H_d (SD)	π (SD) (10^{-4})	UPV	AR	D	Fs
TLR4	CR (7)	5	4	0.6593 (0.1203)	5.86 (1.9)	3	3.80	-0.17	0.48
	MAR (8)	5	4	0.7667 (0.0573)	9.80 (1.2)	4	3.86	2.17	2.07
	COU (7)	8	7	0.6923 (0.1366)	8.32 (2.4)	4	5.29	-0.59	-1.89
	MO (15)	9	14	0.9034 (0.0362)	11.01 (0.8)	8	7.03	0.73	-5.87
	CB (19)	2	2	0.0526 (0.0492)	0.41 (0.4)	1	1.26	-1.49	-0.61
	BR (7)	0	1	0	0	1	1.00	0.00	0.00
	BI (18)	0	1	0	0	1	1.00	0.00	0.00
	MT (5)	0	1	0	0	1	1.00	0.00	0.00
	SIG (5)	4	3	0.5111 (0.1643)	6.01 (2.3)	1	3.00	0.32	1.36
	COP (22)	5	5	0.6342 (0.0511)	7.46 (0.7)	1	3.25	1.63	1.71
	SP (5)	4	3	0.5111 (0.1643)	6.45 (2.1)	1	3.00	0.63	1.52
	BO (6)	5	3	0.5455 (0.1436)	5.94 (2.3)	2	2.97	-0.31	1.60
	GGV (7)	4	2	0.2637 (0.1360)	4.14 (2.1)	1	1.93	-0.53	2.70
	JP (24)	4	3	0.0824 (0.0541)	1.12 (0.7)	1	1.42	-1.56	-0.73
	Total (155)	13	21	0.478 (0.035)	6.1 (0.5)	9	3.45	-0.58	-10.02
TLR5	CR (7)	1	2	0.4396 (0.1120)	1.66 (0.4)	2	2.00	0.84	0.94
	MAR (8)	1	2	0.1250 (0.1064)	0.47 (0.4)	2	1.62	-1.16	-0.70
	COU (7)	15	8	0.8571 (0.0774)	16.09 (2.9)	7	6.28	-0.40	-0.89
	MO (20)	13	17	0.8731 (0.0415)	12.30 (1.3)	14	6.70	0.20	-6.78
	CB (18)	11	17	0.9159 (0.0309)	17.10 (0.7)	12	7.38	2.18	-4.89
	BR (7)	9	6	0.7692 (0.0895)	16.09 (1.9)	5	4.86	1.93	1.05
	BI (12)	5	5	0.4928 (0.1165)	5.25 (1.6)	4	2.88	0.11	-0.03
	MT (5)	9	6	0.8444 (0.1029)	15.72 (2.9)	5	6.00	1.34	-0.05
	SIG (5)	5	3	0.5111 (0.1643)	4.37 (2.3)	3	3.00	-1.39	0.76
	COP (21)	6	6	0.6527 (0.0485)	4.54 (1.0)	6	3.46	-0.37	-0.68
	SP (5)	1	2	0.2000 (0.1541)	0.76 (0.6)	2	2.00	-1.11	-0.34
	BO (6)	5	3	0.3182 (0.1637)	3.15 (2.1)	3	2.67	-1.83	0.32

Cont. next page

Locus	Population (N)	S	H	H_d (SD)	π (SD) (10^{-4})	UPV	AR	D	Fs
	GGV (7)	2	3	0.3846 (0.1494)	1.54 (0.6)	3	2.65	-0.96	-0.85
	JP (24)	1	2	0.1197 (0.0612)	0.45 (0.2)	2	1.51	-0.64	-0.38
	Total (152)	20	46	0.781 (0.022)	14.8 (0.6)	32	5.60	0.60	-23.92
TLR7	CR (7)	0	1	0	0	1	1.00	0.00	0.00
	MAR (8)	0	1	0	0	1	1.00	0.00	0.00
	COU (7)	3	4	0.4945 (0.1506)	2.32 (0.8)	4	2.98	-0.71	-1.12
	MO (18)	2	3	0.3381 (0.0919)	1.13 (0.3)	2	2.27	-0.52	-0.49
	CB (17)	1	2	0.0588 (0.0546)	0.19 (0.2)	2	1.24	-1.14	-1.32
	BR (7)	1	2	0.3626 (0.1302)	1.16 (0.4)	2	1.95	0.32	0.64
	BI (13)	0	1	0	0	1	1.00	0.00	0.00
	MT (4)	1	2	0.4286 (0.1687)	1.37 (0.5)	2	2.00	0.33	0.54
	SIG (5)	2	3	0.6000 (0.1305)	2.13 (0.6)	2	2.80	-0.18	-0.27
	COP (21)	3	4	0.3751 (0.0902)	1.30 (0.3)	3	2.41	-0.90	-1.44
	SP (5)	1	2	0.4667 (0.1318)	1.49 (0.4)	1	2.00	0.82	0.82
	BO (6)	2	3	0.4394 (0.1581)	1.50 (0.6)	2	2.58	-0.85	-0.72
	GGV (7)	2	3	0.3846 (0.1494)	1.30 (0.5)	2	2.41	-0.96	-0.85
	JP (23)	4	5	0.3826 (0.0881)	1.34 (0.3)	4	2.51	-1.25	-2.59
	Total (148)	9	10	0.296 (0.034)	1.0 (0.1)	8	2.14	-1.68	-9.38
HVR1	CR (7)	8	4	0.8570 (0.102)	123.5 (33.0)	N/A	N/A	0.11	0.87
	MAR (8)	5	4	0.7860 (0.113)	68.8 (20.5)	N/A	N/A	-0.17	0.01
	COU (2)	0	1	0	0	N/A	N/A	0.00	0.00
	MO (20)	28	16	0.9580 (0.033)	173.2 (10.0)	N/A	N/A	-0.93	-7.01
	CB (16)	12	10	0.9000 (0.062)	116.4 (15.4)	N/A	N/A	-0.50	-3.58
	BR (7)	5	4	0.8570 (0.102)	77.6 (18.5)	N/A	N/A	0.13	-0.04
	BI (17)	9	8	0.8820 (0.047)	78.2 (13.8)	N/A	N/A	-0.76	-3.71
	MT (5)	4	2	0.6000 (0.175)	88.9 (26.0)	N/A	N/A	1.64	3.02
	SIG (5)	6	5	1.0000 (0.126)	103.7 (18.8)	N/A	N/A	-0.19	-2.37
	COP (21)	22	15	0.9670 (0.024)	151.4 (18.9)	N/A	N/A	-1.27	-6.16
	SP (6)	16	6	1.0000 (0.096)	227.2 (34.2)	N/A	N/A	-0.77	-1.77
	JP (24)	18	15	0.9420 (0.031)	127.2 (18.6)	N/A	N/A	-1.04	-7.43

Cont. next page

Locus	Population (N)	S	H	H_d (SD)	π (SD) (10^{-4})	UPV	AR	D	Fs
	Total (138)	90	79	0.986 (0.003)	744.3 (56.6)	N/A	N/A	0.53	-22.20

Polymorphisms and diversity indices for fourteen colonies of Gentoo penguin at three Toll-like receptor loci and the mitochondrial hypervariable control region. (CR = Crozet Island; MAR = Marion Island; COU = Courbet Peninsula, Kerguelen; MO = Pointe du Morne, Kerguelen; CB = Cow Bay, Falkland/Malvinas Islands; BR = Bull Roads, Falkland/Malvinas Islands; BI = Bird Island, South Georgia; MT = Martillo Island, Tierra del Fuego; SIG = Signy Island, South Orkney Islands; COP = Copacabana (Admiralty Bay), King George Island, South Shetland Islands; SP = Stranger Point, King George Island, South Shetland Islands; BO = Bernardo O'Higgins Base, Western Antarctic Peninsula; GGV = Gabriel González Videla Base, Western Antarctic Peninsula; JP = Jougla Point, Western Antarctic Peninsula). S = number of polymorphic sites; H = number of haplotypes; Hd = haplotype diversity (standard deviation); π = nucleotide diversity; UPV = unique protein variants of translated haplotypes; AR = allelic richness, adjusted for sample size differences; D = Tajima's D test statistic; Fs = Fu's Fs test statistic.

Table A.2: List of primers used for PCR amplification and sequencing of Gentoo penguin TLR genes, along with reaction conditions.

Gene (exon)	Primers	Product (bp)	Taq	Ta (°C)
TLR4 (1)	F: CTCACTCCAGGTCCTTGAGTG	419	Qiagen Taq	55
	R: ACTGCAGTACCCTCAGAGGT			
TLR4 (2)	F: TGCTGCTTTCCTGACAGTGT	358	Qiagen Taq	55
	R: AGCCAAGACCAACAGACACC			
TLR4 (3)	F: AAGGGACAGTGTTGCATGCA	945	Qiagen Taq	55
	R: CACTCAAATCTACAACCTCCAGG			
TLR4 (3)	F: GGCAACGTCTCTAGTATTCGGT	1563/1671	Qiagen Taq	50
	R: TGCCTTCTAGCAGGACTCCT			
	R: ACTTCTGAGCACCAGCCATC			
TLR5	F: AGTGTGACATTAGTCTTTTGTTCTACA	1021	Phusion	64
	R: AGCCAAATCCCGAACCCATT			
TLR5	F: ACCTTAGTGACAACGGCTGG	1948	Phusion	68
	R: GTGTTCATTCCTGCCATGGC			
TLR5	F: TCCCCTGAGCCTGAAGTCTT	962	Phusion	69
	R: GTGTTCATTCCTGCCATGGC			
TLR7	F: AGCCAGTGTGAGAGACAGGA	670	Phusion	62
	R: ACGGTAGTAACAGTTCTGTC			
TLR7	F: GAGCTAGGAAACATAGAAAT	901	Phusion	55
	R: TCATCATACCTGAAATAATG			
TLR7	F: AAATCTAACAGTGTTGGATC	1229	Phusion	50
	R: GTCACTTGAGTCTGATTGAT			
TLR7	F: AATTAAGAAATCTAGCTTCC	984	Phusion	54
	R: TTTAGGGTACAATATCACAA			

Table A.3: *TLR4* pairwise F_{ST} values and pairwise ϕ_{ST} values

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)		0.04	0.43	0.16	0.85	0.78	0.87	0.75	0.54	0.40	0.49	0.56	0.61	0.82
		-0.18	-0.00	-0.01	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
Marion Island (MAR)	0.00		0.20	0.02	0.67	0.55	0.69	0.50	0.32	0.26	0.27	0.34	0.38	0.65
	-0.41		-0.01	-0.24	0.00	0.00	0.00	-0.00	-0.01	-0.00	-0.01	-0.00	-0.00	0.00
Courbet Peninsula (COU)	0.25	0.13		0.05	0.22	0.13	0.26	0.09	0.00	0.07	0.00	0.00	0.00	0.19
	-0.00	-0.15		-0.11	-0.00	-0.04	-0.00	-0.11	-0.41	-0.08	-0.50	-0.40	-0.36	-0.01
Pointe du Morne (MO)	0.08	0.02	0.03		0.40	0.30	0.42	0.27	0.13	0.11	0.09	0.15	0.17	0.39
	-0.01	-0.19	-0.10		0.00	-0.00	0.00	-0.00	-0.03	-0.01	-0.07	-0.01	-0.01	0.00
Cow Bay (CB)	0.71	0.58	0.28	0.38		0.00	0.00	0.00	0.30	0.31	0.34	0.19	0.13	0.00
	0.00	0.00	-0.00	0.00		-1.00	-1.00	-1.00	-0.01	-0.00	-0.01	-0.01	-0.07	-0.61
Bull Roads (BR)	0.62	0.47	0.19	0.30	0.00		0.00	0.00	0.20	0.25	0.23	0.11	0.08	0.00
	0.00	0.00	-0.02	0.00	-1.00		-1.00	-1.00	-0.06	-0.00	-0.05	-0.03	-0.48	-1.00
Bird Island (BI)	0.75	0.62	0.34	0.41	0.00	0.00		0.00	0.38	0.33	0.41	0.25	0.19	0.01
	0.00	0.00	-0.00	0.00	-1.00	-1.00		-1.00	-0.01	0.00	-0.01	-0.00	-0.07	-0.51
Martillo Island (MT)	0.57	0.42	0.15	0.27	0.00	0.00	0.00		0.15	0.22	0.18	0.07	0.04	0.00
	-0.00	-0.00	-0.05	-0.00	-1.00	-1.00	-1.00		-0.21	-0.02	-0.21	-0.32	-0.49	-1.00
Signy Island (SIG)	0.34	0.21	0.00	0.10	0.27	0.20	0.38	0.15		0.00	0.00	0.00	0.00	0.20
	-0.00	-0.01	-0.43	-0.01	-0.01	-0.06	-0.01	-0.21		-0.33	-0.80	-0.68	-0.77	-0.02
Copacabana, KGI (COP)	0.30	0.20	0.06	0.10	0.29	0.23	0.31	0.20	0.00		0.00	0.05	0.07	0.29
	0.00	-0.00	-0.08	-0.00	0.00	-0.00	0.00	-0.01	-0.42		-0.49	-0.15	-0.10	0.00
Stranger Point, KGI (SP)	0.34	0.21	0.00	0.10	0.27	0.20	0.38	0.15	0.00	0.00		0.00	0.00	0.24
	-0.00	-0.01	-0.43	-0.01	-0.01	-0.06	-0.01	-0.22	-1.00	-0.42		-0.61	-0.62	-0.03
Bernardo O'Higgins Base (BO)	0.33	0.21	0.00	0.10	0.29	0.20	0.37	0.16	0.00	0.02	0.00		0.00	0.12
	-0.00	-0.00	-0.37	-0.01	-0.00	-0.03	-0.00	-0.10	-0.70	-0.25	-0.71		-0.68	-0.03
Gabriel González Videla Base (GGV)	0.47	0.33	0.01	0.20	0.10	0.08	0.19	0.04	0.00	0.09	0.00	0.00		0.05
	0.00	-0.00	-0.14	-0.00	-0.07	-0.46	-0.08	-0.50	-0.61	-0.06	-0.62	-0.38		-0.12
Jougla Point (JP)	0.71	0.59	0.27	0.39	0.00	0.00	0.00	0.00	0.23	0.28	0.23	0.26	0.06	
	0.00	0.00	-0.00	0.00	-1.00	-1.00	-0.51	-1.00	-0.03	0.00	-0.02	-0.01	-0.12	

Table A.4: *TLR5* pairwise F_{ST} values and pairwise ϕ_{ST} values

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)		0.10	0.20	0.14	0.42	0.62	0.82	0.65	0.86	0.83	0.94	0.88	0.92	0.96
		-0.16	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Marion Island (MAR)	0.10		0.27	0.19	0.46	0.66	0.84	0.71	0.90	0.85	0.97	0.92	0.95	0.98
	-0.16		-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Courbet Peninsula (COU)	0.13	0.28		0.00	0.18	0.34	0.56	0.32	0.51	0.65	0.58	0.54	0.60	0.79
	-0.02	-0.00		-0.91	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00
Pointe du Morne (MO)	0.10	0.21	0.00		0.25	0.43	0.57	0.41	0.54	0.64	0.59	0.56	0.60	0.72
	-0.01	0.00	-0.56		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cow Bay (CB)	0.28	0.41	0.10	0.10		0.05	0.37	0.05	0.33	0.45	0.38	0.34	0.40	0.55
	0.00	0.00	0.00	0.00		-0.10	0.00	-0.12	0.00	0.00	0.00	0.00	0.00	0.00
Bull Roads (BR)	0.40	0.57	0.18	0.17	0.03		0.50	0.00	0.44	0.58	0.50	0.46	0.53	0.73
	0.00	0.00	0.00	0.00	-0.10		0.00	-0.72	-0.00	0.00	-0.00	-0.00	-0.00	0.00
Bird Island (BI)	0.53	0.66	0.35	0.30	0.13	0.22		0.60	0.01	0.03	0.06	0.01	0.06	0.18
	0.00	0.00	0.00	0.00	0.00	-0.00		0.00	-0.32	-0.13	-0.10	-0.31	-0.11	0.00
Martillo Island (MT)	0.38	0.57	0.14	0.14	0.04	0.00	0.31		0.55	0.68	0.62	0.58	0.65	0.83
	0.00	0.00	-0.00	-0.00	-0.05	-0.81	-0.00		-0.00	0.00	-0.00	-0.00	0.00	0.00
Signy Island (SIG)	0.53	0.72	0.30	0.27	0.10	0.19	0.01	0.27		0.08	0.00	0.00	0.00	0.13
	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.31	-0.00		-0.07	-1.00	-0.68	-0.82	-0.06
Copacabana, KGI (COP)	0.43	0.54	0.26	0.24	0.11	0.19	0.07	0.24	0.09		0.10	0.08	0.11	0.20
	0.00	0.00	0.00	0.00	0.00	-0.00	-0.04	-0.01	-0.06		-0.04	-0.05	-0.01	0.00
Stranger Point, KGI (SP)	0.66	0.85	0.44	0.37	0.20	0.32	0.02	0.43	0.00	0.17		0.00	0.00	0.00
	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.29	-0.00	-0.57	-0.01		-1.00	-1.00	-1.00
Bernardo O'Higgins Base (BO)	0.62	0.79	0.40	0.34	0.17	0.27	0.00	0.37	0.00	0.15	0.00		0.00	0.06
	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.37	-0.00	-0.57	-0.02	-1.00		-0.80	-0.22
Gabriel González Videla Base (GGV)	0.59	0.75	0.38	0.32	0.15	0.26	0.01	0.35	0.00	0.13	0.00	0.00		0.02
	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.32	-0.00	-0.87	-0.02	-0.81	-1.00		-0.19
Jougla Point (JP)	0.79	0.88	0.65	0.52	0.35	0.54	0.12	0.66	0.15	0.31	0.00	0.01	0.05	
	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.06	0.00	-1.00	-0.58	-0.12	

Table A.5: *TLR7* pairwise F_{ST} values and pairwise ϕ_{ST} values

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)		0.00	0.15	0.04	0.00	0.15	0.24	0.00	0.22	0.01	0.28	0.08	0.05	0.00
		-1.00	-0.10	-0.25	-1.00	-0.22	-0.12	-1.00	-0.02	-0.41	-0.06	-0.09	-0.48	-0.62
Marion Island (MAR)	0.00		0.17	0.04	0.00	0.17	0.27	0.00	0.24	0.01	0.30	0.09	0.06	0.00
	-1.00		-0.03	-0.12	-1.00	-0.10	-0.10	-1.00	-0.02	-0.25	-0.04	-0.06	-0.09	-0.26
Courbet Peninsula (COU)	0.13	0.15		0.03	0.25	0.00	0.00	0.24	0.10	0.08	0.17	0.00	0.06	0.11
	-0.10	-0.04		-0.13	-0.01	-0.91	-1.00	-0.01	-0.11	-0.02	-0.02	-0.53	-0.10	-0.01
Pointe du Morne (MO)	0.07	0.08	0.00		0.08	0.00	0.00	0.07	0.14	0.02	0.19	0.00	0.03	0.03
	-0.17	-0.10	-0.58		-0.02	-0.59	-0.60	-0.04	-0.03	-0.12	-0.01	-0.50	-0.20	-0.04
Cow Bay (CB)	0.00	0.00	0.18	0.09		0.21	0.30	0.00	0.32	0.04	0.36	0.14	0.11	0.03
	-1.00	-1.00	-0.01	-0.02		-0.02	-0.03	-1.00	-0.00	-0.03	-0.01	-0.03	-0.02	-0.04
Bull Roads (BR)	0.15	0.17	0.00	0.00	0.19		0.00	0.24	0.11	0.03	0.20	0.00	0.03	0.05
	-0.22	-0.09	-1.00	-0.86	-0.02		-1.00	-0.03	-0.07	-0.14	-0.02	-0.81	-0.24	-0.10
Bird Island (BI)	0.24	0.27	0.00	0.00	0.27	0.00		0.37	0.08	0.03	0.18	0.00	0.03	0.05
	-0.12	-0.10	-1.00	-0.61	-0.03	-1.00		-0.05	-0.21	-0.26	-0.07	-1.00	-0.34	-0.19
Martillo Island (MT)	0.00	0.00	0.22	0.11	0.00	0.24	0.37		0.33	0.03	0.40	0.15	0.11	0.02
	-1.00	-1.00	-0.01	-0.04	-1.00	-0.04	-0.05		-0.00	-0.08	-0.02	-0.03	-0.04	-0.14
Signy Island (SIG)	0.31	0.33	0.02	0.10	0.38	0.06	0.02	0.43		0.04	0.00	0.00	0.00	0.05
	-0.02	-0.01	-0.38	-0.05	-0.00	-0.21	-0.49	-0.00		-0.20	-1.00	-0.44	-0.68	-0.16
Copacabana, KGI (COP)	0.06	0.07	0.00	0.00	0.08	0.00	0.00	0.09	0.03		0.05	0.00	0.00	0.00
	-0.13	-0.10	-0.51	-0.34	-0.01	-0.47	-0.52	-0.02	-0.17		-0.14	-0.83	-0.86	-0.58
Stranger Point, KGI (SP)	0.28	0.30	0.04	0.10	0.33	0.08	0.06	0.40	0.00	0.01		0.06	0.00	0.06
	-0.06	-0.05	-0.23	-0.05	-0.01	-0.15	-0.26	-0.02	-1.00	-0.37		-0.19	-0.61	-0.15
Bernardo O'Higgins Base (BO)	0.14	0.16	0.00	0.00	0.18	0.00	0.00	0.23	0.00	0.00	0.00		0.00	0.00
	-0.08	-0.07	-1.00	-0.76	-0.03	-1.00	-1.00	-0.03	-0.57	-0.82	-0.44		-0.84	-0.49
Gabriel González Videla Base (GGV)	0.10	0.12	0.00	0.00	0.13	0.00	0.00	0.18	0.00	0.00	0.00	0.00		0.00
	-0.22	-0.09	-0.68	-0.46	-0.03	-0.78	-0.58	-0.04	-0.58	-1.00	-0.61	-1.00		-0.85
Jougla Point (JP)	0.05	0.06	0.00	0.01	0.07	0.01	0.00	0.08	0.04	0.00	0.01	0.00	0.00	
	-0.12	-0.09	-0.39	-0.22	-0.01	-0.33	-0.38	-0.02	-0.15	-0.77	-0.36	-0.71	-0.93	

Table A.6: HVRI pairwise F_{ST} values and pairwise ϕ_{ST} values

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	JP
Crozet Island (CR)		0.00	0.98	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.96
		-0.57	-0.03	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00
Marion Island (MAR)	0.00		0.99	0.98	0.97	0.98	0.98	0.98	0.98	0.96	0.96	0.96
	-0.67		-0.02	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00
Courbet Peninsula (COU)	0.35	0.39		0.29	0.94	0.97	0.96	0.96	0.96	0.93	0.91	0.94
	-0.11	-0.02		-0.01	-0.01	-0.03	-0.01	-0.05	-0.05	-0.00	-0.04	-0.00
Pointe du Morne (MO)	0.09	0.12	0.21		0.93	0.93	0.93	0.93	0.91	0.92	0.91	0.92
	-0.00	-0.00	-0.02		0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00
Cow Bay (CB)	0.12	0.15	0.29	0.07		0.11	0.79	0.28	0.74	0.72	0.71	0.74
	-0.00	-0.00	-0.02	-0.00		-0.03	0.00	-0.00	0.00	0.00	0.00	0.00
Bull Roads (BR)	0.14	0.18	0.35	0.09	0.09		0.84	0.28	0.82	0.76	0.75	0.78
	-0.01	-0.01	-0.11	-0.00	-0.03		0.00	-0.03	-0.00	0.00	-0.00	0.00
Bird Island (BI)	0.13	0.16	0.30	0.08	0.11	0.13		0.84	0.33	0.24	0.25	0.22
	-0.01	-0.00	-0.01	0.00	-0.00	-0.01		0.00	-0.00	0.00	-0.00	0.00
Martillo Island (MT)	0.26	0.29	0.55	0.18	0.21	0.11	0.22		0.81	0.74	0.70	0.77
	-0.00	-0.00	-0.10	-0.00	-0.00	-0.13	-0.00		-0.01	0.00	-0.00	0.00
Signy Island (SIG)	0.08	0.12	0.29	0.02	0.06	0.08	0.07	0.20		0.02	0.03	0.01
	-0.11	-0.05	-0.05	-0.20	-0.11	-0.11	-0.08	-0.08		-0.29	-0.26	-0.30
Copacabana, KGI (COP)	0.08	0.11	0.24	0.04	0.07	0.08	0.07	0.17	0.02		0.00	0.02
	-0.00	-0.00	-0.02	0.00	-0.00	-0.00	0.00	-0.00	-0.22		-0.53	-0.16
Stranger Point, KGI (SP)	0.07	0.11	0.27	0.02	0.06	0.07	0.07	0.19	0.00	0.01		0.07
	-0.10	-0.05	-0.03	-0.17	-0.08	-0.10	-0.06	-0.00	-1.00	-0.32		-0.09
Jougla Point (JP)	0.09	0.13	0.25	0.05	0.08	0.09	0.07	0.18	0.02	0.02	0.03	
	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	-0.26	-0.04	-0.14	

A.2.1 Notes on Table A.3, A.4, A.5 and A.6

Pairwise F_{ST} values [630] derived from haplotype frequencies (below diagonal), and pairwise ϕ_{ST} values of genetic distance (above the diagonal), calculated using the Tamura and Nei substitution model [572], with associated P-values, among Gentoo penguin breeding sites based on Toll-like receptor sequence data. mtDNA hypervariable region I ϕ_{ST} values were calculated using the Kimura 2-Parameter (K80) substitution model [310] with a gamma of 0.27.

Table A.7: Geographic distances between sampling locations.

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)	0													
Marion Island (MAR)	960	0												
Courbet Peninsula (COU)	1577	2461	0											
Pointe du Morne (MO)	1565	2449	14	0										
Cow Bay (CB)	7190	6481	7850	7837	0									
Bull Roads (BR)	7185	6504	7833	7819	168	0								
Bird Island (BI)	5958	5204	6786	6773	1359	1432	0							
Martillo Island (MT)	7354	6743	7915	7903	792	639	1892	0						
Signy Island (SIG)	5955	5315	6565	6555	1282	1258	901	1437	0					
Copacabana, KGI(COP)	6419	5848	6978	6970	1275	1173	1501	1010	705	0				
Stranger Point, KGI (SP)	6421	5849	6979	6971	1262	1161	1506	995	711	13	0			
Bernardo O'Higgins Base (BO)	6402	5804	6936	6928	1351	1248	1536	1085	708	130	125	0		
Gabriel González Videla Base (GGV)	6705	6107	7237	7229	1522	1414	1837	1147	1009	372	360	302	0	
Jougla Point (JP)	6719	6138	7264	7256	1543	1435	1861	1170	1035	400	391	328	45	0

Pairwise geographic distances (most direct distance by sea during summer sea ice extent, in km) used for Mantel's test, derived from Google Earth v7.3.2.5776.

Table A.8: Mantel Test results for isolation by distance.

Mantel Test Locus	Correlation Coefficient (R)	P value
TLR4	0.52	0.00
TLR5	0.59	0.00
TLR7	0.03	0.37
HVR1	0.31	0.02

Table A.9: Log-likelihood comparisons for PAML analysis of *TLR4*

Model	lnL	$2(\ln L_1 - \ln L_0)$	df	P-value
M1a	-3605.47	8.33	2	0.02
M2a	-3597.14			
M7	-3605.72	8.34	2	0.02
M8	-3597.38			
M8a	-3605.47	8.09	2	0.02
M8	-3597.38			

Table A.10: Bayes Empirical Bayes analysis of *P. papua TLR4* sites under positive selection (M2a)

Amino acid pos.	Amino acid	$\Pr(\omega > 1)$	Mean $\omega \pm$ SE
12	V	0.93	7.669 ± 2.761
236	I	0.93	7.669 ± 2.761
659	A	0.93	7.671 ± 2.759

Table A.11: Bayes Empirical Bayes analysis of *P. papua TLR4* sites under positive selection (M8)

Amino acid pos.	Amino acid	Pr(ω >1)	Mean ω ± SE
12	V	0.95	1.461 ± 0.163
82	N	0.94	1.485 ± 0.000
236	I	0.95	1.461 ± 0.163
316	C	0.94	1.485 ± 0.000
445	Q	0.94	1.485 ± 0.000
659	A	0.95	1.461 ± 0.163

Table A.12: Log-likelihood comparisons for PAML analysis of *TLR5*

Model	lnL	2($\ln L_1$-$\ln L_0$)	df	P-value
M1a	-3929.20	102.08	2	0.00
M2a	-3878.16			
M7	-3929.32	102.31	2	0.00
M8	-3878.17			
M8a	-3928.47	100.60	2	0.00
M8	-3878.17			

Table A.13: Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection (M2a)

Amino acid pos.	Amino acid	Pr(ω >1)	Mean ω ± SE
10	Y	1.00	9.645 ± 0.504
285	M	0.98	9.472 ± 1.320
442	A	1.00	9.631 ± 0.609
535	S	0.98	9.482 ± 1.286
667	S	1.00	9.643 ± 0.514
698	M	0.98	9.469 ± 1.329
747	L	0.98	9.518 ± 1.163
788	C	0.98	9.492 ± 1.255
845	V	1.00	9.645 ± 0.503

Table A.14: Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection (M8)

Amino acid pos.	Amino acid	Pr(ω >1)	Mean ω ± SE
10	Y	1.00	8.947 ± 1.200
285	M	0.99	8.855 ± 1.487
442	A	1.00	8.941 ± 1.219
535	S	0.99	8.860 ± 1.472
667	S	1.00	8.946 ± 1.202
698	M	0.99	8.853 ± 1.490
747	L	0.99	8.878 ± 1.421
788	C	0.99	8.865 ± 1.459
845	V	1.00	8.947 ± 1.200

Table A.15: Predicted functional consequences of amino acid substitutions using homology-based methods.

TLR	nt site	aa site	variant 1	variant 2	SIFT [432]				Polyphen2 [2]					
					v1→v2		v2→ v1		v1→ v2			v2→ v1		
					score	MSC	score	MSC	score	sensitivity	specificity	score	sensitivity	specificity
TLR4	35	12	V	A	0.15	3.34	0.58	3.38	N/A	N/A	N/A	0.997	0.410	0.980
TLR4	706	236	V	I	1.00	3.20	0.18	3.21	0.001	0.990	0.150	0.049	0.940	0.830
TLR4	1977	659	A	T	0.20	3.04	0.25	3.05	0.810	0.840	0.930	0.000	1.000	0.000
TLR5	29	10	C	Y	0.00	4.32	0.00	4.32	0.004	0.970	0.590	N/A	N/A	N/A
TLR5	854	285	T	M	0.43	3.36	0.10	3.36	0.000	1.000	0.000	0.990	0.720	0.970
TLR5	1325	442	A	V	0.51	3.55	0.72	3.36	0.165	0.920	0.870	0.000	1.000	0.000
TLR5	1604	535	S	T	0.38	3.36	0.82	3.36	0.253	0.910	0.880	0.000	1.000	0.000
TLR5	2000	667	F	S	0.34	3.36	0.01	3.36	0.000	1.000	0.000	0.998	0.270	0.990
TLR5	2094	698	I	M	0.06	3.36	0.45	3.36	0.009	0.960	0.770	0.005	0.970	0.740
TLR5	2239	747	V	L	1.00	3.36	0.00	3.36	0.000	1.000	0.000	0.000	1.000	0.000
TLR5	2362	788	S	C	1.00	3.36	0.00	3.36	0.000	1.000	0.000	0.992	0.700	0.970
TLR5	2533	845	V	I	0.45	3.36	1.00	3.36	0.024	0.950	0.810	0.000	1.000	0.000

Table A.16: Proximity of selected sites to selected and functional sites in other taxa.

TLR	nt site	aa site	variant 1	variant 2	chicken posn	chicken aa	other literature	function
TLR4	35	12	V	A	6	A	[603]	No
TLR4	706	236	V	I	230	I	No	No
TLR4	1977	659	A	T	653	A	No	No
TLR5	29	10	C	Y	N/A	N/A	N/A	N/A
TLR5	854	285	T	M	267	M	No	Yes, adjacent to 2 sites [659, 551]
TLR5	1325	442	A	V	424	A	[603, 220]	No
TLR5	1604	535	S	T	517	S	No	No
TLR5	2000	667	F	S	649	S	No	No
TLR5	2094	698	I	M	680	L	[603]	No
TLR5	2239	747	V	L	729	L	No	No
TLR5	2362	788	S	C	770	C	No	No
TLR5	2533	845	V	I	827	I	No	No

Table A.17: Coordinates and sample sizes of sampling locations.

Sampling Site Name	Region	Latitude	Longitude	Sample size			
				mtDNA	*TLR4*	*TLR5*	*TLR7*
Crozet Island (CR)	Crozet Is.	46.4298 S	50.4046 E	7	7	7	7
Marion Island (MAR)	Prince Edward Is.	46.8872 S	37.8671 E	8	8	8	8
Courbet Peninsula (COU)	Kerguelen Is.	49.2811 S	70.5410 E	2	7	7	7
Pointe du Morne (MO)	Kerguelen Is.	49.3805 S	70.4423 E	20	15	20	18
Cow Bay (CB)	Falkland/Malvinas Is.	51.4343 S	57.8506 W	16	19	18	17
Bull Roads (BR)	Falkland/Malvinas Is.	52.3380 S	59.3633 W	7	7	7	7
Bird Island (BI)	South Georgia	54.0094 S	38.0631 W	17	18	12	13
Martillo Island (MT)	Tierra del Fuego	54.9083 S	67.3833 W	5	5	5	4
Signy Island (SIG)	South Orkney Is.	60.6747 S	45.6278 W	5	5	5	5
Copacabana, KGI (COP)	South Shetland Is.	62.1792 S	58.4472 W	21	22	21	21
Stranger Point, KGI (SP)	South Shetland Is.	62.2375 S	58.5939 W	6	5	5	5
Bernardo O'Higgins Base (BO)	Western Antarctic Peninsula	63.3207 S	57.8987 W	0	6	6	6
Gabriel González Videla Base (GGV)	Western Antarctic Peninsula	64.8238 S	62.8581 W	0	7	7	7
Jougla Point (JP)	Western Antarctic Peninsula	64.8270 S	63.4932 W	24	24	24	23
			Total	138	155	152	148

Table A.18: Census population size (number of breeding pairs) at each location using best available data.

Colony	Census Region	Census Population (breeding pairs)	Year	Reference
CR	Crozet Archipelago	9000	1994	[280]
MAR	Marion Island	600	2012/13	[123]
COU MO	Kerguelen Islands	30000	1988	[629]
CB BR	Falkland/Malvinas Islands	132321	2010	[52]
BI	South Georgia	98867	1996	[586]
MT	Martillo Island	12	2006/07	[203, 387]
SIG	South Orkney Islands	10760	2013	
COP SP BO GGV JP	S. Shet. Is. and WAP	94751	2013	[387]

For estimates of Kerguelen population sizes, Lescroel and Bost, 2006 [353] show significant decrease in 2002 of Kerguelen population, but was localized to the sampled portion of the archipelago, leading us to use 30,000 as the lower range number for Weimerskirch, et al. 1989's estimate of 30-40,000 [629].

Table A.19: Diversity statistics of Gentoo penguin *vs.* avian Toll-like receptors from published studies

Gene	Species	Length (bp)	k	$H_d \pm SD$	$\pi \pm SD$	UPV	A
	Gentoo penguin	2550	1.550	0.478 ± 0.035	0.00061 ± 0.00005	9	21
TLR4	Domestic chicken	2532	6.540	0.947 ± 0.010	0.00258 ± 0.00005	24	70
	Grey partridge	2529	3.063	0.863 ± 0.063	0.00121 ± 0.00018	5	10
	Gentoo penguin	2643	3.919	0.781 ± 0.022	0.00148 ± 0.00006	32	46
TLR5	Domestic chicken	2586	2.866	0.854 ± 0.013	0.00111 ± 0.00004	10	22
	Grey partridge	2589-2592	1.053	0.747 ± 0.076	0.00041 ± 0.00008	4	6
	Gentoo penguin	3126	0.327	0.296 ± 0.034	0.00010 ±0.00001	8	10
TLR7	Domestic chicken	3180	3.363	0.849 ± 0.017	0.00106 ± 0.00005	13	26
	Grey partridge	3141	1.753	0.532 ± 0.100	0.00056 ± 0.00013	3	3

k = average number of nucleotide differences between alleles; H_d = haplotype diversity; SD: standard deviation; π: nucleotide diversity (average number of nucleotide differences per site between two sequences); UPV: number of unique protein variants; A: number of alleles.

To compare diversity levels in our study population with those of other species, we sought datasets that sequenced *TLR4*, *TLR5*, and *TLR7* in birds. Because diversity measures can be biased by gene length and whether sequenced fragments are under selection, we restricted our comparison only to studies that sequenced the full length of these genes: a study in 110 domestic chickens (*Gallus gallus domesticus*, Świderská, *et al.* 2018 [565]) and a study of 10 European grey partridges (*Perdix perdix*, Vinkler, *et al.* 2015 [607]). The difference in gene length among these species is minimal enough to allow comparison (*TLR4*: <0.01%, *TLR5*: 2.15%, *TLR7*: 1.69%). However, sampling strategy differs significantly between studies, and also affects diversity measures (i.e. chicken study sampled few individuals from 25 distinct farmed breeds, partridge study sampled a small geographic subset of the species, while Gentoo penguin study had varying numbers of samples from multiple wild populations). Given the differences in sampling strategy, only relative statements comparing TLRs within a species/study are appropriate, and care should be given to drawing conclusions from interspecific comparisons.

Table A.20: Spearman's rank correlations calculated for haplotype diversity (H_d) *vs.* census population size.

Locus	Spearman's rank correlation	P-value
TLR4	-0.5029245	0.06679
TLR5	0.09462364	0.7476
TLR7	-0.1380377	0.6379
HVR1	N/A	0.3263

A.3 Figures

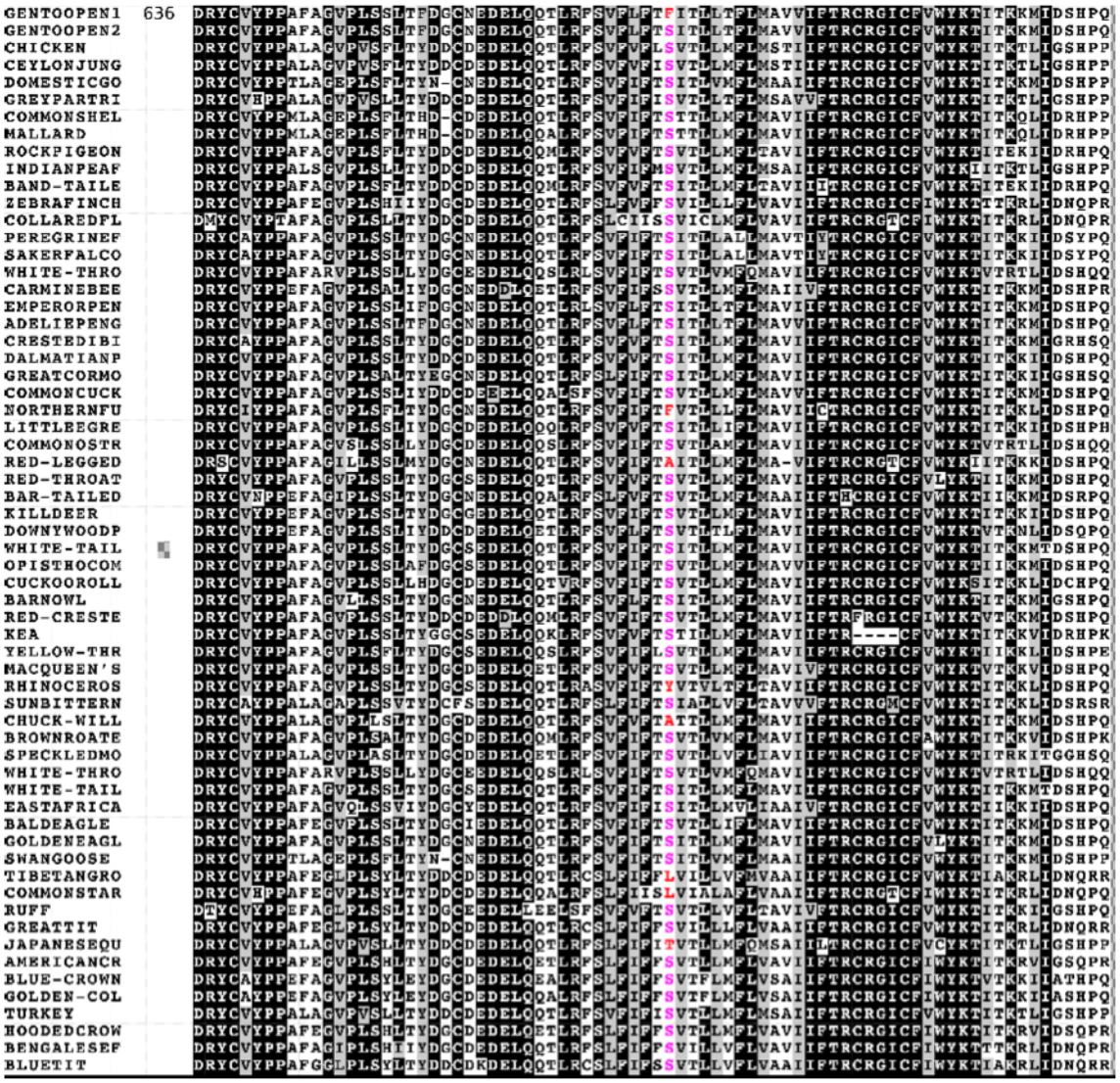


Figure A.3: Alignment of the transmembrane region and surrounding sequence of available avian TLR5 amino acid sequences. The conserved serine residue is highlighted in pink and any variant amino acids are highlighted in red.

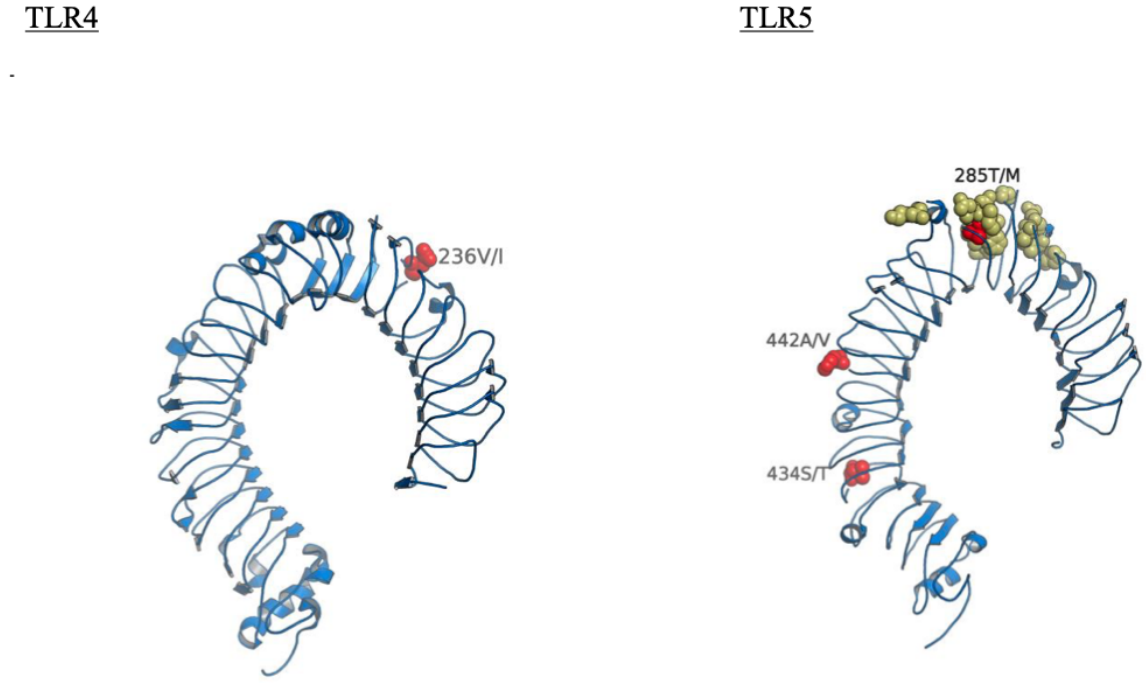


Figure A.4: Positions of positively selected sites in penguin TLR4 and TLR5 ectodomains. Positively selected sites from PAML analysis (highlighted in red) are superimposed on homologous sites from the crystal structure of mouse TLR4 ectodomain (PDB: 2Z64) and a homology model of human TLR5 ectodomain (PDB: 3J0A). Further positively selected sites in the signal peptide of TLR4 (12V/A) and TLR5 (10C/Y), as well as TMD/TIR domain sites (TLR4: 659A/T; TLR5: 667F/S, 698I/M, 747V/L, 788S/C, 845V/I) were omitted because the residues are not included in avaiable model structures or were of low confidence. Relative positions of these residues are depicted schematically in main text Figure 6. Site 285T/M is located within interface B of the flagellin-binding site [659, 551], which is highlighted in yellow. No other positively selected sites were found to be proximal to sites of known function in other vertebrates.

Appendix B

Appendix to Chapter 3

B.1 Figures

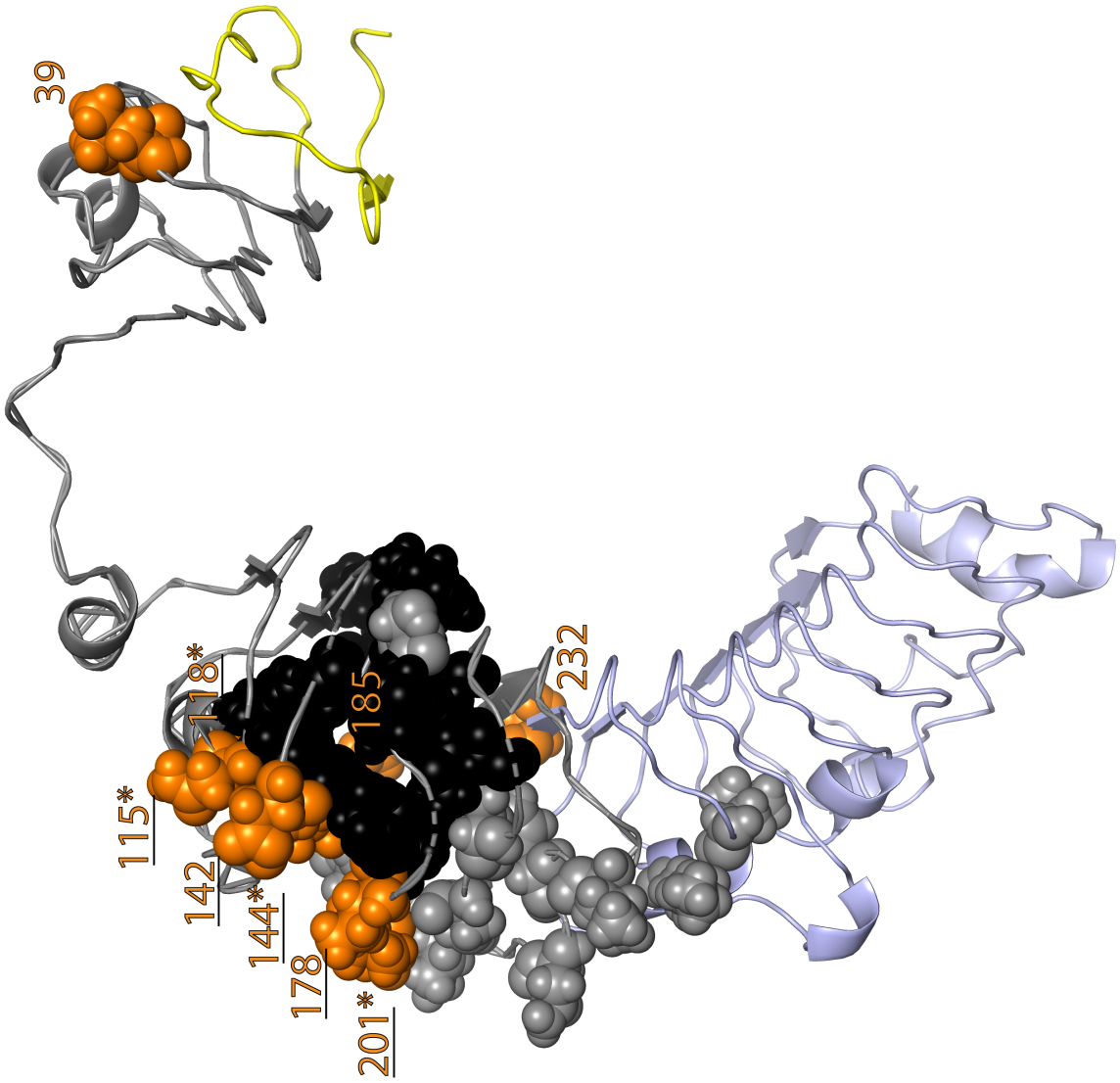


Figure B.1: TLR1B homology model showing positively selected sites.

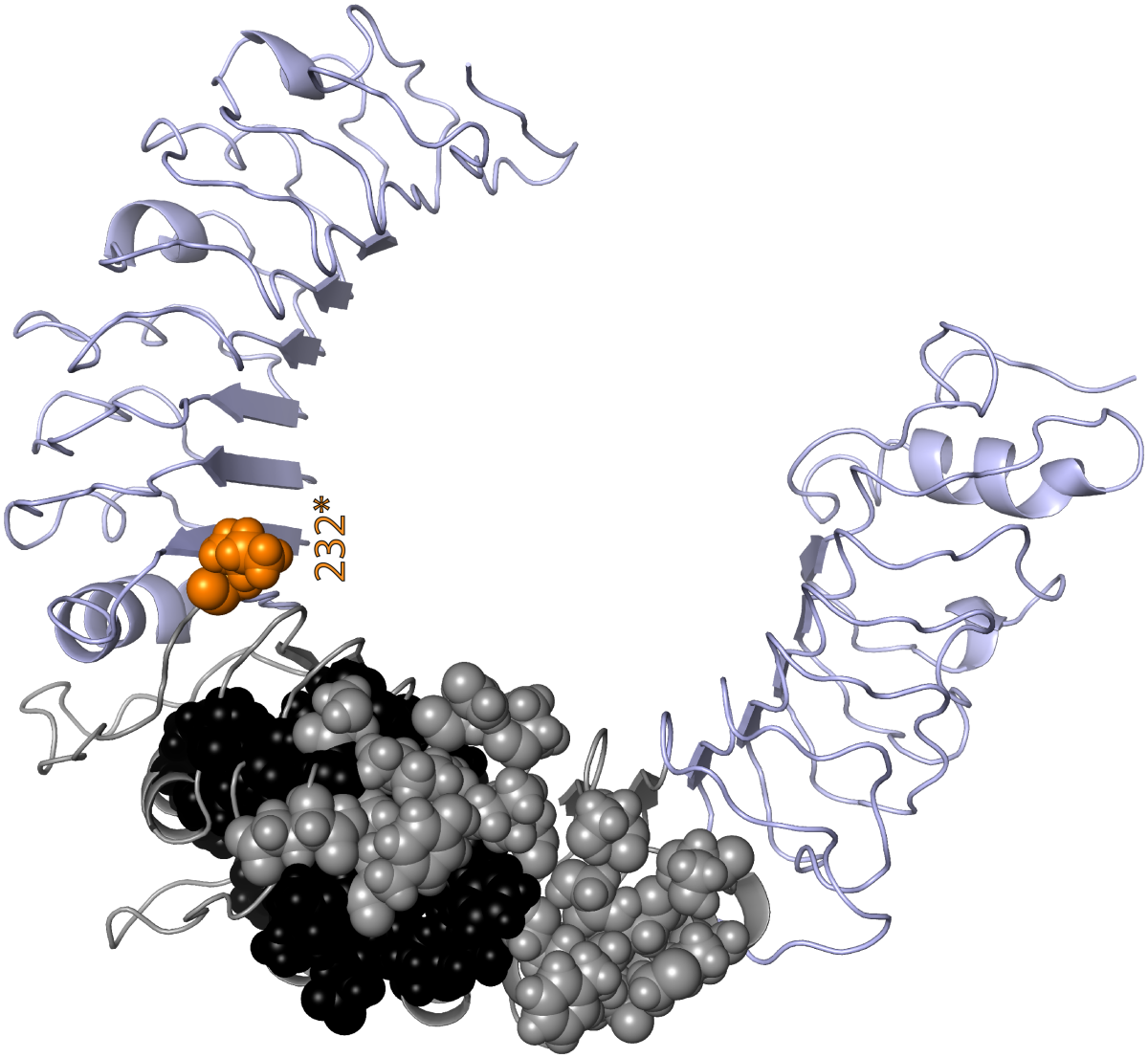


Figure B.2: TLR2B homology model showing positively selected sites.

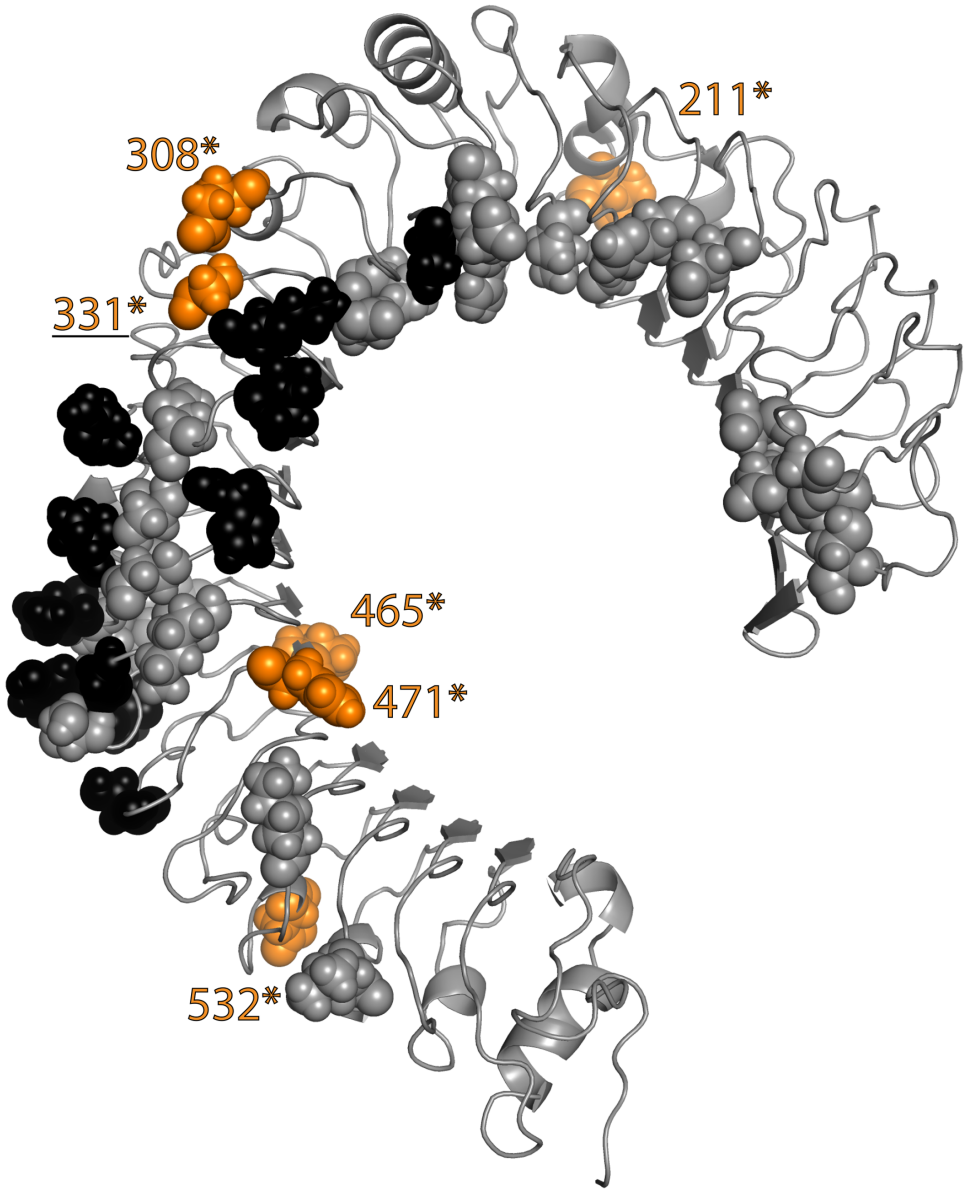


Figure B.3: TLR4 homology model showing positively selected sites.

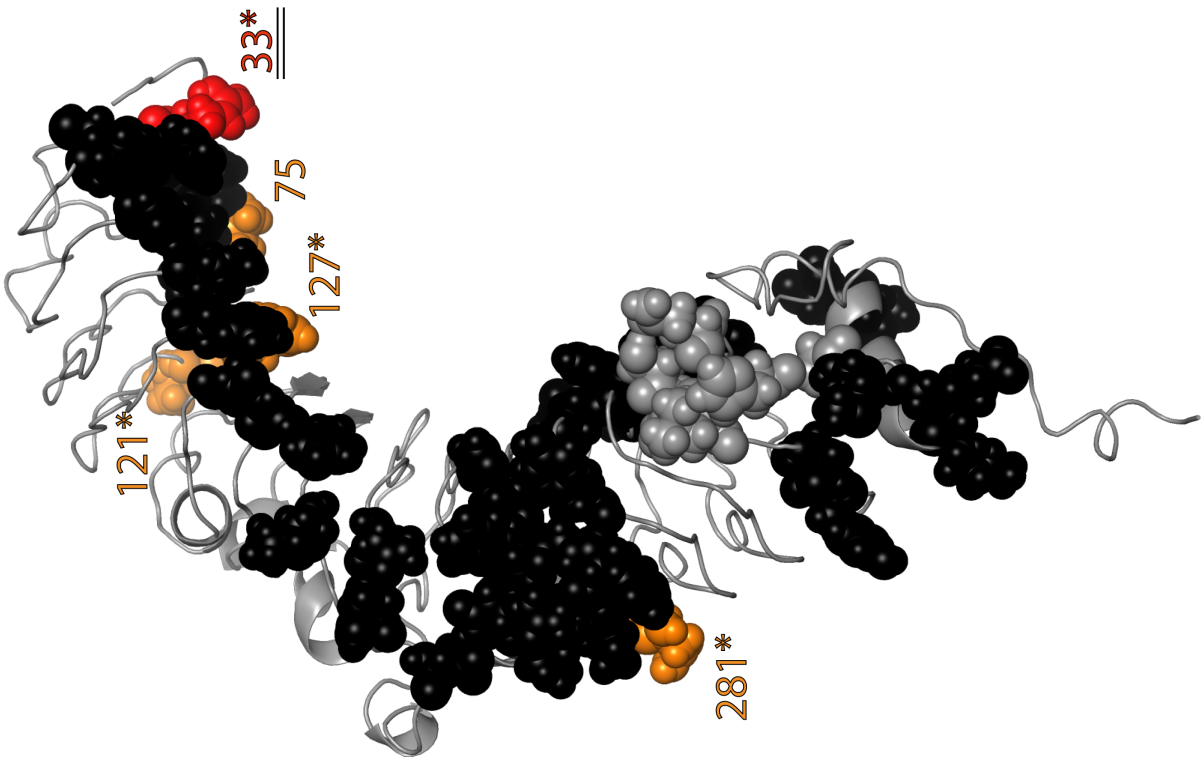


Figure B.4: TLR5 homology model showing positively selected sites.

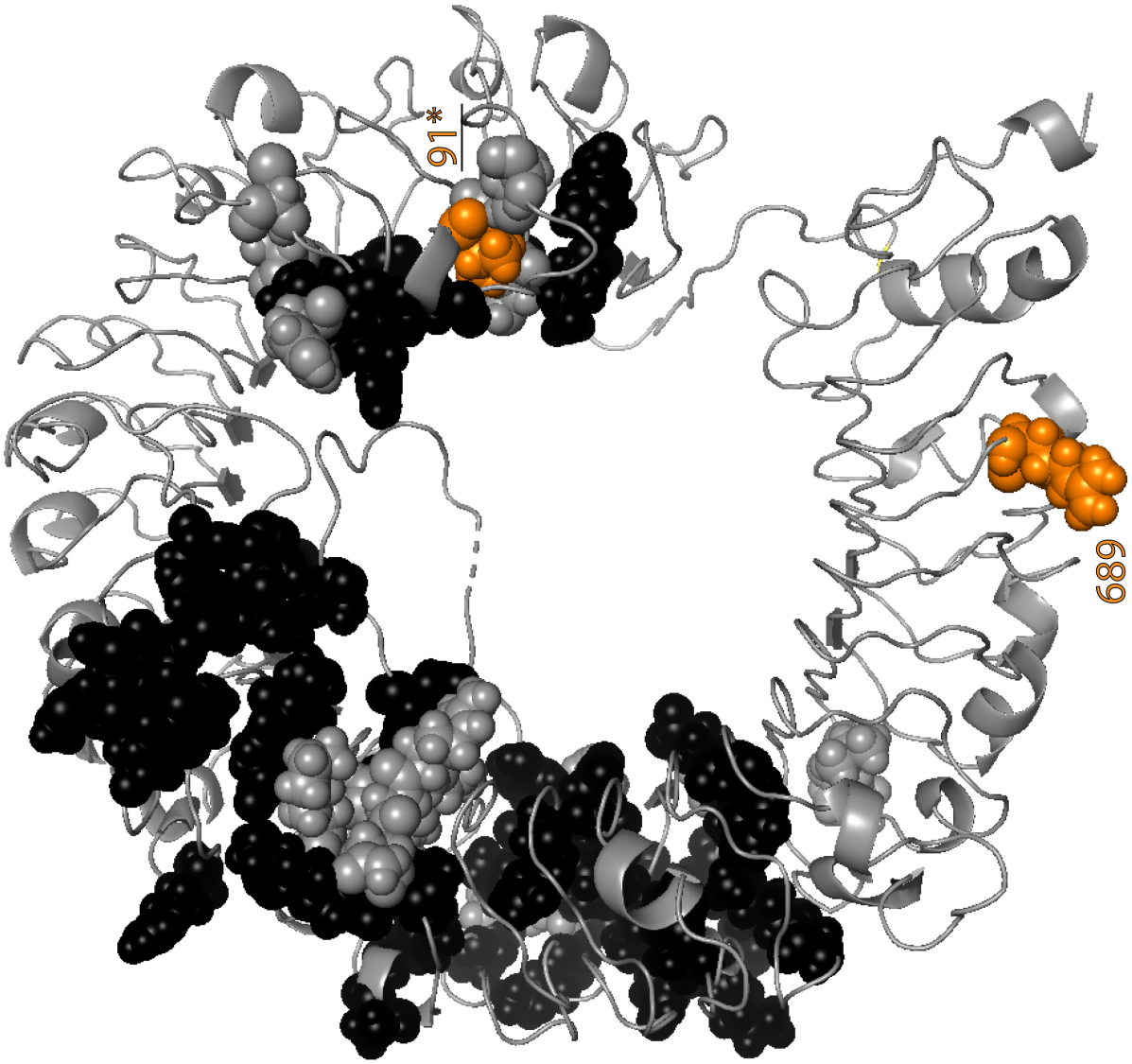


Figure B.5: TLR7 homology model showing positively selected sites.

B.2 Tables

Table B.1: Results from site models comparisons (M2/M2a and M7/M8) in the codeml program in PAML

TLR	Model	lnL	2(ΔlnL)	P-value
1A	M1a	-1821.206256	1e-06	1
1A	M2a	-1821.206255		
1A	M7	-1821.207512	0.00024	1
1A	M8	-1821.207272		
1B	M1a	-1335.556606	13.99	0.000918
1B	M2a	-1328.562989		
1B	M7	-1335.556601	13.99	0.000918
1B	M8	-1328.563326		
2A	M1a	-1063.044738	1.58	0.45
2A	M2a	-1062.25387		
2A	M7	-1063.061381	1.62	0.45
2A	M8	-1062.25387		
2B	M1a	-936.166981	3.97	0.14
2B	M2a	-934.166981		
2B	M7	-936.386588	4.41	0.11
2B	M8	-934.182354		
3	M1a	-4726.812656	4.043078	0.133
3	M2a	-4724.791117		
3	M7	-4726.81527	4.047772	0.132
3	M8	-4724.791384		
4	M1a	-4532.656932	14.015116	0.00091
4	M2a	-4525.649374		
4	M7	-4532.794629	14.29414	0.00079
4	M8	-4525.647559		
5	M1a	-4730.087159	24.504362	4.78e-06
5	M2a	-4717.834978		
5	M7	-4731.145327	26.619834	1.66e-06
5	M8	-4717.83541		
7	M1a	-5183.621165	7.169838	0.0277
7	M2a	-5180.036246		
7	M7	-5183.62674	7.16707	0.0278
7	M8	-5180.043205		
15	M1a	-4816.320022	7.523166	0.0233
15	M2a	-4812.558439		
15	M7	-4816.347303	7.545864	0.023
15	M8	-4812.574371		
21	M1a	-4313.32987	0.870012	0.647
21	M2a	-4312.894864		
21	M7	-4313.570862	1.346234	0.51
21	M8	-4312.897745		

Continued on next page

Likelihood ratio tests were performed by calculating double the difference in log likelihood values between the alternative model (M2a or M8) and the null model (M2 or M7). P-values were obtained from the χ^2 distribution with 2 degrees of freedom (calculated in each case as the difference between the numbers of parameters included in each model). Comparisons with P values < 0.05 were considered to be statistically significant, and are highlighted in green. Note that only the non-gene-converted regions of the TLR1/2 family were included in the analysis.

Table B.2: Positively selected sites in penguin TLRs.

TLR	Site	AA	Pos. prob	Omega	SE	Domain	Chick. Site	AA	PSS in other stud.	Function	Function (ref.)
TLR1B	39	P	0.958	6.353	2.271	ECD	43	L			
TLR1B	115	N	0.923	6.117	2.452	ECD	119	K	[603]	<5Å ABS	[273]
TLR1B	118	T	0.927	6.15	2.429	ECD	122	I	[603]	<5Å ABS	[273]
TLR1B	142	D	0.946	6.276	2.335	ECD	146	N	[264]	<5Å ABS	[273]
TLR1B	144	K	0.999	6.592	2.004	ECD	148	V	[603, 637]	<5Å ABS	[273]
TLR1B	178	I	0.926	6.141	2.435	ECD	182	I		<5Å ABS	[273]
TLR1B	185	A	0.94	6.237	2.366	ECD	189	A			
TLR1B	201	L	0.954	6.326	2.295	ECD	205	S	[25, 637]	<5Å DS	
TLR1B	232	T	0.935	6.2	2.394	ECD	237	F			
TLR2B	232	K	0.96	5.333	2.808	ECD	230	K	[603]		
TLR4	15	L	0.916	4.631	1.482	ECD	9	P	[25, 603]	Not modelled	
TLR4	211	A	0.978	4.872	1.142	ECD	205	T	[637]		
TLR4	308	E	0.979	4.874	1.138	ECD	302	N	[25, 185, 603, 637]		
TLR4	331	E	0.922	4.656	1.453	ECD	325	E	[25, 423, 535, 637, 608]	<5Å ABS/DS	[446, 458]
TLR4	465	V	0.901	4.57	1.546	ECD	459	I	[637]		
TLR4	471	H	0.907	4.595	1.52	ECD	465	Y			
TLR4	532	D	0.91	4.604	1.511	ECD	526	N	[608, 637]		
TLR4	661	V	0.905	4.585	1.531	TMD	655	G	[603, 637]	Not modelled	
TLR4	800	R	0.903	4.576	1.54	TIR	794	R		Not modelled	
TLR5	20	C	0.902	4.72	1.541	ECD	20	C	[603]	Not modelled	
TLR5	33	L	0.985	5.068	1.033	ECD	33	M	[603]	is ABS	[659]
TLR5	75	M	0.986	5.074	1.023	ECD	75	L	[603]		
TLR5	121	Q	0.988	5.082	1.007	ECD	121	Q	[603]		
TLR5	127	R	0.981	5.055	1.059	ECD	127	R			
TLR5	281	D	0.977	5.037	1.092	ECD	281	T	[220, 603, 606]		
TLR5	429	H	0.918	4.789	1.465	ECD	429	H		Not modelled	
TLR5	521	R	0.987	5.079	1.012	ECD	521	Q		Not modelled	
TLR5	660	V	0.917	4.784	1.472	TMD	660	I		Not modelled	

Cont. next page

TLR	Site	AA	Pos. prob	Omega	SE	Domain	Chick. Site	AA	PSS in other stud.	Function	Function (ref.)
TLR5	827	V	0.98	5.048	1.071	TIR	827	I		Not modelled	
TLR5	845	Q	0.93	4.84	1.403	TIR	845	Q	[603]	Not modelled	
TLR7	91	V	0.905	2.132	0.384	ECD	101	V		<5Å DS	
TLR7	689	C	0.901	2.127	0.391	ECD	699	L	[25, 637]	Not modelled	
TLR15	619	V	0.936	2.79	1.169	ECD	620	I		Not modelled	

Abbreviations: Pos. prob, posterior probability; AA, amino acid; PSS, positively selected site; ABS, agonist-binding site; DS, dimerisation site; TIR, Toll/Interleukin-1 Receptor domain; TMD, transmembrane domain; ECD, extracellular domain.

Table B.3: Analysis of positively selected sites in penguin TLRs using MEME and FUBAR

TLR	PAML selected site	MEME (P-value)	FUBAR (post. prob.)
TLR1B	39	0.67	0.49
TLR1B	115	0.18	0.955
TLR1B	118	0.17	0.959
TLR1B	142	0.32	0.87
TLR1B	144	0.08	0.984
TLR1B	178	0.27	0.881
TLR1B	185	0.21	0.886
TLR1B	201	0.02	0.947
TLR1B	232	0.2	0.899
TLR2B	232	0.09	0.984
TLR4	15	0.19	0.928
TLR4	211	0.17	0.964
TLR4	308	0.17	0.967
TLR4	331	0.21	0.94
TLR4	465	0.15	0.935
TLR4	471	0.23	0.927
TLR4	532	0.22	0.925
TLR4	661	0.27	0.902
TLR4	800	0.28	0.907
TLR5	20	0.38	0.878
TLR5	33	0.07	0.986
TLR5	75	0.43	0.818
TLR5	121	0.05	0.99
TLR5	127	0.28	0.947
TLR5	281	0.32	0.924
TLR5	429	0.18	0.926
TLR5	521	0.1	0.979
TLR5	660	0.18	0.931
TLR5	827	0.19	0.964
TLR5	845	0.19	0.931
TLR7	91	0.08	0.979
TLR7	689	0.28	0.953
TLR15	619	0.03	0.986

Alignments used for PAML analysis were also used for MEME and FUBAR analysis and results of MEME and FUBAR analyses are given here with reference to sites that were found to be positively selected in PAML analysis. The significance threshold for MEME was taken to be the default $P \leqslant 0.1$ and the significance threshold for FUBAR was taken to be the default posterior probability > 0.9.

Table B.4: Sampling locations for *Eudyptes* spp. *TLR15* sequences

Sampling Location	Latitude	Longitude	*E. chrysolophus*	*E. schlegeli*	*E. moseleyi*	*E. filholi*	*E. chrysocome*	*E. sclateri*	*E. robustus*	*E. pachyrhynchus*
Antipodes Is.	-49.687838	178.770761						1		
Amsterdam Is.	-37.833381	77.551037			5, 1					
Barnevelt Is.	-55.823192	-66.797728					6, 1			
Bird Island,	-54.006522	-38.04205	5							
Bouvet Island	-54.422308	3.357692	7							
Crozet Islands	-46.310592	50.890745	7			7, 1				
Elephant Island,	-61.112654	-55.140438	7							
Falkland Is.	-51.773942	-59.49515					8			
Harrison Cove	-44.628731	167.908082								1
Kerguelen Is.	-49.338178	69.334367	7			7				
Macquarie Is.	-54.626566	158.859226		6, 1		7				
Marion Is.	-46.911962	37.743137	7, 1			7				
Nightingale Is.	-37.422625	-12.48137			7					
The Snares	-48.025817	166.595973							1	
Terhalten Is.	-55.446769	-67.057671					7			
TOTAL by species			41	7	13	29	22	1	1	1

Locations of samples used in this study. Reference genomes from Pan, et al. (2019) [457] were used for the positive selection analyses, while genomic data from Vianna, et al. (2020) [604] and unpublished data were used for the *TLR15* population-level analysis. Since these sources of data used different seqeuencing technologies and assembly pipelines, only data from Vianna et al. were used for the *TLR15* analysis, rather than a combined dataset. All data used in the analysis are available and details of accession numbers can be found in the 'Data Availability' section of Chapter 3.
Underlined values are from [457]. Non-underlined values are from [604].

B.3 Coverage depth of *TLR15* relative to all known coding sequences

B.3.1 Background

One explanation for the pseudogenization of *TLR15* is that the gene underwent duplication, and one copy experienced functional redundancy and degradation in the genome. Even though this scenario does not fit well with *TLR15* (there were several homozygous pseudogene haplotypes and we did not observe any triallelic sites at the locus), another method of detecting duplication is through analysis of read coverage of the locus.

B.3.2 Methods

BAM files from analysis conducted in Pan, et al. (2019) [457] were obtained for three penguin species (one non-*Eudyptes* spp. and two *Eudyptes* spp.: *Spheniscus humboldti*, *Eudyptes robustus*, *Eudyptes chrysocome*. Using coordinates derived from known coding sequences from Pan, et al. (2019), average depth of coverage was calculated for each region using `samtools depth` and `awk`. Using BLAST, the genomic position of the *TLR15* locus was determined for each species and average coverage was determined for this region. Finally, average depth of coverage was plotted in a pairwise fashion between pairs of the three genomes.

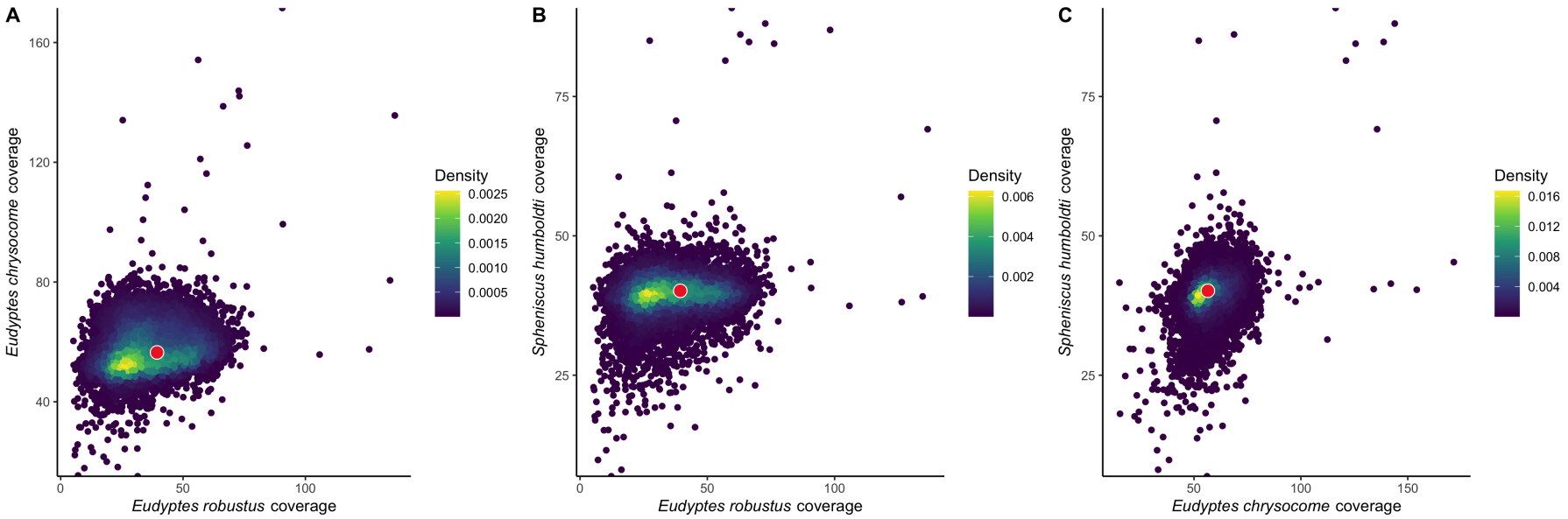


Figure B.6: Pairwise depth of coverage of *TLR15* and all other known coding sequences in the genomes of three penguin species. Depth of coverage for all coding loci are plotted (small points) and coloured according to their density. Depth of coverage for *TLR15* is overlaid (red point).

B.3.3 Results

In each pairwise comparison, depth of coverage for *TLR15* lies well within the high-density cluster of coding loci in the genome.

B.3.4 Interpretation

Since depth of coverage for *TLR15* is comparable to the vast majority of coding sequences in the genome, it is highly likely that *TLR15* has not undergone duplication in the *Eudyptes* lineage. This, along with the presence of homozygous mutations lack of triallelic sites, implies that *TLR15* is a unitary pseudogene in the *Eudyptes* penguins.

B.4 Identification of putative loss-of-function mutations in *Eudyptes* TLR15

B.4.1 Alignment of intact *Eudyptes* TLR15 haplotypes with other birds

TLR15 protein sequences were downloaded from Ensembl or NCBI (chicken, ENSGALP00000013260; northern fulmar, XP_009585200.1, misannotated as TLR2; emu, ENSDNVP00000009442; blue tit, ENSCCEP00000009931; collared flycatcher, ENSFALP00000015961; helmeted guineafowl, ENSNMEP00000012998). These were aligned to TLR15 sequences from other penguins (N = 11) and intact *Eudyptes* haplotypes (N = 45, including the consensus TLR15 used in functional analysis). Next, residues were identified which are well conserved among penguins and other birds, but that are distinct in *Eudyptes* TLR15. These sites are presented in Appendix table B.5.

Table B.5: Amino acid and nucleotide positions of positions which are conserved among penguins and other birds, but are distinct in the *Eudyptes*.

Amino acid site[1]	Amino acid site[2]	Ancestral AA[3]	Derived AA[4]	Chicken coordinates
56	55	T	M	3:3024305
68	67	E	G	3:3024269
161	160	L	P	3:3023990
246	239	I	V	3:3023754
569	560	R/P/N	H	3:3022790
674	665	I/G/V	V	3:3022476
679	670	V/A	A	3:3022448
683	674	L	S	3:3022289
736	727	K/E	G	3:3022173
787	778	Q	R	3:3022136

[1] Relative to alignment.
[2] Relative to chicken (the extra intron-spanning methionine at the start of chicken TLR15 was removed).
[3] Consensus.
[4] Amino acid found in *Eudyptes* spp.
Red text indicates that the position is highly conserved across the whole of vertebrates (see Figure B.7).

B.4.2 Variant conservation across vertebrate genomes

Next, the variant positions were plotted against the conservation scores of TLR15 across 77 vertebrate genomes (data obtained from UCSC; Figure B.7). Amino acid sites 161, 736 and 787 are highly conserved across vertebrates, but are different in *Eudyptes*, which could be indicative of a change in function. The mechanism of TLR15 activation has not been elucidated, so it is unclear which of these polymorphisms is the loss-of-function mutation, or whether multiple mutations resulted in the phenotype.

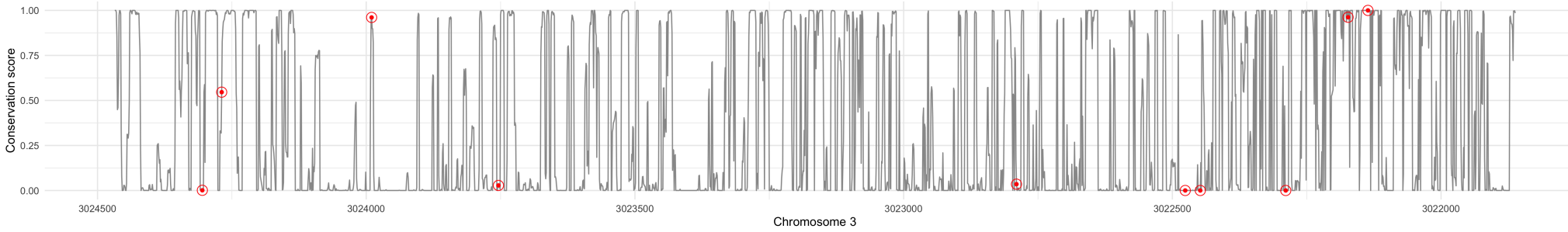


Figure B.7: Per-nucleotide conservation scores for *TLR15* in chicken compared to 77 other vertebrate genomes (data obtained from UCSC). Sites that were identified as being distinct in *Eudyptes* compared to the rest of penguins (see Table B.5) are highlighted in red.

B.4.3 Change-of-function analysis

Next, variants found in the *Eudyptes* were used to replace equivalent residues in the chicken TLR15 sequence, and the functional consequences predicted using Provean and SIFT. Provean analysis predicted L161P (chicken residue 160) and L683S (chicken residue 674) to be deleterious changes (Table B.6), while SIFT also predicted L683S to be a deleterious change (Table B.7). SIFT also predicted deleterious effects for T56M (chicken residue 55) and E68G (chicken residue 67), but since so few sequences were included as comparators for these sites, the prediction is of low confidence.

Table B.6: Provean predictions for changes in TLR15 protein function

Variant	PROVEAN score	Prediction (cutoff = -2.5)
T55M	-1.630	Neutral
E67G	-1.484	Neutral
L160P	-2.555	Deleterious
I239V	-0.236	Neutral
R560H	-0.010	Neutral
I665V	-0.300	Neutral
V670A	-0.668	Neutral
L674S	-4.235	Deleterious
K727G	-1.862	Neutral
E727G	-1.749	Neutral
Q778R	0.337	Neutral

The chicken TLR15 sequence was modified with the indicated changes.

Table B.7: SIFT predictions for changes in TLR15 protein function

Variant	Median conservation	Sequences (N)	Score	Outcome
T55M	4.32	1	0.00	AFFECT PROTEIN FUNCTION[1]
E67G	3.44	2	0.01	AFFECT PROTEIN FUNCTION[1]
L160P	3.44	2	0.06	TOLERATED
I239V	3.34	4	0.57	TOLERATED
R560H	3.14	46	0.13	TOLERATED
I665V	3.03	48	1.00	TOLERATED
V670A	3.03	48	0.22	TOLERATED
L674S	3.09	47	0.00	AFFECT PROTEIN FUNCTION
K727G	3.03	48	0.07	TOLERATED

[1] Low confidence.

As above, the chicken TLR15 sequence was modified with the indicated changes.

B.4.4 Conclusions

In summary, analysis of nucleotide conservation scores indicates that three non-synonymous variants in the *Eudyptes* (amino acid sites 161, 736 and 787) are otherwise highly conserved across vertebrates. This could be evidence of putative change in function at these sites. Moreover, Provean analysis predicts that L161P is a deleterious change, which is consistent with a loss-of-function mutation. Continued work to elucidate the mechanism of action of TLR15 could provide insight into which mutations are the cause of the loss-of-function of *Eudyptes* TLR15.

Appendix C

Appendix to Chapter 4

C.1 Homology modelling

The protein sequences used in the homology modelling of penguin IFIT5 and CD81 are as follows:

>IFIT5
STISKNSLKTSLLELECHFTWTLLKQDVDLDELEETIGDQIEFLIKSNITNYNLLSYVCHLKNSNEEALR
NLQKAEEEIKKNYPDEIARRSLVTWGNYAWIYYHMERYKEAQTYVNKVENSCKKLSSTAQWKIQLPEIYA
EQGWALLKFGGKYYMRAKNCFENALKNEPDNPEFNAGYAIAMYRLEDFSHRQCEDASLSIKPLKRAVQLN
PKDTFIVALLALKLQDFKQVEEGERYIEAAMQKTPDLPYVLRYAAKFYRRKGEVDKALKILKKAIAVTPK
SAFLHHQLGLCYRAKLFQLKKTTRYPPQEQVEEFIRLSIFHFKMATDQKTKFFSAHIDLANMYAQGKRYE
KAEETFQKAFQINILSYNDKQEYYYHYGNFQHFHMKSESEAIRYYLEGLKIEKDSYIRSKCRDALNQLLE
ERIQGGLGDATDFGTLGLIHNLNGEKHKATECYEKAVALCPDNEEYLSALCELRLSI

>CD81
MGVEGCTKCIKYLLFVFNFIFWLAGGIILGVALWLRHDSQTTNILYLQLGDKQAPNTFYVGIYILIAVGA
VMMFVGFLGCYGAIQESQCLLGTFFTCLVILFACEVAAGIWGFVNKDQIAKDVKQFYDQAFQQALMAESD
MNNGKAVVKTFHETLDCCGPDNAVGVITPLWRDDLCPKKEFSLKNVIESSNCHKKIDELFSGKLYLIGIA
AIVVAVIMIFEMILSMVLCCGIRNSSIY

Table C.1: PAML results for all positively selected immune loci

Gene ID	Gene symbol	M1a[1]	M2a[1]	2ΔLnl[2]	P-value	q-value
GALGAL_R00410	CD40L	-1502.786342	-1494.286627	16.99943	0.000203526	0.003630937
GALGAL_R00664	GP174	-1970.850218	-1943.966175	53.768086	2.11e-12	2.33e-10
GALGAL_R00799	WDR61	-1484.80559	-1475.951123	17.708934	0.000142743	0.002689871
GALGAL_R01382	NICA	-3709.969551	-3699.968921	20.00126	4.54e-05	0.001025981
GALGAL_R01527	TLR5	-4745.556405	-4731.16733	28.77815	5.64e-07	2.14e-05
GALGAL_R02292	IFIT5	-3164.010109	-3135.67448	56.671258	4.94e-13	5.98e-11
GALGAL_R02403	NSMA2	-3196.87955	-3186.333319	21.092462	2.63e-05	0.000654005
GALGAL_R03236	ROR2	-4350.678308	-4337.538766	26.279084	1.97e-06	6.56e-05
GALGAL_R04007	CKLF7	-903.66428	-893.25357	20.82142	3.01e-05	0.000730147
GALGAL_R04077	IFI6	-659.985453	-651.163616	17.643674	0.000147477	0.002767101
GALGAL_R04573	IRF1	-1851.195616	-1837.041719	28.307794	7.13e-07	2.64e-05
GALGAL_R05965	CO8B	-3061.995161	-3051.636153	20.718016	3.17e-05	0.000758328
GALGAL_R06181	GPR55	-1579.135249	-1564.711117	28.848264	5.44e-07	2.09e-05
GALGAL_R06228	GP171	-1631.962978	-1622.583938	18.75808	8.45e-05	0.001742774
GALGAL_R06540	F210B	-632.795922	-622.285612	21.02062	2.73e-05	0.000672259
GALGAL_R06619	FUT7	-1932.901798	-1921.502804	22.797988	1.12e-05	0.000308695
GALGAL_R06658	TLR4	-4329.253956	-4320.124161	18.25959	0.000108388	0.0021349
GALGAL_R06833	TCAM1	-4073.736063	-4057.592027	32.288072	9.74e-08	4.46e-06
GALGAL_R07240	DPP4	-4312.432279	-4284.007654	56.84925	4.52e-13	5.62e-11
GALGAL_R07280	CASP8	-2452.03196	-2442.585614	18.892692	7.9e-05	0.001641
GALGAL_R07311	RAB17	-1266.687551	-1256.20107	20.972962	2.79e-05	0.00068257
GALGAL_R07477	CATE	-2100.57277	-2091.354176	18.437188	9.92e-05	0.001975844
GALGAL_R07482	PIGR	-4328.987723	-4302.601889	52.771668	3.47e-12	3.69e-10
GALGAL_R07586	CO5	-8879.426246	-8856.338779	46.174934	9.4e-11	7.58e-09
GALGAL_R08344	I18RA	-2935.385945	-2926.124355	18.52318	9.5e-05	0.001914597
GALGAL_R08947	LRMP	-1833.146882	-1824.025108	18.243548	0.000109261	0.002145715
GALGAL_R08979	UBP18	-1810.131669	-1801.033323	18.196692	0.000111851	0.002183352
GALGAL_R09040	CD4	-2296.864295	-2287.641912	18.444766	9.88e-05	0.001972887

Continued on next page

Gene ID	Gene symbol	M1a	M2a	2ΔLnl	P-value	q-value
GALGAL_R09041	LAG3	-879.054856	-869.784844	18.540024	9.42e-05	0.001902946
GALGAL_R09104	TACT	-923.551914	-904.066294	38.97124	3.45e-09	2.26e-07
GALGAL_R09446	FXDC2	-1742.388498	-1731.226462	22.324072	1.42e-05	0.000380474
GALGAL_R10178	FA7	-2181.283392	-2170.359839	21.847106	1.8e-05	0.000472762
GALGAL_R10507	IFNG	-887.561595	-878.970655	17.18188	0.000185781	0.003355621
GALGAL_R10683	CO4	-9094.117046	-9082.299573	23.634946	7.37e-06	0.000212593
GALGAL_R11083	ZP4	-3282.509785	-3273.901327	17.216916	0.000182555	0.003304208
GALGAL_R11385	CML1	-2098.876453	-2065.282436	67.188034	2.57e-15	3.86e-13
GALGAL_R11530	CD3D	-1261.917669	-1208.217598	107.400142	4.77e-24	1.98e-21
GALGAL_R12274	SPB10	-2527.503023	-2483.485367	88.035312	7.64e-20	2.02e-17
GALGAL_R12478	CAD17	-4301.093349	-4282.689975	36.806748	1.02e-08	5.91e-07
GALGAL_R12525	LY6E	-537.278176	-529.444407	15.667538	0.00039613	0.006387834
GALGAL_R13865	CATS	-1975.546322	-1963.708979	23.674686	7.23e-06	0.0002098
GALGAL_R14254	CD81	-1205.179931	-1193.333779	23.692304	7.17e-06	0.000208655
GALGAL_R14644	TRFE	-3712.844476	-3704.864709	15.959534	0.000342319	0.005655084
GALGAL_R14677	SUCR1	-1834.235629	-1821.011502	26.448254	1.81e-06	6.07e-05
GALGAL_R14919	RN5A	-3978.013733	-3952.077711	51.872044	5.45e-12	5.45e-10
GALGAL_R15280	C4BPA	-2518.088126	-2501.161724	33.852804	4.46e-08	2.17e-06
GALGAL_R15837	AGRF5	-8995.942092	-8837.330038	317.224108	1.31e-69	5.68e-66
GALGAL_R16286	PAR14	-9641.125402	-9625.564338	31.122128	1.75e-07	7.56e-06
GALGAL_R16287	DTX3L	-4726.459411	-4678.481964	95.954894	1.46e-21	4.38e-19
GALGAL_R16478	NMI	-523.175919	-514.498944	17.35395	0.000170466	0.003111273
GALGAL_R17471	PIGR	-2825.209848	-2816.227257	17.965182	0.000125577	0.002418746

[1] Log-likelihood values
[2] Likelihood ratio test

Appendix D

Appendix to Chapter 5

Table D.1: Metadata of chicken genomes used in population genetic analysis of Toll-like receptors

Accession	ID	Species	Breed/Location	Population
ypt2808	Red_bankiva1	*G. g. bankiva*	Java	bankiva
ypt2812	Red_bankiva2	*G. g. bankiva*	Java	bankiva
ypt2819	Red_bankiva3	*G. g. bankiva*	Java	bankiva
SAMN08979067	EthiopiaH1	*G. g. domesticus*	Horro	Africa
SAMN08979068	EthiopiaH2	*G. g. domesticus*	Horro	Africa
SAMN08979069	EthiopiaH3	*G. g. domesticus*	Horro	Africa
SAMN08979070	EthiopiaH4	*G. g. domesticus*	Horro	Africa
SAMN08979071	EthiopiaH5	*G. g. domesticus*	Horro	Africa
SAMN08979072	EthiopiaH6	*G. g. domesticus*	Horro	Africa
SAMN08979073	EthiopiaJ1	*G. g. domesticus*	Jarso	Africa
SAMN08979074	EthiopiaJ2	*G. g. domesticus*	Jarso	Africa
SAMN08979075	EthiopiaJ3	*G. g. domesticus*	Jarso	Africa
SAMN08979076	EthiopiaJ4	*G. g. domesticus*	Jarso	Africa
SAMN08979077	EthiopiaJ5	*G. g. domesticus*	Jarso	Africa
SAMN12623089	ch1957	*G. g. domesticus*	Alberta Meat	Broiler
SAMN12642285	ch1978	*G. g. domesticus*	Alberta Meat	Broiler
SAMN02782188	Broiler01	*G. g. domesticus*	Broiler	Broiler
SAMN02782189	Broiler02	*G. g. domesticus*	Broiler	Broiler
SAMN02782190	Broiler03	*G. g. domesticus*	Broiler	Broiler
SAMN02782191	Broiler04	*G. g. domesticus*	Broiler	Broiler
SAMN02782193	Broiler05	*G. g. domesticus*	Broiler	Broiler
SAMN02782192	Broiler06	*G. g. domesticus*	Broiler	Broiler
SAMN02782194	Broiler07	*G. g. domesticus*	Broiler	Broiler
SAMN02782195	Broiler08	*G. g. domesticus*	Broiler	Broiler
SAMN03009569	Broiler09	*G. g. domesticus*	Broiler	Broiler
SAMN03009570	Broiler10	*G. g. domesticus*	Broiler	Broiler
SAMN03009571	Broiler11	*G. g. domesticus*	Broiler	Broiler
SAMN03009572	Broiler12	*G. g. domesticus*	Broiler	Broiler
SAMN03009573	Broiler13	*G. g. domesticus*	Broiler	Broiler
SAMN03009575	Broiler14	*G. g. domesticus*	Broiler	Broiler
SAMN03009574	Broiler15	*G. g. domesticus*	Broiler	Broiler
SAMN03009576	Broiler16	*G. g. domesticus*	Broiler	Broiler
SAMEA5160235	BRA1	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160244	BRA10	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160245	BRA11	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160246	BRA12	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160247	BRA13	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160248	BRA14	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160249	BRA15	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160250	BRA16	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160251	BRA17	*G. g. domesticus*	BroilerA	Broiler

Accession	ID	Species	Breed/Location	Population
SAMEA5160252	BRA18	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160253	BRA19	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160236	BRA2	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160254	BRA20	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160237	BRA3	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160238	BRA4	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160239	BRA5	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160240	BRA6	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160241	BRA7	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160242	BRA8	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160243	BRA9	*G. g. domesticus*	BroilerA	Broiler
SAMEA5160255	BRB1	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160264	BRB10	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160265	BRB11	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160266	BRB12	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160267	BRB13	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160268	BRB14	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160269	BRB15	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160270	BRB16	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160271	BRB17	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160272	BRB18	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160273	BRB19	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160256	BRB2	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160274	BRB20	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160257	BRB3	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160258	BRB4	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160259	BRB5	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160260	BRB6	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160261	BRB7	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160262	BRB8	*G. g. domesticus*	BroilerB	Broiler
SAMEA5160263	BRB9	*G. g. domesticus*	BroilerB	Broiler
ypt902_L1_I053	Europe_meat1	*G. g. domesticus*	Europe_meat	Broiler
ypt903_L2_I054	Europe_meat2	*G. g. domesticus*	Europe_meat	Broiler
ypt904_L4_I055	Europe_meat3	*G. g. domesticus*	Europe_meat	Broiler
ypt905_L6_I056	Europe_meat4	*G. g. domesticus*	Europe_meat	Broiler
SAMN07327525	REL01	*G. g. domesticus*	REL	Broiler
SAMN07312781	REL02	*G. g. domesticus*	REL	Broiler
SAMN07327526	REL03	*G. g. domesticus*	REL	Broiler
SAMN07327527	REL04	*G. g. domesticus*	REL	Broiler
SAMN07327528	REL05	*G. g. domesticus*	REL	Broiler
SAMN07327529	REL06	*G. g. domesticus*	REL	Broiler
SAMN07327530	REL07	*G. g. domesticus*	REL	Broiler
SAMN07327531	REL08	*G. g. domesticus*	REL	Broiler
SAMN12618230	Ross308	*G. g. domesticus*	Ross 308	Broiler
ypt2163	Indonesia1	*G. g. domesticus*	Bali	EastAsia
ypt2164	Indonesia2	*G. g. domesticus*	Bali	EastAsia
ypt2214	Indonesia28	*G. g. domesticus*	Bali	EastAsia
ypt2215	Indonesia29	*G. g. domesticus*	Bali	EastAsia
ypt2165	Indonesia3	*G. g. domesticus*	Bali	EastAsia
ypt2216	Indonesia30	*G. g. domesticus*	Bali	EastAsia
ypt2217	Indonesia31	*G. g. domesticus*	Bali	EastAsia
ypt2218	Indonesia32	*G. g. domesticus*	Bali	EastAsia
ypt2219	Indonesia33	*G. g. domesticus*	Bali	EastAsia
ypt2220	Indonesia34	*G. g. domesticus*	Bali	EastAsia
ypt2221	Indonesia35	*G. g. domesticus*	Bali	EastAsia
ypt2222	Indonesia36	*G. g. domesticus*	Bali	EastAsia
ypt2166	Indonesia4	*G. g. domesticus*	Bali	EastAsia
ypt2167	Indonesia5	*G. g. domesticus*	Bali	EastAsia
ypt2168	Indonesia6	*G. g. domesticus*	Bali	EastAsia
Y1-f1_L5_I119	Beijing1	*G. g. domesticus*	Beijing	EastAsia
You3_L7_I1000005	Beijing10	*G. g. domesticus*	Beijing	EastAsia
Y1-f2_L6_I121	Beijing2	*G. g. domesticus*	Beijing	EastAsia
Y1-f3_L7_I101	Beijing3	*G. g. domesticus*	Beijing	EastAsia
Y1-m1_L6_I120	Beijing4	*G. g. domesticus*	Beijing	EastAsia
Y1-m2_L4_I122	Beijing5	*G. g. domesticus*	Beijing	EastAsia
Y1-m3_L7_I123	Beijing6	*G. g. domesticus*	Beijing	EastAsia
Y1-m4_I124	Beijing7	*G. g. domesticus*	Beijing	EastAsia
Y1-m5_L7_I102	Beijing8	*G. g. domesticus*	Beijing	EastAsia
You1	Beijing9	*G. g. domesticus*	Beijing	EastAsia
SAMN02486156	Beijing12	*G. g. domesticus*	Beijing	EastAsia
SAMN02486156	BeijingYou01	*G. g. domesticus*	Beijing You	EastAsia

Accession	ID	Species	Breed/Location	Population
SAMN15394828	XichuanBB1	*G. g. domesticus*	Black Bone	EastAsia
SAMN15394829	XichuanBB2	*G. g. domesticus*	Black Bone	EastAsia
SAMN15394830	XichuanBB3	*G. g. domesticus*	Black Bone	EastAsia
SAMN15394831	XichuanBB4	*G. g. domesticus*	Black Bone	EastAsia
SAMN15394832	XichuanBB5	*G. g. domesticus*	Black Bone	EastAsia
SAMN12236815	BlackCochin01	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236816	BlackCochin02	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236813	BlackCochin03	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236814	BlackCochin04	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236819	BlackCochin05	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236820	BlackCochin06	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236817	BlackCochin07	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236818	BlackCochin08	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236821	BlackCochin09	*G. g. domesticus*	Black Cochin	EastAsia
SAMN12236812	BlackCochin10	*G. g. domesticus*	Black Cochin	EastAsia
SAMN14814813	BlackLayer01	*G. g. domesticus*	Black Layer	EastAsia
SAMN14814814	BlackLayer02	*G. g. domesticus*	Black Layer	EastAsia
ypt2189	Indonesia21	*G. g. domesticus*	Borneo	EastAsia
ypt2190	Indonesia22	*G. g. domesticus*	Borneo	EastAsia
ypt2191	Indonesia23	*G. g. domesticus*	Borneo	EastAsia
SAMN02486158	Dongxiang1	*G. g. domesticus*	Dongxiang	EastAsia
production	Europe_egg	*G. g. domesticus*	Europe_egg	EastAsia
Ypt642	Guangzhou1	*G. g. domesticus*	Guangdong	EastAsia
Ypt643	Guangzhou2	*G. g. domesticus*	Guangdong	EastAsia
Ypt645	Guangzhou3	*G. g. domesticus*	Guangdong	EastAsia
Ypt646	Guangzhou4	*G. g. domesticus*	Guangdong	EastAsia
Ypt647	Guangzhou5	*G. g. domesticus*	Guangdong	EastAsia
YPt648	Guangzhou6	*G. g. domesticus*	Guangdong	EastAsia
ypt2831	Guangxi1	*G. g. domesticus*	Guangxi	EastAsia
19297	Guangxi100	*G. g. domesticus*	Guangxi	EastAsia
ypt2832	Guangxi2	*G. g. domesticus*	Guangxi	EastAsia
ypt2842	Guangxi3	*G. g. domesticus*	Guangxi	EastAsia
ypt2843	Guangxi4	*G. g. domesticus*	Guangxi	EastAsia
ypt2849	Guangxi5	*G. g. domesticus*	Guangxi	EastAsia
ypt2859	Guangxi6	*G. g. domesticus*	Guangxi	EastAsia
SAMN09700399	Guangxi01	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700400	Guangxi03	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700401	Guangxi04	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700402	Guangxi05	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700403	Guangxi23	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700404	Guangxi24	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700405	Guangxi25	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700406	Guangxi26	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700407	Guangxi28	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMN09700408	Guangxi32	*G. g. domesticus*	Guangxi Yellow	EastAsia
SAMEA6529977	Hamburgh2486	*G. g. domesticus*	Hamburgh	EastAsia
ypt3210_L6_I051	Henan1	*G. g. domesticus*	Henan	EastAsia
ypt3211_L6_I052	Henan2	*G. g. domesticus*	Henan	EastAsia
ypt3212_L4_I028	Henan3	*G. g. domesticus*	Henan	EastAsia
ypt3213_L8_I029	Henan4	*G. g. domesticus*	Henan	EastAsia
ypt3219_L1_I030	Henan5	*G. g. domesticus*	Henan	EastAsia
ypt3220_L1_I031	Henan6	*G. g. domesticus*	Henan	EastAsia
ypt3221_L1_I032	Henan7	*G. g. domesticus*	Henan	EastAsia
ypt3222_L1_I033	Henan8	*G. g. domesticus*	Henan	EastAsia
SAMN09700409	Hetian01	*G. g. domesticus*	Hetian	EastAsia
SAMN09700410	Hetian02	*G. g. domesticus*	Hetian	EastAsia
SAMN09700411	Hetian10	*G. g. domesticus*	Hetian	EastAsia
SAMN09700412	Hetian13	*G. g. domesticus*	Hetian	EastAsia
SAMN09700413	Hetian15	*G. g. domesticus*	Hetian	EastAsia
SAMN09700414	Hetian17	*G. g. domesticus*	Hetian	EastAsia
SAMN09700415	Hetian18	*G. g. domesticus*	Hetian	EastAsia
SAMN09700416	Hetian22	*G. g. domesticus*	Hetian	EastAsia
SAMN09700417	Hetian25	*G. g. domesticus*	Hetian	EastAsia
SAMN09700418	Hetian26	*G. g. domesticus*	Hetian	EastAsia
SAMN09700479	Huaibeima01	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700480	Huaibeima02	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700481	Huaibeima06	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700482	Huaibeima09	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700483	Huaibeima11	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700484	Huaibeima12	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700485	Huaibeima13	*G. g. domesticus*	Huaibei Partridge	EastAsia

Accession	ID	Species	Breed/Location	Population
SAMN09700486	Huaibeima16	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700487	Huaibeima27	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700488	Huaibeima29	*G. g. domesticus*	Huaibei Partridge	EastAsia
SAMN09700489	Huaixiang01	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700490	Huaixiang03	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700491	Huaixiang05	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700492	Huaixiang13	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700493	Huaixiang14	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700494	Huaixiang18	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700495	Huaixiang20	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700496	Huaixiang25	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700497	Huaixiang28	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700498	Huaixiang31	*G. g. domesticus*	Huaixiang	EastAsia
SAMN09700419	Huanglang02	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700420	Huanglang03	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700421	Huanglang04	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700422	Huanglang06	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700423	Huanglang10	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700424	Huanglang13	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700425	Huanglang16	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700426	Huanglang28	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700427	Huanglang30	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700428	Huanglang46	*G. g. domesticus*	Huanglang	EastAsia
SAMN09700429	Huiyang01	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700430	Huiyang02	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700431	Huiyang03	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700432	Huiyang05	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700433	Huiyang11	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700434	Huiyang12	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700435	Huiyang13	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700436	Huiyang27	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700437	Huiyang29	*G. g. domesticus*	Huiyang Bd	EastAsia
SAMN09700438	Huiyang30	*G. g. domesticus*	Huiyang Bd	EastAsia
ypt2184	Indonesia17	*G. g. domesticus*	Indonesia	EastAsia
ypt2185	Indonesia18	*G. g. domesticus*	Indonesia	EastAsia
ypt2186	Indonesia19	*G. g. domesticus*	Indonesia	EastAsia
ypt2187	Indonesia20	*G. g. domesticus*	Indonesia	EastAsia
ypt2205	Indonesia26	*G. g. domesticus*	Java	EastAsia
ypt2209	Indonesia27	*G. g. domesticus*	Java	EastAsia
SAMN12236864	Java01	*G. g. domesticus*	Java	EastAsia
SAMN12236863	Java02	*G. g. domesticus*	Java	EastAsia
SAMN12236861	Java03	*G. g. domesticus*	Java	EastAsia
SAMN12236862	Java04	*G. g. domesticus*	Java	EastAsia
SAMN12236859	Java05	*G. g. domesticus*	Java	EastAsia
SAMN12236860	Java06	*G. g. domesticus*	Java	EastAsia
SAMN12236857	Java07	*G. g. domesticus*	Java	EastAsia
SAMN12236858	Java08	*G. g. domesticus*	Java	EastAsia
SAMN12236855	Java09	*G. g. domesticus*	Java	EastAsia
SAMN12236856	Java10	*G. g. domesticus*	Java	EastAsia
SAMEA5930916	Java1064	*G. g. domesticus*	Java	EastAsia
SAMN09700439	Jianghan26	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700440	Jianghan27	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700441	Jianghan28	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700442	Jianghan30	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700443	Jianghan34	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700444	Jianghan35	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700445	Jianghan38	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700446	Jianghan43	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700447	Jianghan45	*G. g. domesticus*	Jianghan	EastAsia
SAMN09700448	Jianghan46	*G. g. domesticus*	Jianghan	EastAsia
ypt3224_L8_I043	Jiangsu1	*G. g. domesticus*	Jiangsu	EastAsia
ypt3243_L8_I047	Jiangsu10	*G. g. domesticus*	Jiangsu	EastAsia
ypt3225_L2_I053	Jiangsu2	*G. g. domesticus*	Jiangsu	EastAsia
ypt3226_L3_I044	Jiangsu3	*G. g. domesticus*	Jiangsu	EastAsia
ypt3228-7_L1_I034	Jiangsu4	*G. g. domesticus*	Jiangsu	EastAsia
ypt3229_L8_I045	Jiangsu5	*G. g. domesticus*	Jiangsu	EastAsia
ypt3238_L3_I046	Jiangsu6	*G. g. domesticus*	Jiangsu	EastAsia
ypt3239_L1_I035	Jiangsu7	*G. g. domesticus*	Jiangsu	EastAsia
ypt3240_L6_I054	Jiangsu8	*G. g. domesticus*	Jiangsu	EastAsia
ypt3241_L6_I055	Jiangsu9	*G. g. domesticus*	Jiangsu	EastAsia
S6	Jiangxi1	*G. g. domesticus*	Jiangxi	EastAsia

Accession	ID	Species	Breed/Location	Population
S-m5_I105	Jiangxi10	*G. g. domesticus*	Jiangxi	EastAsia
ypt3245	Jiangxi11	*G. g. domesticus*	Jiangxi	EastAsia
ypt3246_L6_I050	Jiangxi12	*G. g. domesticus*	Jiangxi	EastAsia
ypt3247_L6_I051	Jiangxi13	*G. g. domesticus*	Jiangxi	EastAsia
ypt3257_L1_I056	Jiangxi14	*G. g. domesticus*	Jiangxi	EastAsia
ypt3258_L3_I001	Jiangxi15	*G. g. domesticus*	Jiangxi	EastAsia
ypt3275_L1_I012	Jiangxi16	*G. g. domesticus*	Jiangxi	EastAsia
ypt3276_L7_I021	Jiangxi17	*G. g. domesticus*	Jiangxi	EastAsia
ypt3277_L6_I027	Jiangxi18	*G. g. domesticus*	Jiangxi	EastAsia
ypt3278_L3_I028	Jiangxi19	*G. g. domesticus*	Jiangxi	EastAsia
S-f1_L4_I106	Jiangxi2	*G. g. domesticus*	Jiangxi	EastAsia
Ypt622	Jiangxi20	*G. g. domesticus*	Jiangxi	EastAsia
Ypt623	Jiangxi21	*G. g. domesticus*	Jiangxi	EastAsia
Ypt624	Jiangxi22	*G. g. domesticus*	Jiangxi	EastAsia
Ypt626	Jiangxi23	*G. g. domesticus*	Jiangxi	EastAsia
YPt637	Jiangxi24	*G. g. domesticus*	Jiangxi	EastAsia
Ypt638	Jiangxi25	*G. g. domesticus*	Jiangxi	EastAsia
Ypt640	Jiangxi26	*G. g. domesticus*	Jiangxi	EastAsia
S-f2_I107	Jiangxi3	*G. g. domesticus*	Jiangxi	EastAsia
S-f3_I108	Jiangxi4	*G. g. domesticus*	Jiangxi	EastAsia
S-f4_L4_I109	Jiangxi5	*G. g. domesticus*	Jiangxi	EastAsia
S-m1_I101	Jiangxi6	*G. g. domesticus*	Jiangxi	EastAsia
S-m2_I102	Jiangxi7	*G. g. domesticus*	Jiangxi	EastAsia
S-m3_I103	Jiangxi8	*G. g. domesticus*	Jiangxi	EastAsia
S-m4_L2_I104	Jiangxi9	*G. g. domesticus*	Jiangxi	EastAsia
SAMEA6529932	Shamo1140	*G. g. domesticus*	Ko Shamo	EastAsia
SAMN03940095	Korea01	*G. g. domesticus*	Korea	EastAsia
SAMN03940097	Korea02	*G. g. domesticus*	Korea	EastAsia
SAMN03940096	Korea03	*G. g. domesticus*	Korea	EastAsia
ypt3195_L1_I042	Liaoning1	*G. g. domesticus*	Liaoning	EastAsia
ypt3196_L2_I043	Liaoning2	*G. g. domesticus*	Liaoning	EastAsia
ypt3197_L1_I044	Liaoning3	*G. g. domesticus*	Liaoning	EastAsia
ypt3198_L2_I045	Liaoning4	*G. g. domesticus*	Liaoning	EastAsia
ypt3204_L7_I046	Liaoning5	*G. g. domesticus*	Liaoning	EastAsia
ypt3205_L2_I047	Liaoning6	*G. g. domesticus*	Liaoning	EastAsia
ypt3206_L1_I048	Liaoning7	*G. g. domesticus*	Liaoning	EastAsia
ypt3207_L6_I050	Liaoning8	*G. g. domesticus*	Liaoning	EastAsia
SAMN02712023	LuxiFighting01	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712024	LuxiFighting02	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712025	LuxiFighting03	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712026	LuxiFighting04	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712027	LuxiFighting05	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712028	LuxiFighting06	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712029	LuxiFighting07	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02712030	LuxiFighting08	*G. g. domesticus*	Luxi Fighting	EastAsia
SAMN02486159	LuxiGame1	*G. g. domesticus*	Luxi Game	EastAsia
SAMN14814745	MalaysianMuseum2	*G. g. domesticus*	Malacca	EastAsia
SAMEA6529936	Malay1265	*G. g. domesticus*	Malay	EastAsia
SAMN02486156	Beijing11	*G. g. domesticus*	N/A	EastAsia
SAMN02486162	Shouguang1	*G. g. domesticus*	N/A	EastAsia
SAMN02142123	Silkie1	*G. g. domesticus*	N/A	EastAsia
SAMN09700449	Ningdu02	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700450	Ningdu05	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700451	Ningdu06	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700452	Ningdu09	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700453	Ningdu10	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700454	Ningdu13	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700455	Ningdu18	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700456	Ningdu26	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700457	Ningdu33	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMN09700458	Ningdu50	*G. g. domesticus*	Ningdu Yellow	EastAsia
SAMEA6529964	OhShamo2054	*G. g. domesticus*	Oh Shamo	EastAsia
SAMEA6529940	Ohiki1341	*G. g. domesticus*	Ohiki	EastAsia
SAMEA6529970	Ohiki2285	*G. g. domesticus*	Ohiki	EastAsia
ypt2862	Thailand3	*G. g. domesticus*	Phuket	EastAsia
ypt2863	Thailand4	*G. g. domesticus*	Phuket	EastAsia
ypt2870	Thailand5	*G. g. domesticus*	Phuket	EastAsia
ypt2871	Thailand6	*G. g. domesticus*	Phuket	EastAsia
ypt3179_L8_I025	Shandong1	*G. g. domesticus*	Shandong	EastAsia
Ypt603	Shandong10	*G. g. domesticus*	Shandong	EastAsia
YPt604	Shandong11	*G. g. domesticus*	Shandong	EastAsia

Accession	ID	Species	Breed/Location	Population
Ypt606	Shandong12	*G. g. domesticus*	Shandong	EastAsia
Ypt607	Shandong13	*G. g. domesticus*	Shandong	EastAsia
Ypt608	Shandong14	*G. g. domesticus*	Shandong	EastAsia
Ypt608	Shandong14	*G. g. domesticus*	Shandong	EastAsia
YPt609	Shandong15	*G. g. domesticus*	Shandong	EastAsia
YPt610	Shandong16	*G. g. domesticus*	Shandong	EastAsia
ypt3180	Shandong2	*G. g. domesticus*	Shandong	EastAsia
ypt3181_L7_I038	Shandong3	*G. g. domesticus*	Shandong	EastAsia
ypt3182_L1_I039	Shandong4	*G. g. domesticus*	Shandong	EastAsia
ypt3189_L1_I040	Shandong5	*G. g. domesticus*	Shandong	EastAsia
ypt3190_L8_I026	Shandong6	*G. g. domesticus*	Shandong	EastAsia
ypt3191_L2_I041	Shandong7	*G. g. domesticus*	Shandong	EastAsia
ypt3193_L1_I027	Shandong8	*G. g. domesticus*	Shandong	EastAsia
Ypt602	Shandong9	*G. g. domesticus*	Shandong	EastAsia
ypt3259_L6_I002	Shanxi1	*G. g. domesticus*	Shanxi	EastAsia
ypt3260_L6_I003	Shanxi2	*G. g. domesticus*	Shanxi	EastAsia
ypt3261_L7_I004	Shanxi3	*G. g. domesticus*	Shanxi	EastAsia
ypt3262_L7_I005	Shanxi4	*G. g. domesticus*	Shanxi	EastAsia
ypt3269_L7_I006	Shanxi5	*G. g. domesticus*	Shanxi	EastAsia
ypt3270_L1_I007	Shanxi6	*G. g. domesticus*	Shanxi	EastAsia
ypt3271_L7_I008	Shanxi7	*G. g. domesticus*	Shanxi	EastAsia
ypt3272_L7_I010	Shanxi8	*G. g. domesticus*	Shanxi	EastAsia
SAMN02486162	Shouguang2	*G. g. domesticus*	Shouguang	EastAsia
ypt2172	Indonesia10	*G. g. domesticus*	South Celebes	EastAsia
ypt2173	Indonesia11	*G. g. domesticus*	South Celebes	EastAsia
ypt2169	Indonesia7	*G. g. domesticus*	South Celebes	EastAsia
ypt2170	Indonesia8	*G. g. domesticus*	South Celebes	EastAsia
ypt2171	Indonesia9	*G. g. domesticus*	South Celebes	EastAsia
ypt2174	Indonesia12	*G. g. domesticus*	Sumatra	EastAsia
ypt2175	Indonesia13	*G. g. domesticus*	Sumatra	EastAsia
ypt2176	Indonesia14	*G. g. domesticus*	Sumatra	EastAsia
ypt2177	Indonesia15	*G. g. domesticus*	Sumatra	EastAsia
ypt2178	Indonesia16	*G. g. domesticus*	Sumatra	EastAsia
ypt2200	Indonesia24	*G. g. domesticus*	Sumatra	EastAsia
ypt2201	Indonesia25	*G. g. domesticus*	Sumatra	EastAsia
SAMEA6529953	Sumatra1745	*G. g. domesticus*	Sumatra	EastAsia
SAMN02142124	Taiwan1	*G. g. domesticus*	Taiwan	EastAsia
SAMN02142124	Taiwan2	*G. g. domesticus*	Taiwan	EastAsia
SAMN02142124	Taiwan3	*G. g. domesticus*	Taiwan	EastAsia
SAMN02142124	Taiwan4	*G. g. domesticus*	Taiwan	EastAsia
SAMN02142124	Taiwan5	*G. g. domesticus*	Taiwan	EastAsia
ypt2860	Thailand1	*G. g. domesticus*	Thammarat	EastAsia
ypt2861	Thailand2	*G. g. domesticus*	Thammarat	EastAsia
ypt2437	Tibet1	*G. g. domesticus*	Tibet	EastAsia
ypt2446	Tibet10	*G. g. domesticus*	Tibet	EastAsia
ypt3162_L6_I036	Tibet100	*G. g. domesticus*	Tibet	EastAsia
ypt3289_L8_I021	Tibet101	*G. g. domesticus*	Tibet	EastAsia
SAMN02486164	Tibet102	*G. g. domesticus*	Tibet	EastAsia
ypt2447	Tibet11	*G. g. domesticus*	Tibet	EastAsia
ypt2448	Tibet12	*G. g. domesticus*	Tibet	EastAsia
ypt2449	Tibet13	*G. g. domesticus*	Tibet	EastAsia
ypt2454	Tibet14	*G. g. domesticus*	Tibet	EastAsia
ypt2459	Tibet15	*G. g. domesticus*	Tibet	EastAsia
ypt2468	Tibet16	*G. g. domesticus*	Tibet	EastAsia
ypt2496	Tibet17	*G. g. domesticus*	Tibet	EastAsia
ypt2497	Tibet18	*G. g. domesticus*	Tibet	EastAsia
ypt2498	Tibet19	*G. g. domesticus*	Tibet	EastAsia
ypt2438	Tibet2	*G. g. domesticus*	Tibet	EastAsia
ypt2499	Tibet20	*G. g. domesticus*	Tibet	EastAsia
T2	Tibet200	*G. g. domesticus*	Tibet	EastAsia
T3	Tibet201	*G. g. domesticus*	Tibet	EastAsia
T5	Tibet202	*G. g. domesticus*	Tibet	EastAsia
Tibet	Tibet203	*G. g. domesticus*	Tibet	EastAsia
Yt536	Tibet204	*G. g. domesticus*	Tibet	EastAsia
Yt538	Tibet205	*G. g. domesticus*	Tibet	EastAsia
Yt539	Tibet206	*G. g. domesticus*	Tibet	EastAsia
ZJ1	Tibet207	*G. g. domesticus*	Tibet	EastAsia
ZJ2	Tibet208	*G. g. domesticus*	Tibet	EastAsia
ZJ3	Tibet209	*G. g. domesticus*	Tibet	EastAsia
ypt2500	Tibet21	*G. g. domesticus*	Tibet	EastAsia
ZJ4	Tibet210	*G. g. domesticus*	Tibet	EastAsia

Accession	ID	Species	Breed/Location	Population
SAMN02712045	Tibet211	*G. g. domesticus*	Tibet	EastAsia
SAMN02712047	Tibet212	*G. g. domesticus*	Tibet	EastAsia
SAMN02712048	Tibet212	*G. g. domesticus*	Tibet	EastAsia
SAMN02712048	Tibet213	*G. g. domesticus*	Tibet	EastAsia
SAMN02712049	Tibet214	*G. g. domesticus*	Tibet	EastAsia
SAMN02712050	Tibet215	*G. g. domesticus*	Tibet	EastAsia
SAMN02712051	Tibet216	*G. g. domesticus*	Tibet	EastAsia
SAMN02712052	Tibet217	*G. g. domesticus*	Tibet	EastAsia
SAMN02712053	Tibet218	*G. g. domesticus*	Tibet	EastAsia
SAMN02712054	Tibet219	*G. g. domesticus*	Tibet	EastAsia
ypt2503	Tibet22	*G. g. domesticus*	Tibet	EastAsia
SAMN02712046	Tibet220	*G. g. domesticus*	Tibet	EastAsia
SAMN02712044	Tibet221	*G. g. domesticus*	Tibet	EastAsia
ypt2504	Tibet23	*G. g. domesticus*	Tibet	EastAsia
ypt2507	Tibet24	*G. g. domesticus*	Tibet	EastAsia
ypt2512	Tibet25	*G. g. domesticus*	Tibet	EastAsia
ypt2516	Tibet26	*G. g. domesticus*	Tibet	EastAsia
ypt2521	Tibet27	*G. g. domesticus*	Tibet	EastAsia
ypt2523	Tibet28	*G. g. domesticus*	Tibet	EastAsia
ypt2534	Tibet29	*G. g. domesticus*	Tibet	EastAsia
ypt2439	Tibet3	*G. g. domesticus*	Tibet	EastAsia
ypt2541	Tibet30	*G. g. domesticus*	Tibet	EastAsia
ypt2543	Tibet31	*G. g. domesticus*	Tibet	EastAsia
ypt2544	Tibet32	*G. g. domesticus*	Tibet	EastAsia
ypt2545	Tibet33	*G. g. domesticus*	Tibet	EastAsia
ypt2546	Tibet34	*G. g. domesticus*	Tibet	EastAsia
ypt2547	Tibet35	*G. g. domesticus*	Tibet	EastAsia
ypt2552	Tibet36	*G. g. domesticus*	Tibet	EastAsia
ypt2562	Tibet37	*G. g. domesticus*	Tibet	EastAsia
ypt2563	Tibet38	*G. g. domesticus*	Tibet	EastAsia
ypt2569	Tibet39	*G. g. domesticus*	Tibet	EastAsia
ypt2440	Tibet4	*G. g. domesticus*	Tibet	EastAsia
ypt2570	Tibet40	*G. g. domesticus*	Tibet	EastAsia
ypt2571	Tibet41	*G. g. domesticus*	Tibet	EastAsia
ypt2572	Tibet42	*G. g. domesticus*	Tibet	EastAsia
ypt2573	Tibet43	*G. g. domesticus*	Tibet	EastAsia
ypt2574	Tibet44	*G. g. domesticus*	Tibet	EastAsia
ypt2575	Tibet45	*G. g. domesticus*	Tibet	EastAsia
ypt2579	Tibet46	*G. g. domesticus*	Tibet	EastAsia
ypt2598	Tibet47	*G. g. domesticus*	Tibet	EastAsia
ypt2601	Tibet48	*G. g. domesticus*	Tibet	EastAsia
ypt2606	Tibet49	*G. g. domesticus*	Tibet	EastAsia
ypt2441	Tibet5	*G. g. domesticus*	Tibet	EastAsia
ypt2608	Tibet50	*G. g. domesticus*	Tibet	EastAsia
ypt2615	Tibet51	*G. g. domesticus*	Tibet	EastAsia
ypt2622	Tibet52	*G. g. domesticus*	Tibet	EastAsia
ypt2623	Tibet53	*G. g. domesticus*	Tibet	EastAsia
ypt2624	Tibet54	*G. g. domesticus*	Tibet	EastAsia
ypt2625	Tibet55	*G. g. domesticus*	Tibet	EastAsia
ypt2627	Tibet56	*G. g. domesticus*	Tibet	EastAsia
ypt2628	Tibet57	*G. g. domesticus*	Tibet	EastAsia
ypt2629	Tibet58	*G. g. domesticus*	Tibet	EastAsia
ypt2630	Tibet59	*G. g. domesticus*	Tibet	EastAsia
ypt2442	Tibet6	*G. g. domesticus*	Tibet	EastAsia
ypt2636	Tibet60	*G. g. domesticus*	Tibet	EastAsia
ypt2638	Tibet61	*G. g. domesticus*	Tibet	EastAsia
ypt2640	Tibet62	*G. g. domesticus*	Tibet	EastAsia
ypt2641	Tibet63	*G. g. domesticus*	Tibet	EastAsia
ypt2643	Tibet64	*G. g. domesticus*	Tibet	EastAsia
ypt2644	Tibet65	*G. g. domesticus*	Tibet	EastAsia
ypt2646	Tibet66	*G. g. domesticus*	Tibet	EastAsia
ypt2650	Tibet67	*G. g. domesticus*	Tibet	EastAsia
ypt2653	Tibet68	*G. g. domesticus*	Tibet	EastAsia
ypt2656	Tibet69	*G. g. domesticus*	Tibet	EastAsia
ypt2443	Tibet7	*G. g. domesticus*	Tibet	EastAsia
ypt2657	Tibet70	*G. g. domesticus*	Tibet	EastAsia
ypt2658	Tibet71	*G. g. domesticus*	Tibet	EastAsia
ypt2661	Tibet72	*G. g. domesticus*	Tibet	EastAsia
ypt2664	Tibet73	*G. g. domesticus*	Tibet	EastAsia
ypt2665	Tibet74	*G. g. domesticus*	Tibet	EastAsia
ypt2669	Tibet75	*G. g. domesticus*	Tibet	EastAsia

Accession	ID	Species	Breed/Location	Population
ypt2670	Tibet76	*G. g. domesticus*	Tibet	EastAsia
ypt2671	Tibet77	*G. g. domesticus*	Tibet	EastAsia
ypt2672	Tibet78	*G. g. domesticus*	Tibet	EastAsia
ypt2673	Tibet79	*G. g. domesticus*	Tibet	EastAsia
ypt2444	Tibet8	*G. g. domesticus*	Tibet	EastAsia
ypt2678	Tibet80	*G. g. domesticus*	Tibet	EastAsia
ypt2706	Tibet81	*G. g. domesticus*	Tibet	EastAsia
ypt2722	Tibet82	*G. g. domesticus*	Tibet	EastAsia
ypt2724	Tibet83	*G. g. domesticus*	Tibet	EastAsia
ypt2729	Tibet84	*G. g. domesticus*	Tibet	EastAsia
ypt2733	Tibet85	*G. g. domesticus*	Tibet	EastAsia
ypt2739	Tibet86	*G. g. domesticus*	Tibet	EastAsia
ypt2740	Tibet87	*G. g. domesticus*	Tibet	EastAsia
ypt2745	Tibet88	*G. g. domesticus*	Tibet	EastAsia
ypt2746	Tibet89	*G. g. domesticus*	Tibet	EastAsia
ypt2445	Tibet9	*G. g. domesticus*	Tibet	EastAsia
ypt2760	Tibet90	*G. g. domesticus*	Tibet	EastAsia
ypt2761	Tibet91	*G. g. domesticus*	Tibet	EastAsia
ypt2765	Tibet92	*G. g. domesticus*	Tibet	EastAsia
ypt2766	Tibet93	*G. g. domesticus*	Tibet	EastAsia
ypt2775	Tibet94	*G. g. domesticus*	Tibet	EastAsia
ypt2780	Tibet95	*G. g. domesticus*	Tibet	EastAsia
ypt3149_L2_I034	Tibet96	*G. g. domesticus*	Tibet	EastAsia
ypt3150_L8_I023	Tibet97	*G. g. domesticus*	Tibet	EastAsia
ypt3159_L6_I035	Tibet98	*G. g. domesticus*	Tibet	EastAsia
ypt3160_L8_I024	Tibet99	*G. g. domesticus*	Tibet	EastAsia
SAMN02486164	Tibet103	*G. g. domesticus*	Tibet	EastAsia
SAMEA6529961	Toutenkou1961	*G. g. domesticus*	Toutenkou	EastAsia
ypt2041_L6_I002	Vietnan1	*G. g. domesticus*	Vietnam	EastAsia
ypt2043_L1_I037	Vietnan2	*G. g. domesticus*	Vietnam	EastAsia
ypt2051_L3_I003	Vietnan3	*G. g. domesticus*	Vietnam	EastAsia
ypt2053_L6_I004	Vietnan4	*G. g. domesticus*	Vietnam	EastAsia
ypt2063_L6_I005	Vietnan5	*G. g. domesticus*	Vietnam	EastAsia
ypt2069_L6_I006	Vietnan6	*G. g. domesticus*	Vietnam	EastAsia
ypt2071_L1_I038	Vietnan7	*G. g. domesticus*	Vietnam	EastAsia
SAMN09700499	Wenchang23	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700500	Wenchang25	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700501	Wenchang26	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700502	Wenchang27	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700503	Wenchang28	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700504	Wenchang29	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700505	Wenchang30	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700506	Wenchang31	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700507	Wenchang33	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700508	Wenchang34	*G. g. domesticus*	Wenchang	EastAsia
SAMN02486165	Wenchang1	*G. g. domesticus*	Wenchang	EastAsia
SAMN02486165	Wenchang2	*G. g. domesticus*	Wenchang	EastAsia
SAMN09700459	Wuhua01	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700460	Wuhua04	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700461	Wuhua05	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700462	Wuhua10	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700463	Wuhua12	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700464	Wuhua17	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700465	Wuhua18	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700466	Wuhua29	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700467	Wuhua38	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN09700468	Wuhua39	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572365	WuhuaA	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572366	WuhuaB	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572367	WuhuaC	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572368	WuhuaD	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572369	WuhuaE	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572370	WuhuaF	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572371	WuhuaG	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572372	WuhuaH	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572373	WuhuaI	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572374	WuhuaJ	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572375	WuhuaK	*G. g. domesticus*	Wuhua Yellow	EastAsia
SAMN14572376	WuhuaP	*G. g. domesticus*	Wuhua Yellow	EastAsia
A1	Xinjiang1	*G. g. domesticus*	Xinjiang	EastAsia
N18	Xinjiang10	*G. g. domesticus*	Xinjiang	EastAsia

Accession	ID	Species	Breed/Location	Population
N3	Xinjiang11	*G. g. domesticus*	Xinjiang	EastAsia
N5	Xinjiang12	*G. g. domesticus*	Xinjiang	EastAsia
N7	Xinjiang13	*G. g. domesticus*	Xinjiang	EastAsia
N8	Xinjiang14	*G. g. domesticus*	Xinjiang	EastAsia
TD1	Xinjiang15	*G. g. domesticus*	Xinjiang	EastAsia
TD12	Xinjiang16	*G. g. domesticus*	Xinjiang	EastAsia
TD15	Xinjiang17	*G. g. domesticus*	Xinjiang	EastAsia
TD22	Xinjiang18	*G. g. domesticus*	Xinjiang	EastAsia
TD27	Xinjiang19	*G. g. domesticus*	Xinjiang	EastAsia
A10	Xinjiang2	*G. g. domesticus*	Xinjiang	EastAsia
TD30	Xinjiang20	*G. g. domesticus*	Xinjiang	EastAsia
TD6	Xinjiang21	*G. g. domesticus*	Xinjiang	EastAsia
TD9	Xinjiang22	*G. g. domesticus*	Xinjiang	EastAsia
ypt2902	Xinjiang23	*G. g. domesticus*	Xinjiang	EastAsia
ypt2909	Xinjiang24	*G. g. domesticus*	Xinjiang	EastAsia
ypt2910	Xinjiang25	*G. g. domesticus*	Xinjiang	EastAsia
ypt2911	Xinjiang26	*G. g. domesticus*	Xinjiang	EastAsia
ypt2912	Xinjiang27	*G. g. domesticus*	Xinjiang	EastAsia
ypt2918	Xinjiang28	*G. g. domesticus*	Xinjiang	EastAsia
ypt2929	Xinjiang29	*G. g. domesticus*	Xinjiang	EastAsia
A12	Xinjiang3	*G. g. domesticus*	Xinjiang	EastAsia
ypt2930	Xinjiang30	*G. g. domesticus*	Xinjiang	EastAsia
A14	Xinjiang4	*G. g. domesticus*	Xinjiang	EastAsia
A3	Xinjiang5	*G. g. domesticus*	Xinjiang	EastAsia
A8	Xinjiang6	*G. g. domesticus*	Xinjiang	EastAsia
N1	Xinjiang7	*G. g. domesticus*	Xinjiang	EastAsia
N11	Xinjiang8	*G. g. domesticus*	Xinjiang	EastAsia
N13	Xinjiang9	*G. g. domesticus*	Xinjiang	EastAsia
SAMEA6529958	Yokohama1906	*G. g. domesticus*	Yokohama	EastAsia
ypt3133	Yunnan1	*G. g. domesticus*	Yunnan	EastAsia
ypt3142	Yunnan10	*G. g. domesticus*	Yunnan	EastAsia
18833	Yunnan100	*G. g. domesticus*	Yunnan	EastAsia
19912	Yunnan101	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712031	Yunnan102	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712032	Yunnan103	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712033	Yunnan104	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712034	Yunnan105	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712035	Yunnan106	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712036	Yunnan107	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712037	Yunnan108	*G. g. domesticus*	Yunnan	EastAsia
SAMN02712038	Yunnan109	*G. g. domesticus*	Yunnan	EastAsia
ypt3143	Yunnan11	*G. g. domesticus*	Yunnan	EastAsia
ypt3144_L3_I039	Yunnan12	*G. g. domesticus*	Yunnan	EastAsia
ypt3145	Yunnan13	*G. g. domesticus*	Yunnan	EastAsia
ypt3146	Yunnan14	*G. g. domesticus*	Yunnan	EastAsia
ypt47_L2_I029	Yunnan15	*G. g. domesticus*	Yunnan	EastAsia
ypt50_L2_I030	Yunnan16	*G. g. domesticus*	Yunnan	EastAsia
ypt51_L2_I031	Yunnan17	*G. g. domesticus*	Yunnan	EastAsia
ypt53_L2_I032	Yunnan18	*G. g. domesticus*	Yunnan	EastAsia
ypt55_L1_I033	Yunnan19	*G. g. domesticus*	Yunnan	EastAsia
ypt3134	Yunnan2	*G. g. domesticus*	Yunnan	EastAsia
YPt565	Yunnan20	*G. g. domesticus*	Yunnan	EastAsia
ypt984	Yunnan21	*G. g. domesticus*	Yunnan	EastAsia
yptwj	Yunnan22	*G. g. domesticus*	Yunnan	EastAsia
Yt516	Yunnan23	*G. g. domesticus*	Yunnan	EastAsia
Yt517	Yunnan24	*G. g. domesticus*	Yunnan	EastAsia
Yt518	Yunnan25	*G. g. domesticus*	Yunnan	EastAsia
Yt519	Yunnan26	*G. g. domesticus*	Yunnan	EastAsia
Yt520	Yunnan27	*G. g. domesticus*	Yunnan	EastAsia
Yt521	Yunnan28	*G. g. domesticus*	Yunnan	EastAsia
Yt522	Yunnan29	*G. g. domesticus*	Yunnan	EastAsia
ypt3135	Yunnan3	*G. g. domesticus*	Yunnan	EastAsia
Yt523	Yunnan30	*G. g. domesticus*	Yunnan	EastAsia
Yt524	Yunnan31	*G. g. domesticus*	Yunnan	EastAsia
Yt525	Yunnan32	*G. g. domesticus*	Yunnan	EastAsia
Yt526	Yunnan33	*G. g. domesticus*	Yunnan	EastAsia
Yt527	Yunnan34	*G. g. domesticus*	Yunnan	EastAsia
Yt528	Yunnan35	*G. g. domesticus*	Yunnan	EastAsia
Yt529	Yunnan36	*G. g. domesticus*	Yunnan	EastAsia
Yt530	Yunnan37	*G. g. domesticus*	Yunnan	EastAsia
Yt531	Yunnan38	*G. g. domesticus*	Yunnan	EastAsia

Accession	ID	Species	Breed/Location	Population
ypt3136	Yunnan4	*G. g. domesticus*	Yunnan	EastAsia
ypt3137	Yunnan5	*G. g. domesticus*	Yunnan	EastAsia
ypt3138	Yunnan6	*G. g. domesticus*	Yunnan	EastAsia
ypt3139	Yunnan7	*G. g. domesticus*	Yunnan	EastAsia
ypt3140	Yunnan8	*G. g. domesticus*	Yunnan	EastAsia
ypt3141	Yunnan9	*G. g. domesticus*	Yunnan	EastAsia
SAMN09700469	Zhengyang01	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700470	Zhengyang02	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700471	Zhengyang05	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700472	Zhengyang13	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700473	Zhengyang14	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700474	Zhengyang17	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700475	Zhengyang23	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700476	Zhengyang25	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700477	Zhengyang38	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN09700478	Zhengyang41	*G. g. domesticus*	Zhengyang Yellow	EastAsia
SAMN12236822	BlackMinorca01	*G. g. domesticus*	Black Minorca	Europe
SAMN12236833	BlackMinorca02	*G. g. domesticus*	Black Minorca	Europe
SAMN12236834	BlackMinorca03	*G. g. domesticus*	Black Minorca	Europe
SAMN12236835	BlackMinorca04	*G. g. domesticus*	Black Minorca	Europe
SAMN12236824	BlackMinorca05	*G. g. domesticus*	Black Minorca	Europe
SAMN12236823	BlackMinorca06	*G. g. domesticus*	Black Minorca	Europe
SAMN12236826	BlackMinorca07	*G. g. domesticus*	Black Minorca	Europe
SAMN12236825	BlackMinorca08	*G. g. domesticus*	Black Minorca	Europe
SAMN12236828	BlackMinorca09	*G. g. domesticus*	Black Minorca	Europe
SAMN12236827	BlackMinorca10	*G. g. domesticus*	Black Minorca	Europe
SAMN12236830	BlackMinorca11	*G. g. domesticus*	Black Minorca	Europe
SAMN12236829	BlackMinorca12	*G. g. domesticus*	Black Minorca	Europe
SAMN12236832	BlackMinorca13	*G. g. domesticus*	Black Minorca	Europe
SAMN12236831	BlackMinorca14	*G. g. domesticus*	Black Minorca	Europe
SAMEA5930870	Brabanter249	*G. g. domesticus*	Brabanter	Europe
SAMEA5930871	Brabanter252	*G. g. domesticus*	Brabanter	Europe
SAMEA5930872	Brabanter254	*G. g. domesticus*	Brabanter	Europe
SAMN02712022	Commercial	*G. g. domesticus*	Commercial	Europe
SAMN02486157	Cornish1	*G. g. domesticus*	Cornish	Europe
SAMEA6529937	Marans1283	*G. g. domesticus*	Marans	Europe
SAMEA6529938	Minorca1301	*G. g. domesticus*	Minorca	Europe
SAMEA6529973	Appenzeller2375	*G. g. domesticus*	Appenzeller Pointed Hood	Fancy
SAMEA5930836	Assendelft90	*G. g. domesticus*	Assendelft Fowl Bm	Fancy
SAMEA5930837	Assendelft91	*G. g. domesticus*	Assendelft Fowl Bm	Fancy
SAMEA6529972	BarbueA2311	*G. g. domesticus*	Barbue d'Anvers quail	Fancy
SAMEA6529975	BarbueG2471	*G. g. domesticus*	Barbue du Grubbe	Fancy
SAMEA5930880	Barnevelder317	*G. g. domesticus*	Barnevelder	Fancy
SAMEA5930881	Barnevelder322	*G. g. domesticus*	Barnevelder	Fancy
SAMEA5930882	Barnevelder325	*G. g. domesticus*	Barnevelder	Fancy
SAMEA5930883	Barnevelder334	*G. g. domesticus*	Barnevelder	Fancy
SAMEA5930884	BarBantam344	*G. g. domesticus*	Barnevelder Bm	Fancy
SAMEA5930885	BarBantam346	*G. g. domesticus*	Barnevelder Bm	Fancy
SAMEA5930886	BarBantam348	*G. g. domesticus*	Barnevelder Bm	Fancy
SAMEA6529979	Berg-Schlotter2614	*G. g. domesticus*	Berg-Schlotter	Fancy
SAMEA6529976	Bergische2476	*G. g. domesticus*	Bergische Crower	Fancy
SAMEA6529990	Booted807	*G. g. domesticus*	Booted Bm	Fancy
SAMEA6529991	Booted843	*G. g. domesticus*	Booted Bm	Fancy
SAMEA5930873	BrabanterBantam257	*G. g. domesticus*	Brabanter Bm	Fancy
SAMEA5930874	BrabanterBantam258	*G. g. domesticus*	Brabanter Bm	Fancy
SAMEA5930875	BrabanterBantam261	*G. g. domesticus*	Brabanter Bm	Fancy
SAMEA6529982	Braekel426	*G. g. domesticus*	Braekel	Fancy
SAMEA6529980	Brahma364	*G. g. domesticus*	Brahma	Fancy
SAMEA6529981	Brahma383	*G. g. domesticus*	Brahma	Fancy
SAMEA5930864	BredaFowl203	*G. g. domesticus*	Breda Fowl	Fancy
SAMEA5930865	BredaFowl209	*G. g. domesticus*	Breda Fowl	Fancy
SAMEA5930866	BredaFowl221	*G. g. domesticus*	Breda Fowl	Fancy
SAMEA5930867	BredaFowlBantam235	*G. g. domesticus*	Breda Fowl Bm	Fancy
SAMEA5930868	BredaFowlBantam236	*G. g. domesticus*	Breda Fowl Bm	Fancy
SAMEA5930869	BredaFowlBantam237	*G. g. domesticus*	Breda Fowl Bm	Fancy
SAMN12236844	BuffCochin01	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236843	BuffCochin02	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236841	BuffCochin03	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236842	BuffCochin04	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236836	BuffCochin05	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236837	BuffCochin06	*G. g. domesticus*	Buff Cochin	Fancy

Accession	ID	Species	Breed/Location	Population
SAMN12236838	BuffCochin07	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236839	BuffCochin08	*G. g. domesticus*	Buff Cochin	Fancy
SAMN12236840	BuffCochin09	*G. g. domesticus*	Buff Cochin	Fancy
SAMEA6529950	Carlisle165	*G. g. domesticus*	Carlisle Old English	Fancy
SAMEA6529931	Castilian1125	*G. g. domesticus*	Castilian	Fancy
SAMEA5930903	ChaamFowl406	*G. g. domesticus*	Chaam Fowl	Fancy
SAMEA5930904	ChaamFowl408	*G. g. domesticus*	Chaam Fowl	Fancy
SAMEA6529985	Cochin585	*G. g. domesticus*	Cochin	Fancy
SAMEA5930912	Creeper1058	*G. g. domesticus*	Creeper	Fancy
SAMEA5930913	Creeper1059	*G. g. domesticus*	Creeper	Fancy
SAMEA6529933	Creeper1205	*G. g. domesticus*	Creeper	Fancy
SAMEA6529974	Creeper2457	*G. g. domesticus*	Creeper	Fancy
SAMEA6529989	Dorking742	*G. g. domesticus*	Dorking	Fancy
SAMEA5930834	Drenthe67	*G. g. domesticus*	Drenthe Fowl	Fancy
SAMEA5930835	DrentheBantam76	*G. g. domesticus*	Drenthe Fowl Bm	Fancy
SAMEA5930914	Bolstaart1062	*G. g. domesticus*	Drenthe Fowl Bolstaart	Fancy
SAMEA5930915	Bolstaart1063	*G. g. domesticus*	Drenthe Fowl Bolstaart	Fancy
SAMEA5930876	DutchBantam276	*G. g. domesticus*	Dutch Bm	Fancy
SAMEA5930877	DutchBantam291	*G. g. domesticus*	Dutch Bm	Fancy
SAMEA5930878	DutchBantam294	*G. g. domesticus*	Dutch Bm	Fancy
SAMEA5930879	DutchBootBantam301	*G. g. domesticus*	Dutch Booted Bm	Fancy
SAMEA5930844	DutchFowl116	*G. g. domesticus*	Dutch Fowl	Fancy
SAMEA5930845	DutchFowl117	*G. g. domesticus*	Dutch Fowl	Fancy
SAMEA5930846	DutchFowl118	*G. g. domesticus*	Dutch Fowl	Fancy
SAMEA5930847	DutchFowl119	*G. g. domesticus*	Dutch Fowl	Fancy
SAMEA5930848	DutchBantam127	*G. g. domesticus*	Dutch Fowl Bm	Fancy
SAMEA5930849	DutchBantam128	*G. g. domesticus*	Dutch Fowl Bm	Fancy
SAMEA5930854	DutchOwl147	*G. g. domesticus*	Dutch Owl Bd	Fancy
SAMEA5930855	DutchOwl151	*G. g. domesticus*	Dutch Owl Bd	Fancy
SAMEA5930856	DutchOwl160	*G. g. domesticus*	Dutch Owl Bd	Fancy
SAMEA5930857	DutchOwlBantam165	*G. g. domesticus*	Dutch Owl Bd Bm	Fancy
SAMEA5930858	DutchOwlBantam167	*G. g. domesticus*	Dutch Owl Bd Bm	Fancy
SAMEA5930859	DutchOwlBantam170	*G. g. domesticus*	Dutch Owl Bd Bm	Fancy
SAMEA5930850	DutchPoBeard130	*G. g. domesticus*	Dutch Polish Bd	Fancy
SAMEA5930851	DutchPoBeard131	*G. g. domesticus*	Dutch Polish Bd	Fancy
SAMEA5930852	DutchPoBeardBantam134	*G. g. domesticus*	Dutch Polish Bd Bm	Fancy
SAMEA5930853	DutchPoBeardBantam140	*G. g. domesticus*	Dutch Polish Bd Bm	Fancy
SAMEA5930860	DutchPolishNoBeard176	*G. g. domesticus*	Dutch Polish non Bd	Fancy
SAMEA5930861	DutchPolishNoBeard185	*G. g. domesticus*	Dutch Polish non Bd	Fancy
SAMEA5930862	DutchPoBantam189	*G. g. domesticus*	Dutch Polish non Bd Bm	Fancy
SAMEA5930863	DutchPoBantam190	*G. g. domesticus*	Dutch Polish non Bd Bm	Fancy
SAMEA6529942	EastFrisian1401	*G. g. domesticus*	East Frisian Gulls	Fancy
SAMEA5930907	EikenburgerBantam438	*G. g. domesticus*	Eikenburger Bm	Fancy
SAMEA5930908	EikenburgerBantam440	*G. g. domesticus*	Eikenburger Bm	Fancy
SAMEA5930841	Frisian107	*G. g. domesticus*	Frisian Fowl	Fancy
SAMEA6529995	Frisian888	*G. g. domesticus*	Frisian Fowl	Fancy
SAMEA5930838	Frisian95	*G. g. domesticus*	Frisian Fowl	Fancy
SAMEA5930839	Frisian97	*G. g. domesticus*	Frisian Fowl	Fancy
SAMEA5930842	FrisianBantam110	*G. g. domesticus*	Frisian Fowl Bm	Fancy
SAMEA5930843	FrisianBantam112	*G. g. domesticus*	Frisian Fowl Bm	Fancy
SAMEA6529988	GerBantan706	*G. g. domesticus*	German Bm	Fancy
SAMEA6529986	Faverolles641	*G. g. domesticus*	German Faverolles	Fancy
SAMEA6529987	GerGrey689	*G. g. domesticus*	German Grey	Fancy
SAMEA5930821	Groninger10	*G. g. domesticus*	Groninger Mew	Fancy
SAMEA5930822	Groninger13	*G. g. domesticus*	Groninger Mew	Fancy
SAMEA5930823	Groninger22	*G. g. domesticus*	Groninger Mew	Fancy
SAMEA5930820	Groninger7	*G. g. domesticus*	Groninger Mew	Fancy
SAMEA5930824	GroningerBantam23	*G. g. domesticus*	Groninger Mew Bm	Fancy
SAMEA5930825	GroningerBantam25	*G. g. domesticus*	Groninger Mew Bm	Fancy
SAMEA5930826	GroningerBantam31	*G. g. domesticus*	Groninger Mew Bm	Fancy
SAMEA5930827	GroningerBantam40	*G. g. domesticus*	Groninger Mew Bm	Fancy
SAMEA6529983	JapBantam485	*G. g. domesticus*	Japanese Bm	Fancy
SAMEA6529984	JapBantam523	*G. g. domesticus*	Japanese Bm	Fancy
SAMEA5930897	Kraienkoppe384	*G. g. domesticus*	Kraienkoppe	Fancy
SAMEA5930898	Kraienkoppe385	*G. g. domesticus*	Kraienkoppe	Fancy
SAMEA5930899	Kraienkoppe389	*G. g. domesticus*	Kraienkoppe	Fancy
SAMEA5930900	KraienkoppeBantam398	*G. g. domesticus*	Kraienkoppe Fowl Bm	Fancy
SAMEA5930901	KraienkoppeBantam401	*G. g. domesticus*	Kraienkoppe Fowl Bm	Fancy
SAMEA5930902	KraienkoppeBantam404	*G. g. domesticus*	Kraienkoppe Fowl Bm	Fancy
SAMEA6529934	LaKenvelder1229	*G. g. domesticus*	Lakenvelder	Fancy
SAMEA5930828	Lakenvelder48	*G. g. domesticus*	Lakenvelder	Fancy

Accession	ID	Species	Breed/Location	Population
SAMEA5930829	Lakenvelder50	*G. g. domesticus*	Lakenvelder	Fancy
SAMEA5930830	Lakenvelder52	*G. g. domesticus*	Lakenvelder	Fancy
SAMEA5930831	LakenBantam55	*G. g. domesticus*	Lakenvelder Bm	Fancy
SAMEA5930832	LakenBantam56	*G. g. domesticus*	Lakenvelder Bm	Fancy
SAMEA5930833	LakenBantam58	*G. g. domesticus*	Lakenvelder Bm	Fancy
SAMN12236872	Langshan01	*G. g. domesticus*	Langshan	Fancy
SAMN12236871	Langshan02	*G. g. domesticus*	Langshan	Fancy
SAMN12236868	Langshan03	*G. g. domesticus*	Langshan	Fancy
SAMN12236867	Langshan04	*G. g. domesticus*	Langshan	Fancy
SAMN12236870	Langshan05	*G. g. domesticus*	Langshan	Fancy
SAMN12236869	Langshan06	*G. g. domesticus*	Langshan	Fancy
SAMN12236866	Langshan07	*G. g. domesticus*	Langshan	Fancy
SAMN12236865	Langshan08	*G. g. domesticus*	Langshan	Fancy
SAMN12236877	Langshan09	*G. g. domesticus*	Langshan	Fancy
SAMN12236873	Langshan10	*G. g. domesticus*	Langshan	Fancy
SAMN12236874	Langshan11	*G. g. domesticus*	Langshan	Fancy
SAMN12236875	Langshan12	*G. g. domesticus*	Langshan	Fancy
SAMN12236876	Langshan13	*G. g. domesticus*	Langshan	Fancy
SAMN12236895	LightBrahma01	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236896	LightBrahma02	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236893	LightBrahma03	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236894	LightBrahma04	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236897	LightBrahma05	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236898	LightBrahma06	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236885	LightBrahma07	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236878	LightBrahma08	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236879	LightBrahma09	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236880	LightBrahma10	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236888	LightBrahma11	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236881	LightBrahma12	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236882	LightBrahma13	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236887	LightBrahma14	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236886	LightBrahma15	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236884	LightBrahma16	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236883	LightBrahma17	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236890	LightBrahma18	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236889	LightBrahma19	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236892	LightBrahma20	*G. g. domesticus*	Light Brahma	Fancy
SAMN12236891	LightBrahma21	*G. g. domesticus*	Light Brahma	Fancy
SAMEA5930895	HollandBlue375	*G. g. domesticus*	North Holland Blue	Fancy
SAMEA5930896	HollandBlueBantam379	*G. g. domesticus*	North Holland Blue Bm	Fancy
SAMEA6529962	Orloff1986	*G. g. domesticus*	Orloff	Fancy
SAMEA6529941	Orpington1380	*G. g. domesticus*	Orpington	Fancy
SAMN12236899	PartridgeCochin01	*G. g. domesticus*	Patridge Cochin	Fancy
SAMN12236900	PartridgeCochin02	*G. g. domesticus*	Patridge Cochin	Fancy
SAMN12236901	PartridgeCochin03	*G. g. domesticus*	Patridge Cochin	Fancy
SAMN12236902	PartridgeCochin04	*G. g. domesticus*	Patridge Cochin	Fancy
SAMEA6529959	PekinBantam1929	*G. g. domesticus*	Pekin Bm	Fancy
SAMEA6529960	PekinBantam1940	*G. g. domesticus*	Pekin Bm	Fancy
SAMEA6529944	Phoenix1464	*G. g. domesticus*	Phoenix	Fancy
SAMEA6529945	Plymouth1480	*G. g. domesticus*	Plymouth Rocks	Fancy
SAMEA6529943	Poland1424	*G. g. domesticus*	Poland	Fancy
SAMEA6529996	Poland961	*G. g. domesticus*	Poland	Fancy
SAMEA6529946	Rhinelander1541	*G. g. domesticus*	Rhinelander	Fancy
SAMEA6529947	Rhinelander1567	*G. g. domesticus*	Rhinelander	Fancy
SAMEA6529971	RoseBantam229	*G. g. domesticus*	Rosecomb Bm	Fancy
SAMEA5930905	Schijndelaar430	*G. g. domesticus*	Schijndelaar	Fancy
SAMEA5930906	SchijndBantam437	*G. g. domesticus*	Schijndelaar Bm	Fancy
SAMEA6529948	SebBantam1625	*G. g. domesticus*	Sebright Bm	Fancy
SAMEA6529949	SebBantam1644	*G. g. domesticus*	Sebright Bm	Fancy
SAMN14814819	Silkie8	*G. g. domesticus*	Silkie	Fancy
SAMN14814820	Silkie9	*G. g. domesticus*	Silkie	Fancy
SAMN02142123	Silkie2	*G. g. domesticus*	Silkie	Fancy
SAMN02142123	Silkie3	*G. g. domesticus*	Silkie	Fancy
SAMN02142123	Silkie4	*G. g. domesticus*	Silkie	Fancy
SAMN02142123	Silkie5	*G. g. domesticus*	Silkie	Fancy
SAMN02486163	Silkie6	*G. g. domesticus*	Silkie	Fancy
SAMN02486163	Silkie7	*G. g. domesticus*	Silkie	Fancy
SAMEA6529951	Silkies1687	*G. g. domesticus*	Silkie	Fancy
SAMEA6529952	Silkies1706	*G. g. domesticus*	Silkie	Fancy
SAMEA6529954	Sundheimer1769	*G. g. domesticus*	Sundheimer	Fancy

Accession	ID	Species	Breed/Location	Population
SAMEA6529978	Thuringian2539	*G. g. domesticus*	Thuringian Bd	Fancy
SAMEA6529955	Vorwerk1826	*G. g. domesticus*	Vorwerk Buff columbian	Fancy
SAMEA6529969	Vorwerk22	*G. g. domesticus*	Vorwerk conservation	Fancy
SAMEA5930888	Welsummer352	*G. g. domesticus*	Welsummer	Fancy
SAMEA5930889	Welsummer356	*G. g. domesticus*	Welsummer	Fancy
SAMEA5930890	Welsummer364	*G. g. domesticus*	Welsummer	Fancy
SAMEA5930891	WelsummerBantam367	*G. g. domesticus*	Welsummer Bm	Fancy
SAMEA5930892	WelsummerBantam370	*G. g. domesticus*	Welsummer Bm	Fancy
SAMEA5930893	WelsummerBantam372	*G. g. domesticus*	Welsummer Bm	Fancy
SAMEA6529956	Westphalian1846	*G. g. domesticus*	Westphalian	Fancy
SAMEA6529957	Wyandotte1882	*G. g. domesticus*	Wyandotte	Fancy
SAMEA6529967	Wyandotte2166	*G. g. domesticus*	Wyandotte	Fancy
SAMEA5160275	BL1	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160284	BL10	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160285	BL11	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160286	BL12	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160287	BL13	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160288	BL14	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160289	BL15	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160290	BL16	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160291	BL17	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160292	BL18	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160293	BL19	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160276	BL2	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160294	BL20	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160295	BL21	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160296	BL22	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160297	BL23	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160298	BL24	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160299	BL25	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160277	BL3	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160278	BL4	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160279	BL5	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160280	BL6	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160281	BL7	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160282	BL8	*G. g. domesticus*	BrownLayer	Layer
SAMEA5160283	BL9	*G. g. domesticus*	BrownLayer	Layer
84S_L5_I006	Europe_egg1	*G. g. domesticus*	Europe_egg	Layer
93S_L7_I024	Europe_egg10	*G. g. domesticus*	Europe_egg	Layer
95S_L8_I025	Europe_egg11	*G. g. domesticus*	Europe_egg	Layer
85S_L5_I007	Europe_egg2	*G. g. domesticus*	Europe_egg	Layer
86S_L5_I008	Europe_egg3	*G. g. domesticus*	Europe_egg	Layer
87S_L5_I010	Europe_egg4	*G. g. domesticus*	Europe_egg	Layer
88S_L4_I011	Europe_egg5	*G. g. domesticus*	Europe_egg	Layer
89S_L6_I012	Europe_egg6	*G. g. domesticus*	Europe_egg	Layer
90S_L4_I021	Europe_egg7	*G. g. domesticus*	Europe_egg	Layer
91S_L4_I022	Europe_egg8	*G. g. domesticus*	Europe_egg	Layer
92S_L5_I023	Europe_egg9	*G. g. domesticus*	Europe_egg	Layer
SAMN14814821	WL01	*G. g. domesticus*	White Layer	Layer
SAMN14814822	WL02	*G. g. domesticus*	White Layer	Layer
SAMN14814823	WL03	*G. g. domesticus*	White Layer	Layer
SAMN14814824	WL04	*G. g. domesticus*	White Layer	Layer
SAMN14814825	WL05	*G. g. domesticus*	White Layer	Layer
SAMEA6529929	Leghorn1000	*G. g. domesticus*	White Leghorn	Layer
SAMEA6529930	Leghorn1029	*G. g. domesticus*	White Leghorn	Layer
SAMEA6529935	Leghorn1245	*G. g. domesticus*	White Leghorn	Layer
SAMEA6529992	Leghorn84632	*G. g. domesticus*	White Leghorn	Layer
SAMN16226606	WhiteLeghorn10	*G. g. domesticus*	White Leghorn	Layer
SAMN16226605	WhiteLeghorn11	*G. g. domesticus*	White Leghorn	Layer
SAMN16226604	WhiteLeghorn12	*G. g. domesticus*	White Leghorn	Layer
SAMN16226603	WhiteLeghorn13	*G. g. domesticus*	White Leghorn	Layer
SAMN16226602	WhiteLeghorn14	*G. g. domesticus*	White Leghorn	Layer
SAMN16226601	WhiteLeghorn15	*G. g. domesticus*	White Leghorn	Layer
SAMN16226600	WhiteLeghorn16	*G. g. domesticus*	White Leghorn	Layer
SAMN16226564	WhiteLeghorn17	*G. g. domesticus*	White Leghorn	Layer
SAMN16226599	WhiteLeghorn18	*G. g. domesticus*	White Leghorn	Layer
SAMN16226598	WhiteLeghorn19	*G. g. domesticus*	White Leghorn	Layer
SAMN16226597	WhiteLeghorn20	*G. g. domesticus*	White Leghorn	Layer
SAMN16226596	WhiteLeghorn21	*G. g. domesticus*	White Leghorn	Layer
SAMN03940091	WhiteLeghorn22	*G. g. domesticus*	White Leghorn	Layer
SAMN16226595	WhiteLeghorn22	*G. g. domesticus*	White Leghorn	Layer

Accession	ID	Species	Breed/Location	Population
SAMN03940092	WhiteLeghorn23	*G. g. domesticus*	White Leghorn	Layer
SAMN16226594	WhiteLeghorn23	*G. g. domesticus*	White Leghorn	Layer
SAMN03940093	WhiteLeghorn24	*G. g. domesticus*	White Leghorn	Layer
SAMN16226593	WhiteLeghorn24	*G. g. domesticus*	White Leghorn	Layer
SAMN16226592	WhiteLeghorn25	*G. g. domesticus*	White Leghorn	Layer
SAMN16226591	WhiteLeghorn26	*G. g. domesticus*	White Leghorn	Layer
SAMN16226590	WhiteLeghorn27	*G. g. domesticus*	White Leghorn	Layer
SAMN16226563	WhiteLeghorn28	*G. g. domesticus*	White Leghorn	Layer
SAMN16226589	WhiteLeghorn29	*G. g. domesticus*	White Leghorn	Layer
SAMN16226569	WhiteLeghorn3	*G. g. domesticus*	White Leghorn	Layer
SAMN16226588	WhiteLeghorn30	*G. g. domesticus*	White Leghorn	Layer
SAMN16226587	WhiteLeghorn31	*G. g. domesticus*	White Leghorn	Layer
SAMN16226586	WhiteLeghorn32	*G. g. domesticus*	White Leghorn	Layer
SAMN16226585	WhiteLeghorn33	*G. g. domesticus*	White Leghorn	Layer
SAMN16226584	WhiteLeghorn34	*G. g. domesticus*	White Leghorn	Layer
SAMN16226583	WhiteLeghorn35	*G. g. domesticus*	White Leghorn	Layer
SAMN16226582	WhiteLeghorn36	*G. g. domesticus*	White Leghorn	Layer
SAMN16226581	WhiteLeghorn37	*G. g. domesticus*	White Leghorn	Layer
SAMN16226580	WhiteLeghorn38	*G. g. domesticus*	White Leghorn	Layer
SAMN16226562	WhiteLeghorn39	*G. g. domesticus*	White Leghorn	Layer
SAMN16226568	WhiteLeghorn4	*G. g. domesticus*	White Leghorn	Layer
SAMN16226579	WhiteLeghorn40	*G. g. domesticus*	White Leghorn	Layer
SAMN16226578	WhiteLeghorn41	*G. g. domesticus*	White Leghorn	Layer
SAMN16226577	WhiteLeghorn42	*G. g. domesticus*	White Leghorn	Layer
SAMN16226576	WhiteLeghorn43	*G. g. domesticus*	White Leghorn	Layer
SAMN16226575	WhiteLeghorn44	*G. g. domesticus*	White Leghorn	Layer
SAMN16226574	WhiteLeghorn45	*G. g. domesticus*	White Leghorn	Layer
SAMN16226573	WhiteLeghorn46	*G. g. domesticus*	White Leghorn	Layer
SAMN16226572	WhiteLeghorn47	*G. g. domesticus*	White Leghorn	Layer
SAMN16226571	WhiteLeghorn48	*G. g. domesticus*	White Leghorn	Layer
SAMN16226570	WhiteLeghorn49	*G. g. domesticus*	White Leghorn	Layer
SAMN16226567	WhiteLeghorn5	*G. g. domesticus*	White Leghorn	Layer
SAMN16226561	WhiteLeghorn50	*G. g. domesticus*	White Leghorn	Layer
SAMN16226560	WhiteLeghorn51	*G. g. domesticus*	White Leghorn	Layer
SAMN16226566	WhiteLeghorn6	*G. g. domesticus*	White Leghorn	Layer
SAMN16226565	WhiteLeghorn7	*G. g. domesticus*	White Leghorn	Layer
SAMN16226608	WhiteLeghorn8	*G. g. domesticus*	White Leghorn	Layer
SAMN16226607	WhiteLeghorn9	*G. g. domesticus*	White Leghorn	Layer
SAMN02486166	WhiteLeghorn1	*G. g. domesticus*	White Leghorn	Layer
SAMN02486166	WhiteLeghorn2	*G. g. domesticus*	White Leghorn	Layer
SAMEA5160300	WL1	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160309	WL10	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160310	WL11	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160311	WL12	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160312	WL13	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160313	WL14	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160314	WL15	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160315	WL16	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160316	WL17	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160317	WL18	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160318	WL19	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160301	WL2	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160319	WL20	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160320	WL21	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160321	WL22	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160322	WL23	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160323	WL24	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160324	WL25	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160302	WL3	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160303	WL4	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160304	WL5	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160305	WL6	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160306	WL7	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160307	WL8	*G. g. domesticus*	WhiteLayer	Layer
SAMEA5160308	WL9	*G. g. domesticus*	WhiteLayer	Layer
SAMN12236852	Dominique01	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236851	Dominique02	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236848	Dominique03	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236847	Dominique04	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236846	Dominique05	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236850	Dominique06	*G. g. domesticus*	Dominique	NorthAmerica

Accession	ID	Species	Breed/Location	Population
SAMN12236849	Dominique07	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236845	Dominique08	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236853	Dominique09	*G. g. domesticus*	Dominique	NorthAmerica
SAMN12236854	Dominique10	*G. g. domesticus*	Dominique	NorthAmerica
SAMEA6529939	NewHamp1327	*G. g. domesticus*	New Hampshire	NorthAmerica
SAMEA6529993	NewHamp85780	*G. g. domesticus*	New Hampshire	NorthAmerica
SAMN02486160	RhodeRed1	*G. g. domesticus*	Rhode Island Red	NorthAmerica
SAMEA6529963	RhodeRed2002	*G. g. domesticus*	Rhode Island Red	NorthAmerica
SAMN14814746	MalaysianMuseum1	*G. g. domesticus*	Temerloh	NorthAmerica
SAMN12236790	WhitePlymouthRock10	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236789	WhitePlymouthRock11	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236784	WhitePlymouthRock12	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236783	WhitePlymouthRock13	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236786	WhitePlymouthRock14	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236785	WhitePlymouthRock15	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236753	WhitePlymouthRock16	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236781	WhitePlymouthRock17	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236782	WhitePlymouthRock18	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236779	WhitePlymouthRock19	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236780	WhitePlymouthRock20	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236777	WhitePlymouthRock21	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236778	WhitePlymouthRock22	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236775	WhitePlymouthRock23	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236776	WhitePlymouthRock24	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236773	WhitePlymouthRock25	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236774	WhitePlymouthRock26	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236758	WhitePlymouthRock27	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236760	WhitePlymouthRock28	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236759	WhitePlymouthRock29	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236763	WhitePlymouthRock3	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236754	WhitePlymouthRock30	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236811	WhitePlymouthRock31	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236806	WhitePlymouthRock32	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236805	WhitePlymouthRock33	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236804	WhitePlymouthRock34	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236803	WhitePlymouthRock35	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236810	WhitePlymouthRock36	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236809	WhitePlymouthRock37	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236808	WhitePlymouthRock38	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236807	WhitePlymouthRock39	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236772	WhitePlymouthRock4	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236757	WhitePlymouthRock40	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236756	WhitePlymouthRock41	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236755	WhitePlymouthRock42	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236762	WhitePlymouthRock43	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236761	WhitePlymouthRock44	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236797	WhitePlymouthRock45	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236798	WhitePlymouthRock46	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236799	WhitePlymouthRock47	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236800	WhitePlymouthRock48	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236793	WhitePlymouthRock49	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236771	WhitePlymouthRock5	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236794	WhitePlymouthRock50	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236795	WhitePlymouthRock51	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236796	WhitePlymouthRock52	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236801	WhitePlymouthRock53	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236802	WhitePlymouthRock54	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236770	WhitePlymouthRock55	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236769	WhitePlymouthRock56	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236768	WhitePlymouthRock57	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236767	WhitePlymouthRock58	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236766	WhitePlymouthRock59	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236792	WhitePlymouthRock6	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236765	WhitePlymouthRock60	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236764	WhitePlymouthRock61	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236791	WhitePlymouthRock7	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236788	WhitePlymouthRock8	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN12236787	WhitePlymouthRock9	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN02486167	WhitePlymouthRock1	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN02486167	WhitePlymouthRock2	*G. g. domesticus*	White Plymouth Rock	NorthAmerica
SAMN03292403	Kauai01	*G. g. domesticus*	Polynesian Feral	Polynesia

Accession	ID	Species	Breed/Location	Population
SAMN03292413	Kauai02	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03281075	Kauai03	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292399	Kauai04	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292400	Kauai05	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292401	Kauai06	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292402	Kauai07	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03281076	Kauai08	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292393	Kauai09	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292394	Kauai10	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292395	Kauai11	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292396	Kauai12	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292397	Kauai13	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292398	Kauai14	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292407	Kauai15	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292408	Kauai16	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292405	Kauai17	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292411	Kauai18	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292412	Kauai19	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292404	Kauai20	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292409	Kauai21	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292410	Kauai22	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292406	Kauai23	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03292414	Kauai24	*G. g. domesticus*	Polynesian Feral	Polynesia
SAMN03940099	Araucana2096	*G. g. domesticus*	Araucana	SouthAmerica
SAMN03940100	Araucana2097	*G. g. domesticus*	Araucana	SouthAmerica
SAMN03940098	Araucana2098	*G. g. domesticus*	Araucana	SouthAmerica
SAMEA6529965	Araucana2095	*G. g. domesticus*	Rumpless Araucana	SouthAmerica
SAMEA6529994	Araucana86	*G. g. domesticus*	Rumpless Araucana	SouthAmerica
39S_L6_I053	Afghan1	*G. g. domesticus*	Afghanistan	SouthAsia
ypt943	Afghan10	*G. g. domesticus*	Afghanistan	SouthAsia
ypt944_L3_I030	Afghan11	*G. g. domesticus*	Afghanistan	SouthAsia
40S_L2_I054	Afghan2	*G. g. domesticus*	Afghanistan	SouthAsia
41S_L5_I045	Afghan3	*G. g. domesticus*	Afghanistan	SouthAsia
42S_L3_I055	Afghan4	*G. g. domesticus*	Afghanistan	SouthAsia
43S	Afghan5	*G. g. domesticus*	Afghanistan	SouthAsia
ypt939_L5_I025	Afghan6	*G. g. domesticus*	Afghanistan	SouthAsia
ypt940_L4_I026	Afghan7	*G. g. domesticus*	Afghanistan	SouthAsia
ypt941_L5_I027	Afghan8	*G. g. domesticus*	Afghanistan	SouthAsia
ypt942_L7_I028	Afghan9	*G. g. domesticus*	Afghanistan	SouthAsia
Xcelris_14A	India37	*G. g. domesticus*	Andhra	SouthAsia
Xcelris_14B	India38	*G. g. domesticus*	Andhra	SouthAsia
SAMEA6529966	Aseel2133	*G. g. domesticus*	Aseel	SouthAsia
Xcelris_8A	India50	*G. g. domesticus*	Assam	SouthAsia
Xcelris_8B	India51	*G. g. domesticus*	Assam	SouthAsia
Xcelris_97	India52	*G. g. domesticus*	Bihar	SouthAsia
ypt222_L3_I027	Bengal10	*G. g. domesticus*	Central Bangladesh	SouthAsia
ypt217_L2_I024	Bengal6	*G. g. domesticus*	Central Bangladesh	SouthAsia
ypt219_L3_I025	Bengal7	*G. g. domesticus*	Central Bangladesh	SouthAsia
ypt220_L2_I026	Bengal8	*G. g. domesticus*	Central Bangladesh	SouthAsia
ypt221_L3_I042	Bengal9	*G. g. domesticus*	Central Bangladesh	SouthAsia
ypt2144_L6_I055	Srilanka5	*G. g. domesticus*	Dematagama	SouthAsia
ypt2145_L6_I056	Srilanka6	*G. g. domesticus*	Dematagama	SouthAsia
ypt2146_L6_I001	Srilanka7	*G. g. domesticus*	Dematagama	SouthAsia
Xcelris_4A	India44	*G. g. domesticus*	Gujarat	SouthAsia
Xcelris_4B	India45	*G. g. domesticus*	Gujarat	SouthAsia
India1	India1	*G. g. domesticus*	India	SouthAsia
India27	India10	*G. g. domesticus*	India	SouthAsia
India29	India11	*G. g. domesticus*	India	SouthAsia
india3	India12	*G. g. domesticus*	India	SouthAsia
India31	India13	*G. g. domesticus*	India	SouthAsia
India33	India14	*G. g. domesticus*	India	SouthAsia
India35	India15	*G. g. domesticus*	India	SouthAsia
India37	India16	*G. g. domesticus*	India	SouthAsia
India39	India17	*G. g. domesticus*	India	SouthAsia
India41	India18	*G. g. domesticus*	India	SouthAsia
India43	India19	*G. g. domesticus*	India	SouthAsia
India11	India2	*G. g. domesticus*	India	SouthAsia
India45	India20	*G. g. domesticus*	India	SouthAsia
India47	India21	*G. g. domesticus*	India	SouthAsia
india49	India22	*G. g. domesticus*	India	SouthAsia
India5	India23	*G. g. domesticus*	India	SouthAsia

Accession	ID	Species	Breed/Location	Population
India51	India24	*G. g. domesticus*	India	SouthAsia
India55	India25	*G. g. domesticus*	India	SouthAsia
India57	India26	*G. g. domesticus*	India	SouthAsia
India59	India27	*G. g. domesticus*	India	SouthAsia
India61	India28	*G. g. domesticus*	India	SouthAsia
India7	India29	*G. g. domesticus*	India	SouthAsia
India13	India3	*G. g. domesticus*	India	SouthAsia
India9	India30	*G. g. domesticus*	India	SouthAsia
Xcelris_10A	India31	*G. g. domesticus*	India	SouthAsia
Xcelris_10B	India32	*G. g. domesticus*	India	SouthAsia
India15	India4	*G. g. domesticus*	India	SouthAsia
India17	India5	*G. g. domesticus*	India	SouthAsia
India19	India6	*G. g. domesticus*	India	SouthAsia
India21	India7	*G. g. domesticus*	India	SouthAsia
India23	India8	*G. g. domesticus*	India	SouthAsia
India25	India9	*G. g. domesticus*	India	SouthAsia
SAMEA6529968	IndGame2190	*G. g. domesticus*	Indian Game	SouthAsia
Xcelris_11B	India33	*G. g. domesticus*	Jabalpur	SouthAsia
Xcelris_12A	India35	*G. g. domesticus*	Kashmir	SouthAsia
Xcelris_13B	India36	*G. g. domesticus*	Maharashtra	SouthAsia
Xcelris_2A	India41	*G. g. domesticus*	Nicobar	SouthAsia
Xcelris_2B	India42	*G. g. domesticus*	Nicobar	SouthAsia
ypt169_L3_I011	Bengal1	*G. g. domesticus*	North Bangladesh	SouthAsia
ypt170_L2_I012	Bengal2	*G. g. domesticus*	North Bangladesh	SouthAsia
ypt171_L2_I021	Bengal3	*G. g. domesticus*	North Bangladesh	SouthAsia
ypt173_L2_I022	Bengal4	*G. g. domesticus*	North Bangladesh	SouthAsia
ypt175_L2_I023	Bengal5	*G. g. domesticus*	North Bangladesh	SouthAsia
77S_L4_I029	Pakistan1	*G. g. domesticus*	Northwest Parkistan	SouthAsia
78S_L4_I030	Pakistan2	*G. g. domesticus*	Northwest Parkistan	SouthAsia
79S_L2_I031	Pakistan3	*G. g. domesticus*	Northwest Parkistan	SouthAsia
ypt971	Pakistan36	*G. g. domesticus*	Northwest Parkistan	SouthAsia
ypt972x	Pakistan37	*G. g. domesticus*	Northwest Parkistan	SouthAsia
80S_L3_I004	Pakistan4	*G. g. domesticus*	Northwest Parkistan	SouthAsia
81S_L1_I005	Pakistan5	*G. g. domesticus*	Northwest Parkistan	SouthAsia
ypt2092	Pakistan10	*G. g. domesticus*	Pakistan	SouthAsia
ypt2093	Pakistan11	*G. g. domesticus*	Pakistan	SouthAsia
ypt2094	Pakistan12	*G. g. domesticus*	Pakistan	SouthAsia
ypt2095	Pakistan13	*G. g. domesticus*	Pakistan	SouthAsia
ypt2098	Pakistan14	*G. g. domesticus*	Pakistan	SouthAsia
ypt2100	Pakistan15	*G. g. domesticus*	Pakistan	SouthAsia
ypt2101	Pakistan16	*G. g. domesticus*	Pakistan	SouthAsia
ypt2102	Pakistan17	*G. g. domesticus*	Pakistan	SouthAsia
ypt2103	Pakistan18	*G. g. domesticus*	Pakistan	SouthAsia
ypt2105	Pakistan19	*G. g. domesticus*	Pakistan	SouthAsia
ypt2106	Pakistan20	*G. g. domesticus*	Pakistan	SouthAsia
ypt2107	Pakistan21	*G. g. domesticus*	Pakistan	SouthAsia
ypt2108	Pakistan22	*G. g. domesticus*	Pakistan	SouthAsia
ypt2109	Pakistan23	*G. g. domesticus*	Pakistan	SouthAsia
ypt2110	Pakistan24	*G. g. domesticus*	Pakistan	SouthAsia
ypt2111	Pakistan25	*G. g. domesticus*	Pakistan	SouthAsia
ypt2112	Pakistan26	*G. g. domesticus*	Pakistan	SouthAsia
ypt2114	Pakistan27	*G. g. domesticus*	Pakistan	SouthAsia
ypt2115	Pakistan28	*G. g. domesticus*	Pakistan	SouthAsia
ypt2116	Pakistan29	*G. g. domesticus*	Pakistan	SouthAsia
ypt2117	Pakistan30	*G. g. domesticus*	Pakistan	SouthAsia
ypt2118	Pakistan31	*G. g. domesticus*	Pakistan	SouthAsia
ypt2119	Pakistan32	*G. g. domesticus*	Pakistan	SouthAsia
ypt2120	Pakistan33	*G. g. domesticus*	Pakistan	SouthAsia
ypt2121	Pakistan34	*G. g. domesticus*	Pakistan	SouthAsia
ypt2122	Pakistan35	*G. g. domesticus*	Pakistan	SouthAsia
ypt2082	Pakistan6	*G. g. domesticus*	Pakistan	SouthAsia
ypt2086	Pakistan7	*G. g. domesticus*	Pakistan	SouthAsia
ypt2087	Pakistan8	*G. g. domesticus*	Pakistan	SouthAsia
ypt2091	Pakistan9	*G. g. domesticus*	Pakistan	SouthAsia
Xcelris_6A	India48	*G. g. domesticus*	Punjab	SouthAsia
Xcelris_6B	India49	*G. g. domesticus*	Punjab	SouthAsia
SAMN08979078	Saudi1	*G. g. domesticus*	Saudi-Arabia	SouthAsia
SAMN08979079	Saudi2	*G. g. domesticus*	Saudi-Arabia	SouthAsia
SAMN08979080	Saudi3	*G. g. domesticus*	Saudi-Arabia	SouthAsia
SAMN08979081	Saudi4	*G. g. domesticus*	Saudi-Arabia	SouthAsia
SAMN08979082	Saudi5	*G. g. domesticus*	Saudi-Arabia	SouthAsia

Accession	ID	Species	Breed/Location	Population
ypt258_L3_I028	Bengal11	*G. g. domesticus*	South Bangladesh	SouthAsia
ypt259_L2_I029	Bengal12	*G. g. domesticus*	South Bangladesh	SouthAsia
ypt262_L3_I030	Bengal13	*G. g. domesticus*	South Bangladesh	SouthAsia
ypt263_L3_I031	Bengal14	*G. g. domesticus*	South Bangladesh	SouthAsia
ypt264_L3_I032	Bengal15	*G. g. domesticus*	South Bangladesh	SouthAsia
SAMN08979083	SriLanka100	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979084	SriLanka110	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979085	SriLanka120	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979086	SriLanka130	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979087	SriLanka150	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979088	SriLanka160	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979089	SriLanka20	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979090	SriLanka30	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979091	SriLanka40	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979092	SriLanka50	*G. g. domesticus*	Sri Lanka	SouthAsia
SAMN08979093	SriLanka90	*G. g. domesticus*	Sri Lanka	SouthAsia
ypt2124_L4_I052	Srilanka1	*G. g. domesticus*	Tabbowa	SouthAsia
ypt2126_L3_I053	Srilanka2	*G. g. domesticus*	Tabbowa	SouthAsia
ypt2127_L1_I036	Srilanka3	*G. g. domesticus*	Tabbowa	SouthAsia
ypt2128_L6_I054	Srilanka4	*G. g. domesticus*	Tabbowa	SouthAsia
Xcelris_127	India34	*G. g. domesticus*	Uttar	SouthAsia
Xcelris_394	India43	*G. g. domesticus*	Uttarakhand	SouthAsia
Xcelris_5A	India46	*G. g. domesticus*	Vellore	SouthAsia
Xcelris_5B	India47	*G. g. domesticus*	Vellore	SouthAsia
Xcelris_1A	India39	*G. g. domesticus*	West Bengal	SouthAsia
Xcelris_1B	India40	*G. g. domesticus*	West Bengal	SouthAsia
74S_L4_I026	Iran100	*G. g. domesticus*	Esfahan	WestAsia
ypt957	Iran163	*G. g. domesticus*	Esfahan	WestAsia
ypt958	Iran164	*G. g. domesticus*	Esfahan	WestAsia
ypt959	Iran165	*G. g. domesticus*	Esfahan	WestAsia
ypt960	Iran166	*G. g. domesticus*	Esfahan	WestAsia
ypt961	Iran167	*G. g. domesticus*	Esfahan	WestAsia
67S_L1_I011	Iran92	*G. g. domesticus*	Esfahan	WestAsia
69S	Iran94	*G. g. domesticus*	Esfahan	WestAsia
70S	Iran96	*G. g. domesticus*	Esfahan	WestAsia
71S_L3_I023	Iran97	*G. g. domesticus*	Esfahan	WestAsia
72S_L1_I024	Iran98	*G. g. domesticus*	Esfahan	WestAsia
73S_L1_I025	Iran99	*G. g. domesticus*	Esfahan	WestAsia
124S_L2_I048	Iran26	*G. g. domesticus*	Isfahan	WestAsia
125S_L3_I050	Iran27	*G. g. domesticus*	Isfahan	WestAsia
126S_L3_I051	Iran28	*G. g. domesticus*	Isfahan	WestAsia
64S_L4_I001	Iran89	*G. g. domesticus*	Isfahan	WestAsia
65S_L5_I002	Iran90	*G. g. domesticus*	Isfahan	WestAsia
66S_L2_I003	Iran91	*G. g. domesticus*	Isfahan	WestAsia
68S_L3_I012	Iran93	*G. g. domesticus*	Isfahan	WestAsia
28S_L4_I035	Iran55	*G. g. domesticus*	Khazak	WestAsia
29S_L3_I036	Iran56	*G. g. domesticus*	Khazak	WestAsia
ypt912	Iran126	*G. g. domesticus*	Mahhad	WestAsia
ypt913	Iran127	*G. g. domesticus*	Mahhad	WestAsia
ypt915	Iran128	*G. g. domesticus*	Mahhad	WestAsia
ypt916	Iran129	*G. g. domesticus*	Mahhad	WestAsia
ypt917	Iran130	*G. g. domesticus*	Mahhad	WestAsia
35S_L5_I041	Iran62	*G. g. domesticus*	Mahhad	WestAsia
ypt951	Iran157	*G. g. domesticus*	Marand	WestAsia
ypt952	Iran158	*G. g. domesticus*	Marand	WestAsia
ypt953	Iran159	*G. g. domesticus*	Marand	WestAsia
ypt954	Iran160	*G. g. domesticus*	Marand	WestAsia
ypt955	Iran161	*G. g. domesticus*	Marand	WestAsia
ypt956x	Iran162	*G. g. domesticus*	Marand	WestAsia
ypt973_L4_I037	Iran173	*G. g. domesticus*	Marand	WestAsia
ypt974	Iran174	*G. g. domesticus*	Marand	WestAsia
ypt975_L8_I039	Iran175	*G. g. domesticus*	Marand	WestAsia
ypt976_L7_I040	Iran176	*G. g. domesticus*	Marand	WestAsia
ypt977_L2_I041	Iran177	*G. g. domesticus*	Marand	WestAsia
49S_L1_I005	Iran72	*G. g. domesticus*	Marand	WestAsia
50S_L1_I047	Iran74	*G. g. domesticus*	Marand	WestAsia
51S_L6_I048	Iran75	*G. g. domesticus*	Marand	WestAsia
52S_L2_I050	Iran76	*G. g. domesticus*	Marand	WestAsia
53S_L2_I006	Iran77	*G. g. domesticus*	Marand	WestAsia
54S_L2_I007	Iran78	*G. g. domesticus*	Marand	WestAsia
55S_L1_I008	Iran79	*G. g. domesticus*	Marand	WestAsia

Accession	ID	Species	Breed/Location	Population
56S_L1_I010	Iran80	*G. g. domesticus*	Marand	WestAsia
yp914ypt	Iran111	*G. g. domesticus*	Mashhad	WestAsia
122S_L4_I043	Iran24	*G. g. domesticus*	Mashhad	WestAsia
123S_L4_I044	Iran25	*G. g. domesticus*	Mashhad	WestAsia
30S_L2_I037	Iran58	*G. g. domesticus*	Mashhad	WestAsia
31S_L3_I038	Iran59	*G. g. domesticus*	Mashhad	WestAsia
32S_L1_I039	Iran60	*G. g. domesticus*	Mashhad	WestAsia
33S	Iran61	*G. g. domesticus*	Mashhad	WestAsia
36S_L5_I042	Iran63	*G. g. domesticus*	Mashhad	WestAsia
37S_L2_I043	Iran64	*G. g. domesticus*	Mashhad	WestAsia
38S_L6_I044	Iran65	*G. g. domesticus*	Mashhad	WestAsia
100S	Iran1	*G. g. domesticus*	Sari	WestAsia
109S_L5_I033	Iran10	*G. g. domesticus*	Sari	WestAsia
96S_L4_I026	Iran106	*G. g. domesticus*	Sari	WestAsia
97S_L8_I027	Iran107	*G. g. domesticus*	Sari	WestAsia
98S_L3_I034	Iran108	*G. g. domesticus*	Sari	WestAsia
99S_L2_I035	Iran109	*G. g. domesticus*	Sari	WestAsia
ypt896_L2_I047	Iran114	*G. g. domesticus*	Sari	WestAsia
ypt897_L3_I048	Iran115	*G. g. domesticus*	Sari	WestAsia
ypt898	Iran116	*G. g. domesticus*	Sari	WestAsia
ypt899_L6_I051	Iran117	*G. g. domesticus*	Sari	WestAsia
ypt900_L2_I050	Iran118	*G. g. domesticus*	Sari	WestAsia
ypt901_L1_I052	Iran119	*G. g. domesticus*	Sari	WestAsia
110S_L4_I034	Iran12	*G. g. domesticus*	Sari	WestAsia
111S_L3_I047	Iran13	*G. g. domesticus*	Sari	WestAsia
ypt918_L3_I007	Iran131	*G. g. domesticus*	Sari	WestAsia
ypt919_L4_I008	Iran132	*G. g. domesticus*	Sari	WestAsia
ypt920_L5_I010	Iran133	*G. g. domesticus*	Sari	WestAsia
ypt921_L8_I011	Iran134	*G. g. domesticus*	Sari	WestAsia
ypt922_L4_I012	Iran135	*G. g. domesticus*	Sari	WestAsia
ypt923_L2_I051	Iran136	*G. g. domesticus*	Sari	WestAsia
112S_L3_I035	Iran14	*G. g. domesticus*	Sari	WestAsia
ypt933	Iran145	*G. g. domesticus*	Sari	WestAsia
ypt934x	Iran146	*G. g. domesticus*	Sari	WestAsia
ypt935	Iran147	*G. g. domesticus*	Sari	WestAsia
ypt936	Iran148	*G. g. domesticus*	Sari	WestAsia
ypt937	Iran149	*G. g. domesticus*	Sari	WestAsia
113S_L1_I036	Iran15	*G. g. domesticus*	Sari	WestAsia
ypt938x	Iran150	*G. g. domesticus*	Sari	WestAsia
ypt945_L4_I031	Iran151	*G. g. domesticus*	Sari	WestAsia
ypt946_L4_I032	Iran152	*G. g. domesticus*	Sari	WestAsia
ypt947_L5_I033	Iran153	*G. g. domesticus*	Sari	WestAsia
ypt948_L2_I034	Iran154	*G. g. domesticus*	Sari	WestAsia
ypt949_L3_I035	Iran155	*G. g. domesticus*	Sari	WestAsia
ypt950_L5_I036	Iran156	*G. g. domesticus*	Sari	WestAsia
114S_L3_I037	Iran16	*G. g. domesticus*	Sari	WestAsia
116S_L6_I038	Iran17	*G. g. domesticus*	Sari	WestAsia
117S_L2_I039	Iran18	*G. g. domesticus*	Sari	WestAsia
118S_L4_I040	Iran19	*G. g. domesticus*	Sari	WestAsia
101S_L1_I028	Iran2	*G. g. domesticus*	Sari	WestAsia
119S_L3_I041	Iran20	*G. g. domesticus*	Sari	WestAsia
120S_L1_I041	Iran22	*G. g. domesticus*	Sari	WestAsia
121S_L3_I042	Iran23	*G. g. domesticus*	Sari	WestAsia
102S_L3_I037	Iran3	*G. g. domesticus*	Sari	WestAsia
103S	Iran4	*G. g. domesticus*	Sari	WestAsia
104S_L7_I029	Iran5	*G. g. domesticus*	Sari	WestAsia
105S_L3_I039	Iran6	*G. g. domesticus*	Sari	WestAsia
44S_L4_I001	Iran67	*G. g. domesticus*	Sari	WestAsia
45S_L2_I002	Iran68	*G. g. domesticus*	Sari	WestAsia
46S_L4_I003	Iran69	*G. g. domesticus*	Sari	WestAsia
106S_L6_I030	Iran7	*G. g. domesticus*	Sari	WestAsia
47S_L3_I046	Iran70	*G. g. domesticus*	Sari	WestAsia
48S_L1_I004	Iran71	*G. g. domesticus*	Sari	WestAsia
107S_L2_I031	Iran8	*G. g. domesticus*	Sari	WestAsia
57S_L3_I046	Iran81	*G. g. domesticus*	Sari	WestAsia
58S_L4_I051	Iran82	*G. g. domesticus*	Sari	WestAsia
59S_L5_I052	Iran83	*G. g. domesticus*	Sari	WestAsia
60S_L5_I053	Iran85	*G. g. domesticus*	Sari	WestAsia
61S_L8_I054	Iran86	*G. g. domesticus*	Sari	WestAsia
62S_L8_I055	Iran87	*G. g. domesticus*	Sari	WestAsia
63S_L5_I056	Iran88	*G. g. domesticus*	Sari	WestAsia

Accession	ID	Species	Breed/Location	Population
108S_L6_I032	Iran9	*G. g. domesticus*	Sari	WestAsia
82S	Iran104	*G. g. domesticus*	Shiraz	WestAsia
83S_L3_I033	Iran105	*G. g. domesticus*	Shiraz	WestAsia
ypt909_L5_I004	Iran123	*G. g. domesticus*	Shiraz	WestAsia
ypt910_L3_I005	Iran124	*G. g. domesticus*	Shiraz	WestAsia
ypt911_L6_I006	Iran125	*G. g. domesticus*	Shiraz	WestAsia
ypt924_L3_I021	Iran137	*G. g. domesticus*	Shiraz	WestAsia
ypt925_L2_I022	Iran138	*G. g. domesticus*	Shiraz	WestAsia
ypt926_L2_I023	Iran139	*G. g. domesticus*	Shiraz	WestAsia
127S_L2_I045	Iran29	*G. g. domesticus*	Shiraz	WestAsia
128S_L1_I042	Iran30	*G. g. domesticus*	Shiraz	WestAsia
129S_L3_I043	Iran31	*G. g. domesticus*	Shiraz	WestAsia
130S_L4_I044	Iran33	*G. g. domesticus*	Shiraz	WestAsia
17S_L3_I024	Iran43	*G. g. domesticus*	Shiraz	WestAsia
18S_L7_I025	Iran44	*G. g. domesticus*	Shiraz	WestAsia
19S_L4_I026	Iran45	*G. g. domesticus*	Shiraz	WestAsia
20S_L7_I027	Iran47	*G. g. domesticus*	Shiraz	WestAsia
ypt906_L3_I001	Iran120	*G. g. domesticus*	Yazd	WestAsia
ypt907_L6_I002	Iran121	*G. g. domesticus*	Yazd	WestAsia
ypt908_L1_I003	Iran122	*G. g. domesticus*	Yazd	WestAsia
ypt962	Iran168	*G. g. domesticus*	Yazd	WestAsia
ypt963	Iran169	*G. g. domesticus*	Yazd	WestAsia
ypt964	Iran170	*G. g. domesticus*	Yazd	WestAsia
ypt965	Iran171	*G. g. domesticus*	Yazd	WestAsia
ypt966	Iran172	*G. g. domesticus*	Yazd	WestAsia
131S_L2_I045	Iran34	*G. g. domesticus*	Yazd	WestAsia
132S_L2_I046	Iran35	*G. g. domesticus*	Yazd	WestAsia
133S_L8_I046	Iran36	*G. g. domesticus*	Yazd	WestAsia
134S_L2_I047	Iran37	*G. g. domesticus*	Yazd	WestAsia
135S_L3_I048	Iran38	*G. g. domesticus*	Yazd	WestAsia
75S_L3_I027	Iran101	*G. g. domesticus*	Zabol	WestAsia
76S_L1_I028	Iran102	*G. g. domesticus*	Zabol	WestAsia
ypt890	Iran112	*G. g. domesticus*	Zabol	WestAsia
ypt891	Iran113	*G. g. domesticus*	Zabol	WestAsia
ypt978_L2_I052	Iran178	*G. g. domesticus*	Zabol	WestAsia
ypt979_L6_I042	Iran179	*G. g. domesticus*	Zabol	WestAsia
ypt980_L3_I043	Iran180	*G. g. domesticus*	Zabol	WestAsia
ypt981_L2_I044	Iran181	*G. g. domesticus*	Zabol	WestAsia
ypt982_L4_I045	Iran182	*G. g. domesticus*	Zabol	WestAsia
16S_L5_I023	Iran42	*G. g. domesticus*	Zabol	WestAsia
21S_L6_I028	Iran48	*G. g. domesticus*	Zabol	WestAsia
22S_L6_I029	Iran49	*G. g. domesticus*	Zabol	WestAsia
23S_L2_I030	Iran50	*G. g. domesticus*	Zabol	WestAsia
24S_L4_I031	Iran51	*G. g. domesticus*	Zabol	WestAsia
25S_L3_I032	Iran52	*G. g. domesticus*	Zabol	WestAsia
26S	Iran53	*G. g. domesticus*	Zabol	WestAsia
27S_L4_I034	Iran54	*G. g. domesticus*	Zabol	WestAsia
7S_L4_I007	Iran103	*G. g. domesticus*	Zahedan	WestAsia
10S_L1_I052	Iran11	*G. g. domesticus*	Zahedan	WestAsia
9S	Iran110	*G. g. domesticus*	Zahedan	WestAsia
ypt928	Iran140	*G. g. domesticus*	Zahedan	WestAsia
ypt929	Iran141	*G. g. domesticus*	Zahedan	WestAsia
ypt930	Iran142	*G. g. domesticus*	Zahedan	WestAsia
ypt931	Iran143	*G. g. domesticus*	Zahedan	WestAsia
ypt932	Iran144	*G. g. domesticus*	Zahedan	WestAsia
11S	Iran21	*G. g. domesticus*	Zahedan	WestAsia
12S_L4_I011	Iran32	*G. g. domesticus*	Zahedan	WestAsia
13S_L7_I012	Iran39	*G. g. domesticus*	Zahedan	WestAsia
14S_L2_I021	Iran40	*G. g. domesticus*	Zahedan	WestAsia
15S_L6_I022	Iran41	*G. g. domesticus*	Zahedan	WestAsia
1S	Iran46	*G. g. domesticus*	Zahedan	WestAsia
2S	Iran57	*G. g. domesticus*	Zahedan	WestAsia
3S_L4_I003	Iran66	*G. g. domesticus*	Zahedan	WestAsia
4S_L4_I004	Iran73	*G. g. domesticus*	Zahedan	WestAsia
5S	Iran84	*G. g. domesticus*	Zahedan	WestAsia
6S	Iran95	*G. g. domesticus*	Zahedan	WestAsia
SAMN02712042	RJF91	*G. g. gallus*	Hainan	gallus
SAMEA5930917	Gg.1	*G. g. gallus*	RJF	gallus
SAMN14814753	RJF17	*G. g. gallus*	RJF	gallus
SAMN14814754	RJF18	*G. g. gallus*	RJF	gallus
SAMN14814818	RJF19	*G. g. gallus*	RJF	gallus

Accession	ID	Species	Breed/Location	Population
SAMN14814817	RJF20	*G. g. gallus*	RJF	gallus
SAMN14814816	RJF21	*G. g. gallus*	RJF	gallus
SAMN14814815	RJF22	*G. g. gallus*	RJF	gallus
SAMN14814752	RJF23	*G. g. gallus*	RJF	gallus
SAMN14814812	RJF24	*G. g. gallus*	RJF	gallus
SAMN14814811	RJF25	*G. g. gallus*	RJF	gallus
SAMN14814810	RJF26	*G. g. gallus*	RJF	gallus
SAMN14814809	RJF27	*G. g. gallus*	RJF	gallus
SAMN14814808	RJF28	*G. g. gallus*	RJF	gallus
SAMN14814807	RJF29	*G. g. gallus*	RJF	gallus
SAMN14814806	RJF30	*G. g. gallus*	RJF	gallus
SAMN14814805	RJF31	*G. g. gallus*	RJF	gallus
SAMN14814751	RJF32	*G. g. gallus*	RJF	gallus
SAMN14814804	RJF33	*G. g. gallus*	RJF	gallus
SAMN14814803	RJF34	*G. g. gallus*	RJF	gallus
SAMN14814802	RJF35	*G. g. gallus*	RJF	gallus
SAMN14814801	RJF36	*G. g. gallus*	RJF	gallus
SAMN14814800	RJF37	*G. g. gallus*	RJF	gallus
SAMN14814799	RJF38	*G. g. gallus*	RJF	gallus
SAMN14814798	RJF39	*G. g. gallus*	RJF	gallus
SAMN14814797	RJF40	*G. g. gallus*	RJF	gallus
SAMN14814796	RJF41	*G. g. gallus*	RJF	gallus
SAMN14814795	RJF42	*G. g. gallus*	RJF	gallus
SAMN14814750	RJF43	*G. g. gallus*	RJF	gallus
SAMN14814794	RJF44	*G. g. gallus*	RJF	gallus
SAMN14814793	RJF45	*G. g. gallus*	RJF	gallus
SAMN14814792	RJF46	*G. g. gallus*	RJF	gallus
SAMN14814791	RJF47	*G. g. gallus*	RJF	gallus
SAMN14814790	RJF48	*G. g. gallus*	RJF	gallus
SAMN14814789	RJF49	*G. g. gallus*	RJF	gallus
SAMN14814788	RJF50	*G. g. gallus*	RJF	gallus
SAMN14814787	RJF51	*G. g. gallus*	RJF	gallus
SAMN14814786	RJF52	*G. g. gallus*	RJF	gallus
SAMN14814785	RJF53	*G. g. gallus*	RJF	gallus
SAMN14814749	RJF54	*G. g. gallus*	RJF	gallus
SAMN14814784	RJF55	*G. g. gallus*	RJF	gallus
SAMN14814783	RJF56	*G. g. gallus*	RJF	gallus
SAMN14814782	RJF57	*G. g. gallus*	RJF	gallus
SAMN14814781	RJF58	*G. g. gallus*	RJF	gallus
SAMN14814780	RJF59	*G. g. gallus*	RJF	gallus
SAMN14814779	RJF60	*G. g. gallus*	RJF	gallus
SAMN14814778	RJF61	*G. g. gallus*	RJF	gallus
SAMN14814777	RJF62	*G. g. gallus*	RJF	gallus
SAMN14814776	RJF63	*G. g. gallus*	RJF	gallus
SAMN14814748	RJF64	*G. g. gallus*	RJF	gallus
SAMN14814775	RJF65	*G. g. gallus*	RJF	gallus
SAMN14814774	RJF66	*G. g. gallus*	RJF	gallus
SAMN14814773	RJF67	*G. g. gallus*	RJF	gallus
SAMN14814772	RJF68	*G. g. gallus*	RJF	gallus
SAMN14814771	RJF69	*G. g. gallus*	RJF	gallus
SAMN14814770	RJF70	*G. g. gallus*	RJF	gallus
SAMN14814769	RJF71	*G. g. gallus*	RJF	gallus
SAMN14814768	RJF72	*G. g. gallus*	RJF	gallus
SAMN14814767	RJF73	*G. g. gallus*	RJF	gallus
SAMN14814766	RJF74	*G. g. gallus*	RJF	gallus
SAMN14814765	RJF75	*G. g. gallus*	RJF	gallus
SAMN14814747	RJF76	*G. g. gallus*	RJF	gallus
SAMN14814764	RJF77	*G. g. gallus*	RJF	gallus
SAMN14814763	RJF78	*G. g. gallus*	RJF	gallus
SAMN14814762	RJF79	*G. g. gallus*	RJF	gallus
SAMN14814761	RJF80	*G. g. gallus*	RJF	gallus
SAMN14814760	RJF81	*G. g. gallus*	RJF	gallus
SAMN14814759	RJF82	*G. g. gallus*	RJF	gallus
SAMN14814758	RJF83	*G. g. gallus*	RJF	gallus
SAMN14814757	RJF84	*G. g. gallus*	RJF	gallus
SAMN14814756	RJF85	*G. g. gallus*	RJF	gallus
SAMN14814755	RJF86	*G. g. gallus*	RJF	gallus
SAMN02486161	RJF87	*G. g. gallus*	RJF	gallus
BKL2	Red_g.1	*G. g. gallus*	Sumatra	gallus
ypt2813	Red_g.2	*G. g. gallus*	Sumatra	gallus
ypt2814	Red_g.3	*G. g. gallus*	Sumatra	gallus

Accession	ID	Species	Breed/Location	Population
ypt2815	Red_g.4	*G. g. gallus*	Sumatra	gallus
ypt2816	Red_g.5	*G. g. gallus*	Sumatra	gallus
ypt2817	Red_g.6	*G. g. gallus*	Sumatra	gallus
SAMN02712039	RJF88	*G. g. gallus*	Yunnan	gallus
SAMN02712040	RJF89	*G. g. gallus*	Yunnan	gallus
SAMN02712041	RJF90	*G. g. gallus*	Yunnan	gallus
SAMN02712043	RJF92	*G. g. gallus*	Yunnan	gallus
ypt2781	Red_jabouillei1	*G. g. jabouillei*	Guangxi	jabouillei
ypt2790	Red_jabouillei10	*G. g. jabouillei*	Guangxi	jabouillei
ypt2791	Red_jabouillei11	*G. g. jabouillei*	Guangxi	jabouillei
ypt2792	Red_jabouillei12	*G. g. jabouillei*	Guangxi	jabouillei
ypt2799	Red_jabouillei13	*G. g. jabouillei*	Guangxi	jabouillei
ypt2800	Red_jabouillei14	*G. g. jabouillei*	Guangxi	jabouillei
ypt2821	Red_jabouillei15	*G. g. jabouillei*	Guangxi	jabouillei
ypt2822	Red_jabouillei16	*G. g. jabouillei*	Guangxi	jabouillei
ypt2823	Red_jabouillei17	*G. g. jabouillei*	Guangxi	jabouillei
ypt2824	Red_jabouillei18	*G. g. jabouillei*	Guangxi	jabouillei
ypt2825	Red_jabouillei19	*G. g. jabouillei*	Guangxi	jabouillei
ypt2782	Red_jabouillei2	*G. g. jabouillei*	Guangxi	jabouillei
ypt2826	Red_jabouillei20	*G. g. jabouillei*	Guangxi	jabouillei
ypt2827	Red_jabouillei21	*G. g. jabouillei*	Guangxi	jabouillei
ypt2828	Red_jabouillei22	*G. g. jabouillei*	Guangxi	jabouillei
ypt2783	Red_jabouillei3	*G. g. jabouillei*	Guangxi	jabouillei
ypt2784	Red_jabouillei4	*G. g. jabouillei*	Guangxi	jabouillei
ypt2785	Red_jabouillei5	*G. g. jabouillei*	Guangxi	jabouillei
ypt2786	Red_jabouillei6	*G. g. jabouillei*	Guangxi	jabouillei
ypt2787	Red_jabouillei7	*G. g. jabouillei*	Guangxi	jabouillei
ypt2788	Red_jabouillei8	*G. g. jabouillei*	Guangxi	jabouillei
ypt2789	Red_jabouillei9	*G. g. jabouillei*	Guangxi	jabouillei
ypt3274A_L7_I011	Red_jabouillei23	*G. g. jabouillei*	Yunnan	jabouillei
Xcelris_13	Red_murghi12	*G. g. murghi*	Andhra	murghi
Xcelris_14	Red_murghi14	*G. g. murghi*	Andhra	murghi
Xcelris_113	Red_murghi6	*G. g. murghi*	Andhra	murghi
Xcelris_12	Red_murghi8	*G. g. murghi*	Andhra	murghi
Xcelris_101	Red_murghi2	*G. g. murghi*	Bihar	murghi
Xcelris_104	Red_murghi3	*G. g. murghi*	Bihar	murghi
Xcelris_105	Red_murghi4	*G. g. murghi*	Bihar	murghi
Xcelris_98	Red_murghi59	*G. g. murghi*	Bihar	murghi
Xcelris_MT-99	Red_murghi68	*G. g. murghi*	Bihar	murghi
Xcelris_3	Red_murghi31	*G. g. murghi*	Chattisgarh	murghi
Xcelris_5	Red_murghi52	*G. g. murghi*	Chattisgarh	murghi
Xcelris_331	Red_murghi32	*G. g. murghi*	Haryana	murghi
Xcelris_332	Red_murghi33	*G. g. murghi*	Haryana	murghi
Xcelris_333	Red_murghi34	*G. g. murghi*	Haryana	murghi
Xcelris_335	Red_murghi35	*G. g. murghi*	Haryana	murghi
Xcelris_336	Red_murghi36	*G. g. murghi*	Haryana	murghi
Xcelris_339	Red_murghi37	*G. g. murghi*	Haryana	murghi
Xcelris_340	Red_murghi38	*G. g. murghi*	Haryana	murghi
Xcelris_341	Red_murghi39	*G. g. murghi*	Haryana	murghi
Xcelris_377	Red_murghi41	*G. g. murghi*	Haryana	murghi
Xcelris_378	Red_murghi42	*G. g. murghi*	Haryana	murghi
Xcelris_379	Red_murghi43	*G. g. murghi*	Haryana	murghi
Xcelris_380	Red_murghi44	*G. g. murghi*	Haryana	murghi
Xcelris_381	Red_murghi45	*G. g. murghi*	Haryana	murghi
Xcelris_382	Red_murghi46	*G. g. murghi*	Haryana	murghi
Xcelris_383	Red_murghi47	*G. g. murghi*	Haryana	murghi
Xcelris_384	Red_murghi48	*G. g. murghi*	Haryana	murghi
Xcelris_385	Red_murghi49	*G. g. murghi*	Haryana	murghi
Xcelris_36	Red_murghi40	*G. g. murghi*	Himachal	murghi
Xcelris_44	Red_murghi51	*G. g. murghi*	Himachal	murghi
Xcelris_MT-18	Red_murghi61	*G. g. murghi*	Himachal	murghi
Xcelris_MT-38	Red_murghi62	*G. g. murghi*	Himachal	murghi
Xcelris_MT-46	Red_murghi63	*G. g. murghi*	Himachal	murghi
Xcelris_MT-48	Red_murghi64	*G. g. murghi*	Himachal	murghi
SAMEA5160225	RJFi1	*G. g. murghi*	India	murghi
SAMEA5160234	RJFi10	*G. g. murghi*	India	murghi
SAMEA5160226	RJFi2	*G. g. murghi*	India	murghi
SAMEA5160227	RJFi3	*G. g. murghi*	India	murghi
SAMEA5160228	RJFi4	*G. g. murghi*	India	murghi
SAMEA5160229	RJFi5	*G. g. murghi*	India	murghi
SAMEA5160230	RJFi6	*G. g. murghi*	India	murghi

Accession	ID	Species	Breed/Location	Population
SAMEA5160231	RJFi7	*G. g. murghi*	India	murghi
SAMEA5160232	RJFi8	*G. g. murghi*	India	murghi
SAMEA5160233	RJFi9	*G. g. murghi*	India	murghi
Xcelris_219	Red_murghi19	*G. g. murghi*	Jharkhand	murghi
Xcelris_221	Red_murghi20	*G. g. murghi*	Jharkhand	murghi
Xcelris_222	Red_murghi21	*G. g. murghi*	Jharkhand	murghi
Xcelris_224	Red_murghi22	*G. g. murghi*	Jharkhand	murghi
Xcelris_226	Red_murghi23	*G. g. murghi*	Jharkhand	murghi
Xcelris_227	Red_murghi24	*G. g. murghi*	Jharkhand	murghi
Xcelris_90	Red_murghi58	*G. g. murghi*	Kashmir	murghi
Xcelris_MT-85	Red_murghi66	*G. g. murghi*	Kashmir	murghi
Xcelris_MT-6	Red_murghi65	*G. g. murghi*	Meghalaya	murghi
Xcelris_MT-9	Red_murghi67	*G. g. murghi*	Meghalaya	murghi
Xcelris_112	Red_murghi5	*G. g. murghi*	Mizorum	murghi
Xcelris_392	Red_murghi50	*G. g. murghi*	Mizorum	murghi
Xcelris_114	Red_murghi7	*G. g. murghi*	Mizorum	murghi
Xcelris_275	Red_murghi25	*G. g. murghi*	New Delhi	murghi
Xcelris_276	Red_murghi26	*G. g. murghi*	New Delhi	murghi
Xcelris_278	Red_murghi27	*G. g. murghi*	New Delhi	murghi
Xcelris_279	Red_murghi28	*G. g. murghi*	New Delhi	murghi
Xcelris_280	Red_murghi29	*G. g. murghi*	New Delhi	murghi
Xcelris_282	Red_murghi30	*G. g. murghi*	New Delhi	murghi
Xcelris_10	Red_murghi1	*G. g. murghi*	Orissa	murghi
Xcelris_7	Red_murghi54	*G. g. murghi*	Orissa	murghi
Xcelris_8	Red_murghi56	*G. g. murghi*	Orissa	murghi
Xcelris_141	Red_murghi15	*G. g. murghi*	Tripura	murghi
Xcelris_142	Red_murghi16	*G. g. murghi*	Tripura	murghi
Xcelris_143	Red_murghi17	*G. g. murghi*	Tripura	murghi
Xcelris_144	Red_murghi18	*G. g. murghi*	Tripura	murghi
Xcelris_128	Red_murghi10	*G. g. murghi*	Uttar	murghi
Xcelris_129	Red_murghi11	*G. g. murghi*	Uttar	murghi
Xcelris_120	Red_murghi9	*G. g. murghi*	Uttar	murghi
Xcelris_67	Red_murghi53	*G. g. murghi*	Uttarakhand	murghi
Xcelris_77	Red_murghi55	*G. g. murghi*	Uttarakhand	murghi
Xcelris_84	Red_murghi57	*G. g. murghi*	Uttarakhand	murghi
Xcelris_137	Red_murghi13	*G. g. murghi*	West Bengal	murghi
Xcelris_MT-134	Red_murghi60	*G. g. murghi*	West Bengal	murghi
ypt2893_L3_I025	Red_spadiceus10	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2894	Red_spadiceus11	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2895_L3_I026	Red_spadiceus12	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2887	Red_spadiceus4	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2888	Red_spadiceus5	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2889	Red_spadiceus6	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2890_L3_I040	Red_spadiceus7	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2891_L3_I041	Red_spadiceus8	*G. g. spadiceus*	ChaingMai	spadiceus
ypt2892	Red_spadiceus9	*G. g. spadiceus*	ChaingMai	spadiceus
20736	Red_spadiceus101	*G. g. spadiceus*	Hainan	spadiceus
Xcelris_174	Red_spadiceus1	*G. g. spadiceus*	Manipur	spadiceus
Xcelris_175	Red_spadiceus2	*G. g. spadiceus*	Manipur	spadiceus
Xcelris_176	Red_spadiceus3	*G. g. spadiceus*	Manipur	spadiceus
SAMEA5160200	RJFt1	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160209	RJFt10	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160210	RJFt11	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160211	RJFt12	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160212	RJFt13	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160213	RJFt14	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160214	RJFt15	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160215	RJFt16	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160216	RJFt17	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160217	RJFt18	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160218	RJFt19	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160201	RJFt2	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160219	RJFt20	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160220	RJFt21	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160221	RJFt22	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160222	RJFt23	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160223	RJFt24	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160224	RJFt25	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160202	RJFt3	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160203	RJFt4	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160204	RJFt5	*G. g. spadiceus*	Thailand	spadiceus

Accession	ID	Species	Breed/Location	Population
SAMEA5160205	RJFt6	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160206	RJFt7	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160207	RJFt8	*G. g. spadiceus*	Thailand	spadiceus
SAMEA5160208	RJFt9	*G. g. spadiceus*	Thailand	spadiceus
Yt540	Red_spadiceus100	*G. g. spadiceus*	Yunnan	spadiceus
ypt3001	Red_spadiceus13	*G. g. spadiceus*	Yunnan	spadiceus
ypt3002	Red_spadiceus14	*G. g. spadiceus*	Yunnan	spadiceus
ypt3003	Red_spadiceus15	*G. g. spadiceus*	Yunnan	spadiceus
ypt3004	Red_spadiceus16	*G. g. spadiceus*	Yunnan	spadiceus
ypt3005	Red_spadiceus17	*G. g. spadiceus*	Yunnan	spadiceus
ypt3006	Red_spadiceus18	*G. g. spadiceus*	Yunnan	spadiceus
ypt3007	Red_spadiceus19	*G. g. spadiceus*	Yunnan	spadiceus
ypt3008	Red_spadiceus20	*G. g. spadiceus*	Yunnan	spadiceus
ypt3009	Red_spadiceus21	*G. g. spadiceus*	Yunnan	spadiceus
ypt3010	Red_spadiceus22	*G. g. spadiceus*	Yunnan	spadiceus
ypt3011	Red_spadiceus23	*G. g. spadiceus*	Yunnan	spadiceus
ypt3012	Red_spadiceus24	*G. g. spadiceus*	Yunnan	spadiceus
ypt3016	Red_spadiceus25	*G. g. spadiceus*	Yunnan	spadiceus
ypt3017	Red_spadiceus26	*G. g. spadiceus*	Yunnan	spadiceus
ypt3028	Red_spadiceus27	*G. g. spadiceus*	Yunnan	spadiceus
ypt3038	Red_spadiceus28	*G. g. spadiceus*	Yunnan	spadiceus
ypt3040	Red_spadiceus29	*G. g. spadiceus*	Yunnan	spadiceus
ypt3041	Red_spadiceus30	*G. g. spadiceus*	Yunnan	spadiceus
ypt3042	Red_spadiceus31	*G. g. spadiceus*	Yunnan	spadiceus
ypt3043	Red_spadiceus32	*G. g. spadiceus*	Yunnan	spadiceus
ypt3044	Red_spadiceus33	*G. g. spadiceus*	Yunnan	spadiceus
ypt3045	Red_spadiceus34	*G. g. spadiceus*	Yunnan	spadiceus
ypt3046	Red_spadiceus35	*G. g. spadiceus*	Yunnan	spadiceus
ypt3047	Red_spadiceus36	*G. g. spadiceus*	Yunnan	spadiceus
ypt3050	Red_spadiceus37	*G. g. spadiceus*	Yunnan	spadiceus
ypt3051	Red_spadiceus38	*G. g. spadiceus*	Yunnan	spadiceus
ypt3052	Red_spadiceus39	*G. g. spadiceus*	Yunnan	spadiceus
ypt3061	Red_spadiceus40	*G. g. spadiceus*	Yunnan	spadiceus
ypt3069	Red_spadiceus41	*G. g. spadiceus*	Yunnan	spadiceus
ypt3072	Red_spadiceus42	*G. g. spadiceus*	Yunnan	spadiceus
SAMEA5930918	Glafayetii140	*G. lafayetii*	G. lafayetii	lafayetii
147830	Ceylon1	*G. lafayetii*	Sri Lanka	lafayetii
GLF-denovo	Ceylon2	*G. lafayetii*	Sri Lanka	lafayetii
Ypt4879	Ceylon3	*G. lafayetii*	Sri Lanka	lafayetii
Ypt4880	Ceylon4	*G. lafayetii*	Sri Lanka	lafayetii
SAMEA5930919	Gsonneratii141	*G. sonneratii*	G. sonneratii	sonneratii
147209	Grey1	*G. sonneratii*	India	sonnerattii
ypt4001	Grey2	*G. sonneratii*	India	sonnerattii
BLB1_L2_I007	Green1	*G. varius*	Bali	varius
BLB2_L2_I008	Green2	*G. varius*	Bali	varius
SAMEA5930920	Gvarius226	*G. varius*	G. varius	varius
BYW5_L2_I010	Green10	*G. varius*	Java	varius
GVF	Green11	*G. varius*	Java	varius
ypt4000	Green12	*G. varius*	Java	varius
BLO1	Green3	*G. varius*	Java	varius
BLO11	Green4	*G. varius*	Java	varius
BLO15	Green5	*G. varius*	Java	varius
BLO2	Green6	*G. varius*	Java	varius
BLO3_L4_I022	Green7	*G. varius*	Java	varius
BOJA2	Green8	*G. varius*	Java	varius
BYW3_L3_I033	Green9	*G. varius*	Java	varius

Abbreviations: Bd, bearded; Bm, bantam.
All samples with accession numbers in the form SAMXXXXXXXXX were downloaded from NCBI and all other samples were taken from [620].

Table D.2: HyPhy results using FEL.

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR1A	5	0	0.653	4.506	Pos.p=0.0338	I,T
TLR1A	9	0	0.763	3.638	Pos.p=0.0565	I,V
TLR1A	24	0	0.571	4.099	Pos.p=0.0429	P,T
TLR1A	29	0	0.855	5.042	Pos.p=0.0247	E,K

Continued on next page

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR1A	52	0	5.567	8.601	Pos.p=0.0034	I,V
TLR1A	75	0	0.977	5.329	Pos.p=0.0210	Q,R
TLR1A	83	0	1.426	11.366	Pos.p=0.0007	L,V
TLR1A	126	0	1.121	4.922	Pos.p=0.0265	F,L
TLR1A	158	0	0.897	4.056	Pos.p=0.0440	D,V
TLR1A	163	0	0.57	4.086	Pos.p=0.0432	A,T
TLR1A	169	0	0.522	2.85	Pos.p=0.0914	M,T
TLR1A	199	0	2.586	18.059	Pos.p=0.0000	A,T
TLR1A	218	0	1.719	11.363	Pos.p=0.0007	L,S
TLR1A	227	0	0.989	5.262	Pos.p=0.0218	L,S
TLR1A	237	0	3.195	21.311	Pos.p=0.0000	L,V
TLR1A	245	0	5.348	37.503	Pos.p=0.0000	A,T
TLR1A	249	0	6.538	22.62	Pos.p=0.0000	M,T
TLR1A	255	0	1.705	11.258	Pos.p=0.0008	F,L
TLR1A	307	0	0.602	3.506	Pos.p=0.0612	L,W
TLR1A	309	0	1.602	7.388	Pos.p=0.0066	N,S
TLR1A	359	0	10.175	72.798	Pos.p=0.0000	P,R,T
TLR1A	388	0	3.166	10.575	Pos.p=0.0011	I,L
TLR1A	403	0	1.169	5.402	Pos.p=0.0201	S,T
TLR1A	404	0	7.799	35.854	Pos.p=0.0000	F,L
TLR1A	408	0	0.659	3.295	Pos.p=0.0695	L,R
TLR1A	430	0	0.542	3.558	Pos.p=0.0592	H,R
TLR1A	452	0	1.324	5.43	Pos.p=0.0198	C,F
TLR1A	453	0	3.059	12.895	Pos.p=0.0003	A,E
TLR1A	462	0	3.848	21.511	Pos.p=0.0000	L,V
TLR1A	465	0	5.94	32.255	Pos.p=0.0000	E,Q
TLR1A	471	0	21.326	88.845	Pos.p=0.0000	N,S
TLR1A	476	0	3.218	15.739	Pos.p=0.0001	I,S
TLR1A	477	5.931	15.362	16.475	Pos.p=0.0000	A,T
TLR1A	500	0	2.645	17.168	Pos.p=0.0000	A,T
TLR1A	509	0	4.045	5.777	Pos.p=0.0162	I,T
TLR1A	524	0	1.055	6.671	Pos.p=0.0098	I,T
TLR1A	528	0	3.203	16.727	Pos.p=0.0000	M,V
TLR1A	560	0	7.441	47.444	Pos.p=0.0000	G,S
TLR1A	572	0	2.334	10.687	Pos.p=0.0011	D,V
TLR1A	579	0	11.897	47.774	Pos.p=0.0000	E,K
TLR1A	580	0	1.779	8.141	Pos.p=0.0043	A,D
TLR1A	609	0	2.672	17.305	Pos.p=0.0000	A,V
TLR1A	615	0	1.134	8.156	Pos.p=0.0043	L,P
TLR1A	618	0	1.046	5.931	Pos.p=0.0149	L,P
TLR1A	642	0	6.045	37.566	Pos.p=0.0000	Q,R
TLR1A	645	0	9.888	48.669	Pos.p=0.0000	A,T
TLR1A	669	0	1.22	5.057	Pos.p=0.0245	N,S
TLR1A	700	0	0.702	3.003	Pos.p=0.0831	I,T
TLR1A	718	0	0.697	2.906	Pos.p=0.0882	E,K
TLR1A	720	0	0.767	3.325	Pos.p=0.0682	C,Y
TLR1A	730	0	9.338	44.602	Pos.p=0.0000	K,R
TLR1A	749	0	5.852	38.816	Pos.p=0.0000	P,S
TLR1A	788	0	2.984	14.826	Pos.p=0.0001	A,T,V
TLR1A	794	0.3	2.078	6.428	Pos.p=0.0112	L,P
TLR1A	799	0	0.391	2.733	Pos.p=0.0983	A,T
TLR1A	806	0	0.609	3.752	Pos.p=0.0527	I,V
TLR1A	815	0	12.823	60.306	Pos.p=0.0000	C,R
TLR1A	818	0	8.721	37.737	Pos.p=0.0000	K,T
TLR1B	26	0	1.786	8.347	Pos.p=0.0039	G,I,S,V
TLR1B	36	0	0.851	9.319	Pos.p=0.0023	A,T
TLR1B	39	0	3.218	16.754	Pos.p=0.0000	P,S
TLR1B	46	0	2.286	12.584	Pos.p=0.0004	D,G
TLR1B	48	0	1.272	4.876	Pos.p=0.0272	D,N
TLR1B	55	0	3.217	32.996	Pos.p=0.0000	P,S
TLR1B	59	0	1.522	16.616	Pos.p=0.0000	A,T
TLR1B	69	0	0.728	4.032	Pos.p=0.0446	G,S
TLR1B	77	0	2.368	16.332	Pos.p=0.0001	M,T
TLR1B	86	0	1.23	13.464	Pos.p=0.0002	A,T
TLR1B	87	0	2.724	12.48	Pos.p=0.0004	D,G,N,S
TLR1B	100	0	6.246	26.983	Pos.p=0.0000	F,L
TLR1B	107	0	9.257	90.229	Pos.p=0.0000	Q,R
TLR1B	120	0	1.492	6.835	Pos.p=0.0089	D,N
TLR1B	124	0	3.962	35.087	Pos.p=0.0000	I,V
TLR1B	133	0	2.496	11.649	Pos.p=0.0006	I,V

Continued on next page

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR1B	148	0	0.799	8.22	Pos.p=0.0041	I,V
TLR1B	150	0	4.278	22.833	Pos.p=0.0000	A,T
TLR1B	156	0	2.58	19.639	Pos.p=0.0000	K,N
TLR1B	164	0	0.692	4.912	Pos.p=0.0267	A,V
TLR1B	172	0	3.604	18.507	Pos.p=0.0000	F,L
TLR1B	179	0	5.951	30.669	Pos.p=0.0000	I,V
TLR1B	202	0	3.466	27.181	Pos.p=0.0000	L,R
TLR1B	205	0	0.555	5.073	Pos.p=0.0243	L,S
TLR1B	207	0	0.72	2.753	Pos.p=0.0970	D,E
TLR1B	252	0	12.133	45.351	Pos.p=0.0000	F,L
TLR1B	261	0	0.389	2.786	Pos.p=0.0951	K,R
TLR1B	301	0	2.043	12.713	Pos.p=0.0004	A,E
TLR1B	310	0	2.434	17.836	Pos.p=0.0000	L,V
TLR1B	313	0	8.494	72.2	Pos.p=0.0000	E,Q
TLR1B	319	0	4.474	20.832	Pos.p=0.0000	N,S
TLR1B	324	0	1.105	4.982	Pos.p=0.0256	I,S
TLR1B	336	0	0.576	6.301	Pos.p=0.0121	A,S
TLR1B	351	0	1.089	6.346	Pos.p=0.0118	G,S
TLR1B	352	0	2.38	18.746	Pos.p=0.0000	L,R
TLR1B	370	0	2.619	12.262	Pos.p=0.0005	F,L
TLR1B	376	0	1.362	9.032	Pos.p=0.0027	M,V
TLR1B	397	0	1.045	5.965	Pos.p=0.0146	C,R
TLR1B	408	0	4.689	26.334	Pos.p=0.0000	G,I,S,V
TLR1B	420	0	1.114	4.414	Pos.p=0.0356	D,V
TLR1B	430	0	2.723	10.611	Pos.p=0.0011	H,R
TLR1B	433	0	0.676	4.196	Pos.p=0.0405	E,K
TLR1B	459	1.015	30.661	182.775	Pos.p=0.0000	L,P
TLR1B	490	1.357	3.092	4.506	Pos.p=0.0338	Q,R
TLR1B	578	0	1.987	12.043	Pos.p=0.0005	K,R
TLR1B	597	0	1.22	12.557	Pos.p=0.0004	P,S
TLR1B	621	0	0.54	3.314	Pos.p=0.0687	K,R
TLR1B	643	0	0.489	3.266	Pos.p=0.0707	A,V
TLR1B	644	0	0.849	4.593	Pos.p=0.0321	A,G,S,T
TLR1B	647	0	0.534	2.973	Pos.p=0.0847	E,Q
TLR2A	66	0	18.443	4.597	Pos.p=0.0320	M,T
TLR2A	84	0	11.285	2.941	Pos.p=0.0864	K,Q
TLR2A	85	0	20.376	4.677	Pos.p=0.0306	K,Q
TLR2A	128	0	10.292	3.325	Pos.p=0.0682	P,S
TLR2A	159	0	17.681	4.857	Pos.p=0.0275	L,Q
TLR2A	246	0	15.093	3.017	Pos.p=0.0824	I,V
TLR2A	413	0	18.444	3.882	Pos.p=0.0488	Q,R
TLR2A	505	0	27.54	9.035	Pos.p=0.0026	A,T
TLR2A	551	0	19.87	3.562	Pos.p=0.0591	L,R
TLR2A	616	0	121.899	20.426	Pos.p=0.0000	I,V
TLR2A	645	0	59.019	12.333	Pos.p=0.0004	M,T
TLR2B	16	0	8.429	4.244	Pos.p=0.0394	H,Y
TLR2B	55	0	6.901	3.416	Pos.p=0.0646	M,T
TLR2B	66	0	13.24	8.904	Pos.p=0.0028	L,V
TLR2B	73	0	13.259	6.086	Pos.p=0.0136	K,Q
TLR2B	74	0	7.896	2.957	Pos.p=0.0855	K,Q
TLR2B	80	0	6.742	4.698	Pos.p=0.0302	A,T
TLR2B	118	0	9.779	5.489	Pos.p=0.0191	L,V
TLR2B	130	0	19.674	13.078	Pos.p=0.0003	C,H,R,Y
TLR2B	154	0	7.718	5.154	Pos.p=0.0232	F,L
TLR2B	176	0	6.644	3.343	Pos.p=0.0675	H,N
TLR2B	224	0	6.304	4.121	Pos.p=0.0423	I,T
TLR2B	234	0	9.599	6.397	Pos.p=0.0114	G,S
TLR2B	235	0	4.384	3.026	Pos.p=0.0820	A,V
TLR2B	263	0	4.176	3.081	Pos.p=0.0792	A,T
TLR2B	278	0	8.099	5.971	Pos.p=0.0145	L,V
TLR2B	348	0	7.719	3.648	Pos.p=0.0561	M,V
TLR2B	461	0	19.218	12.764	Pos.p=0.0004	I,V
TLR2B	467	0	19.393	6.007	Pos.p=0.0142	H,Q
TLR2B	481	0	15.034	11.526	Pos.p=0.0007	L,V
TLR2B	493	0	9.188	5.76	Pos.p=0.0164	A,T
TLR2B	547	0	21.925	10.472	Pos.p=0.0012	E,G
TLR2B	576	0	44.913	19.556	Pos.p=0.0000	Q,R
TLR2B	693	0	9.931	5.304	Pos.p=0.0213	R,S
TLR2B	720	0	6.604	4.485	Pos.p=0.0342	C,R
TLR2B	731	0	5.036	3.373	Pos.p=0.0663	F,L

Continued on next page

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR3	17	0	43.594	55.466	Pos.p=0.0000	F,S
TLR3	32	0	4.591	6.816	Pos.p=0.0090	R,S
TLR3	68	0	16.821	18.832	Pos.p=0.0000	D,V
TLR3	92	0	9.969	11.752	Pos.p=0.0006	K,Q
TLR3	98	0	9.416	8.847	Pos.p=0.0029	N,S
TLR3	108	0	2.746	3.351	Pos.p=0.0672	D,E
TLR3	167	0	1.792	2.718	Pos.p=0.0992	H,R
TLR3	188	0	1.695	2.748	Pos.p=0.0974	I,S
TLR3	233	0	11.591	14.859	Pos.p=0.0001	K,T
TLR3	266	0	1.848	3.465	Pos.p=0.0627	G,R
TLR3	303	0	1.876	3.457	Pos.p=0.0630	I,M
TLR3	327	0	3.617	5.445	Pos.p=0.0196	S,T
TLR3	392	0	11.802	19.697	Pos.p=0.0000	R,S
TLR3	409	0	14.568	29.476	Pos.p=0.0000	E,G,K,R
TLR3	410	0	1.852	3.516	Pos.p=0.0608	A,T
TLR3	453	0	3.61	6.897	Pos.p=0.0086	M,V
TLR3	456	0	3.011	6.076	Pos.p=0.0137	Q,R
TLR3	459	0	3.145	5.615	Pos.p=0.0178	S,T
TLR3	480	0	2.598	4.933	Pos.p=0.0263	A,T
TLR3	489	0	2.139	5.261	Pos.p=0.0218	L,M
TLR3	506	0	15.087	25.041	Pos.p=0.0000	K,R
TLR3	521	0	4.287	8.332	Pos.p=0.0039	F,L
TLR3	567	0	2.499	4.073	Pos.p=0.0436	K,R
TLR3	573	0	9.277	11.143	Pos.p=0.0008	E,K
TLR3	587	4.005	15.658	11.47	Pos.p=0.0007	A,V
TLR3	592	0	6.902	10.851	Pos.p=0.0010	D,H
TLR3	639	0	7.549	11.345	Pos.p=0.0008	A,S
TLR3	647	0	1.896	2.979	Pos.p=0.0844	D,N
TLR3	696	0	30.874	45.058	Pos.p=0.0000	A,V
TLR3	760	0	15.769	23.104	Pos.p=0.0000	S,T
TLR3	857	0	3.447	5.108	Pos.p=0.0238	R,S
TLR3	872	0	2.989	2.815	Pos.p=0.0934	N,S
TLR3	877	0	1.887	2.926	Pos.p=0.0871	K,Q
TLR4	2	0	4.265	4.511	Pos.p=0.0337	L,P
TLR4	9	0	1.132	2.726	Pos.p=0.0987	A,V
TLR4	23	6.465	16.074	3.472	Pos.p=0.0624	C,S
TLR4	26	0	14.738	35.741	Pos.p=0.0000	A,V
TLR4	40	0	1.978	4.568	Pos.p=0.0326	I,T
TLR4	83	0	24.48	53.546	Pos.p=0.0000	E,K
TLR4	170	0	2.042	3.936	Pos.p=0.0473	K,R
TLR4	180	0	5.754	6.54	Pos.p=0.0105	A,S
TLR4	225	0	7.871	20.312	Pos.p=0.0000	E,G
TLR4	228	0	1.134	2.79	Pos.p=0.0948	A,V
TLR4	261	0	9.08	21.819	Pos.p=0.0000	K,R
TLR4	301	0	12.282	16.261	Pos.p=0.0001	D,E,G
TLR4	343	0	31.861	71.977	Pos.p=0.0000	K,R
TLR4	383	0	23.814	17.198	Pos.p=0.0000	H,Y
TLR4	410	0	6.024	11.182	Pos.p=0.0008	L,M
TLR4	498	0	5.519	6.052	Pos.p=0.0139	A,D,N,T
TLR4	574	0	1.461	3.341	Pos.p=0.0676	E,K
TLR4	582	0	2.986	5.618	Pos.p=0.0178	F,L
TLR4	611	0	29.319	67.08	Pos.p=0.0000	Q,R
TLR4	827	0	2.537	5.062	Pos.p=0.0245	S,T
TLR5	6	0	8.974	11.097	Pos.p=0.0009	Q,R
TLR5	18	0	8.759	4.57	Pos.p=0.0325	D,Y
TLR5	61	0	6.858	4.451	Pos.p=0.0349	Q,R
TLR5	65	0	7.071	5.211	Pos.p=0.0224	S,T
TLR5	82	0	14.738	16.64	Pos.p=0.0000	H,R
TLR5	83	0	7.637	8.595	Pos.p=0.0034	I,V
TLR5	91	0	6.158	3.483	Pos.p=0.0620	A,E
TLR5	118	0	7.122	8.118	Pos.p=0.0044	A,V
TLR5	122	0	12.869	9.525	Pos.p=0.0020	I,L,R,S
TLR5	127	0	5.486	4.245	Pos.p=0.0394	H,R
TLR5	140	0	4.051	2.825	Pos.p=0.0928	D,E
TLR5	178	0	7.518	5.953	Pos.p=0.0147	D,N
TLR5	204	0	6.083	5.433	Pos.p=0.0198	S,T
TLR5	211	0.053	7.571	3.467	Pos.p=0.0626	F,Y
TLR5	212	0	12.986	9.398	Pos.p=0.0022	K,R
TLR5	217	0	8.126	8.178	Pos.p=0.0042	I,V
TLR5	281	0	9.926	7.171	Pos.p=0.0074	N,T

Continued on next page

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR5	285	0	5.107	4.153	Pos.p=0.0416	S,T
TLR5	302	0	11.415	7.044	Pos.p=0.0080	I,V
TLR5	314	0	4.986	3.52	Pos.p=0.0606	H,R
TLR5	315	0	10.951	5.669	Pos.p=0.0173	N,S
TLR5	341	0	15.334	7.907	Pos.p=0.0049	E,K
TLR5	343	0	5.641	6.462	Pos.p=0.0110	L,V
TLR5	390	0	10.48	4.319	Pos.p=0.0377	I,T
TLR5	409	0	8.57	6.593	Pos.p=0.0102	S,T
TLR5	413	0	23.935	14.649	Pos.p=0.0001	G,S
TLR5	456	0	5.493	5.929	Pos.p=0.0149	F,L
TLR5	528	0	5.553	4.046	Pos.p=0.0443	I,T
TLR5	553	0	14.444	6.854	Pos.p=0.0088	I,T
TLR5	556	0	4.708	3.662	Pos.p=0.0557	N,Y
TLR5	604	0	3.383	2.706	Pos.p=0.1000	A,G
TLR5	608	0	3.772	2.973	Pos.p=0.0846	A,S
TLR5	618	0	4.321	3.649	Pos.p=0.0561	L,V
TLR5	619	0	5.691	6.011	Pos.p=0.0142	A,E
TLR5	623	0	8.552	7.096	Pos.p=0.0077	M,V
TLR5	648	0	12.775	12.384	Pos.p=0.0004	I,L
TLR5	684	0	5.385	3.164	Pos.p=0.0753	H,R
TLR5	707	0	5.919	4.317	Pos.p=0.0377	E,K
TLR5	747	0	4.232	3.171	Pos.p=0.0750	H,R
TLR5	763	0	7.208	4.811	Pos.p=0.0283	H,Q
TLR5	781	0	5.385	4.102	Pos.p=0.0428	C,Y
TLR5	830	0	12.285	6.823	Pos.p=0.0090	F,V
TLR5	834	0	5.639	4.135	Pos.p=0.0420	I,L
TLR5	837	0	10.424	6.232	Pos.p=0.0125	Q,R
TLR5	848	0	5.276	6.151	Pos.p=0.0131	G,V
TLR5	850	0	5.812	4.587	Pos.p=0.0322	G,V
TLR5	859	0	8.787	10.2	Pos.p=0.0014	G,V
TLR7	19	0	6.537	5.663	Pos.p=0.0173	M,T
TLR7	44	0	10.969	6.714	Pos.p=0.0096	F,L
TLR7	121	0	56.345	34.887	Pos.p=0.0000	I,V
TLR7	135	0	13.758	9.392	Pos.p=0.0022	R,S,T
TLR7	148	0	4.474	3.821	Pos.p=0.0506	L,S
TLR7	159	0	7.205	3.576	Pos.p=0.0586	H,R
TLR7	180	0	4.017	3.078	Pos.p=0.0794	R,S
TLR7	248	0	14.342	6.981	Pos.p=0.0082	I,V
TLR7	255	0	5.402	2.96	Pos.p=0.0853	N,S
TLR7	284	0	4.966	2.822	Pos.p=0.0930	I,M
TLR7	298	0	6.649	3.802	Pos.p=0.0512	K,R
TLR7	328	0	6.784	5.19	Pos.p=0.0227	L,P
TLR7	332	0	8.615	4.742	Pos.p=0.0294	F,S
TLR7	335	0	6.097	3.964	Pos.p=0.0465	K,N
TLR7	407	8.119	24.397	4.699	Pos.p=0.0302	F,L,M,S,T
TLR7	447	0	33.382	14.981	Pos.p=0.0001	G,S
TLR7	453	0	8.057	4.369	Pos.p=0.0366	C,W
TLR7	456	0	8.935	5.128	Pos.p=0.0235	H,P
TLR7	499	0	6.298	4.644	Pos.p=0.0312	G,R
TLR7	546	0	8.057	3.355	Pos.p=0.0670	S,T
TLR7	580	0	10.753	4.765	Pos.p=0.0290	I,V
TLR7	663	0	7.28	4.28	Pos.p=0.0386	E,K
TLR7	710	0	14.738	8.4	Pos.p=0.0038	K,R
TLR7	747	0	7.732	7.395	Pos.p=0.0065	Q,R
TLR7	847	0	7.351	5.031	Pos.p=0.0249	K,T
TLR7	877	0	4.164	3.843	Pos.p=0.0500	A,V
TLR7	896	0	7.752	5.191	Pos.p=0.0227	L,S
TLR7	923	0	10.502	5.158	Pos.p=0.0231	I,V
TLR7	944	0	7.53	3.585	Pos.p=0.0583	I,L
TLR15	34	0	7.529	7.68	Pos.p=0.0056	F,S
TLR15	85	0	6.577	6.951	Pos.p=0.0084	L,V
TLR15	87	0	6.828	6.729	Pos.p=0.0095	H,Y
TLR15	124	0	4.378	3.932	Pos.p=0.0474	H,Q
TLR15	125	0	11.955	14.68	Pos.p=0.0001	L,P
TLR15	129	0	12.167	12.276	Pos.p=0.0005	L,P
TLR15	132	0	4.005	2.85	Pos.p=0.0914	G,R
TLR15	148	0	5.068	3.297	Pos.p=0.0694	A,E
TLR15	149	0	11.259	12.032	Pos.p=0.0005	A,T
TLR15	158	0	12.155	5.936	Pos.p=0.0148	E,K
TLR15	192	0	5.806	4.549	Pos.p=0.0329	R,T

Continued on next page

TLR	Codon	α	β	LRT	Selection detected	Variants
TLR15	210	0	6.185	4.663	Pos.p=0.0308	G,R
TLR15	225	0	4.737	4.768	Pos.p=0.0290	A,D
TLR15	309	0	11.456	8.372	Pos.p=0.0038	A,E
TLR15	321	0	10.152	9.885	Pos.p=0.0017	I,V
TLR15	332	0	15.293	15.273	Pos.p=0.0001	F,L
TLR15	337	0	3.236	3.388	Pos.p=0.0657	H,R
TLR15	350	0	4.829	5.09	Pos.p=0.0241	H,R
TLR15	389	0	19.87	11.126	Pos.p=0.0009	K,R
TLR15	421	0	3.809	3.391	Pos.p=0.0655	Q,R
TLR15	430	0	5.783	6.433	Pos.p=0.0112	I,T
TLR15	436	0	7.139	7.036	Pos.p=0.0080	D,V
TLR15	438	0	7.914	7.272	Pos.p=0.0070	I,S
TLR15	444	0	8.905	3.41	Pos.p=0.0648	I,T
TLR15	447	0	4.459	4.93	Pos.p=0.0264	P,S
TLR15	478	0	6.171	3.68	Pos.p=0.0551	M,T
TLR15	560	0	4.705	5.745	Pos.p=0.0165	H,R
TLR15	574	0	7.188	6.827	Pos.p=0.0090	D,N
TLR15	630	0	7.084	7.081	Pos.p=0.0078	C,G
TLR15	696	0	26.928	18.493	Pos.p=0.0000	K,R
TLR15	708	0	5.647	5.025	Pos.p=0.0250	D,N
TLR15	729	0	6.866	7.567	Pos.p=0.0059	L,M
TLR15	731	0	6.395	6.763	Pos.p=0.0093	I,T
TLR15	738	0	10.47	9.388	Pos.p=0.0022	H,Y
TLR15	770	0	7.752	7.086	Pos.p=0.0078	N,S
TLR15	803	0	7.109	7.047	Pos.p=0.0079	D,G
TLR15	836	0	5.292	3.438	Pos.p=0.0637	K,T
TLR15	868	0	7.859	3.858	Pos.p=0.0495	E,K
TLR21	2	0	21.195	5.986	Pos.p=0.0144	M,V
TLR21	9	0	27.797	7.889	Pos.p=0.0050	R,W
TLR21	51	0	3.659	6.674	Pos.p=0.0098	P,R
TLR21	66	0	6.888	8.242	Pos.p=0.0041	L,P
TLR21	73	0	1.705	4.057	Pos.p=0.0440	A,P
TLR21	84	0	1.784	3.684	Pos.p=0.0549	I,K
TLR21	111	0	3.71	2.749	Pos.p=0.0973	A,T
TLR21	148	0	4.242	3.142	Pos.p=0.0763	P,S
TLR21	149	0	4.699	3.214	Pos.p=0.0730	L,M,S,T
TLR21	151	0	5.962	3.44	Pos.p=0.0637	Q,R
TLR21	189	0	39.052	24.979	Pos.p=0.0000	A,V
TLR21	195	0	3.349	8.058	Pos.p=0.0045	A,V
TLR21	218	0	6.946	5.054	Pos.p=0.0246	A,T
TLR21	239	0	24.521	14.803	Pos.p=0.0001	D,E,H,Q
TLR21	244	0	5.737	3.765	Pos.p=0.0523	M,V
TLR21	259	0	11.303	7.871	Pos.p=0.0050	A,V
TLR21	303	0	1.616	3.032	Pos.p=0.0817	A,S
TLR21	321	0	3.696	3.329	Pos.p=0.0681	Q,R
TLR21	328	0	6.385	4.905	Pos.p=0.0268	G,R
TLR21	344	0	3.694	9.198	Pos.p=0.0024	G,R
TLR21	351	0	5.97	3.3	Pos.p=0.0693	H,N,P,T
TLR21	477	0	2.153	4.014	Pos.p=0.0451	H,R
TLR21	517	0	4.401	3.144	Pos.p=0.0762	C,R
TLR21	525	0	2.669	6.573	Pos.p=0.0104	K,R
TLR21	559	0	5.177	3.582	Pos.p=0.0584	C,R
TLR21	625	0	2.593	6.325	Pos.p=0.0119	A,T
TLR21	635	0	5.837	3.251	Pos.p=0.0714	A,D
TLR21	648	0	3.701	3.36	Pos.p=0.0668	Q,R
TLR21	670	0	3.099	2.795	Pos.p=0.0946	Q,R
TLR21	675	0	7.518	4.094	Pos.p=0.0430	H,R
TLR21	691	0	8.401	7.329	Pos.p=0.0068	Q,R
TLR21	756	0	2.795	4.189	Pos.p=0.0407	N,Y
TLR21	757	0	4.571	3.208	Pos.p=0.0733	I,L
TLR21	770	0	6.677	4.421	Pos.p=0.0355	L,V
TLR21	782	0	5.952	7.232	Pos.p=0.0072	L,R
TLR21	818	0	3.127	4.956	Pos.p=0.0260	N,S
TLR21	848	0	5.025	3.209	Pos.p=0.0732	I,S
TLR21	886	0	4.872	4.461	Pos.p=0.0347	Q,R
TLR21	892	0	5.195	9.565	Pos.p=0.0020	C,R
TLR21	917	0	4.331	3.065	Pos.p=0.0800	G,V
TLR21	920	0	4.745	4.5	Pos.p=0.0339	Q,R

Table D.3: HyPhy results using FUBAR.

TLR	Codon	α	β	Selection detected	Variants
TLR1A	5	0.235	1.096	Pos.posterior=0.9449	I,T
TLR1A	9	0.26	1.344	Pos.posterior=0.9435	I,V
TLR1A	24	0.242	0.939	Pos.posterior=0.9315	P,T
TLR1A	29	0.218	1.743	Pos.posterior=0.9696	E,K
TLR1A	52	0.373	15.339	Pos.posterior=0.9989	I,V
TLR1A	75	0.208	2.028	Pos.posterior=0.9733	Q,R
TLR1A	83	0.109	2.804	Pos.posterior=0.9978	L,V
TLR1A	126	0.219	2.309	Pos.posterior=0.9867	F,L
TLR1A	158	0.232	2.027	Pos.posterior=0.9787	D,V
TLR1A	163	0.241	0.965	Pos.posterior=0.9347	A,T
TLR1A	199	0.161	6.91	Pos.posterior=1.0000	A,T
TLR1A	218	0.123	3.426	Pos.posterior=0.9981	L,S
TLR1A	227	0.195	1.542	Pos.posterior=0.9584	L,S
TLR1A	237	0.123	6.959	Pos.posterior=1.0000	L,V
TLR1A	245	0.177	14.202	Pos.posterior=1.0000	A,T
TLR1A	249	0.264	15.518	Pos.posterior=1.0000	M,T
TLR1A	255	0.124	3.385	Pos.posterior=0.9980	F,L
TLR1A	307	0.219	1.026	Pos.posterior=0.9081	L,W
TLR1A	309	0.197	2.923	Pos.posterior=0.9986	N,S
TLR1A	359	0.21	20.336	Pos.posterior=1.0000	P,R,T
TLR1A	363	0.299	1.294	Pos.posterior=0.9008	H,Q
TLR1A	388	0.24	7.121	Pos.posterior=0.9999	I,L
TLR1A	403	0.201	2.716	Pos.posterior=0.9972	S,T
TLR1A	404	0.213	15.519	Pos.posterior=1.0000	F,L
TLR1A	408	0.272	1.106	Pos.posterior=0.9359	L,R
TLR1A	430	0.249	0.934	Pos.posterior=0.9273	H,R
TLR1A	452	0.215	2.54	Pos.posterior=0.9915	C,F
TLR1A	453	0.215	7.121	Pos.posterior=1.0000	A,E
TLR1A	462	0.183	7.152	Pos.posterior=1.0000	L,V
TLR1A	465	0.193	15.44	Pos.posterior=1.0000	E,Q
TLR1A	471	0.37	50	Pos.posterior=1.0000	N,S
TLR1A	476	0.187	7.119	Pos.posterior=1.0000	I,S
TLR1A	477	14.459	29.357	Pos.posterior=1.0000	A,T
TLR1A	500	0.165	7.035	Pos.posterior=1.0000	A,T
TLR1A	509	0.352	7.135	Pos.posterior=0.9955	I,T
TLR1A	524	0.183	2.317	Pos.posterior=0.9924	I,T
TLR1A	528	0.194	7.118	Pos.posterior=1.0000	M,V
TLR1A	560	0.188	15.519	Pos.posterior=1.0000	G,S
TLR1A	572	0.195	6.745	Pos.posterior=0.9999	D,V
TLR1A	579	0.364	29.354	Pos.posterior=1.0000	E,K
TLR1A	580	0.196	3.466	Pos.posterior=0.9989	A,D
TLR1A	609	0.165	7.05	Pos.posterior=1.0000	A,V
TLR1A	615	0.141	2.135	Pos.posterior=0.9888	L,P
TLR1A	618	0.211	2.277	Pos.posterior=0.9894	L,P
TLR1A	642	0.201	15.501	Pos.posterior=1.0000	Q,R
TLR1A	645	0.226	17.115	Pos.posterior=1.0000	A,T
TLR1A	658	0.291	1.028	Pos.posterior=0.9016	D,N
TLR1A	669	0.213	2.607	Pos.posterior=0.9926	N,S
TLR1A	700	0.279	1.195	Pos.posterior=0.9192	I,T
TLR1A	718	0.296	1.283	Pos.posterior=0.9255	E,K
TLR1A	720	0.27	1.369	Pos.posterior=0.9471	C,Y
TLR1A	730	0.242	17.133	Pos.posterior=1.0000	K,R
TLR1A	749	0.176	13.799	Pos.posterior=1.0000	P,S
TLR1A	788	0.191	7.119	Pos.posterior=1.0000	A,T,V
TLR1A	794	0.56	3.667	Pos.posterior=0.9871	L,P
TLR1A	806	0.244	1.029	Pos.posterior=0.9276	I,V
TLR1A	815	0.334	29.356	Pos.posterior=1.0000	C,R
TLR1A	818	0.235	15.723	Pos.posterior=1.0000	K,T
TLR1B	26	0.088	2.806	Pos.posterior=0.9998	G,I,S,V
TLR1B	32	0.248	0.89	Pos.posterior=0.9157	N,Y
TLR1B	36	0.115	1.187	Pos.posterior=0.9929	A,T
TLR1B	39	0.115	5.932	Pos.posterior=1.0000	P,S
TLR1B	46	0.082	2.845	Pos.posterior=0.9999	D,G
TLR1B	48	0.113	2.681	Pos.posterior=0.9982	D,N
TLR1B	55	0.062	5.89	Pos.posterior=1.0000	P,S
TLR1B	59	0.044	2.786	Pos.posterior=1.0000	A,T
TLR1B	69	0.212	1.041	Pos.posterior=0.9667	G,S
TLR1B	73	0.274	2.092	Pos.posterior=0.9698	I,T

Continued on next page

TLR	Codon	α	β	Selection detected	Variants
TLR1B	77	0.075	2.885	Pos.posterior=1.0000	M,T
TLR1B	86	0.056	2.474	Pos.posterior=0.9995	A,T
TLR1B	87	0.123	5.754	Pos.posterior=1.0000	D,G,N,S
TLR1B	100	0.139	7.297	Pos.posterior=1.0000	F,L
TLR1B	107	0.132	15.519	Pos.posterior=1.0000	Q,R
TLR1B	120	0.091	2.791	Pos.posterior=0.9997	D,N
TLR1B	124	0.09	7.125	Pos.posterior=1.0000	I,V
TLR1B	133	0.092	3.279	Pos.posterior=0.9999	I,V
TLR1B	148	0.123	1.143	Pos.posterior=0.9885	I,V
TLR1B	150	0.128	7.125	Pos.posterior=1.0000	A,T
TLR1B	156	0.079	4.715	Pos.posterior=1.0000	K,N
TLR1B	164	0.199	0.976	Pos.posterior=0.9552	A,V
TLR1B	172	0.127	6.861	Pos.posterior=1.0000	F,L
TLR1B	179	0.131	7.304	Pos.posterior=1.0000	I,V
TLR1B	202	0.061	6.506	Pos.posterior=1.0000	L,R
TLR1B	205	0.145	0.787	Pos.posterior=0.9479	L,S
TLR1B	207	0.256	1.047	Pos.posterior=0.9431	D,E
TLR1B	252	0.263	15.546	Pos.posterior=1.0000	F,L
TLR1B	301	0.082	2.828	Pos.posterior=0.9999	A,E
TLR1B	310	0.067	3.018	Pos.posterior=1.0000	L,V
TLR1B	313	0.147	15.519	Pos.posterior=1.0000	E,Q
TLR1B	319	0.139	7.125	Pos.posterior=1.0000	N,S
TLR1B	324	0.154	2.014	Pos.posterior=0.9869	I,S
TLR1B	328	0.264	0.965	Pos.posterior=0.9221	H,R
TLR1B	336	0.139	0.877	Pos.posterior=0.9750	A,S
TLR1B	351	0.122	2.196	Pos.posterior=0.9956	G,S
TLR1B	352	0.043	3.311	Pos.posterior=1.0000	L,R
TLR1B	370	0.086	3.191	Pos.posterior=0.9999	F,L
TLR1B	376	0.087	2.645	Pos.posterior=0.9989	M,V
TLR1B	397	0.154	1.761	Pos.posterior=0.9895	C,R
TLR1B	408	0.131	7.125	Pos.posterior=1.0000	G,I,S,V
TLR1B	420	0.14	2.315	Pos.posterior=0.9931	D,V
TLR1B	430	0.122	4.441	Pos.posterior=0.9998	H,R
TLR1B	433	0.22	0.998	Pos.posterior=0.9518	E,K
TLR1B	459	1.198	50	Pos.posterior=1.0000	L,P
TLR1B	490	2.464	6.971	Pos.posterior=0.9745	Q,R
TLR1B	578	0.082	2.835	Pos.posterior=0.9999	K,R
TLR1B	597	0.072	2.117	Pos.posterior=0.9984	P,S
TLR1B	621	0.235	0.869	Pos.posterior=0.9204	K,R
TLR1B	643	0.221	0.767	Pos.posterior=0.9026	A,V
TLR1B	644	0.202	1.149	Pos.posterior=0.9707	A,G,S,T
TLR1B	647	0.232	0.908	Pos.posterior=0.9038	E,Q
TLR2A	66	0.931	23.422	Pos.posterior=0.9932	M,T
TLR2A	84	0.77	12.722	Pos.posterior=0.9441	K,Q
TLR2A	85	0.949	26.029	Pos.posterior=0.9936	K,Q
TLR2A	91	1.103	10.922	Pos.posterior=0.9534	A,T
TLR2A	104	0.909	11.143	Pos.posterior=0.9640	E,K
TLR2A	128	0.751	12.652	Pos.posterior=0.9784	P,S
TLR2A	159	0.678	20.187	Pos.posterior=0.9752	L,Q
TLR2A	188	0.895	11.018	Pos.posterior=0.9642	M,T
TLR2A	246	1.067	19.485	Pos.posterior=0.9840	I,V
TLR2A	257	0.919	9.323	Pos.posterior=0.9122	I,L
TLR2A	295	2.667	11.552	Pos.posterior=0.9112	I,T
TLR2A	413	0.936	23.164	Pos.posterior=0.9883	Q,R
TLR2A	505	0.771	34.985	Pos.posterior=0.9993	A,T
TLR2A	551	1.175	25.246	Pos.posterior=0.9879	L,R
TLR2A	616	1.246	50	Pos.posterior=0.9976	I,V
TLR2A	645	1.071	49.803	Pos.posterior=0.9989	M,T
TLR2B	16	0.645	14.846	Pos.posterior=0.9944	H,Y
TLR2B	22	0.585	5.082	Pos.posterior=0.9355	A,T
TLR2B	55	0.696	12.715	Pos.posterior=0.9902	M,T
TLR2B	63	0.718	10.717	Pos.posterior=0.9831	G,R
TLR2B	66	0.569	21.705	Pos.posterior=0.9994	L,V
TLR2B	73	0.674	23.313	Pos.posterior=0.9976	K,Q
TLR2B	74	0.799	14.568	Pos.posterior=0.9869	K,Q
TLR2B	80	0.577	12.59	Pos.posterior=0.9956	A,T
TLR2B	118	0.624	16.64	Pos.posterior=0.9967	L,V
TLR2B	130	0.628	33.014	Pos.posterior=0.9999	C,H,R,Y
TLR2B	154	0.578	13.731	Pos.posterior=0.9965	F,L
TLR2B	176	0.647	12.679	Pos.posterior=0.9908	H,N

Continued on next page

TLR	Codon	α	β	Selection detected	Variants
TLR2B	177	0.717	8.673	Pos.posterior=0.9763	M,T
TLR2B	201	0.651	6.635	Pos.posterior=0.9509	H,N
TLR2B	224	0.572	11.823	Pos.posterior=0.9934	I,T
TLR2B	234	0.57	16.699	Pos.posterior=0.9985	G,S
TLR2B	235	0.569	7.979	Pos.posterior=0.9706	A,V
TLR2B	263	0.573	7.754	Pos.posterior=0.9833	A,T
TLR2B	278	0.511	14.343	Pos.posterior=0.9949	L,V
TLR2B	282	0.785	12.149	Pos.posterior=0.9831	D,E
TLR2B	290	0.588	6.988	Pos.posterior=0.9606	D,G
TLR2B	348	0.698	13.847	Pos.posterior=0.9912	M,V
TLR2B	461	0.604	32.485	Pos.posterior=0.9999	I,V
TLR2B	467	0.967	33.964	Pos.posterior=0.9963	H,Q
TLR2B	481	0.519	24.468	Pos.posterior=0.9999	L,V
TLR2B	493	0.607	15.739	Pos.posterior=0.9972	A,T
TLR2B	547	0.758	38.627	Pos.posterior=0.9997	E,G
TLR2B	576	0.793	49.949	Pos.posterior=1.0000	Q,R
TLR2B	693	0.64	17.131	Pos.posterior=0.9969	R,S
TLR2B	720	0.582	12.21	Pos.posterior=0.9951	C,R
TLR2B	731	0.589	9.345	Pos.posterior=0.9875	F,L
TLR2B	757	0.586	5.783	Pos.posterior=0.9465	P,T
TLR3	10	0.437	2.581	Pos.posterior=0.9414	N,S
TLR3	17	0.553	50	Pos.posterior=1.0000	F,S
TLR3	32	0.449	8.969	Pos.posterior=0.9981	R,S
TLR3	68	0.57	36.169	Pos.posterior=1.0000	D,V
TLR3	92	0.554	26.109	Pos.posterior=0.9999	K,Q
TLR3	98	0.597	23.809	Pos.posterior=0.9997	N,S
TLR3	108	0.455	6.486	Pos.posterior=0.9931	D,E
TLR3	130	0.484	6.201	Pos.posterior=0.9887	N,S
TLR3	167	0.443	3.707	Pos.posterior=0.9743	H,R
TLR3	188	0.419	3.693	Pos.posterior=0.9779	I,S
TLR3	232	0.439	3.017	Pos.posterior=0.9306	K,M
TLR3	233	0.541	28.894	Pos.posterior=1.0000	K,T
TLR3	240	0.429	3.154	Pos.posterior=0.9574	A,D
TLR3	245	0.446	2.367	Pos.posterior=0.9074	H,N
TLR3	266	0.416	3.92	Pos.posterior=0.9805	G,R
TLR3	303	0.403	4.157	Pos.posterior=0.9766	I,M
TLR3	327	0.428	7.309	Pos.posterior=0.9975	S,T
TLR3	392	0.48	29.187	Pos.posterior=1.0000	R,S
TLR3	409	0.458	29.533	Pos.posterior=1.0000	E,G,K,R
TLR3	410	0.414	3.858	Pos.posterior=0.9801	A,T
TLR3	416	0.48	3.773	Pos.posterior=0.9414	Q,R
TLR3	453	0.398	7.318	Pos.posterior=0.9987	M,V
TLR3	456	0.389	6.536	Pos.posterior=0.9966	Q,R
TLR3	459	0.405	6.409	Pos.posterior=0.9963	S,T
TLR3	461	0.48	2.819	Pos.posterior=0.9160	F,I
TLR3	480	0.405	6.016	Pos.posterior=0.9956	A,T
TLR3	489	0.348	4.619	Pos.posterior=0.9770	L,M
TLR3	506	0.482	31.366	Pos.posterior=1.0000	K,R
TLR3	517	0.5	4.549	Pos.posterior=0.9742	D,N
TLR3	521	0.38	8.867	Pos.posterior=0.9993	F,L
TLR3	567	0.412	5.856	Pos.posterior=0.9928	K,R
TLR3	573	0.553	23.208	Pos.posterior=0.9999	E,K
TLR3	587	8.395	29.787	Pos.posterior=0.9987	A,V
TLR3	592	0.492	15.635	Pos.posterior=0.9999	D,H
TLR3	637	0.438	2.574	Pos.posterior=0.9411	K,N,S
TLR3	639	0.517	15.696	Pos.posterior=0.9999	A,S
TLR3	647	0.416	4.384	Pos.posterior=0.9840	D,N
TLR3	696	0.535	50	Pos.posterior=1.0000	A,V
TLR3	760	0.513	30.136	Pos.posterior=1.0000	S,T
TLR3	792	0.586	6.202	Pos.posterior=0.9797	I,M
TLR3	857	0.432	7.188	Pos.posterior=0.9971	R,S
TLR3	872	0.481	6.862	Pos.posterior=0.9922	N,S
TLR3	875	0.487	5.729	Pos.posterior=0.9860	H,Y
TLR3	877	0.427	4.296	Pos.posterior=0.9602	K,Q
TLR4	2	0.582	10.008	Pos.posterior=0.9935	L,P
TLR4	23	16.023	39.5	Pos.posterior=0.9167	C,S
TLR4	26	0.462	35.103	Pos.posterior=1.0000	A,V
TLR4	40	0.454	4.275	Pos.posterior=0.9733	I,T
TLR4	83	0.463	50	Pos.posterior=1.0000	E,K
TLR4	100	0.661	5.16	Pos.posterior=0.9642	K,N

Continued on next page

TLR	Codon	α	β	Selection detected	Variants
TLR4	170	0.489	5.059	Pos.posterior=0.9839	K,R
TLR4	180	0.573	14.118	Pos.posterior=0.9980	A,S
TLR4	225	0.458	17.572	Pos.posterior=1.0000	E,G
TLR4	261	0.469	25.549	Pos.posterior=1.0000	K,R
TLR4	301	0.581	29.738	Pos.posterior=1.0000	D,E,G
TLR4	314	0.585	3.688	Pos.posterior=0.9423	I,V
TLR4	343	0.46	50	Pos.posterior=1.0000	K,R
TLR4	346	0.647	4.737	Pos.posterior=0.9334	F,L
TLR4	383	0.694	49.986	Pos.posterior=1.0000	H,Y
TLR4	406	0.482	2.145	Pos.posterior=0.9167	K,R
TLR4	410	0.453	16.908	Pos.posterior=0.9999	L,M
TLR4	498	0.584	14.268	Pos.posterior=0.9980	A,D,N,T
TLR4	574	0.463	2.989	Pos.posterior=0.9501	E,K
TLR4	582	0.45	7.825	Pos.posterior=0.9843	F,L
TLR4	611	0.459	50	Pos.posterior=1.0000	Q,R
TLR4	827	0.454	5.013	Pos.posterior=0.9650	S,T
TLR5	6	0.398	15.27	Pos.posterior=0.9999	Q,R
TLR5	18	0.496	15.3	Pos.posterior=0.9982	D,Y
TLR5	61	0.476	13.12	Pos.posterior=0.9977	Q,R
TLR5	65	0.458	13.173	Pos.posterior=0.9984	S,T
TLR5	82	0.483	25.017	Pos.posterior=1.0000	H,R
TLR5	83	0.402	13.946	Pos.posterior=0.9998	I,V
TLR5	91	0.511	11.729	Pos.posterior=0.9956	A,E
TLR5	118	0.413	13.423	Pos.posterior=0.9996	A,V
TLR5	122	0.494	19.622	Pos.posterior=0.9997	I,L,R,S
TLR5	127	0.469	9.779	Pos.posterior=0.9969	H,R
TLR5	140	0.496	8.013	Pos.posterior=0.9923	D,E
TLR5	178	0.437	14.187	Pos.posterior=0.9993	D,N
TLR5	204	0.437	11.675	Pos.posterior=0.9987	S,T
TLR5	211	0.507	12.77	Pos.posterior=0.9962	F,Y
TLR5	212	0.506	20.713	Pos.posterior=0.9997	K,R
TLR5	217	0.411	14.425	Pos.posterior=0.9997	I,V
TLR5	281	0.455	15.819	Pos.posterior=0.9995	N,T
TLR5	285	0.468	9.239	Pos.posterior=0.9963	S,T
TLR5	302	0.488	17.472	Pos.posterior=0.9993	I,V
TLR5	314	0.485	8.954	Pos.posterior=0.9955	H,R
TLR5	315	0.512	17.103	Pos.posterior=0.9987	N,S
TLR5	341	0.643	25.489	Pos.posterior=0.9993	E,K
TLR5	343	0.412	10.365	Pos.posterior=0.9990	L,V
TLR5	390	0.597	16.189	Pos.posterior=0.9956	I,T
TLR5	409	0.447	14.78	Pos.posterior=0.9993	S,T
TLR5	413	0.684	44.378	Pos.posterior=1.0000	G,S
TLR5	454	0.527	8.307	Pos.posterior=0.9871	S,Y
TLR5	456	0.42	10.103	Pos.posterior=0.9988	F,L
TLR5	528	0.482	10.022	Pos.posterior=0.9962	I,T
TLR5	553	0.661	22.887	Pos.posterior=0.9983	I,T
TLR5	556	0.459	9.258	Pos.posterior=0.9961	N,Y
TLR5	571	0.762	9.58	Pos.posterior=0.9823	M,R
TLR5	604	0.494	6.761	Pos.posterior=0.9886	A,G
TLR5	608	0.488	7.193	Pos.posterior=0.9910	A,S
TLR5	616	0.498	6.575	Pos.posterior=0.9873	P,R
TLR5	618	0.451	8.793	Pos.posterior=0.9852	L,V
TLR5	619	0.432	10.426	Pos.posterior=0.9987	A,E
TLR5	623	0.443	15.03	Pos.posterior=0.9994	M,V
TLR5	648	0.438	19.362	Pos.posterior=1.0000	I,L
TLR5	684	0.506	9.884	Pos.posterior=0.9948	H,R
TLR5	707	0.467	11.462	Pos.posterior=0.9975	E,K
TLR5	747	0.487	7.842	Pos.posterior=0.9942	H,R
TLR5	749	0.52	5.446	Pos.posterior=0.9654	A,T
TLR5	763	0.468	13.837	Pos.posterior=0.9984	H,Q
TLR5	781	0.454	10.265	Pos.posterior=0.9970	C,Y
TLR5	830	0.517	18.536	Pos.posterior=0.9991	F,V
TLR5	834	0.469	10.318	Pos.posterior=0.9968	I,L
TLR5	837	0.497	17.01	Pos.posterior=0.9990	Q,R
TLR5	848	0.413	10.015	Pos.posterior=0.9989	G,V
TLR5	850	0.461	10.99	Pos.posterior=0.9976	G,V
TLR5	859	0.398	15.2	Pos.posterior=0.9999	G,V
TLR7	19	0.546	15.041	Pos.posterior=0.9980	M,T
TLR7	44	0.578	21.657	Pos.posterior=0.9989	F,L
TLR7	121	0.751	50	Pos.posterior=1.0000	I,V

Continued on next page

TLR	Codon	α	β	Selection detected	Variants
TLR7	135	0.669	27.638	Pos.posterior=0.9995	R,S,T
TLR7	148	0.557	10.4	Pos.posterior=0.9803	L,S
TLR7	159	0.641	15.693	Pos.posterior=0.9959	H,R
TLR7	180	0.605	10.74	Pos.posterior=0.9903	R,S
TLR7	248	0.8	30.213	Pos.posterior=0.9989	I,V
TLR7	255	0.613	13.857	Pos.posterior=0.9941	N,S
TLR7	284	0.62	12.666	Pos.posterior=0.9907	I,M
TLR7	298	0.625	15.545	Pos.posterior=0.9954	K,R
TLR7	328	0.547	14.797	Pos.posterior=0.9975	L,P
TLR7	332	0.604	17.561	Pos.posterior=0.9978	F,S
TLR7	334	1.937	12.604	Pos.posterior=0.9468	I,M
TLR7	335	0.592	14.832	Pos.posterior=0.9960	K,N
TLR7	395	0.7	6.917	Pos.posterior=0.9422	E,G
TLR7	407	19.65	48.784	Pos.posterior=0.9531	F,L,M,S,T
TLR7	447	1.02	49.999	Pos.posterior=0.9999	G,S
TLR7	453	0.607	16.854	Pos.posterior=0.9978	C,W
TLR7	456	0.605	18.338	Pos.posterior=0.9984	H,P
TLR7	499	0.576	14.948	Pos.posterior=0.9973	G,R
TLR7	515	1.511	15.966	Pos.posterior=0.9707	I,V
TLR7	546	0.69	17.065	Pos.posterior=0.9957	S,T
TLR7	580	0.716	21.449	Pos.posterior=0.9973	I,V
TLR7	602	2.162	10.901	Pos.posterior=0.9333	M,T
TLR7	663	0.618	16.288	Pos.posterior=0.9969	E,K
TLR7	706	0.66	7.172	Pos.posterior=0.9483	H,R
TLR7	710	0.768	32.874	Pos.posterior=0.9994	K,R
TLR7	737	0.664	10.881	Pos.posterior=0.9841	E,Q
TLR7	747	0.525	16.287	Pos.posterior=0.9993	Q,R
TLR7	847	0.585	15.841	Pos.posterior=0.9973	K,T
TLR7	850	0.688	12.198	Pos.posterior=0.9798	F,L
TLR7	860	0.592	8.026	Pos.posterior=0.9622	I,V
TLR7	877	0.559	10.751	Pos.posterior=0.9906	A,V
TLR7	896	0.549	16.33	Pos.posterior=0.9954	L,S
TLR7	923	0.691	21.002	Pos.posterior=0.9978	I,V
TLR7	944	0.654	15.802	Pos.posterior=0.9947	I,L
TLR7	1059	0.687	11.318	Pos.posterior=0.9839	I,V
TLR15	34	0.425	14.565	Pos.posterior=0.9995	F,S
TLR15	85	0.433	13.073	Pos.posterior=0.9990	L,V
TLR15	87	0.432	13.948	Pos.posterior=0.9992	H,Y
TLR15	99	0.622	7.678	Pos.posterior=0.9625	E,K
TLR15	124	0.484	9.494	Pos.posterior=0.9949	H,Q
TLR15	125	0.418	19.148	Pos.posterior=1.0000	L,P
TLR15	129	0.466	19.06	Pos.posterior=0.9999	L,P
TLR15	132	0.572	8.972	Pos.posterior=0.9851	G,R
TLR15	141	1.402	9.467	Pos.posterior=0.9475	L,M
TLR15	148	0.558	11.165	Pos.posterior=0.9915	A,E
TLR15	149	0.436	17.821	Pos.posterior=0.9999	A,T
TLR15	158	0.682	21.49	Pos.posterior=0.9980	E,K
TLR15	192	0.509	12.033	Pos.posterior=0.9954	R,T
TLR15	210	0.498	13.205	Pos.posterior=0.9966	G,R
TLR15	225	0.454	10.625	Pos.posterior=0.9972	A,D
TLR15	309	0.524	18.33	Pos.posterior=0.9992	A,E
TLR15	321	0.434	16.917	Pos.posterior=0.9998	I,V
TLR15	332	0.504	25.842	Pos.posterior=1.0000	F,L
TLR15	337	0.488	6.853	Pos.posterior=0.9842	H,R
TLR15	350	0.458	9.998	Pos.posterior=0.9973	H,R
TLR15	389	0.805	43.336	Pos.posterior=0.9998	K,R
TLR15	421	0.524	8.096	Pos.posterior=0.9863	Q,R
TLR15	430	0.43	12.127	Pos.posterior=0.9988	I,T
TLR15	436	0.428	14.568	Pos.posterior=0.9994	D,V
TLR15	438	0.434	15.323	Pos.posterior=0.9995	I,S
TLR15	444	0.693	15.782	Pos.posterior=0.9919	I,T
TLR15	447	0.47	8.394	Pos.posterior=0.9967	P,S
TLR15	478	0.571	12.681	Pos.posterior=0.9926	M,T
TLR15	560	0.442	9.984	Pos.posterior=0.9981	H,R
TLR15	574	0.431	14.676	Pos.posterior=0.9994	D,N
TLR15	630	0.428	14.074	Pos.posterior=0.9993	C,G
TLR15	696	0.685	49.215	Pos.posterior=1.0000	K,R
TLR15	708	0.452	12.505	Pos.posterior=0.9980	D,N
TLR15	729	0.42	13.652	Pos.posterior=0.9993	L,M
TLR15	731	0.433	13.1	Pos.posterior=0.9990	I,T

Continued on next page

TLR	Codon	α	β	Selection detected	Variants
TLR15	738	0.445	17.246	Pos.posterior=0.9998	H,Y
TLR15	770	0.435	15.21	Pos.posterior=0.9995	N,S
TLR15	803	0.427	14.578	Pos.posterior=0.9994	D,G
TLR15	829	0.495	5.209	Pos.posterior=0.9472	R,S
TLR15	832	0.665	8.518	Pos.posterior=0.9772	A,E
TLR15	836	0.55	11.616	Pos.posterior=0.9924	K,T
TLR15	868	0.581	15.336	Pos.posterior=0.9951	E,K
TLR21	2	1.277	48.956	Pos.posterior=0.9979	M,V
TLR21	9	1.263	49.977	Pos.posterior=0.9986	R,W
TLR21	48	0.474	7.455	Pos.posterior=0.9882	I,V
TLR21	51	0.327	7.266	Pos.posterior=0.9989	P,R
TLR21	66	0.428	12.095	Pos.posterior=0.9990	L,P
TLR21	73	0.373	3.835	Pos.posterior=0.9363	A,P
TLR21	84	0.376	5.078	Pos.posterior=0.9738	I,K
TLR21	111	0.46	7.359	Pos.posterior=0.9897	A,T
TLR21	145	0.471	7.192	Pos.posterior=0.9880	D,G
TLR21	148	0.462	7.582	Pos.posterior=0.9902	P,S
TLR21	149	0.505	8.056	Pos.posterior=0.9882	L,M,S,T
TLR21	151	0.613	11.296	Pos.posterior=0.9895	Q,R
TLR21	189	0.722	50	Pos.posterior=1.0000	A,V
TLR21	195	0.303	7.198	Pos.posterior=0.9994	A,V
TLR21	218	0.589	14.127	Pos.posterior=0.9966	A,T
TLR21	239	0.752	49.744	Pos.posterior=1.0000	D,E,H,Q
TLR21	244	0.56	10.686	Pos.posterior=0.9911	M,V
TLR21	246	0.441	6.198	Pos.posterior=0.9858	I,L
TLR21	257	0.495	6.785	Pos.posterior=0.9832	A,V
TLR21	259	0.695	23.21	Pos.posterior=0.9989	A,V
TLR21	268	0.545	6.01	Pos.posterior=0.9667	I,V
TLR21	272	0.562	7.428	Pos.posterior=0.9821	H,Y
TLR21	303	0.406	3.672	Pos.posterior=0.9245	A,S
TLR21	321	0.439	7.48	Pos.posterior=0.9915	Q,R
TLR21	328	0.582	13.44	Pos.posterior=0.9957	G,R
TLR21	344	0.303	7.444	Pos.posterior=0.9997	G,R
TLR21	351	0.633	12.42	Pos.posterior=0.9914	H,N,P,T
TLR21	390	0.48	6.656	Pos.posterior=0.9835	A,S
TLR21	443	0.559	7.655	Pos.posterior=0.9833	N,S
TLR21	446	0.577	8.393	Pos.posterior=0.9842	A,E
TLR21	452	0.545	7.267	Pos.posterior=0.9829	N,T
TLR21	466	0.555	7.448	Pos.posterior=0.9826	H,R
TLR21	477	0.357	5.864	Pos.posterior=0.9851	H,R
TLR21	517	0.479	8.153	Pos.posterior=0.9903	C,R
TLR21	525	0.309	6.868	Pos.posterior=0.9988	K,R
TLR21	559	0.513	9.562	Pos.posterior=0.9914	C,R
TLR21	625	0.312	6.624	Pos.posterior=0.9972	A,T
TLR21	628	0.543	7.925	Pos.posterior=0.9851	A,D
TLR21	635	0.609	11.504	Pos.posterior=0.9904	A,D
TLR21	648	0.438	7.489	Pos.posterior=0.9917	Q,R
TLR21	670	0.443	7.048	Pos.posterior=0.9892	Q,R
TLR21	675	0.681	14.82	Pos.posterior=0.9942	H,R
TLR21	686	0.565	7.915	Pos.posterior=0.9836	D,N
TLR21	691	0.59	16.109	Pos.posterior=0.9985	Q,R
TLR21	710	0.551	7.428	Pos.posterior=0.9828	A,D
TLR21	756	0.355	6.769	Pos.posterior=0.9955	N,Y
TLR21	757	0.48	8.306	Pos.posterior=0.9905	I,L
TLR21	770	0.57	12.273	Pos.posterior=0.9937	L,V
TLR21	782	0.4	10.45	Pos.posterior=0.9986	L,R
TLR21	790	0.517	4.371	Pos.posterior=0.9208	Q,R
TLR21	818	0.343	7.132	Pos.posterior=0.9982	N,S
TLR21	835	0.484	6.755	Pos.posterior=0.9812	L,R
TLR21	848	0.534	9.325	Pos.posterior=0.9895	I,S
TLR21	867	0.597	6.046	Pos.posterior=0.9678	R,S
TLR21	886	0.454	8.83	Pos.posterior=0.9937	Q,R
TLR21	892	0.361	9.64	Pos.posterior=0.9995	C,R
TLR21	907	0.564	7.617	Pos.posterior=0.9826	A,E
TLR21	917	0.496	8.227	Pos.posterior=0.9891	G,V
TLR21	920	0.445	8.622	Pos.posterior=0.9939	Q,R

Table D.4: HyPhy results using MEME.

TLR	Codon	α	β+	p+	LRT	Selection detected	#branches	Variants
TLR1A	5	0	0.643	1	4.507	Yes,p=0.0486	5	I,T
TLR1A	9	0	0.759	0.99	3.639	Yes,p=0.0764	5	I,V
TLR1A	24	0	0.575	0.99	4.1	Yes,p=0.0601	5	P,T
TLR1A	29	0	0.874	0.994	5.044	Yes,p=0.0369	5	E,K
TLR1A	52	0	5.671	0.985	8.604	Yes,p=0.0060	46	I,V
TLR1A	75	0	0.981	0.99	5.329	Yes,p=0.0318	4	Q,R
TLR1A	83	0	1.421	0.99	11.366	Yes,p=0.0015	5	L,V
TLR1A	126	0	1.163	0.977	4.923	Yes,p=0.0392	7	F,L
TLR1A	158	0	1.962	0.466	4.082	Yes,p=0.0606	7	D,V
TLR1A	163	0	0.574	0.99	4.086	Yes,p=0.0605	5	A,T
TLR1A	199	0	2.622	0.994	18.057	Yes,p=0.0001	20	A,T
TLR1A	218	0	1.727	0.99	11.363	Yes,p=0.0015	4	L,S
TLR1A	227	0	0.892	0.99	5.291	Yes,p=0.0325	3	L,S
TLR1A	237	0	3.152	1	21.312	Yes,p=0.0000	10	L,V
TLR1A	245	0	5.444	0.994	37.49	Yes,p=0.0000	40	A,T
TLR1A	249	0	6.556	1	22.621	Yes,p=0.0000	45	M,T
TLR1A	255	0	1.714	0.99	11.258	Yes,p=0.0016	5	F,L
TLR1A	307	0	1.195	0.5	3.511	Yes,p=0.0817	2	L,W
TLR1A	309	0	1.66	0.977	7.387	Yes,p=0.0111	14	N,S
TLR1A	359	0	120.731	0.116	79.123	Yes,p=0.0000	70	P,R,T
TLR1A	388	0	2144.696	0.012	31.787	Yes,p=0.0000	24	I,L
TLR1A	403	0	1.183	0.99	5.403	Yes,p=0.0306	11	S,T
TLR1A	404	0	7.767	0.99	35.839	Yes,p=0.0000	37	F,L
TLR1A	408	0	1.312	0.5	3.304	Yes,p=0.0910	6	L,R
TLR1A	430	0	0.543	0.99	3.562	Yes,p=0.0795	5	H,R
TLR1A	452	0	1.33	0.99	5.43	Yes,p=0.0302	7	C,F
TLR1A	453	0	3.07	0.99	12.892	Yes,p=0.0007	22	A,E
TLR1A	462	0	3.844	0.999	21.512	Yes,p=0.0000	21	L,V
TLR1A	465	0	5.944	0.989	32.253	Yes,p=0.0000	23	E,Q
TLR1A	471	0	21.411	1	88.848	Yes,p=0.0000	103	N,S
TLR1A	476	0	3.355	0.961	15.733	Yes,p=0.0002	20	I,S
TLR1A	477	5.923	15.36	1	16.476	Yes,p=0.0001	0	A,T
TLR1A	500	0	2.693	0.994	17.162	Yes,p=0.0001	20	A,T
TLR1A	509	0	4.074	1	5.776	Yes,p=0.0253	29	I,T
TLR1A	524	0	1.044	1	6.672	Yes,p=0.0160	6	I,T
TLR1A	528	0	3.187	1	16.728	Yes,p=0.0001	18	M,V
TLR1A	560	0	7.464	1	47.444	Yes,p=0.0000	60	G,S
TLR1A	572	0	692.614	0.011	20.311	Yes,p=0.0000	18	D,V
TLR1A	579	0	11.95	1	47.774	Yes,p=0.0000	53	E,K
TLR1A	580	0	1.811	0.993	8.139	Yes,p=0.0076	14	A,D
TLR1A	609	0	2.678	0.99	17.305	Yes,p=0.0001	20	A,V
TLR1A	615	0	1.114	1	8.158	Yes,p=0.0075	3	L,P
TLR1A	618	0	1.035	1	5.932	Yes,p=0.0233	8	L,P
TLR1A	642	0	6.126	0.995	37.56	Yes,p=0.0000	39	Q,R
TLR1A	645	0	9.841	1	48.668	Yes,p=0.0000	68	A,T
TLR1A	669	0	1.207	1	5.058	Yes,p=0.0366	6	N,S
TLR1A	720	0	0.759	1	3.326	Yes,p=0.0900	6	C,Y
TLR1A	730	0	9.329	1	44.606	Yes,p=0.0000	61	K,R
TLR1A	749	0	5.847	1	38.817	Yes,p=0.0000	31	P,S
TLR1A	788	0	6.406	0.474	14.878	Yes,p=0.0002	26	A,T,V
TLR1A	794	0.29	2.103	0.996	6.428	Yes,p=0.0181	17	L,P
TLR1A	806	0	0.609	0.99	3.753	Yes,p=0.0720	4	I,V
TLR1A	815	0	13.059	0.986	60.304	Yes,p=0.0000	88	C,R
TLR1A	818	0	75.417	0.142	40.013	Yes,p=0.0000	55	K,T
TLR1B	26	0	16.441	0.11	11.192	Yes,p=0.0016	13	G,I,S,V
TLR1B	36	0	0.844	1	9.321	Yes,p=0.0041	6	A,T
TLR1B	39	0	3.173	0.999	16.751	Yes,p=0.0001	21	P,S
TLR1B	46	0	2.328	0.991	12.579	Yes,p=0.0008	24	D,G
TLR1B	48	0	2.543	0.5	4.878	Yes,p=0.0402	14	D,N
TLR1B	55	0	3.276	0.99	32.992	Yes,p=0.0000	26	P,S
TLR1B	59	0	1.522	0.99	16.614	Yes,p=0.0001	13	A,T
TLR1B	69	0	0.739	0.995	4.031	Yes,p=0.0623	9	G,S
TLR1B	77	0	2.405	0.99	16.329	Yes,p=0.0001	20	M,T
TLR1B	86	0	1.211	0.999	13.461	Yes,p=0.0005	11	A,T
TLR1B	87	0	16.467	0.171	16.696	Yes,p=0.0001	26	D,G,N,S
TLR1B	100	0	6.251	1	26.986	Yes,p=0.0000	26	F,L
TLR1B	107	0	9.668	0.96	90.232	Yes,p=0.0000	46	Q,R
TLR1B	120	0	1.48	1	6.835	Yes,p=0.0147	12	D,N

Continued on next page

TLR	Codon	α	β+	p+	LRT	Selection detected	#branches	Variants
TLR1B	124	0	4.042	0.99	35.075	Yes,p=0.0000	34	I,V
TLR1B	133	0	2.542	0.992	11.645	Yes,p=0.0013	13	I,V
TLR1B	148	0	0.791	1	8.223	Yes,p=0.0072	6	I,V
TLR1B	150	0	4.298	1	22.833	Yes,p=0.0000	31	A,T
TLR1B	156	0	2.596	1	19.639	Yes,p=0.0000	24	K,N
TLR1B	164	0	0.689	1	4.912	Yes,p=0.0394	6	A,V
TLR1B	172	0	3.666	0.992	18.503	Yes,p=0.0000	17	F,L
TLR1B	179	0	6.03	0.995	30.666	Yes,p=0.0000	38	I,V
TLR1B	202	0	3.453	0.99	27.179	Yes,p=0.0000	10	L,R
TLR1B	205	0	0.561	0.967	5.072	Yes,p=0.0363	3	L,S
TLR1B	252	0	12.149	1	45.353	Yes,p=0.0000	62	F,L
TLR1B	301	0	2.09	0.991	12.707	Yes,p=0.0007	20	A,E
TLR1B	310	0	5.273	0.464	17.911	Yes,p=0.0001	19	L,V
TLR1B	313	0	34.848	0.273	72.702	Yes,p=0.0000	46	E,Q
TLR1B	319	0	4.469	1	20.833	Yes,p=0.0000	49	N,S
TLR1B	324	0	1.099	0.998	4.982	Yes,p=0.0381	5	I,S
TLR1B	336	0	0.575	0.99	6.301	Yes,p=0.0193	4	A,S
TLR1B	351	0	10.243	0.113	6.588	Yes,p=0.0167	14	G,S
TLR1B	352	0	2.341	1	18.746	Yes,p=0.0000	8	L,R
TLR1B	370	0	2.662	0.993	12.261	Yes,p=0.0009	13	F,L
TLR1B	376	0	1.393	0.994	9.027	Yes,p=0.0048	11	M,V
TLR1B	397	0	1.045	0.99	5.96	Yes,p=0.0230	9	C,R
TLR1B	408	0	14.725	0.331	29.589	Yes,p=0.0000	57	G,I,S,V
TLR1B	420	0	1.123	0.99	4.415	Yes,p=0.0510	10	D,V
TLR1B	430	0	2.769	0.99	10.608	Yes,p=0.0022	21	H,R
TLR1B	433	0	0.67	1	4.196	Yes,p=0.0571	3	E,K
TLR1B	459	1.017	30.598	1	182.777	Yes,p=0.0000	88	L,P
TLR1B	490	1.342	3.082	1	4.505	Yes,p=0.0487	0	Q,R
TLR1B	578	0	2.018	0.995	12.044	Yes,p=0.0010	16	K,R
TLR1B	597	0	1.208	0.999	12.561	Yes,p=0.0008	9	P,S
TLR1B	621	0	0.537	1	3.315	Yes,p=0.0905	5	K,R
TLR1B	643	0	0.456	0.999	3.279	Yes,p=0.0922	4	A,V
TLR1B	644	0	0.86	0.99	4.593	Yes,p=0.0465	8	A,G,S,T
TLR2A	66	0	18.387	1	4.598	Yes,p=0.0464	4	M,T
TLR2A	77	0	5570.762	0.007	3.926	Yes,p=0.0658	1	L,V
TLR2A	85	0	20.258	0.99	4.676	Yes,p=0.0446	3	K,Q
TLR2A	128	0	21.996	0.477	3.328	Yes,p=0.0899	3	P,S
TLR2A	129	0	3367.457	0.012	5.462	Yes,p=0.0297	2	L,V
TLR2A	159	0	1949.042	0.02	6.866	Yes,p=0.0145	2	L,Q
TLR2A	413	0	18.316	0.99	3.883	Yes,p=0.0672	2	Q,R
TLR2A	505	0	27.471	1	9.038	Yes,p=0.0048	5	A,T
TLR2A	551	0	19.811	1	3.564	Yes,p=0.0794	5	L,R
TLR2A	616	0	285.499	0.516	22.234	Yes,p=0.0000	28	I,V
TLR2A	645	0	58.889	1	12.334	Yes,p=0.0009	15	M,T
TLR2B	16	0	8.644	1	4.242	Yes,p=0.0558	3	H,Y
TLR2B	55	0	6.801	1	3.419	Yes,p=0.0857	5	M,T
TLR2B	66	0	13.507	0.989	8.913	Yes,p=0.0051	7	L,V
TLR2B	73	0	13.414	1	6.088	Yes,p=0.0215	3	K,Q
TLR2B	80	0	6.712	1	4.699	Yes,p=0.0440	1	A,T
TLR2B	118	0	442.359	0.041	6.955	Yes,p=0.0138	6	L,V
TLR2B	130	0	58.95	0.336	14.784	Yes,p=0.0003	15	C,H,R,Y
TLR2B	154	0	7.9	1	5.153	Yes,p=0.0348	4	F,L
TLR2B	176	0	6.782	1	3.34	Yes,p=0.0893	3	H,N
TLR2B	224	0	6.215	1	4.12	Yes,p=0.0594	4	I,T
TLR2B	234	0	9.6	1	6.393	Yes,p=0.0184	2	G,S
TLR2B	278	0	7.766	1	5.969	Yes,p=0.0229	1	L,V
TLR2B	348	0.001	7.639	1	3.649	Yes,p=0.0760	3	M,V
TLR2B	461	0	31.765	0.629	12.8	Yes,p=0.0007	10	I,V
TLR2B	467	0	19.432	1	6.01	Yes,p=0.0224	6	H,Q
TLR2B	481	0	14.802	1	11.529	Yes,p=0.0014	5	L,V
TLR2B	493	0	9.19	1	5.763	Yes,p=0.0255	6	A,T
TLR2B	547	0	22.05	1	10.472	Yes,p=0.0023	13	E,G
TLR2B	576	0	80.555	0.607	20.041	Yes,p=0.0000	20	Q,R
TLR2B	693	0	10.075	0.995	5.3	Yes,p=0.0323	7	R,S
TLR2B	720	0	6.676	1	4.483	Yes,p=0.0492	5	C,R
TLR2B	731	0	4.962	1	3.375	Yes,p=0.0877	3	F,L
TLR3	17	0	94.296	0.512	58.356	Yes,p=0.0000	85	F,S
TLR3	32	0	4.71	0.979	6.818	Yes,p=0.0148	12	R,S
TLR3	68	0	85.117	0.233	20.844	Yes,p=0.0000	40	D,V
TLR3	92	0	74.506	0.152	12.733	Yes,p=0.0007	20	K,Q

Continued on next page

TLR	Codon	α	β+	p+	LRT	Selection detected	#branches	Variants
TLR3	98	0	58.229	0.181	9.579	Yes,p=0.0036	23	N,S
TLR3	108	0	2.798	0.992	3.351	Yes,p=0.0888	7	D,E
TLR3	233	0	24.658	0.494	15.054	Yes,p=0.0002	25	K,T
TLR3	266	0	1.835	1	3.466	Yes,p=0.0836	4	G,R
TLR3	303	0	1.937	0.994	3.456	Yes,p=0.0841	4	I,M
TLR3	327	0	3.718	0.994	5.444	Yes,p=0.0300	9	S,T
TLR3	392	0	11.91	0.997	19.696	Yes,p=0.0000	32	R,S
TLR3	409	0	40.009	0.38	31.336	Yes,p=0.0000	36	E,G,K,R
TLR3	410	0	1.825	1	3.516	Yes,p=0.0814	4	A,T
TLR3	453	0	3.605	0.99	6.894	Yes,p=0.0143	7	M,V
TLR3	456	0	3.072	0.994	6.075	Yes,p=0.0217	5	Q,R
TLR3	459	0	3.146	0.99	5.614	Yes,p=0.0275	5	S,T
TLR3	480	0	2.602	0.99	4.931	Yes,p=0.0391	5	A,T
TLR3	489	0	2.16	0.99	5.262	Yes,p=0.0329	3	L,M
TLR3	506	0	15.102	1	25.042	Yes,p=0.0000	28	K,R
TLR3	521	0	4.263	1	8.333	Yes,p=0.0069	5	F,L
TLR3	567	0	2.45	1	4.073	Yes,p=0.0609	4	K,R
TLR3	573	0	9.321	1	11.143	Yes,p=0.0016	16	E,K
TLR3	587	3.976	31.415	0.488	11.678	Yes,p=0.0013	0	A,V
TLR3	592	0	14.673	0.481	10.887	Yes,p=0.0019	18	D,H
TLR3	637	0	67.679	0.017	8.407	Yes,p=0.0066	3	K,N,S
TLR3	639	0	269.792	0.045	17.275	Yes,p=0.0001	20	A,S
TLR3	696	0	30.972	1	45.058	Yes,p=0.0000	58	A,V
TLR3	760	0	15.902	0.998	23.101	Yes,p=0.0000	34	S,T
TLR3	857	0	3.407	1	5.108	Yes,p=0.0357	8	R,S
TLR4	2	0	4.437	0.96	4.511	Yes,p=0.0485	6	L,P
TLR4	23	6.279	16.566	0.976	3.477	Yes,p=0.0831	24	C,S
TLR4	26	0	55.937	0.298	37.353	Yes,p=0.0000	38	A,V
TLR4	40	0	4.292	0.464	4.592	Yes,p=0.0465	4	I,T
TLR4	83	0	27.233	0.907	53.595	Yes,p=0.0000	51	E,K
TLR4	170	0	2.037	0.99	3.935	Yes,p=0.0654	4	K,R
TLR4	180	0	5.744	1	6.541	Yes,p=0.0171	7	A,S
TLR4	225	0	16.71	0.488	20.67	Yes,p=0.0000	28	E,G
TLR4	261	0	9.011	1	21.819	Yes,p=0.0000	26	K,R
TLR4	301	0	12.273	1	16.263	Yes,p=0.0001	33	D,E,G
TLR4	343	0	32.009	1	71.975	Yes,p=0.0000	79	K,R
TLR4	383	0	23.808	1	17.2	Yes,p=0.0001	37	H,Y
TLR4	410	0	6.16	0.992	11.181	Yes,p=0.0016	4	L,M
TLR4	498	0	61.629	0.077	14.505	Yes,p=0.0003	8	A,D,N,T
TLR4	574	0	1.425	1	3.34	Yes,p=0.0893	4	E,K
TLR4	582	0	3.049	1	5.62	Yes,p=0.0274	3	F,L
TLR4	611	0	29.296	1	67.081	Yes,p=0.0000	62	Q,R
TLR4	827	0	2.507	0.99	5.061	Yes,p=0.0365	3	S,T
TLR5	6	0	8.889	0.993	11.087	Yes,p=0.0017	4	Q,R
TLR5	18	0	8.867	1	4.571	Yes,p=0.0470	4	D,Y
TLR5	61	0	6.898	1	4.452	Yes,p=0.0500	5	Q,R
TLR5	65	0	6.987	0.99	5.206	Yes,p=0.0339	6	S,T
TLR5	82	0	83.769	0.251	19.853	Yes,p=0.0000	19	H,R
TLR5	83	0	7.775	0.999	8.593	Yes,p=0.0060	4	I,V
TLR5	91	0	6.299	1	3.482	Yes,p=0.0829	4	A,E
TLR5	118	0	7.14	1	8.12	Yes,p=0.0076	6	A,V
TLR5	122	0	117.632	0.125	22.715	Yes,p=0.0000	10	I,L,R,S
TLR5	127	0	5.446	1	4.247	Yes,p=0.0557	4	H,R
TLR5	178	0	7.741	1	5.946	Yes,p=0.0232	4	D,N
TLR5	204	0	6.047	1	5.433	Yes,p=0.0302	4	S,T
TLR5	211	0.039	107.222	0.094	5.147	Yes,p=0.0349	6	F,Y
TLR5	212	0	13.145	0.975	9.402	Yes,p=0.0040	8	K,R
TLR5	217	0	8.03	1	8.179	Yes,p=0.0074	6	I,V
TLR5	281	0	9.909	1	7.172	Yes,p=0.0124	10	N,T
TLR5	285	0	58.703	0.115	5.08	Yes,p=0.0362	6	S,T
TLR5	302	0	11.333	1	7.046	Yes,p=0.0132	6	I,V
TLR5	314	0	4.957	1	3.521	Yes,p=0.0812	4	H,R
TLR5	315	0	10.829	1	5.67	Yes,p=0.0267	5	N,S
TLR5	341	0	15.568	0.999	7.906	Yes,p=0.0085	11	E,K
TLR5	343	0	12.18	0.474	6.536	Yes,p=0.0171	4	L,V
TLR5	390	0	10.441	1	4.321	Yes,p=0.0536	6	I,T
TLR5	409	0	18.168	0.506	6.659	Yes,p=0.0161	6	S,T
TLR5	413	0	23.893	1	14.65	Yes,p=0.0003	23	G,S
TLR5	456	0	5.502	1	5.931	Yes,p=0.0234	2	F,L
TLR5	528	0	5.559	1	4.048	Yes,p=0.0617	4	I,T

Continued on next page

TLR	Codon	α	β+	p+	LRT	Selection detected	#branches	Variants
TLR5	553	0	99.798	0.213	8.97	Yes,p=0.0050	16	I,T
TLR5	556	0	69.656	0.088	4.69	Yes,p=0.0442	3	N,Y
TLR5	618	0	9.5	0.463	3.665	Yes,p=0.0753	2	L,V
TLR5	619	0	5.705	0.99	6.009	Yes,p=0.0224	6	A,E
TLR5	623	0	8.964	0.952	7.103	Yes,p=0.0128	4	M,V
TLR5	648	0	13.02	0.998	12.382	Yes,p=0.0009	4	I,L
TLR5	684	0	5.384	1	3.164	Yes,p=0.0979	5	H,R
TLR5	707	0	5.806	1	4.316	Yes,p=0.0537	4	E,K
TLR5	747	0	4.201	0.993	3.168	Yes,p=0.0978	3	H,R
TLR5	763	0	7.029	1	4.808	Yes,p=0.0416	5	H,Q
TLR5	781	0	5.455	1	4.102	Yes,p=0.0600	3	C,Y
TLR5	830	0	12.301	1	6.824	Yes,p=0.0148	6	F,V
TLR5	834	0	5.593	0.99	4.134	Yes,p=0.0590	4	I,L
TLR5	837	0	10.313	1	6.232	Yes,p=0.0200	6	Q,R
TLR5	848	0	5.285	0.99	6.155	Yes,p=0.0208	4	G,V
TLR5	850	0	5.777	1	4.589	Yes,p=0.0466	7	G,V
TLR5	859	0	8.879	1	10.201	Yes,p=0.0027	10	G,V
TLR7	19	0	13.76	0.496	5.69	Yes,p=0.0264	6	M,T
TLR7	44	0	10.593	1	6.71	Yes,p=0.0157	1	F,L
TLR7	121	0	218.133	0.346	41.891	Yes,p=0.0000	36	I,V
TLR7	135	0	13.779	1	9.393	Yes,p=0.0040	8	R,S,T
TLR7	148	0	4.432	1	3.822	Yes,p=0.0694	2	L,S
TLR7	159	0	7.259	1	3.577	Yes,p=0.0789	6	H,R
TLR7	248	0	14.323	1	6.982	Yes,p=0.0136	8	I,V
TLR7	298	0	6.99	0.957	3.804	Yes,p=0.0701	4	K,R
TLR7	328	0	6.789	1	5.192	Yes,p=0.0342	3	L,P
TLR7	332	0.007	8.712	1	4.738	Yes,p=0.0432	2	F,S
TLR7	335	0	6.099	1	3.966	Yes,p=0.0644	3	K,N
TLR7	407	7.455	171.26	0.132	8.049	Yes,p=0.0079	1	F,L,M,S,T
TLR7	447	0	34.132	0.976	14.994	Yes,p=0.0002	30	G,S
TLR7	453	0	17.217	0.48	4.456	Yes,p=0.0499	9	C,W
TLR7	456	0	8.928	1	5.129	Yes,p=0.0353	7	H,P
TLR7	499	0.002	6.208	1	4.641	Yes,p=0.0454	5	G,R
TLR7	546	0	7.914	1	3.353	Yes,p=0.0887	7	S,T
TLR7	580	0	10.641	1	4.765	Yes,p=0.0426	4	I,V
TLR7	663	0	7.211	0.993	4.276	Yes,p=0.0548	3	E,K
TLR7	710	0	14.905	1	8.402	Yes,p=0.0066	8	K,R
TLR7	747	0	7.7	1	7.397	Yes,p=0.0110	6	Q,R
TLR7	847	0	65.638	0.132	5.388	Yes,p=0.0309	3	K,T
TLR7	877	0	4.184	0.99	3.843	Yes,p=0.0687	4	A,V
TLR7	896	0	102.209	0.092	5.571	Yes,p=0.0281	3	L,S
TLR7	923	0	10.486	1	5.159	Yes,p=0.0347	5	I,V
TLR7	944	0	16.299	0.462	3.735	Yes,p=0.0726	3	I,L
TLR15	34	0	7.675	1	7.678	Yes,p=0.0096	6	F,S
TLR15	85	0	234.005	0.05	8.875	Yes,p=0.0052	8	L,V
TLR15	87	0	6.725	1	6.729	Yes,p=0.0155	6	H,Y
TLR15	124	0	4.302	1	3.931	Yes,p=0.0656	5	H,Q
TLR15	125	0	11.838	1	14.681	Yes,p=0.0003	5	L,P
TLR15	129	0	11.997	1	12.276	Yes,p=0.0009	14	L,P
TLR15	148	0	5.099	1	3.298	Yes,p=0.0913	4	A,E
TLR15	149	0	11.273	1	12.034	Yes,p=0.0011	9	A,T
TLR15	158	0	11.991	1	5.952	Yes,p=0.0231	7	E,K
TLR15	192	0	5.924	0.99	4.549	Yes,p=0.0476	7	R,T
TLR15	210	0	6.065	1	4.662	Yes,p=0.0449	2	G,R
TLR15	225	0	4.729	0.99	4.767	Yes,p=0.0425	4	A,D
TLR15	309	0	317.752	0.075	14.474	Yes,p=0.0003	15	A,E
TLR15	321	0	10.118	1	9.886	Yes,p=0.0031	6	I,V
TLR15	332	0	71.305	0.258	16.088	Yes,p=0.0001	14	F,L
TLR15	337	0	3.22	0.99	3.386	Yes,p=0.0872	3	H,R
TLR15	350	0	4.827	1	5.094	Yes,p=0.0359	2	H,R
TLR15	389	0	21.13	1	11.218	Yes,p=0.0016	16	K,R
TLR15	421	0	3.785	1	3.393	Yes,p=0.0869	3	Q,R
TLR15	430	0	5.791	1	6.435	Yes,p=0.0180	2	I,T
TLR15	436	0	57.057	0.149	7.4	Yes,p=0.0110	7	D,V
TLR15	438	0	8.057	0.99	7.274	Yes,p=0.0118	8	I,S
TLR15	444	0	8.867	1	3.41	Yes,p=0.0861	9	I,T
TLR15	447	0	4.313	1	4.926	Yes,p=0.0392	2	P,S
TLR15	478	0	5.999	1	3.679	Yes,p=0.0748	2	M,T
TLR15	560	0	4.912	1	5.747	Yes,p=0.0257	2	H,R
TLR15	574	0	6.976	1	6.823	Yes,p=0.0148	4	D,N

Continued on next page

TLR	Codon	α	β+	p+	LRT	Selection detected	#branches	Variants
TLR15	630	0	7.023	0.99	7.083	Yes,p=0.0130	4	C,G
TLR15	696	0	172.604	0.216	22.02	Yes,p=0.0000	23	K,R
TLR15	708	0	5.633	1	5.026	Yes,p=0.0372	1	D,N
TLR15	729	0	6.77	1	7.567	Yes,p=0.0101	4	L,M
TLR15	731	0	6.418	1	6.764	Yes,p=0.0152	3	I,T
TLR15	738	0	10.455	1	9.389	Yes,p=0.0040	5	H,Y
TLR15	770	0	7.779	1	7.087	Yes,p=0.0129	5	N,S
TLR15	803	0	7.049	1	7.047	Yes,p=0.0132	4	D,G
TLR15	836	0	5.395	0.99	3.438	Yes,p=0.0848	5	K,T
TLR15	868	0	7.912	1	3.859	Yes,p=0.0681	3	E,K
TLR21	2	0	47.62	0.496	8.122	Yes,p=0.0076	32	M,V
TLR21	9	0	108.084	0.333	14.365	Yes,p=0.0003	38	R,W
TLR21	51	0	270.555	0.026	17.575	Yes,p=0.0001	6	P,R
TLR21	66	0	6.744	1	8.242	Yes,p=0.0072	3	L,P
TLR21	73	0	162.102	0.018	7.253	Yes,p=0.0119	3	A,P
TLR21	84	0	1.808	0.99	3.686	Yes,p=0.0745	4	I,K
TLR21	148	0	4.258	1	3.143	Yes,p=0.0990	5	P,S
TLR21	149	0	110.972	0.051	15.296	Yes,p=0.0002	7	L,M,S,T
TLR21	151	0	6.073	1	3.439	Yes,p=0.0848	1	Q,R
TLR21	189	0	81.409	0.557	27.553	Yes,p=0.0000	48	A,V
TLR21	195	0	3.374	1	8.059	Yes,p=0.0079	3	A,V
TLR21	218	0	6.948	1	5.057	Yes,p=0.0366	9	A,T
TLR21	239	0	62.04	0.44	16.094	Yes,p=0.0001	25	D,E,H,Q
TLR21	244	0	5.79	1	3.766	Yes,p=0.0715	5	M,V
TLR21	259	0	11.302	1	7.872	Yes,p=0.0087	7	A,V
TLR21	321	0	3.792	1	3.326	Yes,p=0.0900	4	Q,R
TLR21	328	0	6.393	1	4.906	Yes,p=0.0396	7	G,R
TLR21	344	0	3.701	1	9.2	Yes,p=0.0044	5	G,R
TLR21	351	0	34.508	0.189	5.68	Yes,p=0.0266	10	H,N,P,T
TLR21	477	0	2.136	1	4.015	Yes,p=0.0628	4	H,R
TLR21	517	0	4.39	1	3.144	Yes,p=0.0990	3	C,R
TLR21	525	0	2.649	1	6.573	Yes,p=0.0168	4	K,R
TLR21	559	0	5.193	1	3.584	Yes,p=0.0786	8	C,R
TLR21	625	0	2.571	1	6.328	Yes,p=0.0191	5	A,T
TLR21	635	0	5.91	0.997	3.251	Yes,p=0.0936	8	A,D
TLR21	648	0	3.617	1	3.357	Yes,p=0.0885	1	Q,R
TLR21	675	0	7.551	1	4.094	Yes,p=0.0602	6	H,R
TLR21	691	0	8.404	1	7.33	Yes,p=0.0114	8	Q,R
TLR21	756	0	17.632	0.167	4.847	Yes,p=0.0408	5	N,Y
TLR21	757	0	4.579	1	3.21	Yes,p=0.0956	5	I,L
TLR21	770	0	283.529	0.042	9.504	Yes,p=0.0038	8	L,V
TLR21	782	0	5.926	1	7.234	Yes,p=0.0120	5	L,R
TLR21	818	0	44.56	0.093	6.042	Yes,p=0.0221	8	N,S
TLR21	835	0	228.091	0.024	4.01	Yes,p=0.0629	4	L,R
TLR21	848	0	4.932	1	3.207	Yes,p=0.0958	8	I,S
TLR21	886	0	4.802	1	4.462	Yes,p=0.0498	7	Q,R
TLR21	892	0	5.312	0.994	9.565	Yes,p=0.0037	8	C,R
TLR21	920	0	210.23	0.043	8.614	Yes,p=0.0059	8	Q,R

Table D.5: HyPhy results using SLAC.

TLR	Codon	S	N	dS	dN	Selection	Variants
TLR1A	52	0	50	0	17.828	Pos.p=0.035	I,V
TLR1A	83	0	5	0	4.057	Pos.p=0.012	L,V
TLR1A	199	0	20	0	10	Pos.p=0.000	A,T
TLR1A	218	0	5	0	4.357	Pos.p=0.009	L,S
TLR1A	227	0	3	0	2.406	Pos.p=0.074	L,S
TLR1A	237	0	10	0	8.297	Pos.p=0.000	L,V
TLR1A	245	0	43	0	21.5	Pos.p=0.000	A,T
TLR1A	249	0	52	0	21.59	Pos.p=0.000	M,T
TLR1A	255	0	5	0	4.325	Pos.p=0.010	F,L
TLR1A	309	0	15	0	6.736	Pos.p=0.011	N,S
TLR1A	359	0	70	0	45.655	Pos.p=0.000	P,R,T
TLR1A	388	0	24	0	9.888	Pos.p=0.007	I,L
TLR1A	403	0	11	0	4.943	Pos.p=0.037	S,T
TLR1A	404	0	48	0	21.884	Pos.p=0.000	F,L
TLR1A	408	0	6	0	3	Pos.p=0.088	L,R
TLR1A	452	0	7	0	3.358	Pos.p=0.078	C,F

Continued on next page

TLR	Codon	S	N	dS	dN	Selection	Variants
TLR1A	453	0	22	0	10.312	Pos.p=0.001	A,E
TLR1A	462	0	24	0	12.4	Pos.p=0.000	L,V
TLR1A	465	0	25	0	17.444	Pos.p=0.000	E,Q
TLR1A	471	0	138	0	66.19	Pos.p=0.000	N,S
TLR1A	476	0	20	0	9.391	Pos.p=0.001	I,S
TLR1A	477	17	116	17	58	Pos.p=0.000	A,T
TLR1A	500	0	23	0	11.5	Pos.p=0.000	A,T
TLR1A	524	0	8	0	3.998	Pos.p=0.039	I,T
TLR1A	528	0	21	0	10.39	Pos.p=0.000	M,V
TLR1A	560	0	68	0	33.697	Pos.p=0.000	G,S
TLR1A	572	0	18	0	8.119	Pos.p=0.004	D,V
TLR1A	579	0	73	0	33.889	Pos.p=0.000	E,K
TLR1A	580	0	14	0	6.311	Pos.p=0.015	A,D
TLR1A	609	0	23	0	11.5	Pos.p=0.000	A,V
TLR1A	615	0	4	0	3.228	Pos.p=0.029	L,P
TLR1A	618	0	9	0	4.537	Pos.p=0.024	L,P
TLR1A	642	0	50	0	26.66	Pos.p=0.000	Q,R
TLR1A	645	0	84	0	42	Pos.p=0.000	A,T
TLR1A	669	0	7	0	3.357	Pos.p=0.078	N,S
TLR1A	730	0	76	0	36.487	Pos.p=0.000	K,R
TLR1A	749	0	36	0	20.106	Pos.p=0.000	P,S
TLR1A	788	0	26	0	13	Pos.p=0.000	A,T,V
TLR1A	794	1	18	0.99	9.043	Pos.p=0.004	L,P
TLR1A	815	0	109	0	53.968	Pos.p=0.000	C,R
TLR1A	818	0	55	0	25.914	Pos.p=0.000	K,T
TLR1B	26	0	14	0	6.785	Pos.p=0.005	G,I,S,V
TLR1B	36	0	9	0	4.5	Pos.p=0.026	A,T
TLR1B	39	0	26	0	13	Pos.p=0.000	P,S
TLR1B	46	0	31	0	15.44	Pos.p=0.000	D,G
TLR1B	48	0	14	0	6.236	Pos.p=0.017	D,N
TLR1B	55	0	26	0	13.089	Pos.p=0.000	P,S
TLR1B	59	0	16	0	8	Pos.p=0.002	A,T
TLR1B	69	0	10	0	4.992	Pos.p=0.018	G,S
TLR1B	77	0	22	0	10.779	Pos.p=0.000	M,T
TLR1B	86	0	13	0	6.5	Pos.p=0.005	A,T
TLR1B	87	0	28	0	13.264	Pos.p=0.000	D,G,N,S
TLR1B	100	0	33	0	15.81	Pos.p=0.000	F,L
TLR1B	107	0	62	0	50.924	Pos.p=0.000	Q,R
TLR1B	120	0	15	0	7.103	Pos.p=0.005	D,N
TLR1B	124	0	38	0	17.183	Pos.p=0.000	I,V
TLR1B	133	0	19	0	9.215	Pos.p=0.001	I,V
TLR1B	148	0	7	0	3.454	Pos.p=0.064	I,V
TLR1B	150	0	41	0	20.5	Pos.p=0.000	A,T
TLR1B	156	0	24	0	11.186	Pos.p=0.000	K,N
TLR1B	164	0	6	0	3	Pos.p=0.088	A,V
TLR1B	172	0	22	0	10.844	Pos.p=0.000	F,L
TLR1B	179	0	49	0	24.418	Pos.p=0.000	I,V
TLR1B	202	0	13	0	10.216	Pos.p=0.000	L,R
TLR1B	207	0	8	0	3.566	Pos.p=0.098	D,E
TLR1B	252	0	70	0	31.573	Pos.p=0.000	F,L
TLR1B	301	0	21	0	10.149	Pos.p=0.001	A,E
TLR1B	310	0	19	0	9.982	Pos.p=0.000	L,V
TLR1B	313	0	50	0	34.275	Pos.p=0.000	E,Q
TLR1B	319	0	56	0	26.52	Pos.p=0.000	N,S
TLR1B	324	0	9	0	4.258	Pos.p=0.043	I,S
TLR1B	336	0	6	0	3.016	Pos.p=0.087	A,S
TLR1B	351	0	15	0	7.431	Pos.p=0.003	G,S
TLR1B	352	0	9	0	7.066	Pos.p=0.000	L,R
TLR1B	370	0	13	0	6.462	Pos.p=0.006	F,L
TLR1B	376	0	12	0	5.77	Pos.p=0.012	M,V
TLR1B	397	0	12	0	5.973	Pos.p=0.009	C,R
TLR1B	408	0	62	0	30.431	Pos.p=0.000	G,I,S,V
TLR1B	420	0	12	0	5.379	Pos.p=0.029	D,V
TLR1B	430	0	24	0	10.773	Pos.p=0.001	H,R
TLR1B	433	0	7	0	3.378	Pos.p=0.082	E,K
TLR1B	459	7	114	4.176	86.117	Pos.p=0.000	L,P
TLR1B	490	7	32	6.227	17.506	Pos.p=0.006	Q,R
TLR1B	578	0	18	0	8.641	Pos.p=0.002	K,R
TLR1B	597	0	10	0	5.008	Pos.p=0.017	P,S
TLR1B	644	0	9	0	4.499	Pos.p=0.026	A,G,S,T

Continued on next page

TLR	Codon	S	N	dS	dN	Selection	Variants
TLR2A	505	0	9	0	4.5	Pos.p=0.026	A,T
TLR2A	616	0	30	0	14.912	Pos.p=0.000	I,V
TLR2A	645	0	19	0	8.623	Pos.p=0.003	M,T
TLR2B	66	0	7	0	4.271	Pos.p=0.015	L,V
TLR2B	80	0	6	0	3	Pos.p=0.088	A,T
TLR2B	118	0	6	0	3.189	Pos.p=0.061	L,V
TLR2B	130	0	17	0	8.478	Pos.p=0.001	C,H,R,Y
TLR2B	154	0	6	0	2.988	Pos.p=0.090	F,L
TLR2B	234	0	7	0	3.426	Pos.p=0.068	G,S
TLR2B	278	0	3	0	2.467	Pos.p=0.070	L,V
TLR2B	461	0	12	0	5.985	Pos.p=0.008	I,V
TLR2B	467	0	8	0	5.481	Pos.p=0.027	H,Q
TLR2B	481	0	6	0	4.738	Pos.p=0.006	L,V
TLR2B	493	0	8	0	4	Pos.p=0.039	A,T
TLR2B	547	0	16	0	7.891	Pos.p=0.002	E,G
TLR2B	576	0	24	0	15.038	Pos.p=0.000	Q,R
TLR2B	693	0	8	0	3.676	Pos.p=0.077	R,S
TLR2B	720	0	6	0	2.991	Pos.p=0.091	C,R
TLR3	17	0	95	0	45.235	Pos.p=0.000	F,S
TLR3	32	0	12	0	5.95	Pos.p=0.021	R,S
TLR3	68	0	40	0	18.095	Pos.p=0.000	D,V
TLR3	92	0	20	0	10.023	Pos.p=0.004	K,Q
TLR3	98	0	23	0	9.933	Pos.p=0.003	N,S
TLR3	233	0	25	0	11.75	Pos.p=0.001	K,T
TLR3	327	0	9	0	4.5	Pos.p=0.026	S,T
TLR3	392	0	34	0	15.625	Pos.p=0.000	R,S
TLR3	409	0	40	0	19.917	Pos.p=0.000	E,G,K,R
TLR3	453	0	8	0	3.903	Pos.p=0.048	M,V
TLR3	456	0	5	0	4.5	Pos.p=0.029	Q,R
TLR3	459	0	6	0	3.42	Pos.p=0.067	S,T
TLR3	480	0	7	0	3.5	Pos.p=0.059	A,T
TLR3	489	0	3	0	2.209	Pos.p=0.093	L,M
TLR3	506	0	31	0	15.927	Pos.p=0.000	K,R
TLR3	521	0	5	0	3.973	Pos.p=0.020	F,L
TLR3	573	0	20	0	9.268	Pos.p=0.006	E,K
TLR3	587	5	41	5	20.5	Pos.p=0.000	A,V
TLR3	592	0	18	0	7.976	Pos.p=0.006	D,H
TLR3	639	0	20	0	10	Pos.p=0.000	A,S
TLR3	696	0	73	0	36.5	Pos.p=0.000	A,V
TLR3	760	0	37	0	18.5	Pos.p=0.000	S,T
TLR3	857	0	9	0	4.476	Pos.p=0.053	R,S
TLR4	2	0	6	0	3	Pos.p=0.088	L,P
TLR4	23	4	24	4.09	11.913	Pos.p=0.025	C,S
TLR4	26	0	38	0	19	Pos.p=0.000	A,V
TLR4	83	0	61	0	30.294	Pos.p=0.000	E,K
TLR4	180	0	8	0	4	Pos.p=0.039	A,S
TLR4	225	0	29	0	15.112	Pos.p=0.000	E,G
TLR4	261	0	29	0	14.473	Pos.p=0.000	K,R
TLR4	301	0.5	35.5	0.631	16.429	Pos.p=0.000	D,E,G
TLR4	343	0	87	0	43.15	Pos.p=0.000	K,R
TLR4	383	0	39	0	17.943	Pos.p=0.000	H,Y
TLR4	410	0	6	0	4.367	Pos.p=0.011	L,M
TLR4	498	0	11	0	5.474	Pos.p=0.012	A,D,N,T
TLR4	611	0	69	0	37.842	Pos.p=0.000	Q,R
TLR5	6	0	11	0	6.236	Pos.p=0.005	Q,R
TLR5	18	0	10	0	4.438	Pos.p=0.065	D,Y
TLR5	61	0	6	0	3.862	Pos.p=0.086	Q,R
TLR5	65	0	9	0	4.377	Pos.p=0.033	S,T
TLR5	82	0	21	0	10.138	Pos.p=0.000	H,R
TLR5	83	0	8	0	3.971	Pos.p=0.041	I,V
TLR5	118	0	10	0	5	Pos.p=0.017	A,V
TLR5	122	0	17	0	8.312	Pos.p=0.001	I,L,R,S
TLR5	127	0	8	0	3.96	Pos.p=0.042	H,R
TLR5	204	0	9	0	3.953	Pos.p=0.084	S,T
TLR5	212	0	14	0	6.324	Pos.p=0.024	K,R
TLR5	217	0	8	0	3.794	Pos.p=0.060	I,V
TLR5	281	0	12	0	5.823	Pos.p=0.011	N,T
TLR5	285	0	6	0	3.125	Pos.p=0.081	S,T
TLR5	302	0	11	0	5.107	Pos.p=0.026	I,V
TLR5	314	0	7	0	3.332	Pos.p=0.083	H,R

Continued on next page

TLR	Codon	S	N	dS	dN	Selection	Variants
TLR5	315	0	12	0	5.232	Pos.p=0.040	N,S
TLR5	341	0	16	0	7.29	Pos.p=0.017	E,K
TLR5	343	0	6	0	3	Pos.p=0.088	L,V
TLR5	409	0	10	0	5	Pos.p=0.017	S,T
TLR5	413	0	32	0	14.637	Pos.p=0.000	G,S
TLR5	528	0	7	0	3.283	Pos.p=0.092	I,T
TLR5	553	0	17	0	6.904	Pos.p=0.035	I,T
TLR5	619	0	8	0	3.935	Pos.p=0.050	A,E
TLR5	623	0	10	0	4.363	Pos.p=0.068	M,V
TLR5	648	0	10	0	5.744	Pos.p=0.004	I,L
TLR5	830	0	12	0	5.37	Pos.p=0.029	F,V
TLR5	834	0	6	0	2.955	Pos.p=0.096	I,L
TLR5	837	0	9	0	5.637	Pos.p=0.032	Q,R
TLR5	848	0	6	0	3	Pos.p=0.088	G,V
TLR5	850	0	8	0	4	Pos.p=0.039	G,V
TLR5	859	0	10	0	5	Pos.p=0.017	G,V
TLR7	44	0	4	0	3.182	Pos.p=0.039	F,L
TLR7	121	0	40	0	18.792	Pos.p=0.000	I,V
TLR7	135	0	11	0	5.587	Pos.p=0.011	R,S,T
TLR7	159	0	8	0	3.934	Pos.p=0.045	H,R
TLR7	248	0	12	0	5.962	Pos.p=0.008	I,V
TLR7	447	0	39	0	18.689	Pos.p=0.000	G,S
TLR7	453	0	9	0	4.142	Pos.p=0.081	C,W
TLR7	456	0	8	0	3.556	Pos.p=0.100	H,P
TLR7	499	0	8	0	4.093	Pos.p=0.038	G,R
TLR7	546	0	9	0	4.158	Pos.p=0.053	S,T
TLR7	580	0	7	0	3.405	Pos.p=0.071	I,V
TLR7	710	0	11	0	5.313	Pos.p=0.025	K,R
TLR7	747	0	8	0	4.337	Pos.p=0.029	Q,R
TLR7	847	0	6	0	2.996	Pos.p=0.089	K,T
TLR7	896	0	3	0	2.349	Pos.p=0.095	L,S
TLR7	923	0	9	0	4.476	Pos.p=0.027	I,V
TLR15	34	0	10	0	4.622	Pos.p=0.038	F,S
TLR15	85	0	8	0	4	Pos.p=0.039	L,V
TLR15	87	0	8	0	3.726	Pos.p=0.074	H,Y
TLR15	125	0	10	0	7.567	Pos.p=0.000	L,P
TLR15	129	0	16	0	9.089	Pos.p=0.000	L,P
TLR15	134	5	16	4.345	9.271	Pos.p=0.096	L,S
TLR15	149	0	15	0	7.5	Pos.p=0.002	A,T
TLR15	158	0	11	0	4.885	Pos.p=0.058	E,K
TLR15	192	0	7	0	3.517	Pos.p=0.057	R,T
TLR15	210	0	6	0	3.067	Pos.p=0.077	G,R
TLR15	309	0	15	0	7.415	Pos.p=0.003	A,E
TLR15	321	0	10	0	4.923	Pos.p=0.020	I,V
TLR15	332	0	14	0	8.419	Pos.p=0.000	F,L
TLR15	350	0	6	0	2.984	Pos.p=0.091	H,R
TLR15	389	0	19	0	8.8	Pos.p=0.004	K,R
TLR15	430	0	6	0	2.943	Pos.p=0.098	I,T
TLR15	436	0	7	0	3.261	Pos.p=0.096	D,V
TLR15	438	0	8	0	3.819	Pos.p=0.056	I,S
TLR15	447	0	7	0	3.5	Pos.p=0.059	P,S
TLR15	560	0	6	0	2.975	Pos.p=0.092	H,R
TLR15	630	0	8	0	3.928	Pos.p=0.060	C,G
TLR15	696	0	24	0	12.08	Pos.p=0.000	K,R
TLR15	729	0	6	0	3.765	Pos.p=0.022	L,M
TLR15	731	0	8	0	3.994	Pos.p=0.040	I,T
TLR15	738	0	10	0	4.942	Pos.p=0.026	H,Y
TLR15	770	0	8	0	3.799	Pos.p=0.059	N,S
TLR15	803	0	7	0	3.267	Pos.p=0.095	D,G
TLR21	2	0	36	0	13.784	Pos.p=0.007	M,V
TLR21	9	0	39	0	25.536	Pos.p=0.001	R,W
TLR21	51	0	6	0	3	Pos.p=0.088	P,R
TLR21	66	0	6	0	4.571	Pos.p=0.007	L,P
TLR21	111	0	8	0	4	Pos.p=0.039	A,T
TLR21	145	0	6	0	2.971	Pos.p=0.093	D,G
TLR21	148	0	9	0	4.5	Pos.p=0.026	P,S
TLR21	149	0	9	0	4.705	Pos.p=0.028	L,M,S,T
TLR21	151	0	7	0	4.716	Pos.p=0.043	Q,R
TLR21	189	0	53	0	26.5	Pos.p=0.000	A,V
TLR21	195	0	6	0	3	Pos.p=0.088	A,V

Continued on next page

TLR	Codon	S	N	dS	dN	Selection	Variants
TLR21	218	0	13	0	6.5	Pos.p=0.005	A,T
TLR21	239	0	32	0	15.727	Pos.p=0.000	D,E,H,Q
TLR21	244	0	8	0	3.883	Pos.p=0.049	M,V
TLR21	259	0	15	0	7.5	Pos.p=0.002	A,V
TLR21	321	0	6	0	3.375	Pos.p=0.053	Q,R
TLR21	328	0	12	0	5.986	Pos.p=0.008	G,R
TLR21	344	0	8	0	4.051	Pos.p=0.041	G,R
TLR21	351	0	12	0	5.476	Pos.p=0.023	H,N,P,T
TLR21	443	0	8	0	3.59	Pos.p=0.093	N,S
TLR21	446	0	7	0	3.411	Pos.p=0.096	A,E
TLR21	517	0	8	0	3.992	Pos.p=0.041	C,R
TLR21	559	0	10	0	4.973	Pos.p=0.021	C,R
TLR21	635	0	10	0	4.585	Pos.p=0.041	A,D
TLR21	648	0	6	0	3.357	Pos.p=0.052	Q,R
TLR21	670	0	5	0	2.82	Pos.p=0.086	Q,R
TLR21	675	0	12	0	5.472	Pos.p=0.023	H,R
TLR21	691	0	13	0	7.328	Pos.p=0.002	Q,R
TLR21	757	0	6	0	2.998	Pos.p=0.088	I,L
TLR21	770	0	8	0	4.193	Pos.p=0.027	L,V
TLR21	782	0	5	0	3.944	Pos.p=0.013	L,R
TLR21	818	0	8	0	3.657	Pos.p=0.080	N,S
TLR21	848	0	9	0	4.366	Pos.p=0.034	I,S
TLR21	886	0	8	0	4.456	Pos.p=0.019	Q,R
TLR21	892	0	10	0	4.946	Pos.p=0.020	C,R
TLR21	917	0	6	0	3	Pos.p=0.088	G,V
TLR21	920	0	8	0	4.387	Pos.p=0.019	Q,R

Appendix E

Appendix to Chapter 6

Table E.1: Accession numbers for *TLR11*, *TLR12*, *TLR13* and *TLR21* sequences

Gene	Accession	Common name	Binomial name	Taxon	Short name [1]
TLR11	ENSMUSG00000051969	Mouse	*Mus musculus*	Mammal	N/A
TLR11	ENSEASP00005025576	Donkey	*Equus asinus*	Mammal	N/A
TLR11	ENSECAP00000004749	Horse	*Equus caballus*	Mammal	N/A
TLR11	ENSPSMP00000005910	Greater bamboo lemur	*Prolemur simus*	Mammal	N/A
TLR11	ENSSTOP00000017221	Thirteen-lined ground squirrel	*Ictidomys tridecemlineatus*	Mammal	N/A
TLR11	ENSUPAP00010010234	Arctic ground squirrel	*Urocitellus parryii*	Mammal	N/A
TLR11	ENSSVLP00005003665	Eurasian red squirrel	*Sciurus vulgaris*	Mammal	N/A
TLR11	ENSLAFP00000023990	African savanna elephant	*Loxodonta africana*	Mammal	N/A
TLR11	ENSCLAP00000001340	Long-tailed chinchilla	*Chinchilla lanigera*	Mammal	N/A
TLR11	ENSHGLP00000016079	Naked mole-rat	*Heterocephalus glaber*	Mammal	N/A
TLR11	ENSOCUP00000018317	Rabbit	*Oryctolagus cuniculus*	Mammal	N/A
TLR11	ENSRNOP00000051449	Norway rat	*Rattus norvegicus*	Mammal	N/A
TLR11	MGP_CAROLIEiJ_P0036456	Ryukyu mouse	*Mus caroli*	Mammal	N/A
TLR11	MGP_SPRETEiJ_P0038170	Algerian mouse	*Mus spretus*	Mammal	N/A
TLR11	ENSMSIP00000022378	Steppe mouse	*Mus spicilegus*	Mammal	N/A
TLR11	ENSMUSP00000138814	Mouse	*Mus musculus*	Mammal	N/A
TLR11	MGP_PahariEiJ_P0084567	Shrew mouse	*Mus pahari*	Mammal	N/A
TLR11	ENSCGRP00001010033	Chinese hamster	*Cricetulus griseus*	Mammal	N/A
TLR11	ENSPEMP00000000345	Northern American deer mouse	*Peromyscus maniculatus*	Mammal	N/A
TLR11	ENSMAUP00000021298	Golden hamster	*Mesocricetus auratus*	Mammal	N/A
TLR11	ENSNGAP00000008079	Blind mole rat	*Spalax galili*	Mammal	N/A
TLR12	ENSMUSG00000062545	Mouse	*Mus musculus*	Mammal	N/A
TLR12	ENSMLUP00000018360	Little brown bat	*Myotis lucifugus*	Mammal	N/A
TLR12	ENSEASP00005017139	Donkey	*Equus asinus*	Mammal	N/A
TLR12	ENSECAP00000003731	Horse	*Equus caballus*	Mammal	N/A
TLR12	ENSPSMP00000030700	Greater bamboo lemur	*Prolemur simus*	Mammal	N/A
TLR12	ENSMMMP00000009453	Alpine marmot	*Marmota marmota*	Mammal	N/A
TLR12	ENSUPAP00010007221	Arctic ground squirrel	*Urocitellus parryii*	Mammal	N/A

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR12	ENSSVLP00005000681	Eurasian red squirrel	*Sciurus vulgaris*	Mammal	N/A
TLR12	ENSLAFP00000020999	African savanna elephant	*Loxodonta africana*	Mammal	N/A
TLR12	ENSDNOP00000028847	Nine-banded armadillo	*Dasypus novemcinctus*	Mammal	N/A
TLR12	ENSOCUP00000022559	Rabbit	*Oryctolagus cuniculus*	Mammal	N/A
TLR12	ENSRNOP00000062596	Norway rat	*Rattus norvegicus*	Mammal	N/A
TLR12	MGP_PahariEiJ_P0079599	Shrew mouse	*Mus pahari*	Mammal	N/A
TLR12	MGP_SPRETEiJ_P0067019	Algerian mouse	*Mus spretus*	Mammal	N/A
TLR12	ENSMUSP00000074381	Mouse	*Mus musculus*	Mammal	N/A
TLR12	MGP_CAROLIEiJ_P0064506	Ryukyu mouse	*Mus caroli*	Mammal	N/A
TLR12	ENSCGRP00001002659	Chinese hamster	*Cricetulus griseus*	Mammal	N/A
TLR12	ENSPEMP00000032849	Northern American deer mouse	*Peromyscus maniculatus*	Mammal	N/A
TLR13	ENSSHAP00000015466	Tasmanian devil	*Sarcophilus harrisii*	Mammal	N/A
TLR13	ENSPCIP00000036481	Koala	*Phascolarctos cinereus*	Mammal	N/A
TLR13	ENSNGAP00000010238	Blind mole rat	*Nannospalax galili*	Mammal	N/A
TLR13	ENSMOCP00000010504	Vole	*Microtus ochrogaster*	Mammal	N/A
TLR13	ENSPEMP00000012577	Northern American deer mouse	*Peromyscus maniculatus*	Mammal	N/A
TLR13	ENSJJAP00000000619	Lesser Egyptian jerboa	*Jaculus jaculus*	Mammal	N/A
TLR13	ENSOCUP00000017543	Rabbit	*Oryctolagus cuniculus*	Mammal	N/A
TLR13	ENSLAFP00000026404	African savanna elephant	*Loxodonta africana*	Mammal	N/A
TLR13	ENSSVLP00005014962	Eurasian red squirrel	*Sciurus vulgaris*	Mammal	N/A
TLR13	ENSMMMP00000014965	Alpine marmot	*Marmota marmota*	Mammal	N/A
TLR13	ENSUPAP00010025565	Arctic ground squirrel	*Urocitellus parryii*	Mammal	N/A
TLR13	ENSSDAP00000017966	Daurian ground squirrel	*Spermophilus dauricus*	Mammal	N/A
TLR13	ENSMICP00000021023	Gray mouse lemur	*Microcebus murinus*	Mammal	N/A
TLR13	ENSPSMP00000036864	Greater bamboo lemur	*Prolemur simus*	Mammal	N/A
TLR13	ENSPCOP00000013822	Coquerel's sifaka	*Propithecus coquereli*	Mammal	N/A
TLR13	ENSMLUP00000017330	Little brown bat	*Myotis lucifugus*	Mammal	N/A
TLR13	ENSCWAP00000002250	Chacoan peccary	*Catagonus wagneri*	Mammal	N/A
TLR13	ENSOANP00000036363	Platypus	*Ornithorhynchus anatinus*	Mammal	N/A
TLR13	ENSMUSG00000033777	Mouse	*Mus musculus*	Mammal	N/A

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSLLEP00000020491	Leishan spiny toad	*Leptobrachium leishanense*	Amphibian	sp_toad
TLR21	ENSXETP00000089415	Tropical clawed frog	*Xenopus tropicalis*	Amphibian	cl_frog
TLR21	ENSCPUP00000001706	Ruff	*Philomachus pugnax*	Bird	ruff
TLR21	ENSCPGP00000010344	Spoon-billed sandpiper	*Calidris pygmaea*	Bird	sandpiper
TLR21	ENSLCOP00000013021	Blue-crowned manakin	*Lepidothrix coronata*	Bird	manakin
TLR21	ENSFALP00000015278	Collared flycatcher	*Ficedula albicollis*	Bird	flycatcher
TLR21	ENSCUSP00005012212	Swainson's thrush	*Catharus ustulatus*	Bird	sw_thrush
TLR21	ENSCMUP00000013557	New Caledonian crow	*Corvus moneduloides*	Bird	crow
TLR21	ENSPMJP00000020959	Great tit	*Parus major*	Bird	gr_tit
TLR21	ENSCCEP00000003666	Blue tit	*Cyanistes caeruleus*	Bird	bl_tit
TLR21	ENSCRFP00000018034	Rufous-capped babbler	*Cyanoderma ruficeps*	Bird	babbler
TLR21	ENSZLMP00000010892	Silver-eye	*Zosterops lateralis*	Bird	silvereye
TLR21	ENSSCAP00000007179	Common canary	*Serinus canaria*	Bird	canary
TLR21	ENSGFOP00000006629	Medium ground-finch	*Geospiza fortis*	Bird	gr_finch
TLR21	ENSCPVP00000007862	Small tree finch	*Camarhynchus parvulus*	Bird	tr_finch
TLR21	ENSJHYP00000009982	Dark-eyed junco	*Junco hyemalis*	Bird	junco
TLR21	ENSZALP00000009324	White-throated sparrow	*Zonotrichia albicollis*	Bird	wt_sparrow
TLR21	ENSZALP00000009326	White-throated sparrow	*Zonotrichia albicollis*	Bird	wt_sparrow
TLR21	ENSTGUP00000018795	Zebra finch	*Taeniopygia guttata*	Bird	ze_finch
TLR21	ENSTGUP00000029518	Zebra finch	*Taeniopygia guttata*	Bird	ze_finch
TLR21	ENSEGOP00005015553	Gouldian finch	*Chloebia gouldiae*	Bird	go_finch
TLR21	ENSEGOP00005020751	Gouldian finch	*Chloebia gouldiae*	Bird	go_finch
TLR21	ENSAZOP00000010184	Eastern spot-billed duck	*Anas zonorhyncha*	Bird	ea_duck
TLR21	ENSABRP00000005550	Pink-footed goose	*Anser brachyrhynchus*	Bird	pf_goose
TLR21	ENSACDP00005009358	Swan goose	*Anser cygnoides*	Bird	sw_goose
TLR21	ENSNMEP00000013996	Helmeted guineafowl	*Numida meleagris*	Bird	he_gfowl
TLR21	ENSCJPP00005014891	Japanese quail	*Coturnix japonica*	Bird	ja_quail
TLR21	ENSMGAP00000016520	Domestic turkey	*Meleagris gallopavo*	Bird	turkey
TLR21	ENSPCLP00000023325	Ring-necked pheasant	*Phasianus colchicus*	Bird	phesant
TLR21	ENSGALP00000072066	Chicken	*Gallus gallus*	Bird	chicken

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSPSTP00000019449	Indian peafowl	*Pavo cristatus*	Bird	peafowl
TLR21	ENSDNVP00000004916	Emu	*Dromaius novaehollandiae*	Bird	emu
TLR21	ENSAHAP00000026008	Great spotted kiwi	*Apteryx haastii*	Bird	gs_kiwi
TLR21	ENSARWP00000008712	Okarito brown kiwi	*Apteryx rowi*	Bird	ob_kiwi
TLR21	ENSAOWP00000008582	Little spotted kiwi	*Apteryx owenii*	Bird	ls_kiwi
TLR21	ENSFTIP00000020005	Common kestrel	*Falco tinnunculus*	Bird	kestrel
TLR21	ENSANIP00000014229	Eurasian sparrowhawk	*Accipiter nisus*	Bird	sphawk
TLR21	ENSSOCP00000013224	Northern spotted owl	*Strix occidentalis caurina*	Bird	ns_owl
TLR21	ENSEBUP00000008448	Inshore hagfish	*Eptatretus burgeri*	Fish	hagfish
TLR21	ENSECRP00000024209	Reedfish	*Erpetoichthys calabaricus*	Fish	reedfish
TLR21	ENSACIP00000026852	Midas cichlid	*Amphilophus citrinellus*	Fish	m_cichlid
TLR21	ENSACIP00000026841	Midas cichlid	*Amphilophus citrinellus*	Fish	m_cichlid
TLR21	ENSACIP00000026810	Midas cichlid	*Amphilophus citrinellus*	Fish	m_cichlid
TLR21	ENSPMAP00000005491	Sea lamprey gene	*Petromyzon marinus*	Fish	lamprey
TLR21	ENSLACP00000001202	Coelacanth	*Latimeria chalumnae*	Fish	coelacanth
TLR21	ENSPMAP00000011162	Sea lamprey gene	*Petromyzon marinus*	Fish	lamprey
TLR21	ENSDCDP00000009648	Denticle herring	*Denticeps clupeoides*	Fish	de_herring
TLR21	ENSPMAP00000011080	Sea lamprey gene	*Petromyzon marinus*	Fish	lamprey
TLR21	ENSELUP00000019757	Northern pike	*Esox lucius*	Fish	pike
TLR21	ENSOKIP00005001517	Coho salmon	*Oncorhynchus kisutch*	Fish	co_salmon
TLR21	ENSHHUP00000088563	Huchen	*Hucho hucho*	Fish	huchen
TLR21	ENSSSAP00000063294	Atlantic salmon	*Salmo salar*	Fish	at_salmon
TLR21	ENSOMYP00000001865	Rainbow trout	*Oncorhynchus mykiss*	Fish	r_trout
TLR21	ENSOTSP00005004165	Chinook salmon	*Oncorhynchus tshawytscha*	Fish	ch_salmon
TLR21	ENSOKIP00005001505	Coho salmon	*Oncorhynchus kisutch*	Fish	co_salmon
TLR21	ENSHHUP00000074461	Huchen	*Hucho hucho*	Fish	huchen
TLR21	ENSSTUP00000073379	Brown trout	*Salmo trutta*	Fish	br_trout
TLR21	ENSCHAP00000032036	Atlantic herring	*Clupea harengus*	Fish	at_herring
TLR21	ENSNMLP00000017734	Round goby	*Neogobius melanostomus*	Fish	goby
TLR21	ENSGMOP00000019648	Atlantic cod	*Gadus morhua*	Fish	cod

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSOSIP00000021748	Chinese medaka	*Oryzias sinensis*	Fish	ch_medaka
TLR21	ENSORLP00015026176	Japanese medaka	*Oryzias latipes*	Fish	ja_medaka
TLR21	ENSORLP00000016852	Japanese medaka	*Oryzias latipes*	Fish	ja_medaka
TLR21	ENSORLP00020022336	Japanese medaka	*Oryzias latipes*	Fish	ja_medaka
TLR21	ENSOMEP00000031396	Indian medaka	*Oryzias dancena*	Fish	in_medaka
TLR21	ENSOJAP00000041542	Javanese ricefish	*Oryzias javanicus*	Fish	ja_rfish
TLR21	ENSGWIP00000019955	Blunt-snouted clingfish	*Gouania willdenowi*	Fish	bs_clfish
TLR21	ENSNFUP00015006333	Turquoise killifish	*Nothobranchius furzeri*	Fish	tu_kfish
TLR21	ENSHCOP00000009215	Tiger tail seahorse	*Hippocampus comes*	Fish	seahorse
TLR21	ENSTNIP00000011033	Spotted green pufferfish	*Dichotomyctere nigroviridis*	Fish	sg_pfish
TLR21	ENSTRUP00000009723	Fugu	*Takifugu rubripes*	Fish	fugu
TLR21	ENSCVAP00000006226	Sheepshead minnow	*Cyprinodon variegatus*	Fish	sh_minnow
TLR21	ENSFHEP00000025152	Mummichog	*Fundulus heteroclitus*	Fish	mummichog
TLR21	ENSPREP00000020999	Guppy	*Poecilia reticulata*	Fish	guppy
TLR21	ENSPLAP00000023104	Sailfin molly	*Poecilia latipinna*	Fish	sa_molly
TLR21	ENSPFOP00000020442	Amazon molly	*Poecilia formosa*	Fish	am_molly
TLR21	ENSPMEP00000010332	Shortfin molly	*Poecilia mexicana*	Fish	sh_molly
TLR21	ENSGAFP00000015561	Western mosquitofish	*Gambusia affinis*	Fish	w_mofish
TLR21	ENSXCOP00000025614	Monterrey platyfish	*Xiphophorus couchianus*	Fish	mo_pfish
TLR21	ENSXMAP00000010790	Southern platyfish	*Xiphophorus maculatus*	Fish	so_pfish
TLR21	ENSSFAP00005003714	Jewelled blenny	*Salarias fasciatus*	Fish	je_blenny
TLR21	ENSSORP00005054894	Orbiculate cardinalfish	*Sphaeramia orbicularis*	Fish	cardfish
TLR21	ENSOABP00000023054	Blue tilapia	*Oreochromis aureus*	Fish	tilapia
TLR21	ENSHBUP00000018839	Burton's mouthbrooder	*Astatotilapia burtoni*	Fish	mbrooder
TLR21	ENSMZEP00005003421	Zebra mbuna	*Maylandia zebra*	Fish	mbuna
TLR21	ENSPNYP00000020645	Makobe island cichlid	*Pundamilia Nyererei*	Fish	ma_cichlid
TLR21	ENSACLP00000039073	Eastern happy	*Astatotilapia calliptera*	Fish	ea_happy
TLR21	ENSKMAP00000014068	Mangrove rivulus	*Kryptolebias marmoratus*	Fish	ma_rivulus
TLR21	ENSMMDP00005043289	Pinecone soldierfish	*Myripristis murdjan*	Fish	solfish
TLR21	ENSBSLP00000008575	Siamese fighting fish	*Betta splendens*	Fish	siam_ffish

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSPRNP00000056807	Indian glassy fish	*Parambassis ranga*	Fish	In_glafish
TLR21	ENSLBEP00000028582	Ballan wrasse	*Labris bergylta*	Fish	ba_wrasse
TLR21	ENSSAUP00010035869	Gilthead seabream	*Sparus aurata*	Fish	gh_sbream
TLR21	ENSSPAP00000005826	Bicolor damselfish	*Stegastes partitus*	Fish	bi_damsel
TLR21	ENSAPOP00000020373	Spiny chromis	*Acanthochromis polyacanthus*	Fish	sp_chrom
TLR21	ENSAPEP00000013685	Orange clownfish	*Amphiprion percula*	Fish	or_clfish
TLR21	ENSAOCP00000006753	Clown anemonefish	*Amphiprion ocellaris*	Fish	cl_anfish
TLR21	ENSENLP00000042002	Live sharksucker	*Echeneis naucrates*	Fish	shsucker
TLR21	ENSGACP00000012342	Three-spined stickleback	*Gasterosteus aculeatus*	Fish	stickback
TLR21	ENSLCRP00005027006	Large yellow croaker	*Larimichthys crocea*	Fish	croaker
TLR21	ENSATEP00000027694	Climbing perch	*Anabas testudineus*	Fish	perch
TLR21	ENSMAMP00000007222	Zig-zag eel	*Mastacembelus armatus*	Fish	zz_eel
TLR21	ENSMALP00000020543	Swamp eel	*Monopterus albus*	Fish	swamp_eel
TLR21	ENSSMAP00000014138	Turbot	*Scophthalmus maximus*	Fish	turbot
TLR21	ENSCGOP00000020109	Channel bull blenny	*Cottoperca gobio*	Fish	bu_blenny
TLR21	ENSLCAP00010000155	Barramundi perch	*Lates calcarifer*	Fish	b_perch
TLR21	ENSSLDP00000000331	Yellowtail amberjack	*Seriola lalandi*	Fish	yt_amber
TLR21	ENSSLDP00000001480	Yellowtail amberjack	*Seriola lalandi*	Fish	yt_amber
TLR21	ENSDLAP00005001729	European seabass	*Dicentrarchus labrax*	Fish	eu_seabass
TLR21	ENSSLUP00000046320	Pike-perch	*Sander lucioperca*	Fish	pike_perch
TLR21	ENSCLMP00005029014	Lumpfish	*Cyclopterus lumpus*	Fish	lumpfish
TLR21	ENSPKIP00000003185	Elephantfish	*Paramormyrops kingsleyae*	Fish	elfish
TLR21	ENSSFOP00015034568	Asian bonytongue	*Scleropages formosus*	Fish	bonytongue
TLR21	ENSDARP00000098992	Zebrafish	*Danio rerio*	Fish	zebrafish
TLR21	ENSCCRP00015036516	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00010018808	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00000074184	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00020086209	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00000056794	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00020053509	Common carp	*Cyprinus carpio*	Fish	co_carp

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSCCRP00015053584	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00010023822	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSCCRP00000010711	Common carp	*Cyprinus carpio*	Fish	co_carp
TLR21	ENSSANP00000102392	Blind barbel	*Sinocyclocheilus anshuiensis*	Fish	bl_barbel
TLR21	ENSSRHP00000073110	Horned golden-line barbel	*Sinocyclocheilus rhinocerous*	Fish	hgl_barbel
TLR21	ENSSGRP00000068258	Golden-line barbel	*Sinocyclocheilus grahami*	Fish	gl_barbel
TLR21	ENSSGRP00000042547	Golden-line barbel	*Sinocyclocheilus grahami*	Fish	gl_barbel
TLR21	ENSCARP00000074642	Goldfish	*Carassius auratus*	Fish	goldfish
TLR21	ENSCARP00000074623	Goldfish	*Carassius auratus*	Fish	goldfish
TLR21	ENSSGRP00000013558	Golden-line barbel	*Sinocyclocheilus grahami*	Fish	gl_barbel
TLR21	ENSSRHP00000045239	Horned golden-line barbel	*Sinocyclocheilus rhinocerous*	Fish	hgl_barbel
TLR21	ENSSANP00000043363	Blind barbel	*Sinocyclocheilus anshuiensis*	Fish	bl_barbel
TLR21	ENSAMXP00000025842	Mexican tetra	*Astyanax mexicanus*	Fish	mex_tetra
TLR21	ENSAMXP00005004233	Mexican tetra	*Astyanax mexicanus*	Fish	mex_tetra
TLR21	ENSIPUP00000009226	Channel catfish	*Ictalurus punctatus*	Fish	ch_catfish
TLR21	ENSPNAP00000032422	Red-bellied piranha	*Pygocentrus nattereri*	Fish	rb_piranha
TLR21	ENSEEEP00000007284	Electric eel	*Electrophorus electricus*	Fish	el_eel
TLR21	ENSVKKP00000009554	Komodo dragon	*Varanus komodoensis*	Reptile	k_dragon
TLR21	ENSNNAP00000021111	Indian cobra	*Naja naja*	Reptile	cobra
TLR21	ENSLLTP00000002568	Blue-ringed sea krait	*Laticauda laticaudata*	Reptile	krait
TLR21	ENSNSUP00000013205	Mainland tiger snake	*Notechis scutatus*	Reptile	ti_snake
TLR21	ENSPTXP00000027730	Eastern brown snake	*Pseudonaja textilis*	Reptile	br_snake
TLR21	ENSSMRP00000007455	Argentine black and white tegu	*Salvator merianae*	Reptile	tegu
TLR21	ENSPMRP00000012481	Common wall lizard	*Podarcis muralis*	Reptile	wa_lizard
TLR21	ENSPVIP00000003831	Central bearded bragon	*Pogona vitticeps*	Reptile	b_dragon
TLR21	ENSACAP00000014787	Green anole	*Anolis carolinensis*	Reptile	anole
TLR21	ENSPSIP00000015712	Chinese softshell turtle	*Pelodiscus sinensis*	Reptile	ssh_turtle
TLR21	ENSPCEP00000012878	West African mud turtle	*Pelusios castaneus*	Reptile	mu_turtle
TLR21	ENSCSRP00000010377	Common snapping turtle	*Chelydra serpentina*	Reptile	sn_turtle
TLR21	ENSCSRP00000010509	Common snapping turtle	*Chelydra serpentina*	Reptile	sn_turtle

Cont. on next page

Gene	Accession	Common name	Binomial name	Taxon	Short name
TLR21	ENSCPBP00000010867	Western painted turtle	*Chrysemys picta*	Reptile	wp_turtle
TLR21	ENSTMTP00000019067	Three-toed box turtle	*Terrapene carolina triunguis*	Reptile	tt_turtle
TLR21	ENSTMTP00000019087	Three-toed box turtle	*Terrapene carolina triunguis*	Reptile	tt_turtle
TLR21	ENSCABP00000026918	Abingdon island giant tortoise	*Chelonoidis abingdonii*	Reptile	ai_tortoi
TLR21	ENSGAGP00000008333	Agassiz's desert tortoise	*Gopherus agassizii*	Reptile	ad_tortoi

[1] Species names which were annotated in trees were given a shortened name.

Table E.2: Results from PAML branch-site analysis of *TLR21*.

Site	Chicken AA	Posterior probability
161	SER	0.991**
224	GLN	0.999**
236	GLY	1.000**
264	HIS	1.000**
285	ASP	0.954*
361	ARG	1.000**
599	GLN	0.989*
689	ALA	0.983*
693	SER	1.000**

Appendix F

Appendix to Chapter 7

Table F.1: Variants in swan candidate genes with respect to the swan goose (*Anser cygnoides*).

Gene	Exon	Position (in CDS)	Change	S/NS	AA Change
RIG-I	1	3	G→T	S	-
RIG-I	1	12	A→G	S	-
RIG-I	2	308	A→G	NS	D→G
RIG-I	2	316	A→G	NS	K→E
RIG-I	4	372	A→C	S	-
RIG-I	4	375	T→C	S	-
RIG-I	4	434	C→A	NS	A→D
RIG-I	5	446	A→G	NS	D→G
RIG-I	6	687	A→G	S	-
RIG-I	7	864	C→T	S	-
RIG-I	7	943	C→T	NS	P→S
RIG-I	10	1229	G→A	NS	G→D
RIG-I	10	1332	G→A	S	-
RIG-I	10	1374	T→C	S	-
RIG-I	11	1395	T→C	S	-
RIG-I	11	1404	T→C	S	-
RIG-I	11	1452	C→T	S	-
RIG-I	12	1557	C→T	S	-
RIG-I	13	1702	A→G	NS	I→V
RIG-I	13	1759	A→G	NS	T→A
RIG-I	14	1833	A→T	S	-
RIG-I	15	1952	G→A	NS	G→D
RIG-I	16	2089	C→T	NS	H→Y
RIG-I	16	2166	G→A	S	-
RIG-I	16	2196	A→G	S	-
RIG-I	16	2474	G→A	NS	R→K

Continued on next page

Gene	Exon	Position (in CDS)	Change	S/NS	AA Change
TLR3	1	30	C→G	NS	S→R
TLR3	1	75	T→C	S	-
TLR3	1	81	A→G	S	-
TLR3	1	192	G→A	S	-
TLR3	1	201	G→A	S	-
TLR3	1	246	T→C	S	-
TLR3	1	342	A→G	S	-
TLR3	1	354	T→A	S	-
TLR3	1	355	G→A	NS	A→T
TLR3	1	366	C→T	S	-
TLR3	2	497	T→G	NS	L→W
TLR3	2	511	A→C	NS	K→Q
TLR3	2	532	C→T	S	-
TLR3	2	600	T→A	NS	N→K
TLR3	2	618	G→A	S	-
TLR3	3	682	C→T	S	-
TLR3	3	709	T→C	NS	C→R
TLR3	3	732	A→G	S	-
TLR3	3	776	A→T	NS	K→M
TLR3	3	826	A→C	NS	I→L
TLR3	3	834	T→C	S	-
TLR3	3	879	T→A	S	-
TLR3	3	993	A→G	S	-
TLR3	3	998	T→C	NS	I→T
TLR3	3	1026	C→T	S	-
TLR3	3	1062	C→T	S	-
TLR3	3	1170	G→A	S	-
TLR3	3	1179	A→G	S	-
TLR3	3	1256	G→T	NS	C→F
TLR3	3	1278	C→A	S	-
TLR3	3	1395	A→G	S	-
TLR3	3	1425	C→G	S	-
TLR3	3	1506	T→C	S	-
TLR3	3	1531	A→G	NS	I→V
TLR3	3	1590	C→T	S	-
TLR3	3	1623	G→A	S	-
TLR3	3	1853-1854	AA→TG	NS	E→V
TLR3	3	1958	C→T	NS	A→V
TLR3	3	1972	A→G	NS	I→V
TLR3	3	2010	G→T	S	-
TLR3	3	2106	T→A	S	-
TLR3	3	2124	T→C	S	-
TLR3	3	2182	A→G	NS	I→V
TLR3	3	2257	T→G	NS	S→A
TLR3	3	2268	T→C	S	-

Continued on next page

Gene	Exon	Position (in CDS)	Change	S/NS	AA Change
TLR3	3	2389, 2341	AGT→TGC	NS	S→C
TLR7	2	237	T→C	S	-
TLR7	2	249	A→G	S	-
TLR7	2	294	A→G	S	-
TLR7	2	1009	A→G	NS	N→D
TLR7	2	1113	G→A	S	-
TLR7	2	1545	C→T	S	-
TLR7	2	1593	T→C	S	-
TLR7	2	1676	G→A	NS	S→N
TLR7	2	1821	G→A	S	-
TLR7	2	1861	T→C	S	-
TLR7	2	1907	C→A	NS	S→Y
TLR7	2	1910	C→G	NS	A→G
TLR7	2	1934	A→C	NS	E→A
TLR7	2	2119	T→C	S	-
TLR7	2	2491	T→C	S	-
TLR7	2	2553	A→G	S	-
TLR7	2	2613	G→A	S	-
TLR7	2	2673	A→G	S	-
TLR7	2	2904	A→G	S	-
TLR7	2	2968	T→A	NS	S→T

Appendix G

Additional publications

G.1 Additional publications and manuscripts in preparation

A list of publications and manuscripts in preparation to which I have contributed during my DPhil is as follows:

G.1.1 Published manuscripts and book chapter

1. **Steven R Fiddaman**, Michal Vinkler, Simon G Spiro, Hila Levy, Christopher A Emerling, Amy C Boyd, Evangelos A. Dimopoulos, Juliana A Vianna, Theresa L Cole, Hailin Pan, Miaoquan Fang, Guojie Zhang, Tom Hart, Laurent AF Frantz, Adrian L Smith, Adaptation and cryptic pseudogenization in penguin Toll-like receptors, *Molecular Biology and Evolution*, msab354, https://doi.org/10.1093/molbev/msab354.

2. Robert Dixon, Stephen G Preston, Stefan Dascalu, Patrik G Flammer, **Steven R Fiddaman**, Kirstie McLoughlin, Amy Boyd, Jiri Volf, Ivan Rychlik, Michael B Bonsall, Bernd Kaspers, Adrian L Smith, Repertoire analysis of γδ T cells in the chicken enables functional annotation of the genomic region revealing highly variable pan-tissue TCR gamma V gene usage as well as identifying public and private repertoires, *BMC Genomics* 22, 719 (2021). https://doi.org/10.1186/s12864-021-08036-9.

3. Adrian Smith, **Steven R Fiddaman**, Pattern Recognition Receptors, in: *Avian Immunology* (3rd Edition), 2021, Bernd Kaspers, Karel Schat, Thomas Göbel, Lonneke Vervelde, Elsevier Academic Press.

4. Levy, H.; **Fiddaman, S.R.**; Djurhuus, A.; Black, C.E.; Kraberger, S.; Smith, A.L.; Hart, T.; Varsani, A. Identification of Circovirus Genome in a Chinstrap Penguin (*Pygoscelis antarcticus*) and Adélie Penguin (*Pygoscelis adeliae*) on the Antarctic Peninsula. *Viruses* 2020, 12, 858. https://doi.org/10.3390/v12080858

5. Hila Levy*, **Steven R Fiddaman***, Juliana A Vianna, Daly Noll, Gemma V Clucas, Jasmine K H Sidhu, Michael J Polito, Charles A Bost, Richard A Phillips, Sarah Crofts, Gary D Miller, Pierre Pistorius, Francesco Bonnadonna, Céline Le Bohec, Andrés Barbosa, Phil Trathan, Andrea Raya Rey, Laurent A F Frantz, Tom Hart, Adrian L Smith, Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird, *Molecular Biology and Evolution*, Volume 37, Issue 6, June 2020, Pages 1708–1726, https://doi.org/10.1093/molbev/msaa040

6. Theresa L. Cole, Ludovic Dutoit, Nicolas Dussex, Tom Hart, Alana Alexander, Jane L. Younger, Gemma V. Clucas, María José Frugone, Yves Cherel, Richard Cuthbert, Ursula Ellenberg, **Steven R. Fiddaman**, Johanna Hiscock, David Houston, Pierre Jouventin, Thomas Mattern, Gary Miller, Colin Miskelly, Paul Nolan, Michael J. Polito, Petra Quillfeldt, Peter G. Ryan, Adrian Smith, Alan J. D. Tennyson, David Thompson, Barbara Wienecke, Juliana A. Vianna, Jonathan M. Waters, Receding ice drove parallel expansions in Southern Ocean penguins, *Proceedings of the National Academy of Sciences*, Dec 2019, 116 (52) 26690-26696; DOI: 10.1073/pnas.1904048116

7. Hailin Pan, Theresa L Cole, Xupeng Bi, Miaoquan Fang, Chengran Zhou, Zhengtao Yang, Daniel T Ksepka, Tom Hart, Juan L Bouzat, Lisa S Argilla, Mads F Bertelsen, P Dee Boersma, Charles-André Bost, Yves Cherel, Peter Dann, **Steven R Fiddaman**, Pauline Howard, Kim Labuschagne, Thomas Mattern, Gary Miller, Patricia Parker, Richard A Phillips, Petra Quillfeldt, Peter G Ryan, Helen Taylor, David R Thompson, Melanie J Young, Martin R Ellegaard, M Thomas P Gilbert, Mikkel-Holger S Sinding, George Pacheco, Lara D Shepherd, Alan J D Tennyson, Stefanie Grosser, Emily Kay, Lisa J Nupen, Ursula Ellenberg, David M Houston, Andrew Hart Reeve, Kathryn Johnson, Juan F Masello, Thomas Stracke, Bruce McKinlay, Pablo García Borboroglu, De-Xing Zhang, Guojie Zhang, High-coverage genomes to elucidate the evolution of penguins, *GigaScience*, Volume 8, Issue 9, September 2019, giz117, https://doi.org/10.1093/gigascience/giz117

8. Theresa L Cole, Daniel T Ksepka, Kieren J Mitchell, Alan J D Tennyson, Daniel B Thomas, Hailin Pan, Guojie Zhang, Nicolas J Rawlence, Jamie R Wood, Pere Bover, Juan L Bouzat, Alan Cooper, **Steven R Fiddaman**, Tom Hart, Gary Miller, Peter G Ryan, Lara D Shepherd, Janet M Wilmshurst, Jonathan M Waters, Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation, *Molecular Biology and Evolution*, Volume 36, Issue 4, April 2019, Pages 784–797, https://doi.org/10.1093/molbev/msz017

9. Hill Sarah C., Hansen Rowena, Watson Samantha, Coward Vivien, Russell Christine, Cooper Jayne, Essen Steve, Everest Holly, Parag Kris V., **Fiddaman Steven**, Reid Scott, Lewis Nicola, Brookes Sharon M., Smith Adrian

L., Sheldon Ben, Perrins Christopher M., Brown Ian H. and Pybus Oliver G., Comparative micro-epidemiology of pathogenic avian influenza virus outbreaks in a wild bird population, *Phil. Trans. R. Soc. B*, 2019, 374:20180259 http://doi.org/10.1098/rstb.2018.0259

G.1.2 Manuscript in review

1. Theresa L. Cole†, Chengran Zhou†, Miaoquan Fang†, Hailin Pan*, Daniel T. Ksepka*, Daniel Thomas*, **Steven R. Fiddaman***, Christopher Emerling*, Xupeng Bi, Qi Fang, Martin R. Ellegaard, Adrian L. Smith, Tracy Heath, Alan J. D. Tennyson, Pablo García Borboroglu, Jamie R. Wood, Charles-André Bost, Peter Hadden, Misha Vorobyev, Jason Turuwhenua, Yves Cherel, Thomas Mattern, Tom Hart, Mikkel-Holger S. Sinding, Lara D. Shepherd, Richard A. Phillips, Petra Quillfeldt, Juan F. Masello, Stefanie Grosser, Juan L. Bouzat, Peter Ryan, David Thomas, Ursula Ellenberg, Peter Dann, Gary Miller, P. Dee Boersma, M. Thomas P. Gilbert, De-Xing Zhang, Guojie Zhang, A comprehensive genomic history of the evolution of penguins. *In Review: Nature Communications*.

G.1.3 Pre-submission manuscript

1. Nick A. Ciccone, **Steven R. Fiddaman**, Karen Staines, Amy Boyd, Holly Shelton, Colin Butter, Adrian L. Smith, Avian Influenza virus disrupts macrophage responses to Lipopolysaccharide through mechanisms involving repression of host genomic 5-hydroxymethylcytosine levels. Manuscript in preparation.

Bibliography

[1] Abbas, Y. M., Pichlmair, A., Górna, M. W., Superti-Furga, G., and Nagar, B. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature 494*, 7435 (2013), 60–4.

[2] Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7* (2013), Unit7.20.

[3] Ainsworth, G., and Rewell, R. The Incidence of Aspergillosis in Captive Wild Birds. *Journal of Comparative Pathology and Therapeutics 59* (1949), 213–217.

[4] Ajibade, A. A., Wang, H. Y., and Wang, R. F. Cell type-specific function of TAK1 in innate immune signaling. *Trends Immunol 34*, 7 (2013), 307–316.

[5] Akashi-Takamura, S., and Miyake, K. TLR accessory molecules. *Curr Opin Immunol 20*, 4 (2008), 420–425.

[6] Akira, S., and Takeda, K. Toll-like receptor signalling. *Nature Reviews Immunology 4*, 7 (2004), 499–511.

[7] Akira, S., Takeda, K., and Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature Immunology 2*, 8 (2001), 675–680.

[8] Akira, S., Uematsu, S., and Takeuchi, O. Pathogen recognition and innate immunity. *Cell 124*, 4 (2006), 783–801.

[9] Albrecht, J. C., Nicholas, J., Cameron, K. R., Newman, C., Fleckenstein, B., and Honess, R. W. Herpesvirus saimiri has a gene specifying a homologue of the cellular membrane glycoprotein CD59. *Virology 190*, 1 (1992), 527–30.

[10] Alcaide, M., and Edwards, S. V. Molecular Evolution of the Toll-Like Receptor Multigene Family in Birds. *Molecular Biology and Evolution 28*, 5 (2011), 1703–1715.

[11] Alders, R. G. Making Newcastle disease vaccines available at village level. *Vet Rec 174*, 20 (2014), 502–3.

[12] Aljanabi, S. M., and Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research 25*, 22 (1997), 4692–4693.

[13] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *J Mol Biol 215*, 3 (1990), 403–410.

[14] Andersen-Nissen, E., Smith, K. D., Strobe, K. L., Barrett, S. L. R., Cookson, B. T., Logan, S. M., and Aderem, A. Evasion of Toll-like receptor 5 by flagellated bacteria. *Proceedings of the National Academy of Sciences of the United States of America 102*, 26 (2005), 9247–9252.

[15] Anderson, K. V., Bokla, L., and Nüsslein-Volhard, C. Establishment of dorsal-ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product. *Cell 42*, 3 (1985), 791–798.

[16] Andrade, W. A., Souza, M. C., Ramos-Martinez, E., Nagpal, K., Dutra, M. S., Melo, M. B., Bartholomeu, D. C., Ghosh, S., Golenbock, D. T., and Gazzinelli, R. T. Combined action of nucleic acid-sensing Toll-like receptors and TLR11/TLR12 heterodimers imparts resistance to Toxoplasma gondii in mice. *Cell Host Microbe 13*, 1 (2013), 42–53.

[17] Antonopoulou, A., Baziaka, F., Tsaganos, T., Raftogiannis, M., Koutoukas, P., Spyridaki, A., Mouktaroudi, M., Kotsaki, A., Savva, A., Georgitsi, M., and Giamarellos-Bourboulis, E. J. Role of tumor necrosis factor gene single nucleotide polymorphisms in the natural course of 2009 influenza A H1N1 virus infection. *Int J Infect Dis 16*, 3 (2012), e204–e208.

[18] APHA. Situation assessment following detection and spread of H5 HPAI in EU Member States since October 2016, 2017.

[19] APHA. Rapid Risk Assessment on the finding of H5N6 HPAI in wild birds in England and Wales, JAN 2018.

[20] APHA. Situation Assessment: Findings of H5N6 HPAI in wild birds, JAN 2018.

[21] APHA. Highly pathogenic avian influenza (HPAI) in the UK, and Europe: Updated Outbreak Assessment no. 10 (VITT/1200 HPAI in the UK and Europe), 2020.

[22] APHA. Avian influenza in wild birds: 2020, 2021.

[23] APHA. Updated Outbreak Assessment no. 12: Highly pathogenic avian influenza (HPAI) in the UK, and Europe, JAN 2021.

[24] Arbour, N. C., Lorenz, E., Schutte, B. C., Zabner, J., Kline, J. N., Jones, M., Frees, K., Watt, J. L., and Schwartz, D. A. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet 25*, 2 (2000), 187–191.

[25] Areal, H., Abrantes, J., and Esteves, P. J. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol 11* (2011), 368.

[26] Ariel, E., Nicolajsen, N., Christophersen, M.-B., Holopainen, R., Tapiovaara, H., and Bang Jensen, B. Propagation and isolation of ranaviruses in cell culture. *Aquaculture 294*, 3 (2009), 159–164.

[27] Armstrong, J., Hickey, G., Diekhans, M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Haussler, D., Alföldi, J., Lindblad-Toh, K., Karlsson, E., Jarvis, E. D., Zhang, G., and Paten, B. Progressive alignment with Cactus: a multiple-genome aligner for the thousand-genome era. *bioRxiv* (2019).

[28] Asea, A., Rehli, M., Kabingu, E., Boch, J. A., Bare, O., Auron, P. E., Stevenson, M. A., and Calderwood, S. K. Novel signal transduction pathway utilized by extracellular HSP70: role of toll-like receptor (TLR) 2 and TLR4. *J Biol Chem 277*, 17 (2002), 15028–15034.

[29] Astill, J., Dara, R. A., Fraser, E. D. G., and Sharif, S. Detecting and Predicting Emerging Disease in Poultry With the Implementation of New Technologies and Big Data: A Focus on Avian Influenza Virus. *Front Vet Sci 5* (2018), 263.

[30] Austin, G., Collier, M., Calbrade, N., Hall, C., and Musgrove, A. Waterbirds in the UK 2006/07: The Wetland Bird Survey. Report, BTO/WWT/RSPB/JNCC, 2008.

[31] Azoulay, E., Russell, L., Van de Louw, A., Metaxa, V., Bauer, P., Povoa, P., Montero, J., Loeches, I. M., Mehta, S., Puxty, K., Schellongowski, P., Rello, J., Mokart, D., Lemiale, V., Mirouse, A., and the Nine-i Investigators. Diagnosis of severe respiratory infections in immunocompromised patients. *Intensive Care Medicine 46*, 2 (2020), 298–314.

[32] Baas, T., Baskin, C. R., Diamond, D. L., García-Sastre, A., Bielefeldt-Ohmann, H., Tumpey, T. M., Thomas, M. J., Carter, V. S., Teal, T. H., Van Hoeven, N., Proll, S., Jacobs, J. M., Caldwell, Z. R., Gritsenko, M. A., Hukkanen, R. R., Camp, D. G., Smith, R. D., and Katze, M. G. Integrated molecular signature of disease: analysis of influenza virus-infected macaques through functional genomics and proteomics. *J Virol 80*, 21 (2006), 10813–10828.

[33] Babik, W., Dudek, K., Fijarczyk, A., Pabijan, M., Stuglik, M., Szkotak, R., and Zieliński, P. Constraint and adaptation in newt toll-like receptor genes. *Genome Biol Evol 7*, 1 (2014), 81–95.

[34] Backström, N., Zhang, Q., and Edwards, S. V. Evidence from a house finch (Haemorhous mexicanus) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds. *Mol Biol Evol 30*, 5 (2013), 1046–50.

[35] Bae, Y. S., Lee, J. H., Choi, S. H., Kim, S., Almazan, F., Witztum, J. L., and Miller, Y. I. Macrophages generate reactive oxygen species in response to minimally oxidized low-density lipoprotein: toll-like receptor 4- and

spleen tyrosine kinase-dependent activation of NADPH oxidase 2. *Circ Res 104*, 2 (2009), 210–218.

[36] Bainová, H., Králová, T., Bryjová, A., Albrecht, T., Bryja, J., and Vinkler, M. First evidence of independent pseudogenization of toll-like receptor 5 in passerine birds. *Dev Comp Immunol 45*, 1 (2014), 151–5.

[37] Baker, A. J., Pereira, S. L., Haddrath, O. P., and Edge, K. A. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B-Biological Sciences 273*, 1582 (2006), 11–17.

[38] Balachandran, S., Roberts, P. C., Brown, L. E., Truong, H., Pattnaik, A. K., Archer, D. R., and Barber, G. N. Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection. *Immunity 13*, 1 (2000), 129–141.

[39] Balakirev, E. S., and Ayala, F. J. Is esterase-P encoded by a cryptic pseudogene in Drosophila melanogaster? *Genetics 144*, 4 (1996), 1511–1518.

[40] Bandelt, H. J., Forster, P., and Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution 16*, 1 (1999), 37–48.

[41] Barber, M. F., and Elde, N. C. Escape from bacterial iron piracy through rapid evolution of transferrin. *Science 346*, 6215 (2014), 1362–6.

[42] Barber, M. R., Aldridge, J. R., Webster, R. G., and Magor, K. E. Association of RIG-I with innate immunity of ducks to influenza. *Proc Natl Acad Sci U S A 107*, 13 (2010), 5913–5918.

[43] Barbosa, A., Benzal, J., Vidal, V., D'Amico, V., Coria, N., Diaz, J., Motas, M., Palacios, M. J., Cuervo, J. J., Ortiz, J., and Chitimia, L. Seabird ticks (Ixodes uriae) distribution along the Antarctic Peninsula. *Polar Biology 34*, 10 (2011), 1621–1624.

[44] Barbosa, A., De Mas, E., Benzal, J., Diaz, J. I., Motas, M., Jerez, S., Pertierra, L., Benayas, J., Justel, A., Lauzurica, P., Javier Garcia-Pena, F., and Serrano, T. Pollution and physiological variability in gentoo penguins at two rookeries with different levels of human visitation. *Antarctic Science 25*, 2 (2013), 329–338.

[45] Barbosa, A., and Jose Palacios, M. Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology 32*, 8 (2009), 1095–1115.

[46] Barreiro, L. B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J. K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., Kidd, J. R., Kidd, K. K., Alcais, A., Ragimbeau, J., Pellegrini, S., Abel, L., Casanova, J.-L., and Quintana-Murci, L. Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. *Plos Genetics 5*, 7 (2009).

[47] Barreiro, L. B., and Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet 11*, 1 (2010), 17–30.

[48] Barton, G. M., and Kagan, J. C. A cell biological view of Toll-like receptor function: regulation through compartmentalization. *Nat Rev Immunol 9*, 8 (2009), 535–542.

[49] Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A., and Cosset, F. L. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem 278*, 43 (2003), 41624–30.

[50] Bauer, S., Kirschning, C. J., Häcker, H., Redecke, V., Hausmann, S., Akira, S., Wagner, H., and Lipford, G. B. Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc Natl Acad Sci U S A 98*, 16 (2001), 9237–9242.

[51] Baxter, A. G., and Hodgkin, P. D. Activation rules: the two-signal theories of immune activation. *Nat Rev Immunol 2*, 6 (2002), 439–446.

[52] Baylis, A., Crofts, S., and Wolfaardt, A. Population trends of gentoo penguins pygoscelis papua breeding at the falkland islands. *Marine Ornithology 41*, 1 (2013), 1–5.

[53] Bayly-Jones, C., Bubeck, D., and Dunstone, M. A. The mystery behind membrane insertion: a review of the complement membrane attack complex. *Philos Trans R Soc Lond B Biol Sci 372*, 1726 (2017).

[54] Beernaert, L. A., Pasmans, F., Van Waeyenberghe, L., Haesebrouck, F., and Martel, A. Aspergillus infections in birds: a review. *Avian Pathology 39*, 5 (2010), 325–331.

[55] Bell, J. K., Mullen, G. E. D., Leifer, C. A., Mazzoni, A., Davies, D. R., and Segal, D. M. Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends in Immunology 24*, 10 (2003), 528–533.

[56] Berland, R., Fernandez, L., Kari, E., Han, J. H., Lomakin, I., Akira, S., Wortis, H. H., Kearney, J. F., Ucci, A. A., and Imanishi-Kari, T. Toll-like receptor 7-dependent loss of B cell tolerance in pathogenic autoantibody knockin mice. *Immunity 25*, 3 (2006), 429–440.

[57] Biragyn, A., Ruffini, P. A., Leifer, C. A., Klyushnenkova, E., Shakhov, A., Chertov, O., Shirakawa, A. K., Farber, J. M., Segal, D. M., Oppenheim, J. J., and Kwak, L. W. Toll-like receptor 4-dependent activation of dendritic cells by beta-defensin 2. *Science 298*, 5595 (2002), 1025–1029.

[58] Birney, E., Clamp, M., and Durbin, R. GeneWise and Genomewise. *Genome Res 14*, 5 (2004), 988–95.

[59] Biswas, S., and Akey, J. M. Genomic insights into positive selection. *Trends Genet 22*, 8 (2006), 437–446.

[60] Blander, J. M., and Medzhitov, R. Toll-dependent selection of microbial antigens for presentation by dendritic cells. *Nature 440*, 7085 (2006), 808–812.

[61] Blaustein, A. R., and Bancroft, B. A. Amphibian Population Declines: Evolutionary Considerations. *BioScience 57*, 5 (2007), 437–444.

[62] Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., and Hochreiter, S. msa: an R package for multiple sequence alignment. *Bioinformatics 31*, 24 (2015), 3997–3999.

[63] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res 31*, 1 (2003), 365–70.

[64] Bollmer, J. L., Vargas, F. H., and Parker, P. G. Low MHC variation in the endangered Galapagos penguin (Spheniscus mendiculus). *Immunogenetics 59*, 7 (2007), 593–602.

[65] Bonneaud, C., Balenger, S. L., Zhang, J., Edwards, S. V., and Hill, G. E. Innate immunity and the evolution of resistance to an emerging infectious disease in a wild bird. *Mol Ecol 21*, 11 (2012), 2628–2639.

[66] Bonnedahl, J. *Antibiotic resistance in Enterobacteriaceae isolated from wild birds*. Thesis, Faculty of Medicine, 2011.

[67] Borboroglu, P. G., and Boersma, P. D. *Penguins : natural history and conservation [electronic resource]*. Ebook central. Seattle, 2013.

[68] Botos, I., Liu, L., Wang, Y., Segal, D. M., and Davies, D. R. The toll-like receptor 3:dsRNA signaling complex. *Biochim Biophys Acta 1789*, 9-10 (2009), 667–674.

[69] Botos, I., Segal, D. M., and Davies, D. R. The structural biology of Toll-like receptors. *Structure 19*, 4 (2011), 447–459.

[70] Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kuhnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Muller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C. H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. J. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *Plos Computational Biology 15*, 4 (2019).

[71] Boyd, A., Philbin, V. J., and Smith, A. L. Conserved and distinct aspects of the avian Toll-like receptor (TLR) system: implications for transmission and control of bird-borne zoonoses. *Biochemical Society Transactions 35* (2007), 1504–1507.

[72] Boyd, A. C., Peroval, M. Y., Hammond, J. A., Prickett, M. D., Young, J. R., and Smith, A. L. TLR15 is unique to avian and reptilian lineages and recognizes a yeast-derived agonist. *J Immunol 189*, 10 (2012), 4930–8.

[73] Brand, M. D., Hill, R. D., Brenes, R., Chaney, J. C., Wilkes, R. P., Grayfer, L., Miller, D. L., and Gray, M. J. Water Temperature Affects Susceptibility to Ranavirus. *Ecohealth 13*, 2 (2016), 350–359.

[74] Brand, S. P., and Keeling, M. J. The impact of temperature changes on vector-borne disease transmission:. *J R Soc Interface 14*, 128 (2017).

[75] Bravo García-Morato, M., Calvo Apalategi, A., Bravo-Gallego, L. Y., Blázquez Moreno, A., Simón-Fuentes, M., Garmendia, J. V., Méndez Echevarría, A., Del Rosal Rabes, T., Domínguez-Soto, Á., López-Granados, E., Reyburn, H. T., and Rodríguez Pena, R. Impaired control of multiple viral infections in a family with complete IRF9 deficiency. *J Allergy Clin Immunol 144*, 1 (2019), 309–312.

[76] Brennan, J. J., and Gilmore, T. D. Evolutionary Origins of Toll-like Receptor Signaling. *Mol Biol Evol 35*, 7 (2018), 1576–1587.

[77] Brown, I., Mulatti, P., Smietanka, K., Staubach, C., Willeberg, P., Adlhoch, C., Candiani, D., Fabris, C., Zancanaro, G., Morgado, J., Verdonck, F., Authority, E. F. S., Control, E. C. f. D. P., , and influenza, E. U. R. L. f. A. Avian influenza overview October 2016-August 2017. *EFSA J 15*, 10 (2017), e05018.

[78] Brown, J. D., Stallknecht, D. E., and Swayne, D. E. Experimental infection of swans and geese with highly pathogenic avian influenza virus (H5N1) of Asian lineage. *Emerg Infect Dis 14*, 1 (2008), 136–142.

[79] Browning, B. L., Zhou, Y., and Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics 103*, 3 (2018), 338–348.

[80] Browning, S. R., and Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics 81*, 5 (2007), 1084–1097.

[81] Brownlie, R., and Allan, B. Avian toll-like receptors. *Cell and Tissue Research 343*, 1 (2011), 121–130.

[82] Brownlie, R., Zhu, J., Allan, B., Mutwiri, G. K., Babiuk, L. A., Potter, A., and Griebel, P. Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. *Mol Immunol 46*, 15 (2009), 3163–70.

[83] Buckley, K. M., and Rast, J. P. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev Comp Immunol 49*, 1 (2015), 179–189.

[84] Buwitt-Beckmann, U., Heine, H., Wiesmüller, K. H., Jung, G., Brock, R., Akira, S., and Ulmer, A. J. TLR1- and TLR6-independent recognition of bacterial lipopeptides. *J Biol Chem 281*, 14 (2006), 9049–9057.

[85] Califano, D., Furuya, Y., Roberts, S., Avram, D., McKenzie, A. N. J., and Metzger, D. W. IFN-γ increases susceptibility to influenza A infection through suppression of group II innate lymphoid cells. *Mucosal Immunol 11*, 1 (2018), 209–219.

[86] Carpenter, S., and O'Neill, L. A. Recent insights into the structure of Toll-like receptors and post-translational modifications of their associated signalling proteins. *Biochem J 422*, 1 (2009), 1–10.

[87] Carpenter-Kling, T., Handley, J. M., Connan, M., Crawford, R. J. M., Makhado, A. B., Dyer, B. M., Froneman, W., Lamont, T., Wolfaardt, A. C., Landman, M., Sigqala, M., and Pistorius, P. A. Gentoo penguins as sentinels of climate change at the sub-Antarctic Prince Edward Archipelago, Southern Ocean. *Ecological Indicators 101* (2019), 163–172.

[88] Carrasco, L., Lima, J. S., Halfen, D. C., Salguero, F. J., Sánchez-Cordon, P., and Becker, G. Systemic aspergillosis in an oiled magallanic penguin (Spheniscus magellanicus). *J Vet Med B Infect Dis Vet Public Health 48*, 7 (2001), 551–554.

[89] Carvajal-Rodriguez, A. Myriads: P-value-based multiple testing correction. *Bioinformatics 34*, 6 (2018), 1043–1045.

[90] Carvajal-Rodriguez, A., and de Una-Alvarez, J. Assessing Significance in High-Throughput Experiments by Sequential Goodness of Fit and q-Value Estimation. *Plos One 6*, 9 (2011).

[91] Casrouge, A., Zhang, S. Y., Eidenschenk, C., Jouanguy, E., Puel, A., Yang, K., Alcais, A., Picard, C., Mahfoufi, N., Nicolas, N., Lorenzo, L., Plancoulaine, S., Sénéchal, B., Geissmann, F., Tabeta, K., Hoebe, K., Du, X., Miller, R. L., Héron, B., Mignot, C., de Villemeur, T. B., Lebon, P., Dulac, O., Rozenberg, F., Beutler, B., Tardieu, M., Abel, L., and Casanova, J. L. Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science 314*, 5797 (2006), 308–312.

[92] Cassat, J. E., and Skaar, E. P. Iron in infection and immunity. *Cell Host Microbe 13*, 5 (2013), 509–519.

[93] Cerda-Cuellar, M., More, E., Ayats, T., Aguilera, M., Munoz-Gonzalez, S., Antilles, N., Ryan, P. G., and Gonzalez-Solis, J. Do humans spread zoonotic enteric bacteria in Antarctic? *Science of the Total Environment 654* (2019), 190–196.

[94] Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S., Valvano, M. A., Foster, S. J., Mak, T. W., Nuñez, G., and Inohara, N. An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat Immunol 4*, 7 (2003), 702–707.

[95] Chapman, J. R., Hill, T., and Unckless, R. L. Balancing Selection Drives the Maintenance of Genetic Variation in Drosophila Antimicrobial Peptides. *Genome Biol Evol 11*, 9 (2019), 2691–2701.

[96] Chen, L. Co-inhibitory molecules of the B7-CD28 family in the control of T-cell immunity. *Nat Rev Immunol 4*, 5 (2004), 336–347.

[97] Chen, Z. J. Ubiquitination in signaling to and activation of IKK. *Immunol Rev 246*, 1 (2012), 95–106.

[98] Cheng, J., Sun, Y., Zhang, X., Zhang, F., Zhang, S., Yu, S., Qiu, X., Tan, L., Song, C., Gao, S., Wu, Y., and Ding, C. Toll-like receptor 3 inhibits Newcastle disease virus replication through activation of pro-inflammatory cytokines and the type-1 interferon pathway. *Arch Virol 159*, 11 (2014), 2937–48.

[99] Cheng, N., He, R., Tian, J., Ye, P. P., and Ye, R. D. Cutting edge: TLR2 is a functional receptor for acute-phase serum amyloid A. *J Immunol 181*, 1 (2008), 22–26.

[100] Choe, J., Kelker, M. S., and Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science 309*, 5734 (2005), 581–5.

[101] Choi, Y., and Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics 31*, 16 (2015), 2745–2747.

[102] Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J., and Gusovsky, F. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *Journal of Biological Chemistry 274*, 16 (1999), 10689–10692.

[103] Chown, S. L., Clarke, A., Fraser, C. I., Cary, S. C., Moon, K. L., and McGeoch, M. A. The changing form of Antarctic biodiversity. *Nature 522*, 7557 (2015), 431–8.

[104] Christensen, S. R., Kashgarian, M., Alexopoulou, L., Flavell, R. A., Akira, S., and Shlomchik, M. J. Toll-like receptor 9 controls anti-DNA autoantibody production in murine lupus. *J Exp Med 202*, 2 (2005), 321–331.

[105] Christensen, S. R., Shupe, J., Nickerson, K., Kashgarian, M., Flavell, R. A., and Shlomchik, M. J. Toll-like receptor 7 and TLR9 dictate autoantibody specificity and have opposing inflammatory and regulatory roles in a murine model of lupus. *Immunity 25*, 3 (2006), 417–428.

[106] Chuang, T. H., Lee, J., Kline, L., Mathison, J. C., and Ulevitch, R. J. Toll-like receptor 9 mediates CpG-DNA signaling. *J Leukoc Biol 71*, 3 (2002), 538–544.

[107] Chuang, Y. C., Tseng, J. C., Yang, J. X., Liu, Y. L., Yeh, D. W., Lai, C. Y., Yu, G. Y., Hsu, L. C., Huang, C. M., and Chuang, T. H. Toll-Like Receptor 21 of Chicken and Duck Recognize a Broad Array of Immunostimulatory CpG-oligodeoxynucleotide Sequences. *Vaccines (Basel) 8*, 4 (2020).

[108] Ciancanelli, M. J., Huang, S. X., Luthra, P., Garner, H., Itan, Y., Volpi, S., Lafaille, F. G., Trouillet, C., Schmolke, M., Albrecht, R. A., Israelsson, E., Lim, H. K., Casadio, M., Hermesh, T., Lorenzo, L., Leung, L. W., Pedergnana, V., Boisson, B., Okada, S., Picard, C., Ringuier, B., Troussier, F., Chaussabel, D., Abel, L., Pellier, I., Notarangelo, L. D., García-Sastre, A., Basler, C. F., Geissmann, F., Zhang, S. Y., Snoeck, H. W., and Casanova, J. L. Infectious disease. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science 348*, 6233 (2015), 448–453.

[109] Clark, R., and Kupper, T. Old meets new: the interaction between innate and adaptive immunity. *J Invest Dermatol 125*, 4 (2005), 629–637.

[110] Clarke, J., and Kerry, K. Diseases and parasites of penguins. *Penguin Conservation 13*, 1 (2000), 5–24.

[111] Clausen, A., and Putz, K. Winter diet and foraging range of gentoo penguins (Pygoscelis papua) from Kidney Cove, Falkland Islands. *Polar Biology 26*, 1 (2003), 32–40.

[112] Clements, J., Schulenberg, T., Iliff, M., Billerman, S., Fredericks, T., Sullivan, B., and Wood, C. *The eBird/Clements Checklist of Birds of the World: v2019.* Downloaded from https://www.birds.cornell.edu/clementschecklist/download/, 2019.

[113] Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., Polito, M. J., Pybus, O. G., Rogers, A. D., and Hart, T. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Scientific Reports 4* (2014).

[114] Clucas, G. V., Younger, J. L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., Rogers, A. D., Bost, C.-A., Miller, G. D., Polito, M. J., Lelliott, P., Handley, J., Crofts, S., Phillips, R. A., Dunn, M. J., Miller, K. J., and Hart, T. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Molecular Ecology 27*, 23 (2018), 4680–4697.

[115] Coffman, R. L., Sher, A., and Seder, R. A. Vaccine adjuvants: putting innate immunity to work. *Immunity 33*, 4 (2010), 492–503.

[116] Cohen, J. M., Sauer, E. L., Santiago, O., Spencer, S., and Rohr, J. R. Divergent impacts of warming weather on wildlife disease risk across climates. *Science 370*, 6519 (2020).

[117] Cole, T. L., Dutoit, L., Dussex, N., Hart, T., Alexander, A., Younger, J. L., Clucas, G. V., Frugone, M. J., Cherel, Y., Cuthbert, R., Ellenberg, U., Fiddaman, S. R., Hiscock, J., Houston, D., Jouventin, P., Mattern, T., Miller, G., Miskelly, C., Nolan, P., Polito, M. J., Quillfeldt, P., Ryan, P. G., Smith, A., Tennyson, A. J. D., Thompson, D., Wienecke, B., Vianna, J. A., and Waters, J. M. Receding ice drove parallel expansions in Southern Ocean penguins. *Proc Natl Acad Sci U S A* (2019).

[118] Cole, T. L., Ksepka, D. T., Mitchell, K. J., Tennyson, A. J. D., Thomas, D. B., Pan, H., Zhang, G., Rawlence, N. J., Wood, J. R., Bover, P., Bouzat, J. L., Cooper, A., Fiddaman, S. R., Hart, T., Miller, G., Ryan, P. G., Shepherd, L. D., Wilmshurst, J. M., and Waters, J. M. Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation. *Mol Biol Evol 36*, 4 (2019), 784–797.

[119] Coleman, J., Jones, A., Collier, L., Richardson, M., and Bright-Thomas, R. M1 Do climate changes influence environmental aspergillus fumigatus load at the manchester university NHS foundation trust adult cystic fibrosis centre? *Thorax 74*, Suppl 2 (2019), A235–A235.

[120] Cooper, J. C., Leonard, C. J., Pedersen, B. S., Carey, C. M., Quinlan, A. R., Elde, N. C., and Phadnis, N. Endless Conflicts: Detecting Molecular Arms Races in Mammalian Genomes. *bioRxiv* (2019).

[121] Cormican, P., Lloyd, A. T., Downing, T., Connell, S. J., Bradley, D., and O'Farrelly, C. The avian Toll-Like receptor pathway–subtle differences amidst general conformity. *Dev Comp Immunol 33*, 9 (2009), 967–973.

[122] Crawford, R. J. M., Cooper, J., Du Toit, M., Greyling, M. D., Hanise, B., Holness, C. L., Keith, D. G., Nel, J. L., Petersen, S. L., Spencer, K., Tshingana, D., and Wolfaardt, A. C. Population and breeding of the gentoo penguin Pygoscelis papua at Marion Island, 1994/95-2002/03. *African Journal of Marine Science 25* (2003), 463–474.

[123] Crawford, R. J. M., Dyer, B. M., Upfold, L., and Makhado, A. B. Congruent, decreasing trends of gentoo penguins and Crozet shags at sub-Antarctic Marion Island suggest food limitation through common environmental forcing. *African Journal of Marine Science 36*, 2 (2014), 225–231.

[124] Crawford, R. J. M., Whittington, P. A., Upfold, L., Ryan, P. G., Petersen, S. L., Dyer, B. M., and Cooper, J. Recent trends in numbers of four species of penguins at the Prince Edward Islands. *African Journal of Marine Science 31*, 3 (2009), 419–426.

[125] Cumming, G. S., and Guegan, J. F. Food webs and disease: Is pathogen diversity limited by vector diversity? *Ecohealth 3*, 3 (2006), 163–170.

[126] Cummings, S. M., McMullan, M., Joyce, D. A., and van Oosterhout, C. Solutions for PCR, cloning and sequencing errors in population genetic analysis. *Conservation Genetics 11*, 3 (2010), 1095–1097.

[127] Curran, C. S., Demick, K. P., and Mansfield, J. M. Lactoferrin activates macrophages via TLR4-dependent and -independent signaling pathways. *Cell Immunol 242*, 1 (2006), 23–30.

[128] Curry, C. H., McCarthy, J. S., Darragh, H. M., Wake, R. A., Todhunter, R., and Terris, J. Could tourist boots act as vectors for disease transmission in Antarctica? *J Travel Med 9*, 4 (2002), 190–3.

[129] Dalton, D. L., Vermaak, E., Roelofse, M., and Kotze, A. Diversity in the Toll-Like Receptor Genes of the African Penguin (Spheniscus demersus). *Plos One 11*, 10 (2016).

[130] Dalton, D. L., Vermaak, E., Smit-Robinson, H. A., and Kotze, A. Lack of diversity at innate immunity Toll-like receptor genes in the Critically Endangered White-winged Flufftail (Sarothrura ayresi). *Scientific Reports 6* (2016).

[131] 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., and Group, . G. P. A. The variant call format and VCFtools. *Bioinformatics 27*, 15 (2011), 2156–8.

[132] Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. Twelve years of SAMtools and BCFtools. *Gigascience 10*, 2 (2021).

[133] Dannemann, M., Andrés, A. M., and Kelso, J. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am J Hum Genet 98*, 1 (2016), 22–33.

[134] Darfour-Oduro, K. A., Megens, H.-J., Roca, A., Groenen, M. A. M., and Schook, L. B. Evidence for adaptation of porcine Toll-like receptors. *Immunogenetics 68*, 3 (2016), 179–189.

[135] Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods 9*, 8 (2012), 772–772.

[136] Davidson, C. R., Best, N. M., Francis, J. W., Cooper, E. L., and Wood, T. C. Toll-like receptor genes (TLRs) from Capitella capitata and Helobdella robusta (Annelida). *Dev Comp Immunol 32*, 6 (2008), 608–612.

[137] de Dinechin, M., Dobson, F. S., Zehtindjiev, P., Metcheva, R., Couchoux, C., Martin, A., Quillfeldt, P., and Jouventin, P. The biogeography of Gentoo Penguins (Pygoscelis papua). *Canadian Journal of Zoology-Revue Canadienne De Zoologie 90*, 3 (2012), 352–360.

[138] de Zoete, M. R., Bouwman, L. I., Keestra, A. M., and van Putten, J. P. Cleavage and activation of a Toll-like receptor by microbial proteases. *Proc Natl Acad Sci U S A 108*, 12 (2011), 4968–73.

[139] Delaneau, O., Marchini, J., and Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods 9*, 2 (2011), 179–81.

[140] Delport, W., Poon, A. F. Y., Frost, S. D. W., and Pond, S. L. K. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics 26*, 19 (2010), 2455–2457.

[141] Denton, A. E., Doherty, P. C., Turner, S. J., and La Gruta, N. L. IL-18, but not IL-12, is required for optimal cytokine production by influenza virus-specific CD8+ T cells. *Eur J Immunol 37*, 2 (2007), 368–375.

[142] DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics 43*, 5 (2011), 491–+.

[143] Devaney, J. M., Greene, C. M., Taggart, C. C., Carroll, T. P., O'Neill, S. J., and McElvaney, N. G. Neutrophil elastase up-regulates interleukin-8 via toll-like receptor 4. *FEBS Lett 544*, 1-3 (2003), 129–132.

[144] Dewar, M. L., Arnould, J. P. Y., Dann, P., Trathan, P., Groscolas, R., and Smith, S. Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen 2*, 1 (2013), 195–204.

[145] Dewar, M. L., Arnould, J. P. Y., Krause, L., Dann, P., and Smith, S. C. Interspecific variations in the faecal microbiota of Procellariiform seabirds. *Fems Microbiology Ecology 89*, 1 (2014), 47–55.

[146] Dey, S., Pathak, D. C., Ramamurthy, N., Maity, H. K., and Chellappa, M. M. Infectious bursal disease virus in chickens: prevalence, impact, and management strategies. *Vet Med (Auckl) 10* (2019), 85–97.

[147] Dimitrov, K., Metcheva, R., and Kenarova, A. Salmonella presence - an indicator of direct and indirect human impact on Gentoo in Antarctica. *Biotechnology and Biotechnological Equipment 23* (2009), 246–249.

[148] Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., and Bernatchez, L. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution 61*, 9 (2007), 2154–2164.

[149] Dolby, G. A., Morales, M., Webster, T. H., DeNardo, D. F., Wilson, M. A., and Kusumi, K. Discovery of a New TLR Gene and Gene Expansion Event through Improved Desert Tortoise Genome Assembly with Chromosome-Scale Scaffolds. *Genome Biol Evol 12*, 2 (2020), 3917–3925.

[150] Döring, Y., Hurst, J., Lorenz, M., Prinz, N., Clemens, N., Drechsler, M. D., Bauer, S., Chapman, J., Shoenfeld, Y., Blank, M., Lackner, K. J., and von Landenberg, P. Human antiphospholipid antibodies induce TNFalpha in monocytes via Toll-like receptor 8. *Immunobiology 215*, 3 (2010), 230–241.

[151] Downing, T., Lloyd, A. T., O'Farrelly, C., and Bradley, D. G. The differential evolutionary dynamics of avian cytokine and TLR gene classes. *J Immunol 184*, 12 (2010), 6993–7000.

[152] Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. Intrinsically disordered protein. *J Mol Graph Model 19*, 1 (2001), 26–59.

[153] Dunnett, C. W. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association 50*, 272 (1955), 1096–1121.

[154] Duschene, K. S., and Broderick, J. B. Viperin: a radical response to viral infection. *Biomol Concepts 3*, 3 (2012), 255–266.

[155] Dziarski, R., and Gupta, D. Staphylococcus aureus peptidoglycan is a toll-like receptor 2 activator: a reevaluation. *Infect Immun 73*, 8 (2005), 5212–5216.

[156] Ebel, E. R., Telis, N., Venkataram, S., Petrov, D. A., and Enard, D. High rate of adaptation of mammalian proteins that interact with Plasmodium and related parasites. *PLoS Genet 13*, 9 (2017), e1007023.

[157] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research 32*, 5 (2004), 1792–1797.

[158] Eisenbarth, S. C., Piggott, D. A., Huleatt, J. W., Visintin, I., Herrick, C. A., and Bottomly, K. Lipopolysaccharide-enhanced, toll-like receptor 4-dependent T helper cell type 2 responses to inhaled antigen. *J Exp Med 196*, 12 (2002), 1645–1651.

[159] El-Zayat, S. R., Sibaii, H., and Mannaa, F. A. Toll-like receptors activation, signaling, and targeting: an overview. *Bulletin of the National Research Centre 43*, 1 (2019), 187.

[160] Ellis, J. S., Turner, L. M., and Knight, M. E. Patterns of selection and polymorphism of innate immunity genes in bumblebees (Hymenoptera: Apidae). *Genetica 140*, 4-6 (2012), 205–217.

[161] Enard, D., Cai, L., Gwennap, C., and Petrov, D. A. Viruses are a dominant driver of protein adaptation in mammals. *Elife 5* (2016).

[162] Epstein, C. J. Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature 215*, 5099 (1967), 355–9.

[163] Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Stromstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., and Andersson, L. Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken. *PLOS Genetics 4*, 2 (2008).

[164] Esposito, S., Molteni, C. G., Giliani, S., Mazza, C., Scala, A., Tagliaferri, L., Pelucchi, C., Fossali, E., Plebani, A., and Principi, N. Toll-like receptor 3 gene polymorphisms and severity of pandemic A/H1N1/2009 influenza in otherwise healthy children. *Virol J 9* (2012), 270.

[165] Evseev, D., and Magor, K. E. Innate Immune Responses to Avian Influenza Viruses in Ducks and Chickens. *Vet Sci 6*, 1 (2019).

[166] Ewald, S. E., Lee, B. L., Lau, L., Wickliffe, K. E., Shi, G. P., Chapman, H. A., and Barton, G. M. The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. *Nature 456*, 7222 (2008), 658–662.

[167] Excoffier, L., and Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources 10*, 3 (2010), 564–567.

[168] Excoffier, L., Smouse, P. E., and Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics 131*, 2 (1992), 479–491.

[169] Faber, E., Tedin, K., Speidel, Y., Brinkmann, M. M., and Josenhans, C. Functional expression of TLR5 of different vertebrate species and diversification in intestinal pathogen recognition. *Scientific Reports 8* (2018).

[170] Faber, J., Henninger, N., Finn, A., Zenz, W., Zepp, F., and Knuf, M. A toll-like receptor 4 variant is associated with fatal outcome in children with invasive meningococcal disease. *Acta Paediatr 98*, 3 (2009), 548–552.

[171] Faller, P., Hureau, C., and La Penna, G. Metal ions and intrinsically disordered proteins and peptides: from Cu/Zn amyloid-β to general principles. *Acc Chem Res 47*, 8 (2014), 2252–2259.

[172] Fan, J., Li, L., Han, J., Ming, H., Li, J., Na, G., and Chen, J. Diversity and structure of bacterial communities in Fildes Peninsula, King George Island. *Polar Biology 36*, 10 (2013), 1385–1399.

[173] Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonet, T., Ramírez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J., and Casals, F. Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol 181*, 2 (2008), 1315–22.

[174] Ferwerda, B., McCall, M. B. B., Alonso, S., Mouktaroudi, M., Giamarellos-Bourboulis, E. J., Izagirre, N., Syafruddin, D., Kibiki, G., Cristea, T., Hijmans, A., Hamann, L., Israel, S., Eighazali, G., Troye-Blomberg, M., Kumpf, O., Maiga, B., Dolo, A., Doumbo, O., Hermsen, C. C., Stalenhoef, A. F. H., van Crevel, R., Brunner, H. G., Oh, D. Y., Schumann, R. R., de la Rua, C., Sauerwein, R., Kullberg, B. J., van der Ven, A., van der Meer, J. W. M., and Netea, M. G. TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. *Proceedings of the National Academy of Sciences of the United States of America 104*, 42 (2007), 16645–16650.

[175] Fitzgerald, K. A., Palsson-McDermott, E. M., Bowie, A. G., Jefferies, C. A., Mansell, A. S., Brady, G., Brint, E., Dunne, A., Gray, P., Harte, M. T., McMurray, D., Smith, D. E., Sims, J. E., Bird, T. A., and O'Neill, L. A. Mal (MyD88-adapter-like) is required for Toll-like receptor-4 signal transduction. *Nature 413*, 6851 (2001), 78–83.

[176] Fitzgerald, K. A., Rowe, D. C., Barnes, B. J., Caffrey, D. R., Visintin, A., Latz, E., Monks, B., Pitha, P. M., and Golenbock, D. T. LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF. *J Exp Med 198*, 7 (2003), 1043–1055.

[177] Fix, A. S., Waterhouse, C., Greiner, E. C., and Stoskopf, M. K. Plasmodium relictum as a cause of avian malaria in wild-caught magellanic penguins (Spheniscus magellanicus). *Journal of Wildlife Diseases 24*, 4 (1988), 610–619.

[178] Flach, E. J., Stevenson, M. F., and Henderson, G. M. Aspergillosis in Gentoo penguins (Pygoscelis papua) at Edinburgh Zoo, 1964 to 1988. *Veterinary Record 126*, 4 (1990), 81–85.

[179] Flacke, G., Becker, P., Cooper, D., Gunther, M. S., Robertson, I., Holyoake, C., Donaldson, R., and Warren, K. An infectious disease and mortality survey in a population of free-ranging African wild dogs and sympatric domestic dogs. *International Journal of Biodiversity 2013* (2013), 497623.

[180] Fleming-Canepa, X., Brusnyk, C., Aldridge, J. R., Ross, K. L., Moon, D., Wang, D., Xia, J., Barber, M. R., Webster, R. G., and

Magor, K. E. Expression of duck CCL19 and CCL21 and CCR7 receptor in lymphoid and influenza-infected tissues. *Mol Immunol 48*, 15-16 (2011), 1950–1957.

[181] Flo, T. H., Ryan, L., Latz, E., Takeuchi, O., Monks, B. G., Lien, E., Halaas, O., Akira, S., Skjåk-Braek, G., Golenbock, D. T., and Espevik, T. Involvement of toll-like receptor (TLR) 2 and TLR4 in cell activation by mannuronic acid polymers. *J Biol Chem 277*, 38 (2002), 35489–35495.

[182] Foell, D., Wittkowski, H., Vogl, T., and Roth, J. S100 proteins expressed in phagocytes: a novel group of damage-associated molecular pattern molecules. *J Leukoc Biol 81*, 1 (2007), 28–37.

[183] Ford, M. J. Molecular evolution of transferrin: evidence for positive selection in salmonids. *Mol Biol Evol 18*, 4 (2001), 639–47.

[184] Forni, D., Cagliani, R., Pontremoli, C., Pozzoli, U., Vertemara, J., De Gioia, L., Clerici, M., and Sironi, M. Evolutionary Analysis Provides Insight Into the Origin and Adaptation of HCV. *Front Microbiol 9* (2018), 854.

[185] Fornuskova, A., Vinkler, M., Pages, M., Galan, M., Jousselin, E., Cerqueira, F., Morand, S., Charbonnel, N., Bryja, J., and Cosson, J. F. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). *Bmc Evolutionary Biology 13* (2013).

[186] Forsbach, A., Nemorin, J. G., Montino, C., Müller, C., Samulowitz, U., Vicari, A. P., Jurk, M., Mutwiri, G. K., Krieg, A. M., Lipford, G. B., and Vollmer, J. Identification of RNA sequence motifs stimulating sequence-specific TLR8-dependent immune responses. *J Immunol 180*, 6 (2008), 3729–3738.

[187] Frantz, L. A., 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., 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., and Larson, G. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science 352*, 6290 (2016), 1228–31.

[188] Freeman, N. M., Lovenduski, N. S., and Gent, P. R. Temporal variability in the Antarctic Polar Front (2002-2014). *Journal of Geophysical Research-Oceans 121*, 10 (2016), 7263–7276.

[189] Frugone, M. J., Lopez, M. E., Segovia, N. I., Cole, T. L., Lowther, A., Pistorius, P., Dantas, G. P. M., Petry, M. V., Bonadonna, F., Trathan, P., Polanowski, A., Wienecke, B., Bi, K., Wang-Claypool, C. Y., Waters, J. M., Bowie, R. C. K., Poulin, E., and Vianna, J. A. More than the eye can see: Genomic insights into the drivers of

genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Molecular Phylogenetics and Evolution 139* (2019).

[190] Frugone, M. J., Lowther, A., Noll, D., Ramos, B., Pistorius, P., Dantas, G. P. M., Petry, M. V., Bonadonna, F., Steinfurth, A., Polanowski, A., Rey, A. R., Lois, N. A., Putz, K., Trathan, P., Wienecke, B., Poulin, E., and Vianna, J. A. Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. *Scientific Reports 8* (2018).

[191] Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics 147*, 2 (1997), 915–925.

[192] Funderburg, N., Lederman, M. M., Feng, Z., Drage, M. G., Jadlowsky, J., Harding, C. V., Weinberg, A., and Sieg, S. F. Human -defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc Natl Acad Sci U S A 104*, 47 (2007), 18631–18635.

[193] Galtier, N., Roux, C., Rousselle, M., Romiguier, J., Figuet, E., Glémin, S., Bierne, N., and Duret, L. Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion. *Mol Biol Evol 35*, 5 (2018), 1092–1103.

[194] Gangloff, M. Different dimerisation mode for TLR4 upon endosomal acidification? *Trends Biochem Sci 37*, 3 (2012), 92–8.

[195] Garamszegi, L. Climate change increases the risk of malaria in birds. *Global Change Biology 17*, 5 (2011), 1751–1759.

[196] Garate, J. A., and Oostenbrink, C. Lipid A from lipopolysaccharide recognition: structure, dynamics and cooperativity by molecular dynamics simulations. *Proteins 81*, 4 (2013), 658–674.

[197] Garcia-Pena, F. J., Llorente, M. T., Serrano, T., Ruano, M. J., Belliure, J., Benzal, J., Herrera-Leon, S., Vidal, V., D'Amico, V., Perez-Boto, D., and Barbosa, A. Isolation of Campylobacter spp. from Three Species of Antarctic Penguins in Different Geographic Locations. *Ecohealth 14*, 1 (2017), 78–87.

[198] García-Ramírez, R. A., Ramírez-Venegas, A., Quintana-Carrillo, R., Camarena, Á., Falfán-Valencia, R., and Mejía-Aranguré, J. M. TNF, IL6, and IL1B Polymorphisms Are Associated with Severe Influenza A (H1N1) Virus Infection in the Mexican Population. *PLoS One 10*, 12 (2015), e0144832.

[199] Gauthier, M. E., Du Pasquier, L., and Degnan, B. M. The genome of the sponge Amphimedon queenslandica provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. *Evol Dev 12*, 5 (2010), 519–533.

[200] Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., and Drummond, A. J. Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology 66*, 1 (2017), 57–73.

[201] Geijtenbeek, T. B., and Gringhuis, S. I. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol 9*, 7 (2009), 465–479.

[202] Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J., and Madara, J. L. Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol 167*, 4 (2001), 1882–5.

[203] Ghys, M. I., Rey, A. R., and Schiavini, A. Population Trend and Breeding Biology of Gentoo Penguin in Martillo Island, Tierra Del Fuego, Argentina. *Waterbirds 31*, 4 (2008), 625–631.

[204] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, Clyde A., I., and Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 6*, 5 (2009), 343–U41.

[205] Gillespie, J. H. *The causes of molecular evolution*. Oxford University Press, New York ; Oxford, 1991.

[206] Gilroy, D. L., van Oosterhout, C., Komdeur, J., and Richardson, D. S. Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. *Animal Conservation 20*, 3 (2017), 235–250.

[207] Goldberg, T. L., Sibley, S. D., Pinkerton, M. E., Dunn, C. D., Long, L. J., White, L. C., and Strom, S. M. Multidecade Mortality and a Homolog of Hepatitis C Virus in Bald Eagles (Haliaeetus leucocephalus), the National Bird of the USA. *Sci Rep 9*, 1 (2019), 14953.

[208] Goletić, T., Gagić, A., Residbegović, E., Kustura, A., Kavazović, A., Savić, V., Harder, T., Starick, E., and Prasović, S. Highly pathogenic avian influenza virus subtype H5N1 in mute swans (Cygnus olor) in Central Bosnia. *Avian Dis 54*, 1 Suppl (2010), 496–501.

[209] Gomi, K., Kawasaki, K., Kawai, Y., Shiozaki, M., and Nishijima, M. Toll-like receptor 4-MD-2 complex mediates the signal transduction induced by flavolipin, an amino acid-containing lipid unique to Flavobacterium meningosepticum. *J Immunol 168*, 6 (2002), 2939–2943.

[210] Gonzalez-Acuna, D., Hernandez, J., Moreno, L., Herrmann, B., Palma, R., Latorre, A., Medina-Vogel, G., Kinsella, M. J., Martin, N., Araya, K., Torres, I., Fernandez, N., and Olsen, B. Health evaluation of wild gentoo penguins (Pygoscelis papua) in the Antarctic Peninsula. *Polar Biology 36*, 12 (2013), 1749–1760.

[211] Gonzalez-Quevedo, C., Spurgin, L. G., Carlos Illera, J., and Richardson, D. S. Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology 24*, 23 (2015), 5852–5863.

[212] Goudet, J. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity 86*, 6 (1995), 485–486.

[213] Graczyk, T. K., and Cockrem, J. F. Aspergillus spp Seropositivity in New Zealand penguins. *Mycopathologia 131*, 3 (1995), 179–184.

[214] Grantham, R. Amino acid difference formula to help explain protein evolution. *Science 185*, 4154 (1974), 862–4.

[215] Grassin-Delyle, S., Abrial, C., Salvator, H., Brollo, M., Naline, E., and Devillier, P. The Role of Toll-Like Receptors in the Production of Cytokines by Human Lung Macrophages. *J Innate Immun 12*, 1 (2020), 63–73.

[216] Gratz, N. G. Emerging and resurging vector-borne diseases. *Annu Rev Entomol 44* (1999), 51–75.

[217] Grilo, M. L., Vanstreels, R. E. T., Wallace, R., Garcia-Parraga, D., Braga, E. M., Chitty, J., Catao-Dias, J. L., and Madeira de Carvalho, L. M. Malaria in penguins - current perceptions. *Avian Pathology 45*, 4 (2016), 393–407.

[218] Grimaldi, W. W., Seddon, P. J., Lyver, P. O., Nakagawa, S., and Tompkins, D. M. Infectious diseases of Antarctic penguins: current status and future threats. *Polar Biology 38*, 5 (2015), 591–606.

[219] Grueber, C. E., Wallis, G. P., and Jamieson, I. G. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology 22*, 17 (2013), 4470–4482.

[220] Grueber, C. E., Wallis, G. P., and Jamieson, I. G. Episodic Positive Selection in the Evolution of Avian Toll-Like Receptor Innate Immunity Genes. *Plos One 9*, 3 (2014).

[221] Grueber, C. E., Wallis, G. P., King, T. M., and Jamieson, I. G. Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. *Plos One 7*, 9 (2012).

[222] Grundberg, E., Small, K. S., Hedman, Å., Nica, A. C., Buil, A., Keildson, S., Bell, J. T., Yang, T. P., Meduri, E., Barrett, A., Nisbett, J., Sekowska, M., Wilk, A., Shin, S. Y., Glass, D., Travers, M., Min, J. L., Ring, S., Ho, K., Thorleifsson, G., Kong, A., Thorsteindottir, U., Ainali, C., Dimas, A. S., Hassanali, N., Ingle, C., Knowles, D., Krestyaninova, M., Lowe, C. E., Di Meglio, P., Montgomery, S. B., Parts, L., Potter, S., Surdulescu, G., Tsaprouni, L., Tsoka, S., Bataille, V., Durbin, R., Nestle, F. O., O'Rahilly, S., Soranzo, N., Lindgren, C. M., Zondervan, K. T., Ahmadi, K. R., Schadt, E. E., Stefansson, K., Smith, G. D., McCarthy, M. I., Deloukas, P., Dermitzakis, E. T., Spector, T. D., and Consortium, M. T. H. E. R. M. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet 44*, 10 (2012), 1084–1089.

[223] Guan, Y., and Smith, G. J. The emergence and diversification of panzootic H5N1 influenza viruses. *Virus Res 178*, 1 (2013), 35–43.

[224] Guégan, J.-F., Prugnolle, F., and Thomas, F. *Global spatial patterns of infectious diseases and human evolution*. Oxford University Press, Oxford, 2007, book section 2, pp. 1–17.

[225] Guernier, V., Hochberg, M. E., and Guegan, J. F. Ecology drives the worldwide distribution of human diseases. *Plos Biology 2*, 6 (2004), 740–746.

[226] Guillot, L., Le Goffic, R., Bloch, S., Escriou, N., Akira, S., Chignard, M., and Si-Tahar, M. Involvement of toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus. *J Biol Chem 280*, 7 (2005), 5571–80.

[227] Guo, Y., Lillie, M., Zan, Y., Beranger, J., Martin, A., Honaker, C. F., Siegel, P. B., and Carlborg, Ö. A genomic inference of the White Plymouth Rock genealogy. *Poult Sci 98*, 11 (2019), 5272–5280.

[228] Gupta, S. K., Bajwa, P., Deb, R., Chellappa, M. M., and Dey, S. Flagellin a toll-like receptor 5 agonist as an adjuvant in chicken vaccines. *Clin Vaccine Immunol 21*, 3 (2014), 261–70.

[229] Häcker, H., Mischak, H., Miethke, T., Liptay, S., Schmid, R., Sparwasser, T., Heeg, K., Lipford, G. B., and Wagner, H. CpG-DNA-specific activation of antigen-presenting cells requires stress kinase activity and is preceded by non-specific endocytosis and endosomal maturation. *EMBO J 17*, 21 (1998), 6230–6240.

[230] Haddas, R., Meir, R., Perk, S., Horowitz, I., Lapin, E., Rosenbluth, E., and Lublin, A. Newcastle disease virus in little owls (Athene noctua) and African penguins (Spheniscus demersus) in an Israeli zoo. *Transbound Emerg Dis 61*, 6 (2014), e79–82.

[231] Hars, J., Ruette, S., Benmergui, M., Fouque, C., Fournier, J. Y., Legouge, A., Cherbonnel, M., Daniel, B., Dupuy, C., and Jestin, V. The epidemiology of the highly pathogenic H5N1 avian influenza in Mute Swan (Cygnus olor) and other Anatidae in the Dombes region (France), 2006. *J Wildl Dis 44*, 4 (2008), 811–823.

[232] Hartmann, S. A., Schaefer, H. M., and Segelbacher, G. Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Molecular Ecology 23*, 23 (2014), 5712–5725.

[233] Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S., and Samuel, M. D. Climate warming and disease risks for terrestrial and marine biota. *Science 296*, 5576 (2002), 2158–2162.

[234] Hasegawa, M., Kishino, H., and Yano, T. A. Dating the human ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution 22*, 2 (1985), 160–174.

[235] Hashimoto, C., Hudson, K. L., and Anderson, K. V. The Toll gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell 52*, 2 (1988), 269–279.

[236] Hawn, T. R., Verbon, A., Lettinga, K. D., Zhao, L. P., Li, S. S., Laws, R. J., Skerrett, S. J., Beutler, B., Schroeder, L., Nachman, A., Ozinsky, A., Smith, K. D., and Aderem, A. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med 198*, 10 (2003), 1563–1572.

[237] Hawn, T. R., Wu, H., Grossman, J. M., Hahn, B. H., Tsao, B. P., and Aderem, A. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc Natl Acad Sci U S A 102*, 30 (2005), 10593–10597.

[238] Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi, F., Goto, S., Kameoka, S., Fujikura, D., Yamada, T., Mizutani, T., Kazumata, M., Sato, M., Tanaka, J., Asaka, M., Ohba, Y., Miyazaki, T., Imamura, M., and Takaoka, A. ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat Immunol 12*, 1 (2011), 37–44.

[239] Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M., and Aderem, A. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature 410*, 6832 (2001), 1099–103.

[240] Hayashi, T., Gray, C. S., Chan, M., Tawatao, R. I., Ronacher, L., McGargill, M. A., Datta, S. K., Carson, D. A., and Corr, M. Prevention of autoimmune disease by induction of tolerance to Toll-like receptor 7. *Proc Natl Acad Sci U S A 106*, 8 (2009), 2764–2769.

[241] Hayman, T. J., Hsu, A. C., Kolesnik, T. B., Dagley, L. F., Willemsen, J., Tate, M. D., Baker, P. J., Kershaw, N. J., Kedzierski, L., Webb, A. I., Wark, P. A., Kedzierska, K., Masters, S. L., Belz, G. T., Binder, M., Hansbro, P. M., Nicola, N. A., and Nicholson, S. E. RIPLET, and not TRIM25, is required for endogenous RIG-I-dependent antiviral responses. *Immunol Cell Biol 97*, 9 (2019), 840–852.

[242] He, H., Crippen, T. L., Farnell, M. B., and Kogut, M. H. Identification of CpG oligodeoxynucleotide motifs that stimulate nitric oxide and cytokine production in avian macrophage and peripheral blood mononuclear cells. *Dev Comp Immunol 27*, 6-7 (2003), 621–627.

[243] He, R. L., Zhou, J., Hanson, C. Z., Chen, J., Cheng, N., and Ye, R. D. Serum amyloid A induces G-CSF expression and neutrophilia via Toll-like receptor 2. *Blood 113*, 2 (2009), 429–437.

[244] He, X., Chen, Y., Kang, S., Chen, G., and Wei, P. Differential Regulation of chTLR3 by Infectious Bursal Disease Viruses with Different Virulence In Vitro and In Vivo. *Viral Immunol 30*, 7 (2017), 490–499.

[245] Heegaard, P. M., Dedieu, L., Johnson, N., Le Potier, M. F., Mockey, M., Mutinelli, F., Vahlenkamp, T., Vascellari, M., and Sørensen, N. S. Adjuvants and delivery systems in veterinary vaccinology: current state and future developments. *Arch Virol 156*, 2 (2011), 183–202.

[246] Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science 303*, 5663 (2004), 1526–1529.

[247] Hemmi, H., Kaisho, T., Takeuchi, O., Sato, S., Sanjo, H., Hoshino, K., Horiuchi, T., Tomizawa, H., Takeda, K., and Akira, S. Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat Immunol 3*, 2 (2002), 196–200.

[248] Hentschel, U., Piel, J., Degnan, S. M., and Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol 10*, 9 (2012), 641–654.

[249] Hernandez, N., Melki, I., Jing, H., Habib, T., Huang, S. S. Y., Danielson, J., Kula, T., Drutman, S., Belkaya, S., Rattina, V., Lorenzo-Diaz, L., Boulai, A., Rose, Y., Kitabayashi, N., Rodero, M. P., Dumaine, C., Blanche, S., Lebras, M. N., Leung, M. C., Mathew, L. S., Boisson, B., Zhang, S. Y., Boisson-Dupuis, S., Giliani, S., Chaussabel, D., Notarangelo, L. D., Elledge, S. J., Ciancanelli, M. J., Abel, L., Zhang, Q., Marr, N., Crow, Y. J., Su, H. C., and Casanova, J. L. Life-threatening influenza pneumonitis in a child with inherited IRF9 deficiency. *J Exp Med 215*, 10 (2018), 2567–2585.

[250] Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., Buckley, K. M., Brockton, V., Nair, S. V., Berney, K., Fugmann, S. D., Anderson, M. K., Pancer, Z., Cameron, R. A., Smith, L. C., and Rast, J. P. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol 300*, 1 (2006), 349–365.

[251] Hidmark, A., von Saint Paul, A., and Dalpke, A. H. Cutting edge: TLR13 is a receptor for bacterial RNA. *J Immunol 189*, 6 (2012), 2717–2721.

[252] Higgs, R., Cormican, P., Cahalane, S., Allan, B., Lloyd, A. T., Meade, K., James, T., Lynn, D. J., Babiuk, L. A., and O'farrelly, C. Induction of a novel chicken Toll-like receptor following Salmonella enterica serovar Typhimurium infection. *Infect Immun 74*, 3 (2006), 1692–1698.

[253] Higuchi, M., Matsuo, A., Shingai, M., Shida, K., Ishii, A., Funami, K., Suzuki, Y., Oshiumi, H., Matsumoto, M., and Seya, T. Combinational recognition of bacterial lipoproteins and peptidoglycan by chicken Toll-like receptor 2 subfamily. *Dev Comp Immunol 32*, 2 (2008), 147–155.

[254] Hill, S. C. *Genomic approaches to virus discovery and molecular epidemiology.* PhD thesis, University of Oxford, 2017.

[255] Hill, S. C., Manvell, R. J., Schulenburg, B., Shell, W., Wikramaratna, P. S., Perrins, C., Sheldon, B. C., Brown, I. H., and Pybus, O. G. Antibody responses to avian influenza viruses in wild birds broaden with age. *Proc Biol Sci 283*, 1845 (2016).

[256] Hinke, J. T., Cossio, A. M., Goebel, M. E., Reiss, C. S., Trivelpiece, W. Z., and Watters, G. M. Identifying Risk: Concurrent Overlap of the Antarctic Krill Fishery with Krill-Dependent Predators in the Scotia Sea. *Plos One 12*, 1 (2017).

[257] Hiratsuka, S., Watanabe, A., Sakurai, Y., Akashi-Takamura, S., Ishibashi, S., Miyake, K., Shibuya, M., Akira, S., Aburatani, H., and Maru, Y. The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. *Nat Cell Biol 10*, 11 (2008), 1349–1355.

[258] Hocken, A. G. Cause of death in blue penguins (Eudyptula m. minor) in North Otago, New Zealand. *New Zealand Journal of Zoology 27*, 4 (2000), 305–309.

[259] Holm, C. K., Rahbek, S. H., Gad, H. H., Bak, R. O., Jakobsen, M. R., Jiang, Z., Hansen, A. L., Jensen, S. K., Sun, C., Thomsen, M. K., Laustsen, A., Nielsen, C. G., Severinsen, K., Xiong, Y., Burdette, D. L., Hornung, V., Lebbink, R. J., Duch, M., Fitzgerald, K. A., Bahrami, S., Mikkelsen, J. G., Hartmann, R., and Paludan, S. R. Influenza A virus targets a cGAS-independent STING pathway that controls enveloped RNA viruses. *Nat Commun 7* (2016), 10680.

[260] HOLMQUIST, G. U., WALKER, H. W., and STAHR, H. M. Influence of Temperature, pH, Water Activity and Antifungal Agents on Growth of Aspergillus flavus and A. parasiticus. *Journal of Food Science 48*, 3 (1983), 778–782.

[261] Horng, T., Barton, G. M., Flavell, R. A., and Medzhitov, R. The adaptor molecule TIRAP provides signalling specificity for Toll-like receptors. *Nature 420*, 6913 (2002), 329–333.

[262] Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Huang, H., Dong, M., Yu, C., Chen, S., and Xu, A. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res 18*, 7 (2008), 1112–1126.

[263] Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., Yi, K., Feng, H., Zhu, P., Li, B., Liu, Q., Fairley, S., Magor, K. E., Du, Z., Hu, X., Goodman, L., Tafer, H., Vignal, A., Lee, T., Kim, K. W., Sheng, Z., An, Y., Searle, S., Herrero, J., Groenen, M. A. M., Crooijmans, R. P. M.,

Faraut, T., Cai, Q., Webster, R. G., Aldridge, J. R., Warren, W. C., Bartschat, S., Kehr, S., Marz, M., Stadler, P. F., Smith, J., Kraus, R. H. S., Ren, L., Fei, J., Morisson, M., Kaiser, P., Griffin, D. K., Rao, M., Pitel, F., Wang, J., and Li, N. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet 45*, 7 (2013), 776–783.

[264] Huang, Y., Temperley, N. D., Ren, L., Smith, J., Li, N., and Burt, D. W. Molecular evolution of the vertebrate TLR1 gene family–a complex history of gene duplication, gene conversion, positive selection and co-evolution. *BMC Evol Biol 11* (2011), 149.

[265] Hughes, K. A., and Convey, P. The protection of Antarctic terrestrial ecosystems from inter- and intra-continental transfer of non-indigenous species by human activities: A review of current systems and practices. *Global Environmental Change 20*, 1 (2010), 96–112. Adaptive Capacity to Global Change in Latin America.

[266] Humphreys, J. M., Ramey, A. M., Douglas, D. C., Mullinax, J. M., Soos, C., Link, P., Walther, P., and Prosser, D. J. Waterfowl occurrence and residence time as indicators of H5 and H7 avian influenza in North American Poultry. *Sci Rep 10*, 1 (2020), 2592.

[267] Hurst, J., Prinz, N., Lorenz, M., Bauer, S., Chapman, J., Lackner, K. J., and von Landenberg, P. TLR7 and TLR8 ligands and antiphospholipid antibodies show synergistic effects on the induction of IL-1beta and caspase-1 in monocytes and dendritic cells. *Immunobiology 214*, 8 (2009), 683–691.

[268] IAATO. IAATO Antarctic visitor figures 2019-2020, 2020.

[269] Jaeger, A., Lebarbenchon, C., Bourret, V., Bastien, M., Lagadec, E., Thiebot, J.-B., Boulinier, T., Delord, K., Barbraud, C., Marteau, C., Dellagi, K., Tortosa, P., and Weimerskirch, H. Avian cholera outbreaks threaten seabird species on Amsterdam Island. *Plos One 13*, 5 (2018).

[270] Janeway, C. A. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol 54 Pt 1* (1989), 1–13.

[271] Janeway, C. A. The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunol Today 13*, 1 (1992), 11–16.

[272] Jeong, O. M., Kim, M. C., Kim, M. J., Kang, H. M., Kim, H. R., Kim, Y. J., Joh, S. J., Kwon, J. H., and Lee, Y. J. Experimental infection of chickens, ducks and quails with the highly pathogenic H5N1 avian influenza virus. *J Vet Sci 10*, 1 (2009), 53–60.

[273] Jin, M. S., Kim, S. E., Heo, J. Y., Lee, M. E., Kim, H. M., Paik, S. G., Lee, H., and Lee, J. O. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell 130*, 6 (2007), 1071–1082.

[274] Jin, W., Wu, D. D., Zhang, X., Irwin, D. M., and Zhang, Y. P. Positive selection on the gene RNASEL: correlation between patterns of evolution and function. *Mol Biol Evol 29*, 10 (2012), 3161–8.

[275] Johnson, C. M., Lyle, E. A., Omueti, K. O., Stepensky, V. A., Yegin, O., Alpsoy, E., Hamann, L., Schumann, R. R., and Tapping, R. I. Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol 178*, 12 (2007), 7520–7524.

[276] Johnson, G. B., Brunn, G. J., Kodaira, Y., and Platt, J. L. Receptor-mediated monitoring of tissue well-being via detection of soluble heparan sulfate by Toll-like receptor 4. *J Immunol 168*, 10 (2002), 5233–5239.

[277] Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinformatics 30*, 9 (2014), 1236–40.

[278] Joosten, L. A., Abdollahi-Roodsaz, S., Dinarello, C. A., O'Neill, L., and Netea, M. G. Toll-like receptors and chronic inflammation in rheumatic diseases: new developments. *Nat Rev Rheumatol 12*, 6 (2016), 344–357.

[279] Jørgensen, S. E., Christiansen, M., Ryø, L. B., Gad, H. H., Gjedsted, J., Staeheli, P., Mikkelsen, J. G., Storgaard, M., Hartmann, R., and Mogensen, T. H. Defective RNA sensing by RIG-I in severe influenza virus infection. *Clin Exp Immunol 192*, 3 (2018), 366–376.

[280] Jouventin, P. Les populations d'oiseaux marins des T.A.A.F.: résumé de 20 années de recherche. *Alauda 62* (1994), 44–47.

[281] Kall, L., Krogh, A., and Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology 338*, 5 (2004), 1027–1036.

[282] Kall, L., Krogh, A., and Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Research 35* (2007), W429–W432.

[283] Kalthoff, D., Breithaupt, A., Teifke, J. P., Globig, A., Harder, T., Mettenleiter, T. C., and Beer, M. Highly pathogenic avian influenza virus (H5N1) in experimentally infected adult mute swans. *Emerg Infect Dis 14*, 8 (2008), 1267–1270.

[284] Kaminski, M. M., Ohnemus, A., Cornitescu, M., and Staeheli, P. Plasmacytoid dendritic cells and Toll-like receptor 7-dependent signalling promote efficient protection of mice against highly virulent influenza A virus. *J Gen Virol 93*, Pt 3 (2012), 555–559.

[285] Kamiya, T., O'Dwyer, K., Nakagawa, S., and Poulin, R. Host diversity drives parasite diversity: meta-analytical insights into patterns and causal mechanisms. *Ecography 37*, 7 (2014), 689–697.

[286] Kane, O. J., Uhart, M. M., Rago, V., Pereda, A. J., Smith, J. R., Van Buren, A., Clark, J. A., and Boersma, P. D. Avian Pox in Magellanic Penguins (Spheniscus magellanicus). *Journal of Wildlife Diseases 48*, 3 (2012), 790–794.

[287] Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res 47*, D1 (2019), D590–D595.

[288] Kang, J. Y., Nan, X., Jin, M. S., Youn, S. J., Ryu, Y. H., Mah, S., Han, S. H., Lee, H., Paik, S. G., and Lee, J. O. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity 31*, 6 (2009), 873–884.

[289] Karikó, K., Ni, H., Capodici, J., Lamphier, M., and Weissman, D. mRNA is an endogenous ligand for Toll-like receptor 3. *J Biol Chem 279*, 13 (2004), 12542–12550.

[290] Karpala, A. J., Lowenthal, J. W., and Bean, A. G. Activation of the TLR3 pathway regulates IFNbeta production in chickens. *Dev Comp Immunol 32*, 4 (2008), 435–444.

[291] Katibah, G. E., Lee, H. J., Huizar, J. P., Vogan, J. M., Alber, T., and Collins, K. tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Mol Cell 49*, 4 (2013), 743–50.

[292] Katibah, G. E., Qin, Y., Sidote, D. J., Yao, J., Lambowitz, A. M., and Collins, K. Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. *Proc Natl Acad Sci U S A 111*, 33 (2014), 12025–30.

[293] Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T. S., Fujita, T., and Akira, S. Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med 205*, 7 (2008), 1601–1610.

[294] Katoh, K., Misawa, K., Kuma, K., and Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res 30*, 14 (2002), 3059–3066.

[295] Kawai, T., and Akira, S. TLR signaling. *Cell Death and Differentiation 13*, 5 (2006), 816–825.

[296] Kawai, T., and Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol 11*, 5 (2010), 373–384.

[297] Kawasaki, K., Akashi, S., Shimazu, R., Yoshida, T., Miyake, K., and Nishijima, M. Mouse toll-like receptor 4.MD-2 complex mediates lipopolysaccharide-mimetic signal transduction by Taxol. *J Biol Chem 275*, 4 (2000), 2251–2254.

[298] Kawasaki, T., and Kawai, T. Toll-like receptor signaling pathways. *Front Immunol 5* (2014), 461.

[299] Keawcharoen, J., van Riel, D., van Amerongen, G., Bestebroer, T., Beyer, W. E., van Lavieren, R., Osterhaus, A. D., Fouchier, R. A., and Kuiken, T. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). *Emerg Infect Dis 14*, 4 (2008), 600–607.

[300] Keestra, A. M., de Zoete, M. R., Bouwman, L. I., Vaezirad, M. M., and van Putten, J. P. Unique features of chicken Toll-like receptors. *Dev Comp Immunol 41*, 3 (2013), 316–323.

[301] Keestra, A. M., de Zoete, M. R., Bouwman, L. I., and van Putten, J. P. M. Chicken TLR21 Is an Innate CpG DNA Receptor Distinct from Mammalian TLR9. *Journal of Immunology 185*, 1 (2010), 460–467.

[302] Ketscher, L., Hannss, R., Morales, D. J., Basters, A., Guerra, S., Goldmann, T., Hausmann, A., Prinz, M., Naumann, R., Pekosz, A., Utermöhlen, O., Lenschow, D. J., and Knobeloch, K. P. Selective inactivation of USP18 isopeptidase activity in vivo enhances ISG15 conjugation and viral resistance. *Proc Natl Acad Sci U S A 112*, 5 (2015), 1577–1582.

[303] Keynan, Y., Juno, J., Meyers, A., Ball, T. B., Kumar, A., Rubinstein, E., and Fowke, K. R. Chemokine receptor 5 δ32 allele in patients with severe pandemic (H1N1) 2009. *Emerg Infect Dis 16*, 10 (2010), 1621–1622.

[304] Khan, Z. U., Pal, M., Paliwal, D. K., and Damodaran, V. N. Aspergillosis in imported penguins. *Sabouraudia 15*, 1 (1977), 43–45.

[305] Killip, M. J., Fodor, E., and Randall, R. E. Influenza virus activation of the interferon system. *Virus Res 209* (2015), 11–22.

[306] Kilpatrick, A. M., Chmura, A. A., Gibbons, D. W., Fleischer, R. C., Marra, P. P., and Daszak, P. Predicting the global spread of H5N1 avian influenza. *Proc Natl Acad Sci U S A 103*, 51 (2006), 19368–19373.

[307] Kim, K. S., Jung, H., Shin, I. K., Choi, B. R., and Kim, D. H. Induction of interleukin-1 beta (IL-1β) is a critical component of lung inflammation during influenza A (H1N1) virus infection. *J Med Virol 87*, 7 (2015), 1104–1112.

[308] Kim, S., Takahashi, H., Lin, W. W., Descargues, P., Grivennikov, S., Kim, Y., Luo, J. L., and Karin, M. Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature 457*, 7225 (2009), 102–106.

[309] Kimura, M. Evolutionary rate at the molecular level. *Nature 217*, 5129 (1968), 624–626.

[310] Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution 16*, 2 (1980), 111–120.

[311] Kimura, M. The Neutral Theory of Molecular Evolution, 1983.

[312] King, J. L., and Jukes, T. H. Non-Darwinian evolution. *Science 164*, 3881 (1969), 788–798.

[313] Kloch, A., Wenzel, M. A., Laetsch, D. R., Michalski, O., Bajer, A., Behnke, J. M., Welc-Faleciak, R., and Piertney, S. B. Signatures of balancing selection in toll-like receptor (TLRs) genes - novel insights from a free-living rodent. *Sci Rep 8*, 1 (2018), 8361.

[314] Klune, J. R., Dhupar, R., Cardinal, J., Billiar, T. R., and Tsung, A. HMGB1: endogenous danger signaling. *Mol Med 14*, 7-8 (2008), 476–484.

[315] Knafler, G. J., Clark, J. A., Boersma, P. D., and Bouzat, J. L. MHC Diversity and Mate Choice in the Magellanic Penguin, Spheniscus magellanicus. *Journal of Heredity 103*, 6 (2012), 759.

[316] Ko, J. H., Jin, H. K., Asano, A., Takada, A., Ninomiya, A., Kida, H., Hokiyama, H., Ohara, M., Tsuzuki, M., Nishibori, M., Mizutani, M., and Watanabe, T. Polymorphisms and the differential antiviral activity of the chicken Mx gene. *Genome Res 12*, 4 (2002), 595–601.

[317] Koblansky, A. A., Jankovic, D., Oh, H., Hieny, S., Sungnak, W., Mathur, R., Hayden, M. S., Akira, S., Sher, A., and Ghosh, S. Recognition of profilin by Toll-like receptor 12 is critical for host resistance to Toxoplasma gondii. *Immunity 38*, 1 (2013), 119–130.

[318] Koethe, S., Ulrich, L., Ulrich, R., Amler, S., Graaf, A., Harder, T. C., Grund, C., Mettenleiter, T. C., Conraths, F. J., Beer, M., and Globig, A. Modulation of lethal HPAIV H5N8 clade 2.3.4.4B infection in AIV pre-exposed mallards. *Emerg Microbes Infect 9*, 1 (2020), 180–193.

[319] Koliopoulos, M. G., Lethier, M., van der Veen, A. G., Haubrich, K., Hennig, J., Kowalinski, E., Stevens, R. V., Martin, S. R., Reis e Sousa, C., Cusack, S., and Rittinger, K. Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. *Nat Commun 9*, 1 (2018), 1820.

[320] Kollewe, C., Mackensen, A. C., Neumann, D., Knop, J., Cao, P., Li, S., Wesche, H., and Martin, M. U. Sequential autophosphorylation steps in the interleukin-1 receptor-associated kinase-1 regulate its availability as an adapter in interleukin-1 signaling. *J Biol Chem 279*, 7 (2004), 5227–5236.

[321] Korbie, D. J., and Mattick, J. S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols 3*, 9 (2008), 1452–1456.

[322] Kosakovsky Pond, S. L., and Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol 22*, 5 (2005), 1208–1222.

[323] Kosiol, C., Vinar, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., and Siepel, A. Patterns of positive selection in six Mammalian genomes. *PLoS Genet 4*, 8 (2008), e1000144.

[324] Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics (Oxford, England)* (2019).

[325] Králová, T., Albrecht, T., Bryja, J., Hořák, D., Johnsen, A., Lifjeld, J. T., Novotný, M., Sedláček, O., Velová, H., and Vinkler, M. Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an evolutionary context. *Mol Ecol 27*, 13 (2018), 2871–2883.

[326] Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., Palese, P., Shaw, M. L., Treanor, J., Webster, R. G., and García-Sastre, A. Influenza. *Nat Rev Dis Primers 4*, 1 (2018), 3.

[327] Krieg, A. M. CpG motifs in bacterial DNA and their immune effects. *Annu Rev Immunol 20* (2002), 709–760.

[328] Krieg, A. M., Yi, A. K., Matson, S., Waldschmidt, T. J., Bishop, G. A., Teasdale, R., Koretzky, G. A., and Klinman, D. M. CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature 374*, 6522 (1995), 546–549.

[329] Krol, L., Moore, R. P., Mutlow, A. G., Brady, S. M., and Dorsa, D. A retrospective analysis of mortality in captive Magellanic penguins (Spheniscus magellanicus) in the United States, 2008-2018. *Zoo Biology 39*, 6 (2020), 405–410.

[330] Ksepka, D. T., Bertelli, S., and Giannini, N. P. The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics 22*, 5 (2006), 412–441.

[331] Kurt-Jones, E. A., Popova, L., Kwinn, L., Haynes, L. M., Jones, L. P., Tripp, R. A., Walsh, E. E., Freeman, M. W., Golenbock, D. T., Anderson, L. J., and Finberg, R. W. Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nat Immunol 1*, 5 (2000), 398–401.

[332] Laayouni, H., Oosting, M., Luisi, P., Ioana, M., Alonso, S., Ricaño-Ponce, I., Trynka, G., Zhernakova, A., Plantinga, T. S., Cheng, S. C., van der Meer, J. W., Popp, R., Sood, A., Thelma, B. K., Wijmenga, C., Joosten, L. A., Bertranpetit, J., and Netea, M. G.

Convergent evolution in European and Rroma populations reveals pressure exerted by plague on Toll-like receptors. *Proc Natl Acad Sci U S A 111*, 7 (2014), 2668–2673.

[333] Lai, C. Y., Yu, G. Y., Luo, Y., Xiang, R., and Chuang, T. H. Immunostimulatory Activities of CpG-Oligodeoxynucleotides in Teleosts: Toll-Like Receptors 9 and 21. *Front Immunol 10* (2019), 179.

[334] Lande, R., Gregorio, J., Facchinetti, V., Chatterjee, B., Wang, Y. H., Homey, B., Cao, W., Su, B., Nestle, F. O., Zal, T., Mellman, I., Schröder, J. M., Liu, Y. J., and Gilliet, M. Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide. *Nature 449*, 7162 (2007), 564–569.

[335] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. Initial sequencing and analysis of the human genome. *Nature 409*, 6822 (2001), 860–921.

[336] Lang, T., and Mansell, A. The negative regulation of Toll-like receptor and associated pathways. *Immunol Cell Biol 85*, 6 (2007), 425–434.

[337] Larson, G., and Burger, J. A population genetics view of animal domestication. *Trends in Genetics 29*, 4 (2013), 197–205.

[338] Larson, G., and Fuller, D. Q. The Evolution of Animal Domestication. 115–136.

[339] Lawal, R. A., Martin, S. H., Vanmechelen, K., Vereijken, A., Silva, P., Al-Atiyat, R. M., Aljumaah, R. S., Mwacharo, J. M., Wu, D.-D., Zhang, Y.-P., Hocking, P. M., Smith, J., Wragg, D., and Hanotte,

O. The wild species genome ancestry of domestic chickens. *BMC Biology 18*, 1 (2020).

[340] Le Goffic, R., Pothlichet, J., Vitour, D., Fujita, T., Meurs, E., Chignard, M., and Si-Tahar, M. Cutting Edge: Influenza A virus activates TLR3-dependent inflammatory and RIG-I-dependent antiviral responses in human lung epithelial cells. *J Immunol 178*, 6 (2007), 3368–3372.

[341] Leadbetter, E. A., Rifkin, I. R., Hohlbaum, A. M., Beaudette, B. C., Shlomchik, M. J., and Marshak-Rothstein, A. Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature 416*, 6881 (2002), 603–607.

[342] Lee, B. L., Moon, J. E., Shu, J. H., Yuan, L., Newman, Z. R., Schekman, R., and Barton, G. M. UNC93B1 mediates differential trafficking of endosomal TLRs. *Elife 2* (2013), e00291.

[343] Lee, C. C., Avalos, A. M., and Ploegh, H. L. Accessory molecules for Toll-like receptors and their function. *Nat Rev Immunol 12*, 3 (2012), 168–179.

[344] Lee, N., Cao, B., Ke, C., Lu, H., Hu, Y., Tam, C. H. T., Ma, R. C. W., Guan, D., Zhu, Z., Li, H., Lin, M., Wong, R. Y. K., Yung, I. M. H., Hung, T. N., Kwok, K., Horby, P., Hui, D. S. C., Chan, M. C. W., and Chan, P. K. S. IFITM3, TLR3, and CD55 Gene SNPs and Cumulative Genetic Risks for Severe Outcomes in Chinese Patients With H7N9/H1N1pdm09 Influenza. *J Infect Dis 216*, 1 (2017), 97–104.

[345] Lee, S. M., Yip, T. F., Yan, S., Jin, D. Y., Wei, H. L., Guo, R. T., and Peiris, J. S. M. Recognition of Double-Stranded RNA and Regulation of Interferon Pathway by Toll-Like Receptor 10. *Front Immunol 9* (2018), 516.

[346] Leifer, C. A., Kennedy, M. N., Mazzoni, A., Lee, C., Kruhlak, M. J., and Segal, D. M. TLR9 is localized in the endoplasmic reticulum prior to stimulation. *J Immunol 173*, 2 (2004), 1179–1183.

[347] Leigh, J. W., and Bryant, D. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution 6*, 9 (2015), 1110–1116.

[348] Lemaho, Y., Karmann, H., Briot, D., Handrich, Y., Robin, J. P., Mioskowski, E., Cherel, Y., and Farni, J. STRESS IN BIRDS DUE TO ROUTINE HANDLING AND A TECHNIQUE TO AVOID IT. *American Journal of Physiology 263*, 4 (1992), R775–R781.

[349] Lemaitre, B., and Hoffmann, J. The host defense of Drosophila melanogaster. *Annu Rev Immunol 25* (2007), 697–743.

[350] Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., and Hoffmann, J. A. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell 86*, 6 (1996), 973–983.

[351] Lenstra, J. A., Ajmone-Marsan, P., Beja-Pereira, A., Bollongino, R., Bradley, D. G., Colli, L., Gaetano, A. d., Edwards, C. J., Felius, M., Ferretti, L., Ginja, C., Hristov, P., Kantanen, J., Lirón, J. P., Magee, D. A., Negrini, R., and Radoslavov, G. A. Meta-analysis of mitochondrial DNA reveals several population bottlenecks. *Diversity 6*, 1 (2014), 178–187.

[352] Lepage, D., Vaidya, G., and Guralnick, R. Avibase - a database system for managing and organizing taxonomic concepts. *Zookeys*, 420 (2014), 117–135.

[353] Lescroel, A., and Bost, C. A. Recent decrease in gentoo penguin populations at Iles Kerguelen. *Antarctic Science 18*, 2 (2006), 171–174.

[354] Letunic, I., and Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* (2021).

[355] Leulier, F., and Lemaitre, B. Toll-like receptors–taking an evolutionary approach. *Nat Rev Genet 9*, 3 (2008), 165–178.

[356] Levin, I. I., Zwiers, P., Deem, S. L., Geest, E. A., Higashiguchi, J. M., Iezhova, T. A., Jimenez-Uzcategui, G., Kim, D. H., Morton, J. P., Perlut, N. G., Renfrew, R. B., Sari, E. H. R., Valkiunas, G., and Parker, P. G. Multiple Lineages of Avian Malaria Parasites (Plasmodium) in the Galapagos Islands and Evidence for Arrival via Migratory Birds. *Conservation Biology 27*, 6 (2013), 1366–1377.

[357] Levy, H., Clucas, G. V., Rogers, A. D., Leache, A. D., Ciborowski, K. L., Polito, M. J., Lynch, H. J., Dunn, M. J., and Hart, T. Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc. *Ecology and Evolution 6*, 6 (2016), 1834–1853.

[358] Levy, H., Fiddaman, S. R., Vianna, J. A., Noll, D., Clucas, G. V., Sidhu, J. K. H., Polito, M. J., Bost, C. A., Phillips, R. A., Crofts, S., Miller, G. D., Pistorius, P., Bonnadonna, F., Bohec, C. L., Barbosa, A. A., Trathan, P., Rey, A. R., Frantz, L. A. F., Hart, T., and Smith, A. L. Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Mol Biol Evol* (2020).

[359] Li, C., Zhang, Y., Li, J., Kong, L., Hu, H., Pan, H., Xu, L., Deng, Y., Li, Q., Jin, L., Yu, H., Chen, Y., Liu, B., Yang, L., Liu, S., Lang, Y., Xia, J., He, W., Shi, Q., Subramanian, S., Millar, C. D., Meader, S., Rands, C. M., Fujita, M. K., Greenwold, M. J., Castoe, T. A., Pollock, D. D., Gu, W., Nam, K., Ellegren, H., Ho, S. Y., Burt, D. W., Ponting, C. P., Jarvis, E. D., Gilbert, M. T., Yang, H., Wang, J., Lambert, D. M., and Zhang, G. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience 3*, 1 (2014), 27.

[360] Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics 27*, 5 (2011), 718–719.

[361] Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *AarXiv 1303.3997v2* (2013).

[362] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data, P. The Sequence Alignment/Map format and SAMtools. *Bioinformatics 25*, 16 (2009), 2078–2079.

[363] Li, S., Min, J. Y., Krug, R. M., and Sen, G. C. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology 349*, 1 (2006), 13–21.

[364] Li, S., Wang, G., Liu, D., Liu, Q., and Hu, G. Cloning and expression analysis of a Toll-like receptor 21 (TLR21) gene from turbot, Scophthalmus maximus. *Dev Comp Immunol 73* (2017), 163–168.

[365] Li, W. H., Gojobori, T., and Nei, M. Pseudogenes as a paradigm of neutral evolution. *Nature 292*, 5820 (1981), 237–239.

[366] Li, X. D., and Chen, Z. J. Sequence specific detection of bacterial 23S ribosomal RNA by TLR13. *Elife 1* (2012), e00102.

[367] Li, Y., Chen, R., Zhou, Q., Xu, Z., Li, C., Wang, S., Mao, A., Zhang, X., He, W., and Shu, H. B. LSm14A is a processing body-associated sensor of viral nucleic acids that initiates cellular antiviral response in the early phase of viral infection. *Proc Natl Acad Sci U S A 109*, 29 (2012), 11770–11775.

[368] Lim, H. K., Huang, S. X. L., Chen, J., Kerner, G., Gilliaux, O., Bastard, P., Dobbs, K., Hernandez, N., Goudin, N., Hasek, M. L., García Reino, E. J., Lafaille, F. G., Lorenzo, L., Luthra, P., Kochetkov, T., Bigio, B., Boucherit, S., Rozenberg, F., Vedrinne, C., Keller, M. D., Itan, Y., García-Sastre, A., Celard, M., Orange, J. S., Ciancanelli, M. J., Meyts, I., Zhang, Q., Abel, L., Notarangelo, L. D., Snoeck, H. W., Casanova, J. L., and Zhang, S. Y. Severe influenza pneumonitis in children with inherited TLR3 deficiency. *J Exp Med 216*, 9 (2019), 2038–2056.

[369] Lin, S. C., Lo, Y. C., and Wu, H. Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature 465*, 7300 (2010), 885–890.

[370] Lindgren, E., Tälleklint, L., and Polfeldt, T. Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick Ixodes ricinus. *Environ Health Perspect 108*, 2 (2000), 119–123.

[371] Liniger, M., Summerfield, A., Zimmer, G., McCullough, K. C., and Ruggli, N. Chicken cells sense influenza A virus infection through MDA5 and CARDIF signaling involving LGP2. *J Virol 86*, 2 (2012), 705–717.

[372] Liu, B., Mori, I., Hossain, M. J., Dong, L., Takeda, K., and Kimura, Y. Interleukin-18 improves the early defence system against influenza virus infection by augmenting natural killer cell-mediated cytotoxicity. *J Gen Virol 85*, Pt 2 (2004), 423–428.

[373] Liu, G., Zhang, H., Sun, G., Zhao, C., Shang, S., Gao, X., Xia, T., and Yang, X. Characterization of the peripheral blood transcriptome and adaptive evolution of the MHC I and TLR gene families in the wolf (Canis lupus). *BMC Genomics 18*, 1 (2017), 584.

[374] Liu, G., Zhang, H., and Zhao, C. Evolutionary History of the Toll-Like Receptor Gene Family across Vertebrates. *Genome Biol Evol 12*, 1 (2020), 3615–3634.

[375] Liu, J., Wei, Y., Lu, Y., Li, Y., and Chen, Q. The discovery of potent immunostimulatory CpG-ODNs widely distributed in bacterial genomes. *J Microbiol 58*, 2 (2020), 153–162.

[376] Liu, J., Xu, C., Hsu, L. C., Luo, Y., Xiang, R., and Chuang, T. H. A five-amino-acid motif in the undefined region of the TLR8 ectodomain is required for species-specific ligand recognition. *Mol Immunol 47*, 5 (2010), 1083–1090.

[377] Liu, L., Botos, I., Wang, Y., Leonard, J. N., Shiloach, J., Segal, D. M., and Davies, D. R. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science 320*, 5874 (2008), 379–381.

[378] Liu, Y., Li, S., Zhang, G., Nie, G., Meng, Z., Mao, D., Chen, C., Chen, X., Zhou, B., and Zeng, G. Genetic variants in IL1A and IL1B contribute to the susceptibility to 2009 pandemic H1N1 influenza A virus. *BMC Immunol 14* (2013), 37.

[379] Liu, Y., Wu, G., Yao, Y., Miao, Y., Luikart, G., Baig, M., Beja-Pereira, A., Ding, Z., Palanichamy, M., and Zhang, Y. Multiple maternal origins of chickens: Out of the Asian jungles. *Molecular Phylogenetics and Evolution 38*, 1 (2006), 12–19.

[380] Liu, Y., Yang, Q., and Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annual Review of Biochemistry 90*, 1 (2021), 375–401. PMID: 33441035.

[381] Liu, Y., Yin, H., Zhao, M., and Lu, Q. TLR2 and TLR4 in autoimmune diseases: a comprehensive review. *Clin Rev Allergy Immunol 47*, 2 (2014), 136–147.

[382] Loo, Y. M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M. A., García-Sastre, A., Katze, M. G., and Gale, M. Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol 82*, 1 (2008), 335–345.

[383] Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol 1079* (2014), 155–70.

[384] Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Barker-Collo, S., Bartels, D. H., Bell, M. L., Benjamin, E. J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S. S., Coffeng, L. E., Colan, S. D., Colquhoun, S., Colson, K. E., Condon, J., Connor, M. D., Cooper, L. T., Corriere, M., Cortinovis, M., de Vaccaro, K. C., Couser, W., Cowie, B. C., Criqui, M. H., Cross, M., Dabhadkar, K. C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D. C., Dharmaratne, S. D., Dorsey, E. R., Driscoll, T., Duber, H., Ebel, B., Erwin, P. J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A. D., Forouzanfar, M. H., Fowkes, F. G., Franklin, R., Fransen, M., Freeman, M. K., Gabriel, S. E., Gakidou, E., Gaspari, F., Gillum, R. F., Gonzalez-Medina, D., Halasa, Y. A., Haring, D., Harrison, J. E., Havmoeller, R., Hay, R. J., Hoen, B., Hotez, P. J., Hoy, D., Jacobsen, K. H., James, S. L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J. P., Knowlton, L. M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S. E., Ohno, S. L., et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet 380*, 9859 (2012), 2095–2128.

[385] Lund, J. M., Alexopoulou, L., Sato, A., Karow, M., Adams, N. C., Gale, N. W., Iwasaki, A., and Flavell, R. A. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A 101*, 15 (2004), 5598–5603.

[386] Luo, J., Obmolova, G., Malia, T. J., Wu, S. J., Duffy, K. E., Marion, J. D., Bell, J. K., Ge, P., Zhou, Z. H., Teplyakov, A., Zhao, Y., Lamb, R. J., Jordan, J. L., San Mateo, L. R., Sweet, R. W., and Gilliland, G. L. Lateral clustering of TLR3:dsRNA signaling units revealed by TLR3ecd:3Fabs quaternary structure. *J Mol Biol 421*, 1 (2012), 112–124.

[387] Lynch, H. *The Gentoo Penguin (Pygoscelis papua)*. University of Washington Press, Seattle, USA, 2013.

[388] Lynch, H. J., Naveen, R., Trathan, P. N., and Fagan, W. F. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology 93*, 6 (2012), 1367–1377.

[389] Ma, D., Zhu, R., Ding, W., Shen, C., Chu, H., and Lin, X. Ex-situ enzyme activity and bacterial community diversity through soil depth profiles

in penguin and seal colonies on Vestfold Hills, East Antarctica. *Polar Biology 36*, 9 (2013), 1347–1361.

[390] Ma, X., Liu, Y., Gowen, B. B., Graviss, E. A., Clark, A. G., and Musser, J. M. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS One 2*, 12 (2007), e1318.

[391] Mac Donald, J. W., and Conroy, J. W. H. Virus disease resembling puffinosis in the Gentoo penguin (Pygoscelis papua) on Signy Island (South Orkney Islands). *British Antarctic Survey Bulletin*, 26 (1971), 80–82.

[392] Malmstrøm, M., Jentoft, S., Gregers, T. F., and Jakobsen, K. S. Unraveling the evolution of the Atlantic cod's (Gadus morhua L.) alternative immune strategy. *PLoS One 8*, 9 (2013), e74004.

[393] Manry, J., Laval, G., Patin, E., Fornarino, S., Itan, Y., Fumagalli, M., Sironi, M., Tichit, M., Bouchier, C., Casanova, J. L., Barreiro, L. B., and Quintana-Murci, L. Evolutionary genetic dissection of human interferons. *J Exp Med 208*, 13 (2011), 2747–2759.

[394] Marsden, C. D., Ortega-Del Vecchyo, D., O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R. D., Wayne, R. K., and Lohmueller, K. E. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci U S A 113*, 1 (2016), 152–157.

[395] Marshak-Rothstein, A. Toll-like receptors in systemic autoimmune disease. *Nat Rev Immunol 6*, 11 (2006), 823–835.

[396] Matzinger, P. Tolerance, danger, and the extended family. *Annu Rev Immunol 12* (1994), 991–1045.

[397] Mayewski, P. State of the Antarctic Climate system - Excerpts from SASOCS (Mayewski et al. 2009). *Anales del Instituto de la Patagonia 40*, 1 (2012), 25–30.

[398] McGettrick, A. F., and O'Neill, L. A. The expanding family of MyD88-like adaptors in Toll-like receptor signal transduction. *Mol Immunol 41*, 6-7 (2004), 577–582.

[399] McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., and Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol 17*, 1 (2016), 122.

[400] Means, T. K., Latz, E., Hayashi, F., Murali, M. R., Golenbock, D. T., and Luster, A. D. Human lupus autoantibody-DNA complexes activate DCs through cooperation of CD32 and TLR9. *J Clin Invest 115*, 2 (2005), 407–417.

[401] Medvedev, A. E. Toll-like receptor polymorphisms, inflammatory and infectious diseases, allergies, and cancer. *J Interferon Cytokine Res 33*, 9 (2013), 467–484.

[402] Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature 388*, 6640 (1997), 394–397.

[403] Melo, A. M., Stevens, D. A., Tell, L. A., Veríssimo, C., Sabino, R., and Xavier, M. O. Aspergillosis, Avian Species and the One Health Perspective: The Possible Importance of Birds in Azole Resistance. *Microorganisms 8*, 12 (2020).

[404] Meyerson, N. R., Zhou, L., Guo, Y. R., Zhao, C., Tao, Y. J., Krug, R. M., and Sawyer, S. L. Nuclear TRIM25 Specifically Targets Influenza Virus Ribonucleoproteins to Block the Onset of RNA Chain Elongation. *Cell Host Microbe 22*, 5 (2017), 627–638.

[405] Miao, Y.-W., Peng, M.-S., Wu, G.-S., Ouyang, Y.-N., Yang, Z.-Y., Yu, N., Liang, J.-P., Pianchou, G., Beja-Pereira, A., Mitra, B., Palanichamy, M. G., Baig, M., Chaudhuri, T. K., Shen, Y.-Y., Kong, Q.-P., Murphy, R. W., Yao, Y.-G., and Zhang, Y.-P. Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity 110*, 3 (2013), 277–282.

[406] Michel, O. Role of lipopolysaccharide (LPS) in asthma and other pulmonary conditions. *J Endotoxin Res 9*, 5 (2003), 293–300.

[407] Mikami, T., Miyashita, H., Takatsuka, S., Kuroki, Y., and Matsushima, N. Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene 503*, 2 (2012), 235–243.

[408] Miller, Y. I., Viriyakosol, S., Binder, C. J., Feramisco, J. R., Kirkland, T. N., and Witztum, J. L. Minimally modified LDL binds to CD14, induces macrophage spreading via TLR4/MD-2, and inhibits phagocytosis of apoptotic cells. *J Biol Chem 278*, 3 (2003), 1561–1568.

[409] Mineev, K. S., Goncharuk, S. A., and Arseniev, A. S. Toll-like receptor 3 transmembrane domain is able to perform various homotypic interactions: an NMR structural study. *FEBS Lett 588*, 21 (2014), 3802–3807.

[410] Miyata, T., Miyazawa, S., and Yasunaga, T. Two types of amino acid substitutions in protein evolution. *J Mol Evol 12*, 3 (1979), 219–36.

[411] Mockenhaupt, F. P., Hamann, L., von Gaertner, C., Bedu-Addo, G., von Kleinsorgen, C., Schumann, R. R., and Bienzle, U. Common polymorphisms of toll-like receptors 4 and 9 are associated with the clinical manifestation of malaria during pregnancy. *J Infect Dis 194*, 2 (2006), 184–188.

[412] Modhiran, N., Watterson, D., Muller, D. A., Panetta, A. K., Sester, D. P., Liu, L., Hume, D. A., Stacey, K. J., and Young, P. R. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. *Sci Transl Med 7*, 304 (2015), 304ra142.

[413] Molini, U., Aikukutu, G., Roux, J. P., Kemper, J., Ntahonshikira, C., Marruchella, G., Khaiseb, S., Cattoli, G., and Dundon, W. G. Avian Influenza H5N8 Outbreak in African Penguins (. *J Wildl Dis 56*, 1 (2020), 214–218.

[414] Moore, J. K., Abbott, M. R., and Richman, J. G. Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. *Journal of Geophysical Research-Oceans 104*, C2 (1999), 3059–3073.

[415] Morgan, I., and Westbury, H. Studies of viruses in penguins in the Vestfold hills. *Hydrobiologia 165* (1988), 263–269.

[416] Mukherjee, S., Ganguli, D., and Majumder, P. P. Global Footprints of Purifying Selection on Toll-Like Receptor Genes Primarily Associated with Response to Bacterial Infections in Humans. *Genome Biology and Evolution 6*, 3 (2014), 551–558.

[417] Mulla, M. J., Brosens, J. J., Chamley, L. W., Giles, I., Pericleous, C., Rahman, A., Joyce, S. K., Panda, B., Paidas, M. J., and Abrahams, V. M. Antiphospholipid antibodies induce a pro-inflammatory response in first trimester trophoblast via the TLR4/MyD88 pathway. *Am J Reprod Immunol 62*, 2 (2009), 96–111.

[418] Murakami, S., Iwaki, D., Mitsuzawa, H., Sano, H., Takahashi, H., Voelker, D. R., Akino, T., and Kuroki, Y. Surfactant protein A inhibits peptidoglycan-induced tumor necrosis factor-alpha secretion in U937 cells and alveolar macrophages by direct interaction with toll-like receptor 2. *J Biol Chem 277*, 9 (2002), 6830–6837.

[419] Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., and Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution 30*, 5 (2013), 1196–1205.

[420] Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Pond, S. L. K. Detecting Individual Sites Subject to Episodic Diversifying Selection. *Plos Genetics 8*, 7 (2012).

[421] Nagy, A., Machova, J., Hornickova, J., Tomci, M., Nagl, I., Horyna, B., and Holko, I. Highly pathogenic avian influenza virus subtype H5N1 in Mute swans in the Czech Republic. *Vet Microbiol 120*, 1-2 (2007), 9–16.

[422] Nahori, M. A., Fournie-Amazouz, E., Que-Gewirth, N. S., Balloy, V., Chignard, M., Raetz, C. R. H., Saint Girons, I., and Werts, C. Differential TLR recognition of leptospiral lipid A and lipopolysaccharide in muine and human cells. *Journal of Immunology 175*, 9 (2005), 6022–6031.

[423] Nakajima, T., Ohtani, H., Satta, Y., Uno, Y., Akari, H., Ishida, T., and Kimura, A. Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics 60*, 12 (2008), 727–35.

[424] Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J. B., Backström, N., Künstner, A., Balakrishnan, C. N., Heger, A., Ponting, C. P., Clayton, D. F., and Ellegren, H. Molecular evolution of genes in avian genomes. *Genome Biol 11*, 6 (2010), R68.

[425] Napier, J. E., Hinrichs, S. H., Lampen, F., Iwen, P. C., Wickert, R. S., Garrett, J. L., Aden, T. A., Restis, E., Curro, T. G., Simmons, L. G., and Armstrong, D. L. An outbreak of avian mycobacteriosis caused by Mycobacterium intracellulare in little blue penguins (Eudyptula minor). *J Zoo Wildl Med 40*, 4 (2009), 680–6.

[426] Nei, M. *Molecular evolutionary genetics*. Columbia University Press, New York, 1987.

[427] Neira, V., Tapia, R., Verdugo, C., Barriga, G., Mor, S., Ng, T. F. F., García, V., Del Río, J., Rodrigues, P., Briceño, C., Medina, R. A., and González-Acuña, D. Novel Avulaviruses in Penguins, Antarctica. *Emerg Infect Dis 23*, 7 (2017), 1212–1214.

[428] Nembrini, C., Kisielow, J., Shamshiev, A. T., Tortola, L., Coyle, A. J., Kopf, M., and Marsland, B. J. The kinase activity of Rip2 determines its stability and consequently Nod1- and Nod2-mediated immune responses. *J Biol Chem 284*, 29 (2009), 19183–19188.

[429] Netea, M. G., Van der Meer, J. W., Sutmuller, R. P., Adema, G. J., and Kullberg, B. J. From the Th1/Th2 paradigm towards a Toll-like receptor/T-helper bias. *Antimicrob Agents Chemother 49*, 10 (2005), 3991–3996.

[430] Newell, D. G., Elvers, K. T., Dopfer, D., Hansson, I., Jones, P., James, S., Gittins, J., Stern, N. J., Davies, R., Connerton, I., Pearson, D., Salvat, G., and Allen, V. M. Biosecurity-based interventions and strategies to reduce Campylobacter spp. on poultry farms. *Appl Environ Microbiol 77*, 24 (2011), 8605–8614.

[431] Newman, S. H., Iverson, S. A., Takekawa, J. Y., Gilbert, M., Prosser, D. J., Batbayar, N., Natsagdorj, T., and Douglas, D. C. Migration of whooper swans and outbreaks of highly pathogenic avian influenza H5N1 virus in eastern Asia. *PLoS One 4*, 5 (2009), e5729.

[432] Ng, P. C., and Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res 31*, 13 (2003), 3812–4.

[433] Nie, L., Cai, S.-Y., Shao, J.-Z., and Chen, J. Toll-Like Receptors, Associated Biological Roles, and Signaling Networks in Non-Mammals. *Frontiers in Immunology 9* (2018), 1523.

[434] Noden, B. H., Kent, M. D., and Beier, J. C. The impact of variations in temperature on early Plasmodium falciparum development in Anopheles stephensi. *Parasitology 111 (Pt 5)* (1995), 539–545.

[435] Nomura, F., Akashi, S., Sakao, Y., Sato, S., Kawai, T., Matsumoto, M., Nakanishi, K., Kimoto, M., Miyake, K., Takeda, K., and Akira, S. Cutting edge: endotoxin tolerance in mouse peritoneal macrophages correlates with down-regulation of surface toll-like receptor 4 expression. *J Immunol 164*, 7 (2000), 3476–3479.

[436] Norderhaug, I. N., Johansen, F. E., Krajci, P., and Brandtzaeg, P. Domain deletions in the human polymeric Ig receptor disclose differences between its dimeric IgA and pentameric IgM interaction. *Eur J Immunol 29*, 10 (1999), 3401–9.

[437] Novak, K., Bjelka, M., Samake, K., and Valcikova, T. Potential of TLR-gene diversity in Czech indigenous cattle for resistance breeding as revealed by hybrid sequencing. *Archives Animal Breeding 62*, 2 (2019), 477–490.

[438] Nozawa, M., Suzuki, Y., and Nei, M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci U S A 106*, 16 (2009), 6700–6705.

[439] Nutsford, A. N., Galvin, H. D., Ahmed, F., and Husain, M. The Class IV human deacetylase, HDAC11, exhibits anti-influenza A virus properties via its involvement in host innate antiviral response. *Cell Microbiol 21*, 4 (2019), e12989.

[440] Nyman, T., Stenmark, P., Flodin, S., Johansson, I., Hammarström, M., and Nordlund, P. The crystal structure of the human toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. *J Biol Chem 283*, 18 (2008), 11861–11865.

[441] Obendorf, D. L., and McColl, K. Mortality in little penguins (Eudyptula minor) along the coast of Victoria, Australia. *J Wildl Dis 16*, 2 (1980), 251–259.

[442] O'Connor, B. D., and Auwera, G. v. d. *Genomics in the cloud : using Docker, GATK, and WDL in Terra*. O'Reilly, Beijing, 2020.

[443] O'Connor, E. A., Hasselquist, D., Nilsson, J., Westerdahl, H., and Cornwallis, C. K. Wetter climates select for higher immune gene diversity in resident, but not migratory, songbirds. *Proc Biol Sci 287*, 1919 (2020), 20192675.

[444] Ohashi, K., Burkart, V., Flohé, S., and Kolb, H. Cutting edge: heat shock protein 60 is a putative endogenous ligand of the toll-like receptor-4 complex. *J Immunol 164*, 2 (2000), 558–561.

[445] Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature 246*, 5428 (1973), 96–98.

[446] Ohto, U., Yamakawa, N., Akashi-Takamura, S., Miyake, K., and Shimizu, T. Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *J Biol Chem 287*, 48 (2012), 40611–40617.

[447] Ohya, M., Nishitani, C., Sano, H., Yamada, C., Mitsuzawa, H., Shimizu, T., Saito, T., Smith, K., Crouch, E., and Kuroki, Y. Human pulmonary surfactant protein D binds the extracellular domains of Toll-like receptors 2 and 4 through the carbohydrate recognition domain by a mechanism different from its binding to phosphatidylinositol and lipopolysaccharide. *Biochemistry 45*, 28 (2006), 8657–8664.

[448] Okamura, Y., Watari, M., Jerud, E. S., Young, D. W., Ishizaka, S. T., Rose, J., Chow, J. C., and Strauss, J. F. The extra domain A of fibronectin activates Toll-like receptor 4. *J Biol Chem 276*, 13 (2001), 10229–10233.

[449] Oldenburg, M., Krüger, A., Ferstl, R., Kaufmann, A., Nees, G., Sigmund, A., Bathke, B., Lauterbach, H., Suter, M., Dreher, S., Koedel, U., Akira, S., Kawai, T., Buer, J., Wagner, H., Bauer, S., Hochrein, H., and Kirschning, C. J. TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification. *Science 337*, 6098 (2012), 1111–1115.

[450] Olivares, F., Tapia, R., Gálvez, C., Meza, F., Barriga, G. P., Borras-Chavez, R., Mena-Vasquez, J., Medina, R. A., and Neira, V. Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the Antarctic Peninsula. *Transbound Emerg Dis* (2019).

[451] Oliveira-Nascimento, L., Massari, P., and Wetzler, L. M. The Role of TLR2 in Infection and Immunity. *Front Immunol 3* (2012), 79.

[452] Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D., and Fouchier, R. A. Global patterns of influenza a virus in wild birds. *Science 312*, 5772 (2006), 384–388.

[453] Oshiumi, H., Miyashita, M., Inoue, N., Okabe, M., Matsumoto, M., and Seya, T. The ubiquitin ligase Riplet is essential for RIG-I-dependent innate immune responses to RNA virus infection. *Cell Host Microbe 8*, 6 (2010), 496–509.

[454] Osland, M. J., and Feher, L. C. Winter climate change and the poleward range expansion of a tropical invasive tree (Brazilian pepper-Schinus terebinthifolius). *Glob Chang Biol 26*, 2 (2020), 607–615.

[455] Oven, I., Resman Rus, K., Dušanić, D., Benčina, D., Keeler, C. L., and Narat, M. Diacylated lipopeptide from Mycoplasma synoviae mediates TLR15 induced innate immune responses. *Vet Res 44* (2013), 99.

[456] Pálmai, N., Erdélyi, K., Bálint, A., Márton, L., Dán, A., Deim, Z., Ursu, K., Löndt, B. Z., Brown, I. H., and Glávits, R. Pathobiology of highly pathogenic avian influenza virus (H5N1) infection in mute swans (Cygnus olor). *Avian Pathol 36*, 3 (2007), 245–249.

[457] Pan, H., Cole, T. L., Bi, X., Fang, M., Zhou, C., Yang, Z., Ksepka, D. T., Hart, T., Bouzat, J. L., Argilla, L. S., Bertelsen, M. F., Boersma, P. D., Bost, C. A., Cherel, Y., Dann, P., Fiddaman, S. R., Howard, P., Labuschagne, K., Mattern, T., Miller, G., Parker, P., Phillips, R. A., Quillfeldt, P., Ryan, P. G., Taylor, H., Thompson, D. R., Young, M. J., Ellegaard, M. R., Gilbert, M. T. P., Sinding, M. S., Pacheco, G., Shepherd, L. D., Tennyson, A. J. D., Grosser, S., Kay, E., Nupen, L. J., Ellenberg, U., Houston, D. M., Reeve, A. H., Johnson, K., Masello, J. F., Stracke, T., McKinlay, B., Borboroglu, P. G., Zhang, D. X., and Zhang, G. High-coverage genomes to elucidate the evolution of penguins. *Gigascience 8*, 9 (2019).

[458] Paramo, T., Piggot, T. J., Bryant, C. E., and Bond, P. J. The Structural Basis for Endotoxin-induced Allosteric Regulation of the Toll-like Receptor 4 (TLR4) Innate Immune Receptor. *Journal of Biological Chemistry 288*, 51 (2013), 36215–36225.

[459] Park, B., Brinkmann, M. M., Spooner, E., Lee, C. C., Kim, Y. M., and Ploegh, H. L. Proteolytic cleavage in an endolysosomal compartment is required for activation of Toll-like receptor 9. *Nat Immunol 9*, 12 (2008), 1407–1414.

[460] Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., and Lee, J.-O. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature 458*, 7242 (2009), 1191–U130.

[461] Park, J. S., Gamboni-Robertson, F., He, Q., Svetkauskaite, D., Kim, J. Y., Strassheim, D., Sohn, J. W., Yamada, S., Maruyama, I., Banerjee, A., Ishizaka, A., and Abraham, E. High mobility group box 1 protein interacts with multiple Toll-like receptors. *Am J Physiol Cell Physiol 290*, 3 (2006), 917–924.

[462] Park, J. S., Svetkauskaite, D., He, Q., Kim, J. Y., Strassheim, D., Ishizaka, A., and Abraham, E. Involvement of toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J Biol Chem 279*, 9 (2004), 7370–7377.

[463] Pelka, K., Bertheloot, D., Reimer, E., Phulphagar, K., Schmidt, S. V., Christ, A., Stahl, R., Watson, N., Miyake, K., Hacohen, N., Haas, A., Brinkmann, M. M., Marshak-Rothstein, A., Meissner, F., and Latz, E. The Chaperone UNC93B1 Regulates Toll-like Receptor Stability Independently of Endosomal TLR Transport. *Immunity 48*, 5 (2018), 911–922.

[464] Pena M, F., Poulin, E., Dantas, G. P. M., Gonzalez-Acuna, D., Petry, M. V., and Vianna, J. A. Have Historical Climate Changes Affected Gentoo Penguin (Pygoscelis papua) Populations in Antarctica? *Plos One 9*, 4 (2014).

[465] Peroval, M. Y., Boyd, A. C., Young, J. R., and Smith, A. L. A critical role for MAPK signalling pathways in the transcriptional regulation of toll like receptors. *PLoS One 8*, 2 (2013), e51243.

[466] Pertierra, L. R., Segovia, I, N., Noll, D., Martinez, P. A., Pliscoff, P., Barbosa, A., Aragon, P., Rey, A. R., Pistorius, P., Trathan, P., Polanowski, A., Bonadonna, F., Le Bohec, C., Bi, K., Wang-Claypool, C. Y., Gonzalez-Acuna, D., Dantas, G. P. M., Bowie, R. C. K., Poulin, E., and Vianna, J. A. Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Diversity and Distributions 26*, 8 (2020), 958–975.

[467] Peters, J., Lebrasseur, O., Deng, H., and Larson, G. Holocene cultural history of Red jungle fowl (Gallus gallus) and its domestic descendant in East Asia. *Quaternary Science Reviews 142* (2016), 102–119.

[468] Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol 31*, 7 (2014), 1929–1936.

[469] Philbin, V. J., Iqbal, M., Boyd, Y., Goodchild, M. J., Beal, R. K., Bumstead, N., Young, J., and Smith, A. L. Identification and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens. *Immunology 114*, 4 (2005), 507–521.

[470] Philipp, E. E., Kraemer, L., Melzner, F., Poustka, A. J., Thieme, S., Findeisen, U., Schreiber, S., and Rosenstiel, P. Massively parallel RNA sequencing identifies a complex immune gene repertoire in the lophotrochozoan Mytilus edulis. *PLoS One 7*, 3 (2012), e33091.

[471] Pierangeli, S. S., Vega-Ostertag, M. E., Raschi, E., Liu, X., Romay-Penabad, Z., De Micheli, V., Galli, M., Moia, M., Tincani, A., Borghi, M. O., Nguyen-Oghalai, T., and Meroni, P. L. Toll-like receptor and antiphospholipid mediated thrombosis: in vivo studies. *Ann Rheum Dis 66*, 10 (2007), 1327–1333.

[472] Pisetsky, D. S. Mechanisms of immune stimulation by bacterial DNA. *Springer Semin Immunopathol 22*, 1-2 (2000), 21–33.

[473] Pisitkun, P., Deane, J. A., Difilippantonio, M. J., Tarasenko, T., Satterthwaite, A. B., and Bolland, S. Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. *Science 312*, 5780 (2006), 1669–1672.

[474] Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science 282*, 5396 (1998), 2085–2088.

[475] Pond, S. L., and Frost, S. D. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics 21*, 10 (2005), 2531–2533.

[476] Pond, S. L., Frost, S. D., and Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics 21*, 5 (2005), 676–679.

[477] Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., and Banks, E. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* (2018).

[478] Posada, D., and Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics 14*, 9 (1998), 817–818.

[479] Poulin, R. Parasite biodiversity revisited: frontiers and constraints. *International Journal for Parasitology 44*, 9 (2014), 581–589.

[480] Prabhu, S. S., Chakraborty, T. T., Kumar, N., and Banerjee, I. Association between IFITM3 rs12252 polymorphism and influenza susceptibility and severity: A meta-analysis. *Gene 674* (2018), 70–79.

[481] Price, S. J., Leung, W. T. M., Owen, C. J., Puschendorf, R., Sergeant, C., Cunningham, A. A., Balloux, F., Garner, T. W. J., and Nichols, R. A. Effects of historic and projected climate change on the range and impacts of an emerging wildlife disease. *Glob Chang Biol 25*, 8 (2019), 2648–2660.

[482] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., and Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet 81*, 3 (2007), 559–575.

[483] Pybus, O. G., Perrins, C. M., Choudhury, B., Manvell, R. J., Nunez, A., Schulenburg, B., Sheldon, B. C., and Brown, I. H. The ecology and age structure of a highly pathogenic avian influenza virus outbreak in wild mute swans. *Parasitology 139*, 14 (2012), 1914–1923.

[484] Qin, M., Li, C., Li, Z., Chen, W., and Zeng, Y. Genetic Diversities and Differentially Selected Regions Between Shandong Indigenous Pig Breeds and Western Pig Breeds. *Front Genet 10* (2019), 1351.

[485] Quintana-Murci, L., and Clark, A. G. Population genetic tools for dissecting innate immunity in humans. *Nat Rev Immunol 13*, 4 (2013), 280–93.

[486] Qureshi, S. T., Larivière, L., Leveque, G., Clermont, S., Moore, K. J., Gros, P., and Malo, D. Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J Exp Med 189*, 4 (1999), 615–625.

[487] Rajsbaum, R., Albrecht, R. A., Wang, M. K., Maharaj, N. P., Versteeg, G. A., Nistal-Villán, E., García-Sastre, A., and Gack, M. U. Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein. *PLoS Pathog 8*, 11 (2012), e1003059.

[488] Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology 67*, 5 (2018), 901–904.

[489] Ranjith-Kumar, C. T., Miller, W., Xiong, J., Russell, W. K., Lamb, R., Santos, J., Duffy, K. E., Cleveland, L., Park, M., Bhardwaj, K., Wu, Z., Russell, D. H., Sarisky, R. T., Mbow, M. L., and Kao, C. C. Biochemical and functional analyses of the human Toll-like receptor 3 ectodomain. *J Biol Chem 282*, 10 (2007), 7668–7678.

[490] Rankin, R., Pontarollo, R., Ioannou, X., Krieg, A. M., Hecker, R., Babiuk, L. A., and van Drunen Littel-van den Hurk, S. CpG motif identification for veterinary and laboratory species demonstrates that sequence recognition is highly conserved. *Antisense Nucleic Acid Drug Dev 11*, 5 (2001), 333–340.

[491] Raven, N., Lisovski, S., Klaassen, M., Lo, N., Madsen, T., Ho, S. Y. W., and Ujvari, B. Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. *Infect Genet Evol 53* (2017), 135–145.

[492] Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S. C., Lama, T. M., Grutzner, F., Warren, W. C., Balakrishnan, C. N., Burt, D., George, J. M., Biegler, M. T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A. F., Franchini, P., Detrich, H. W., Svardal, H., Wagner, M., Naylor, G. J. P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B. T., Pesout, T., Houck, M., Misuraca, A., Kingan, S. B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R. E., Putnam, N. H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R. E., Guan, D., London, S. E., Clayton, D. F., Mello, C. V., Friedrich, S. R., Lovell, P. V., Osipova, E., Al-Ajli, F. O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R. S., Makova, K. D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, K., Flicek, P., Walenz, B. P., Kwak, W., Clawson, H., et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature 592*, 7856 (2021), 737–746.

[493] Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., and Aderem, A. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci U S A 102*, 27 (2005), 9577–9582.

[494] Robinson, M. J., Osorio, F., Rosas, M., Freitas, R. P., Schweighoffer, E., Gross, O., Verbeek, J. S., Ruland, J., Tybulewicz, V.,

Brown, G. D., Moita, L. F., Taylor, P. R., and Reis e Sousa, C. Dectin-2 is a Syk-coupled pattern recognition receptor crucial for Th17 responses to fungal infection. *J Exp Med 206*, 9 (2009), 2037–2051.

[495] Robinson, R. A., Lawson, B., Toms, M. P., Peck, K. M., Kirkwood, J. K., Chantrey, J., Clatworthy, I. R., Evans, A. D., Hughes, L. A., Hutchinson, O. C., John, S. K., Pennycott, T. W., Perkins, M. W., Rowley, P. S., Simpson, V. R., Tyler, K. M., and Cunningham, A. A. Emerging infectious disease leads to rapid population declines of common British birds. *PLoS One 5*, 8 (2010), e12215.

[496] Roeder, A. D., Ritchie, P. A., and Lambert, D. M. New DNA markers for penguins. *Conservation Genetics 3*, 3 (2002), 341–344.

[497] Roelofs, M. F., Boelens, W. C., Joosten, L. A., Abdollahi-Roodsaz, S., Geurts, J., Wunderink, L. U., Schreurs, B. W., van den Berg, W. B., and Radstake, T. R. Identification of small heat shock protein B8 (HSP22) as a novel TLR4 ligand and potential involvement in the pathogenesis of rheumatoid arthritis. *J Immunol 176*, 11 (2006), 7021–7027.

[498] Rohaim, M. A., Santhakumar, D., Naggar, R. F. E., Iqbal, M., Hussein, H. A., and Munir, M. Chickens Expressing IFIT5 Ameliorate Clinical Outcome and Pathology of Highly Pathogenic Avian Influenza and Velogenic Newcastle Disease Viruses. *Front Immunol 9* (2018), 2025.

[499] Rohde, K., and Heap, M. Latitudinal differences in species and community richness and in community structure of metazoan endo- and ectoparasites of marine teleost fish. *International Journal for Parasitology 28*, 3 (1998), 461–474.

[500] Roman, M., Martin-Orozco, E., Goodman, J. S., Nguyen, M. D., Sato, Y., Ronaghy, A., Kornbluth, R. S., Richman, D. D., Carson, D. A., and Raz, E. Immunostimulatory DNA sequences function as T helper-1-promoting adjuvants. *Nat Med 3*, 8 (1997), 849–854.

[501] Romani, L. Immunity to fungal infections. *Nat Rev Immunol 11*, 4 (2011), 275–288.

[502] Ropert-Coudert, Y., Hindell, M. A., Phillips, R. A., Charrassin, Jean-Benoit, Trudelle, L., and Raymond, B. *Biogeographic patterns of birds and mammals*. The Scientific Committee on Antarctic Research, Cambridge, 2014, book section 8, pp. 364–387.

[503] Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B. G., Schoenemeyer, A., Yamamoto, M., Akira, S., and Fitzgerald, K. A. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J Immunol 175*, 8 (2005), 5260–5268.

[504] Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., and Keller, L. Patterns of positive selection in seven ant genomes. *Mol Biol Evol 31*, 7 (2014), 1661–1685.

[505] Roy, A., Kucukural, A., and Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc 5*, 4 (2010), 725–38.

[506] Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution 34*, 12 (2017), 3299–3302.

[507] Rushing, C. S., Royle, J. A., Ziolkowski, D. J., and Pardieck, K. L. Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. *Proc Natl Acad Sci U S A 117*, 23 (2020), 12897–12903.

[508] Rutz, M., Metzger, J., Gellert, T., Luppa, P., Lipford, G. B., Wagner, H., and Bauer, S. Toll-like receptor 9 binds single-stranded CpG-DNA in a sequence- and pH-dependent manner. *Eur J Immunol 34*, 9 (2004), 2541–2550.

[509] S, N. Cheaper DNA sequencing unlocks secrets of rare diseases, 2018.

[510] Sabbaghi, A., Miri, S. M., Keshavarz, M., Mahooti, M., Zebardast, A., and Ghaemi, A. Role of γδ T cells in controlling viral infections with a focus on influenza virus: implications for designing novel therapeutic approaches. *Virol J 17*, 1 (2020), 174.

[511] Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D., and Group, I. S. M. W. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature 409*, 6822 (2001), 928–933.

[512] Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., and Clark, A. G. Dynamic evolution of the innate immune system in Drosophila. *Nat Genet 39*, 12 (2007), 1461–1468.

[513] Saito, T., Watanabe, C., Takemae, N., Chaisingh, A., Uchida, Y., Buranathai, C., Suzuki, H., Okamatsu, M., Imada, T., Parchariyanon, S., Traiwanatam, N., and Yamaguchi, S. Pathogenicity of highly pathogenic avian influenza viruses of H5N1 subtype isolated in Thailand for different poultry species. *Vet Microbiol 133*, 1-2 (2009), 65–74.

[514] Salentin, R., Gemsa, D., Sprenger, H., and Kaufmann, A. Chemokine receptor expression and chemotactic responsiveness of human monocytes after influenza A virus infection. *J Leukoc Biol 74*, 2 (2003), 252–259.

[515] Sallaberry-Pincheira, N., Gonzalez-Acuna, D., Herrera-Tello, Y., Dantas, G. P. M., Luna-Jorquera, G., Frere, E., Valdes-Velasquez, A., Simeone, A., and Vianna, J. A. Molecular Epidemiology of Avian Malaria in Wild Breeding Colonies of Humboldt and Magellanic Penguins in South America. *Ecohealth 12*, 2 (2015), 267–277.

[516] Sallaberry-Pincheira, N., González-Acuña, D., Padilla, P., Dantas, G. P. M., Luna-Jorquera, G., Frere, E., Valdés-Velásquez, A., and Vianna, J. A. Contrasting patterns of selection between MHC I and II across populations of Humboldt and Magellanic penguins. *Ecol Evol 6*, 20 (2016), 7498–7510.

[517] Sandri, S., Rodriguez, D., Gomes, E., Monteiro, H. P., Russo, M., and Campa, A. Is serum amyloid A an endogenous TLR4 agonist? *J Leukoc Biol 83*, 5 (2008), 1174–1180.

[518] Santhakumar, D., Iqbal, M., Nair, V., and Munir, M. Chicken IFN Kappa: A Novel Cytokine with Antiviral Activities. *Sci Rep 7*, 1 (2017), 2719.

[519] Sato, M., Sano, H., Iwaki, D., Kudo, K., Konishi, M., Takahashi, H., Takahashi, T., Imaizumi, H., Asai, Y., and Kuroki, Y. Direct binding of Toll-like receptor 2 to zymosan, and zymosan-induced NF-kappa B activation and TNF-alpha secretion are down-regulated by lung collectin surfactant protein A. *J Immunol 171*, 1 (2003), 417–425.

[520] Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura, T., Fujita, T., Akira, S., and Takeuchi, O. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proc Natl Acad Sci U S A 107*, 4 (2010), 1512–1517.

[521] Satta, N., Dunoyer-Geindre, S., Reber, G., Fish, R. J., Boehlen, F., Kruithof, E. K., and de Moerloose, P. The role of TLR2 in the inflammatory activation of mouse fibroblasts by human antiphospholipid antibodies. *Blood 109*, 4 (2007), 1507–1514.

[522] Schaefer, L., Babelova, A., Kiss, E., Hausser, H. J., Baliova, M., Krzyzankova, M., Marsche, G., Young, M. F., Mihalik, D., Götte, M., Malle, E., Schaefer, R. M., and Gröne, H. J. The matrix component biglycan is proinflammatory and signals through Toll-like receptors 4 and 2 in macrophages. *J Clin Invest 115*, 8 (2005), 2223–2233.

[523] Schaeffler, A., Gross, P., Buettner, R., Bollheimer, C., Buechler, C., Neumeier, M., Kopp, A., Schoelmerich, J., and Falk, W. Fatty acid-induced induction of Toll-like receptor-4/nuclear factor-kappaB pathway in adipocytes links nutritional signalling with innate immunity. *Immunology 126*, 2 (2009), 233–245.

[524] Scheele, B. C., Pasmans, F., Skerratt, L. F., Berger, L., Martel, A., Beukema, W., Acevedo, A. A., Burrowes, P. A., Carvalho, T., Catenazzi, A., De la Riva, I., Fisher, M. C., Flechas, S. V., Foster, C. N., Frías-Álvarez, P., Garner, T. W. J., Gratwicke, B., Guayasamin, J. M., Hirschfeld, M., Kolby, J. E., Kosch, T. A., La Marca, E., Lindenmayer, D. B., Lips, K. R., Longo, A. V., Maneyro, R., McDonald, C. A., Mendelson, J., Palacios-Rodriguez, P., Parra-Olea, G., Richards-Zawacki, C. L., Rödel, M. O., Rovito, S. M., Soto-Azat, C., Toledo, L. F., Voyles, J., Weldon, C., Whitfield, S. M., Wilkinson, M., Zamudio, K. R., and Canessa, S. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science 363*, 6434 (2019), 1459–1463.

[525] Scheu, A., Powell, A., Bollongino, R., Vigne, J.-D., Tresset, A., Cakirlar, C., Benecke, N., and Burger, J. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genetics 16* (2015).

[526] Schlenke, T. A., and Begun, D. J. Natural selection drives Drosophila immune system evolution. *Genetics 164*, 4 (2003), 1471–1480.

[527] Schneider, T., Olsen, D., Dykstra, C., Huettner, S., Branch, S., Sirpenski, G., Sarro, S., Waterfall, K., and Henry, L. *Penguin (Spheniscidae) Care Manual*. Association of Zoos and Aquariums, 2014.

[528] Schröder, N. W., Morath, S., Alexander, C., Hamann, L., Hartung, T., Zähringer, U., Göbel, U. B., Weber, J. R., and Schumann, R. R. Lipoteichoic acid (LTA) of Streptococcus pneumoniae and Staphylococcus aureus activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide-binding protein (LBP), and CD14, whereas TLR-4 and MD-2 are not involved. *J Biol Chem 278*, 18 (2003), 15587–15594.

[529] Schwarz, H., Schneider, K., Ohnemus, A., Lavric, M., Kothlow, S., Bauer, S., Kaspers, B., and Staeheli, P. Chicken toll-like receptor 3 recognizes its cognate ligand when ectopically expressed in human cells. *J Interferon Cytokine Res 27*, 2 (2007), 97–101.

[530] Scior, T., Lozano-Aponte, J., Figueroa-Vazquez, V., Yunes-Rojas, J. A., Zähringer, U., and Alexander, C. Three-dimensional mapping of differential amino acids of human, murine, canine and equine TLR4/MD-2 receptor complexes conferring endotoxic activation by lipid A, antagonism by Eritoran and species-dependent activities of Lipid IVA in the mammalian LPS sensor system. *Comput Struct Biotechnol J 7* (2013), e201305003.

[531] Sell-Kubiak, E., Wimmers, K., Reyer, H., and Szwaczkowski, T. Genetic aspects of feed efficiency and reduction of environmental footprint in broilers: a review. *J Appl Genet 58*, 4 (2017), 487–498.

[532] Seppey, M., Manni, M., and Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol 1962* (2019), 227–245.

[533] Sharma, V., Walther, F., Hecker, N., Stuckas, H., and Hiller, M. Convergent Losses of TLR5 Suggest Altered Extracellular Flagellin Detection in Four Mammalian Lineages. *Molecular Biology and Evolution 37*, 7 (2020), 1847–1854.

[534] Shen, H., Tesar, B. M., Walker, W. E., and Goldstein, D. R. Dual signaling of MyD88 and TRIF is critical for maximal TLR4-induced dendritic cell maturation. *J Immunol 181*, 3 (2008), 1849–1858.

[535] Shen, T., Xu, S. X., Wang, X. H., Yu, W. H., Zhou, K. Y., and Yang, G. Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. *BMC Evolutionary Biology 12* (2012).

[536] Shi, H., Kokoeva, M. V., Inouye, K., Tzameli, I., Yin, H., and Flier, J. S. TLR4 links innate immunity and fatty acid-induced insulin resistance. *J Clin Invest 116*, 11 (2006), 3015–3025.

[537] Shi, Z., Cai, Z., Sanchez, A., Zhang, T., Wen, S., Wang, J., Yang, J., Fu, S., and Zhang, D. A novel Toll-like receptor that recognizes vesicular stomatitis virus. *J Biol Chem 286*, 6 (2011), 4517–4524.

[538] Shiah, F. K., and Ducklow, H. W. Regulation of bacterial abundance and production by substrate supply and bacterivory: A mesocosm study. *Microb Ecol 30*, 3 (1995), 239–255.

[539] Shimada, S., Yano, O., Inoue, H., Kuramoto, E., Fukuda, T., Yamamoto, H., Kataoka, T., and Tokunaga, T. Antitumor activity of the DNA fraction from Mycobacterium bovis BCG. II. Effects on various syngeneic mouse tumors. *J Natl Cancer Inst 74*, 3 (1985), 681–688.

[540] Shimazu, R., Akashi, S., Ogata, H., Nagai, Y., Fukudome, K., Miyake, K., and Kimoto, M. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J Exp Med 189*, 11 (1999), 1777–1782.

[541] Shinya, K., Ito, M., Makino, A., Tanaka, M., Miyake, K., Eisfeld, A. J., and Kawaoka, Y. The TLR4-TRIF pathway protects against H5N1 influenza virus infection. *J Virol 86*, 1 (2012), 19–24.

[542] Shultz, A. J., and Sackton, T. B. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife 8* (2019).

[543] Si-Tahar, M., Blanc, F., Furio, L., Chopy, D., Balloy, V., Lafon, M., Chignard, M., Fiette, L., Langa, F., Charneau, P., and Pothlichet, J. Protective role of LGP2 in influenza virus pathogenesis. *J Infect Dis 210*, 2 (2014), 214–223.

[544] Siegel, P. B. Evolution of the modern broiler and feed efficiency. *Annu Rev Anim Biosci 2* (2014), 375–385.

[545] Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E., Arnason, U., and Penny, D. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution 23*, 6 (2006), 1144–1155.

[546] Smiley, S. T., King, J. A., and Hancock, W. W. Fibrinogen stimulates macrophage chemokine secretion through toll-like receptor 4. *J Immunol 167*, 5 (2001), 2887–2894.

[547] Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol 32*, 5 (2015), 1342–1353.

[548] Smith, S. A., Jann, O. C., Haig, D., Russell, G. C., Werling, D., Glass, E. J., and Emes, R. D. Adaptive evolution of Toll-like receptor 5 in domesticated mammals. *Bmc Evolutionary Biology 12* (2012).

[549] Sneath, P. H. Relations between chemical structure and biological activity in peptides. *J Theor Biol 12*, 2 (1966), 157–95.

[550] Solbakken, M. H., Tørresen, O. K., Nederbragt, A. J., Seppola, M., Gregers, T. F., Jakobsen, K. S., and Jentoft, S. Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like receptor repertoire by gene losses and expansions. *Sci Rep 6* (2016), 25211.

[551] Song, W. S., Jeon, Y. J., Namgung, B., Hong, M., and Yoon, S.-i. A conserved TLR5 binding and activation hot spot on flagellin. *Scientific Reports 7* (2017).

[552] Spence, J. S., He, R., Hoffmann, H. H., Das, T., Thinon, E., Rice, C. M., Peng, T., Chandran, K., and Hang, H. C. IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nat Chem Biol 15*, 3 (2019), 259–268.

[553] Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T. H., Plachetzki, D. C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S. F., Goodstein, D. M., Harris, C., Jackson, D. J., Leys, S. P., Shu, S., Woodcroft, B. J., Vervoort, M., Kosik, K. S., Manning, G., Degnan, B. M., and Rokhsar, D. S. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature 466*, 7307 (2010), 720–726.

[554] Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics 30*, 9 (2014), 1312–1313.

[555] Star, B., and Jentoft, S. Why does the immune system of Atlantic cod lack MHC II? *Bioessays 34*, 8 (2012), 648–651.

[556] Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzén, A., Winer, R., Knight, J., Vogel, J. H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjøen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C., and Jakobsen, K. S. The genome sequence of Atlantic cod reveals a unique immune system. *Nature 477*, 7363 (2011), 207–210.

[557] Steig, E. J., Schneider, D. P., Rutherford, S. D., Mann, M. E., Comiso, J. C., and Shindell, D. T. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature 457*, 7228 (2009), 459–62.

[558] Stephens, M., Smith, N. J., and Donnelly, P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics 68*, 4 (2001), 978–989.

[559] Stewart, C. R., Bagnaud-Baule, A., Karpala, A. J., Lowther, S., Mohr, P. G., Wise, T. G., Lowenthal, J. W., and Bean, A. G. Toll-Like Receptor 7 Ligands Inhibit Influenza A Infection in Chickens. *Journal of Interferon and Cytokine Research 32*, 1 (2012), 46–51.

[560] Stewart, C. R., Stuart, L. M., Wilkinson, K., van Gils, J. M., Deng, J., Halle, A., Rayner, K. J., Boyer, L., Zhong, R., Frazier, W. A., Lacy-Hulbert, A., El Khoury, J., Golenbock, D. T., and Moore, K. J. CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat Immunol 11*, 2 (2010), 155–161.

[561] Stonehouse, B. Geographic Variation in Gentoo Penguins Pygoscelis papua. *IBIS 112*, 1 (1970), 52–57.

[562] Subramanian, S., Beans-Picon, G., Swaminathan, S. K., Millar, C. D., and Lambert, D. M. Evidence for a recent origin of penguins. *Biology Letters 9*, 6 (2013).

[563] Subramanian, S., Tus, K., Li, Q. Z., Wang, A., Tian, X. H., Zhou, J., Liang, C., Bartov, G., McDaniel, L. D., Zhou, X. J., Schultz, R. A., and Wakeland, E. K. A Tlr7 translocation accelerates systemic autoimmunity in murine lupus. *Proc Natl Acad Sci U S A 103*, 26 (2006), 9970–9975.

[564] Sun, Y., Zhu, Z., Wang, R., and Xu, T. Miiuy croaker transferrin gene and evidence for positive selection events reveal different evolutionary patterns. *PLoS One 7*, 9 (2012), e43936.

[565] Swiderska, Z., Smidova, A., Buchtova, L., Bryjova, A., Fabianova, A., Munclinger, P., and Vinkler, M. Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds. *Scientific Reports 8* (2018).

[566] Tabeta, K., Hoebe, K., Janssen, E. M., Du, X., Georgel, P., Crozat, K., Mudd, S., Mann, N., Sovath, S., Goode, J., Shamel, L., Herskovits, A. A., Portnoy, D. A., Cooke, M., Tarantino, L. M., Wiltshire, T., Steinberg, B. E., Grinstein, S., and Beutler, B. The Unc93b1 mutation 3d disrupts exogenous antigen presentation and signaling via Toll-like receptors 3, 7 and 9. *Nat Immunol 7*, 2 (2006), 156–164.

[567] Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics 123*, 3 (1989), 585–595.

[568] Takeda, K., Kaisho, T., and Akira, S. Toll-like receptors. *Annual Review of Immunology 21* (2003), 335–376.

[569] Takeuchi, O., and Akira, S. Pattern recognition receptors and inflammation. *Cell 140*, 6 (2010), 805–820.

[570] Takeuchi, O., Kawai, T., Mühlradt, P. F., Morr, M., Radolf, J. D., Zychlinsky, A., Takeda, K., and Akira, S. Discrimination of bacterial lipoproteins by Toll-like receptor 6. *Int Immunol 13*, 7 (2001), 933–940.

[571] Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., Modlin, R. L., and Akira, S. Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *J Immunol 169*, 1 (2002), 10–14.

[572] Tamura, K., and Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution 10*, 3 (1993), 512–526.

[573] Tanji, H., Ohto, U., Shibata, T., Miyake, K., and Shimizu, T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science 339*, 6126 (2013), 1426–1429.

[574] Tassia, M. G., Whelan, N. V., and Halanych, K. M. Toll-like receptor pathway evolution in deuterostomes. *Proc Natl Acad Sci U S A 114*, 27 (2017), 7055–7060.

[575] Tate, M. D., and Mansell, A. An update on the NLRP3 inflammasome and influenza: the road to redemption or perdition? *Curr Opin Immunol 54* (2018), 80–85.

[576] Tattoli, I., Travassos, L. H., Carneiro, L. A., Magalhaes, J. G., and Girardin, S. E. The Nodosome: Nod1 and Nod2 control bacterial infections and inflammation. *Semin Immunopathol 29*, 3 (2007), 289–301.

[577] Teifke, J. P., Klopfleisch, R., Globig, A., Starick, E., Hoffmann, B., Wolf, P. U., Beer, M., Mettenleiter, T. C., and Harder, T. C. Pathology of natural infections by H5N1 highly pathogenic avian influenza virus in mute (Cygnus olor) and whooper (Cygnus cygnus) swans. *Vet Pathol 44*, 2 (2007), 137–143.

[578] Tekaia, F., Yeramian, E., and Dujon, B. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. *Gene 297*, 1-2 (2002), 51–60.

[579] Temperley, N. D., Berlin, S., Paton, I. R., Griffin, D. K., and Burt, D. W. Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss. *Bmc Genomics 9* (2008).

[580] Termeer, C., Benedix, F., Sleeman, J., Fieber, C., Voith, U., Ahrens, T., Miyake, K., Freudenberg, M., Galanos, C., and Simon, J. C. Oligosaccharides of Hyaluronan activate dendritic cells via toll-like receptor 4. *J Exp Med 195*, 1 (2002), 99–111.

[581] Terregino, C., Milani, A., Capua, I., Marino, A. M., and Cavaliere, N. Highly pathogenic avian influenza H5N1 subtype in mute swans in Italy. *Vet Rec 158*, 14 (2006), 491.

[582] Těšický, M., Velová, H., Novotný, M., Kreisinger, J., Beneš, V., and Vinkler, M. Positive selection and convergent evolution shape molecular phenotypic traits of innate immunity receptors in tits (Paridae). *Mol Ecol 29*, 16 (2020), 3056–3070.

[583] Thomas, C. D. Climate, climate change and range boundaries. *Diversity and Distributions 16*, 3 (2010), 488–495.

[584] Thomsen, M. M., Jørgensen, S. E., Gad, H. H., Storgaard, M., Gjedsted, J., Christiansen, M., Hartmann, R., and Mogensen, T. H. Defective interferon priming and impaired antiviral responses in a patient with an IRF7 variant and severe influenza. *Med Microbiol Immunol 208*, 6 (2019), 869–876.

[585] Tokunaga, T., Yamamoto, H., Shimada, S., Abe, H., Fukuda, T., Fujisawa, Y., Furutani, Y., Yano, O., Kataoka, T., and Sudo, T. Antitumor activity of deoxyribonucleic acid fraction from Mycobacterium bovis BCG. I. Isolation, physicochemical characterization, and antitumor activity. *J Natl Cancer Inst 72*, 4 (1984), 955–962.

[586] Trathan, P., Daunt, F., and Murphy, E. *South Georgia: An Ecological Atlas*. British Antarctic Survey, Cambridge, UK, 1996.

[587] Trathan, P. N., Forcada, J., and Murphy, E. J. Environmental forcing and Southern Ocean marine predator populations: effects of climate change and variability. *Philosophical Transactions of the Royal Society B-Biological Sciences 362*, 1488 (2007), 2351–2365.

[588] Trivelpiece, W. Z., Trivelpiece, S. G., and Volkman, N. J. Ecological segregation of Adelie, Gentoo, and Chinstrap penguins at King George Island, Antarctica. *Ecology 68*, 2 (1987), 351–361.

[589] Tsuda, T. T., Tsuda, M., Naruse, T., Kawata, H., Ando, A., Shiina, T., Fukuda, M., Kurita, M., LeMaho, I., Kulski, J. K., and Inoko, H. Phylogenetic analysis of penguin (Spheniscidae) species based on sequence variation in MHC class II genes. *Immunogenetics 53*, 8 (2001), 712–716.

[590] Tyler, J., Bonfitto, M. T., Clucas, G. V., Reddy, S., and Younger, J. L. Morphometric and genetic evidence for four species of gentoo penguin. *Ecol Evol 10*, 24 (2020), 13836–13846.

[591] Uhart, M., Vanstreels, R. E. T., Gallo, L., Cook, R. A., and Karesh, W. B. Serological survey for select infectious agents in wild Magellanic penguins (Spheniscus magellanicus) in Argentina, 1994-2008. *Journal of Wildlife Diseases 56*, 1 (2020), 66–81.

[592] Underhill, D. M., Ozinsky, A., Hajjar, A. M., Stevens, A., Wilson, C. B., Bassetti, M., and Aderem, A. The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature 401*, 6755 (1999), 811–815.

[593] Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. Primer3-new capabilities and interfaces. *Nucleic Acids Research 40*, 15 (2012).

[594] Urbina, D., Tang, B., and Higgs, P. G. The response of amino acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. *J Mol Evol 62*, 3 (2006), 340–61.

[595] Uversky, V. N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins 1*, 1 (2013), e24684.

[596] Vabulas, R. M., Ahmad-Nejad, P., Ghose, S., Kirschning, C. J., Issels, R. D., and Wagner, H. HSP70 as endogenous stimulus of the Toll/interleukin-1 receptor signal pathway. *J Biol Chem 277*, 17 (2002), 15107–15112.

[597] Vabulas, R. M., Braedel, S., Hilf, N., Singh-Jasuja, H., Herter, S., Ahmad-Nejad, P., Kirschning, C. J., Da Costa, C., Rammensee, H. G., Wagner, H., and Schild, H. The endoplasmic reticulum-resident heat shock protein Gp96 activates dendritic cells via the Toll-like receptor 2/4 pathway. *J Biol Chem 277*, 23 (2002), 20847–20853.

[598] van de Veerdonk, F. L., Marijnissen, R. J., Kullberg, B. J., Koenen, H. J., Cheng, S. C., Joosten, I., van den Berg, W. B., Williams, D. L., van der Meer, J. W., Joosten, L. A., and Netea, M. G. The

macrophage mannose receptor induces IL-17 in response to Candida albicans. *Cell Host Microbe 5*, 4 (2009), 329–340.

[599] van den Brand, J. M. A., Verhagen, J. H., Kroeze, E. J. B. V., van de Bildt, M. W. G., Bodewes, R., Herfst, S., Richard, M., Lexmond, P., Bestebroer, T. M., Fouchier, R. A. M., and Kuiken, T. Wild ducks excrete highly pathogenic avian influenza virus H5N8 (2014–2015) without clinical or pathological evidence of disease. *Emerging Microbes & Infections 7*, 1 (2018), 1–10. PMID: 29670093.

[600] van den Brand, J. M. A., Verhagen, J. H., Veldhuis Kroeze, E. J. B., van de Bildt, M. W. G., Bodewes, R., Herfst, S., Richard, M., Lexmond, P., Bestebroer, T. M., Fouchier, R. A. M., and Kuiken, T. Wild ducks excrete highly pathogenic avian influenza virus H5N8 (2014-2015) without clinical or pathological evidence of disease. *Emerg Microbes Infect 7*, 1 (2018), 67.

[601] Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics 43* (2013), 11.10.1–11.10.33.

[602] van Nimwegen, K. J. M., van Soest, R. A., Veltman, J. A., Nelen, M. R., van der Wilt, G. J., Vissers, L. E. L. M., and Grutters, J. P. C. Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clinical Chemistry 62*, 11 (2016), 1458–1464.

[603] Velova, H., Gutowska-Ding, M. W., Burt, D. W., and Vinkler, M. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. *Molecular Biology and Evolution 35*, 9 (2018), 2170–2184.

[604] Vianna, J. A., Fernandes, F. A. N., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., Bi, K., Wang-Claypool, C. Y., Lowther, A., Parker, P., Le Bohec, C., Bonadonna, F., Wienecke, B., Pistorius, P., Steinfurth, A., Burridge, C. P., Dantas, G. P. M., Poulin, E., Simison, W. B., Henderson, J., Eizirik, E., Nery, M. F., and Bowie, R. C. K. Genome-wide analyses reveal drivers of penguin diversification. *Proc Natl Acad Sci U S A 117*, 36 (2020), 22303–22310.

[605] Vianna, J. A., Noll, D., Dantas, G. P. M., Petry, M. V., Barbosa, A., Gonzalez-Acuna, D., Le Bohec, C., Bonadonna, F., and Poulin, E. Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Molecular Phylogenetics and Evolution 107* (2017), 486–498.

[606] Vinkler, M., Bainova, H., and Bryja, J. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. *Genetics Selection Evolution 46* (2014).

[607] Vinkler, M., Bainova, H., Bryjova, A., Tomasek, O., Albrecht, T., and Bryja, J. Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. *Genetica 143*, 1 (2015), 101–112.

[608] Vinkler, M., Bryjova, A., Albrecht, T., and Bryja, J. Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (Taeniopygia guttata). *Tissue Antigens 74*, 1 (2009), 32–41.

[609] Vogl, T., Tenbrock, K., Ludwig, S., Leukert, N., Ehrhardt, C., van Zoelen, M. A., Nacken, W., Foell, D., van der Poll, T., Sorg, C., and Roth, J. Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nat Med 13*, 9 (2007), 1042–1049.

[610] Vollmer, J., Tluk, S., Schmitz, C., Hamm, S., Jurk, M., Forsbach, A., Akira, S., Kelly, K. M., Reeves, W. H., Bauer, S., and Krieg, A. M. Immune stimulation mediated by autoantigen binding sites within small nuclear RNAs involves Toll-like receptors 7 and 8. *J Exp Med 202*, 11 (2005), 1575–1585.

[611] Voogdt, C. G., Bouwman, L. I., Kik, M. J., Wagenaar, J. A., and van Putten, J. P. Reptile Toll-like receptor 5 unveils adaptive evolution of bacterial flagellin recognition. *Sci Rep 6* (2016), 19046.

[612] Voogdt, C. G. P., Merchant, M. E., Wagenaar, J. A., and van Putten, J. P. M. Evolutionary Regression and Species-Specific Codon Usage of TLR15. *Frontiers in Immunology 9* (2018).

[613] Vreman, S., McCaffrey, J., Popma-de Graaf, D. J., Nauwynck, H., Savelkoul, H. F. J., Moore, A., Rebel, J. M. J., and Stockhofe-Zurwieden, N. Toll-like receptor agonists as adjuvants for inactivated porcine reproductive and respiratory syndrome virus (PRRSV) vaccine. *Vet Immunol Immunopathol 212* (2019), 27–37.

[614] Wallensten, A., Munster, V. J., Osterhaus, A., Waldenstrom, J., Bonnedahl, J., Broman, T., Fouchier, R. A. M., and Olsen, B. Mounting evidence for the presence of influenza A virus in the avifauna of the Antarctic region. *Antarctic Science 18*, 3 (2006), 353–356.

[615] Walsh, S. A., and Suarez, M. E. New penguin remains from the Pliocene of northern Chile. *Historical Biology 18* (2006), 119–130.

[616] Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. A novel coronavirus outbreak of global health concern. *The Lancet 395*, 10223 (2020), 470–473.

[617] Wang, J., Zhang, Z., Chang, F., and Yin, D. Bioinformatics analysis of the structural and evolutionary characteristics for toll-like receptor 15. *PeerJ 4* (2016), e2079.

[618] Wang, J., Zhang, Z., Liu, J., Zhao, J., and Yin, D. Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. *Sci Rep 6* (2016), 26705.

[619] Wang, J., Zhang, Z., Liu, J., Zhao, J., and Yin, D. Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. *Scientific Reports 6* (2016).

[620] Wang, M. S., Thakur, M., Peng, M. S., Jiang, Y., Frantz, L. A. F., Li, M., Zhang, J. J., Wang, S., Peters, J., Otecko, N. O., Suwannapoom, C., Guo, X., Zheng, Z. Q., Esmailizadeh, A., Hirimuthugoda, N. Y., Ashari, H., Suladari, S., Zein, M. S. A., Kusza, S., Sohrabi, S., Kharrati-Koopaee, H., Shen, Q. K., Zeng, L., Yang, M. M., Wu, Y. J., Yang, X. Y., Lu, X. M., Jia, X. Z., Nie, Q. H., Lamont, S. J., Lasagna, E., Ceccobelli, S., Gunwardana, H. G. T., Senasige, T. M., Feng, S. H., Si, J. F., Zhang, H., Jin, J. Q., Li, M. L., Liu, Y. H., Chen, H. M., Ma, C., Dai, S. S., Bhuiyan, A. K. F., Khan, M. S., Silva, G. L. L., Le, T. T., Mwai, O. A., Ibrahim, M. N. M., Supple, M., Shapiro, B., Hanotte, O., Zhang, G., Larson, G., Han, J. L., Wu, D. D., and Zhang, Y. P. 863 genomes reveal the origin and domestication of chicken. *Cell Res 30*, 8 (2020), 693–701.

[621] Wang, R., Zhu, Y., Lin, X., Ren, C., Zhao, J., Wang, F., Gao, X., Xiao, R., Zhao, L., Chen, H., Jin, M., Ma, W., and Zhou, H. Influenza M2 protein regulates MAVS-mediated signaling pathway through interacting with MAVS and increasing ROS production. *Autophagy 15*, 7 (2019), 1163–1181.

[622] Wang, X., Hinson, E. R., and Cresswell, P. The interferon-inducible protein viperin inhibits influenza virus release by perturbing lipid rafts. *Cell Host Microbe 2*, 2 (2007), 96–105.

[623] Wang, Y., Brahmakshatriya, V., Lupiani, B., Reddy, S., Okimoto, R., Li, X., Chiang, H., and Zhou, H. Associations of chicken Mx1 polymorphism with antiviral responses in avian influenza virus infected embryos and broilers. *Poult Sci 91*, 12 (2012), 3019–3024.

[624] Waterhouse, R. M., Kriventseva, E. V., Meister, S., Xi, Z., Alvarez, K. S., Bartholomay, L. C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B. M., Dong, Y., Jiang, H., Kanost, M. R., Koutsos, A. C., Levashina, E. A., Li, J., Ligoxygakis, P., Maccallum, R. M., Mayhew, G. F., Mendes, A., Michel, K., Osta, M. A., Paskewitz, S., Shin, S. W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D. W., Raikhel, A. S., Kafatos, F. C., Dimopoulos, G., Zdobnov, E. M., and Christophides, G. K. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science 316*, 5832 (2007), 1738–1743.

[625] Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Pond, S. L. K. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution 35*, 3 (2018), 773–777.

[626] Weber, T. P., and Stilianakis, N. I. Ecologic immunology of avian influenza (H5N1) in migratory birds. *Emerg Infect Dis 13*, 8 (2007), 1139–1143.

[627] Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol Rev 56*, 1 (1992), 152–179.

[628] Wei, T., Gong, J., Jamitzky, F., Heckl, W. M., Stark, R. W., and Rössle, S. C. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci 18*, 8 (2009), 1684–1691.

[629] Weimerskirch, H., Zotier, R., and Jouventin, P. The avifauna of the Kerguelen Islands. *EMU 89*, 1 (1989), 15–29.

[630] Weir, B. S., and Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution 38*, 6 (1984), 1358–1370.

[631] Wellington, D., Laurenson-Schafer, H., Abdel-Haq, A., and Dong, T. IFITM3: How genetics influence influenza infection demographically. *Biomed J 42*, 1 (2019), 19–26.

[632] Wen, F., Guo, J., Li, Z., and Huang, S. Sex-specific patterns of gene expression following influenza vaccination. *Sci Rep 8*, 1 (2018), 13517.

[633] Wheeler, D. S., Chase, M. A., Senft, A. P., Poynter, S. E., Wong, H. R., and Page, K. Extracellular Hsp72, an endogenous DAMP, is released by virally infected airway epithelial cells and activates neutrophils via Toll-like receptor (TLR)-4. *Respir Res 10* (2009), 31.

[634] Wille, M., Aban, M., Wang, J., Moore, N., Shan, S., Marshall, J., González-Acuña, D., Vijaykrishna, D., Butler, J., Hall, R. J., Williams, D. T., and Hurt, A. C. Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *J Virol 93*, 11 (2019).

[635] Wilson, R. P., Alvarrez, B., Latorre, L., Adelung, D., Culik, B., and Bannasch, R. The movements of gentoo penguins Pygoscelis papua from Ardley Island, Antarctica. *Polar Biology 19*, 6 (1998), 407–413.

[636] Wlasiuk, G., Khan, S., Switzer, W. M., and Nachman, M. W. A history of recurrent positive selection at the toll-like receptor 5 in primates. *Mol Biol Evol 26*, 4 (2009), 937–949.

[637] Wlasiuk, G., and Nachman, M. W. Adaptation and Constraint at Toll-Like Receptors in Primates. *Molecular Biology and Evolution 27*, 9 (2010), 2172–2186.

[638] Wolfe, A. D., and dePamphilis, C. W. The effect of relaxed functional constraints on the photosynthetic gene rbcL in photosynthetic and nonphotosynthetic parasitic plants. *Mol Biol Evol 15*, 10 (1998), 1243–1258.

[639] Wong, G. K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Li, R., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Dong, L., Lan, F., Sun, Y., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L. W., Yang, S. P., Warren, W. C., Wilson, R. K., Brandström, M., Ellegren, H., Crooijmans, R. P., van der Poel, J. J., Bovenhuis, H., Groenen, M. A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J. R., Rogers, S., Walker, B. A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S. J., Zhou, H., Hocking, P. M., Morrice, D., de Koning, D. J., Law, A., Bartley, N., Burt, D. W., Hunt, H., Cheng, H. H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M. T., Andersson, B., Webber, C., Ponting, C. P., Overton, I. M., et al. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature 432*, 7018 (2004), 717–722.

[640] Woods, R., Jones, H. I., Watts, J., Miller, G. D., and Shellam, G. R. Diseases of Antarctic seabirds. *Health of Antarctic wildlife: a challenge for science and policy.* (2009), 35–55.

[641] Wu, H., Wang, H., Jiang, W., and Lian, Z. The evolutionary characteristics and structural biology of Gallus toll-like receptor 21. *J Mol Recognit 31*, 6 (2018), e2696.

[642] Xavier, M. O., Soares, M. P., Meinerz, A. R. M., Nobre, M. O., Osorio, L. G., da Silva, R. P., and Meireles, M. C. A. Aspergillosis: A limiting factor during recovery of captive Magellanic penguins. *Brazilian Journal of Microbiology 38*, 3 (2007), 480–484.

[643] Xie, H., Raybourne, R. B., Babu, U. S., Lillehoj, H. S., and Heckert, R. A. CpG-induced immunomodulation and intracellular bacterial killing in a chicken macrophage cell line. *Dev Comp Immunol 27*, 9 (2003), 823–834.

[644] Xu, X. H., Shah, P. K., Faure, E., Equils, O., Thomas, L., Fishbein, M. C., Luthringer, D., Xu, X. P., Rajavashisth, T. B., Yano, J., Kaul, S., and Arditi, M. Toll-like receptor-4 is expressed by macrophages in murine and human lipid-rich atherosclerotic plaques and upregulated by oxidized LDL. *Circulation 104*, 25 (2001), 3103–3108.

[645] Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., and Akira,

S. Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science 301*, 5633 (2003), 640–643.

[646] Yamamoto, M., Sato, S., Hemmi, H., Sanjo, H., Uematsu, S., Kaisho, T., Hoshino, K., Takeuchi, O., Kobayashi, M., Fujita, T., Takeda, K., and Akira, S. Essential role for TIRAP in activation of the signalling cascade shared by TLR2 and TLR4. *Nature 420*, 6913 (2002), 324–329.

[647] Yampolsky, L. Y., and Stoltzfus, A. The exchangeability of amino acids in proteins. *Genetics 170*, 4 (2005), 1459–1472.

[648] Yang, D., Chen, Q., Su, S. B., Zhang, P., Kurosaka, K., Caspi, R. R., Michalek, S. M., Rosenberg, H. F., Zhang, N., and Oppenheim, J. J. Eosinophil-derived neurotoxin acts as an alarmin to activate the TLR2-MyD88 signal pathway in dendritic cells and enhances Th2 immune responses. *J Exp Med 205*, 1 (2008), 79–90.

[649] Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods 12*, 1 (2015), 7–8.

[650] Yang, L., and Gui, J. F. Positive selection on multiple antique allelic lineages of transferrin in the polyploid Carassius auratus. *Mol Biol Evol 21*, 7 (2004), 1264–77.

[651] Yang, S., Sugawara, S., Monodane, T., Nishijima, M., Adachi, Y., Akashi, S., Miyake, K., Hase, S., and Takada, H. Micrococcus luteus teichuronic acids activate human and murine monocytic cells in a CD14- and toll-like receptor 4-dependent manner. *Infect Immun 69*, 4 (2001), 2025–2030.

[652] Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution 24*, 8 (2007), 1586–1591.

[653] Yang, Z., Nielsen, R., and Goldman, N. In defense of statistical methods for detecting positive selection. *Proc Natl Acad Sci U S A 106*, 36 (2009), E95; author reply E96.

[654] Yang, Z. H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences 13*, 5 (1997), 555–556.

[655] Yang, Z. H., Wong, W. S. W., and Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution 22*, 4 (2005), 1107–1118.

[656] Yarovinsky, F., Zhang, D., Andersen, J. F., Bannenberg, G. L., Serhan, C. N., Hayden, M. S., Hieny, S., Sutterwala, F. S., Flavell, R. A., Ghosh, S., and Sher, A. TLR11 activation of dendritic cells by a protozoan profilin-like protein. *Science 308*, 5728 (2005), 1626–1629.

[657] Yeh, D. W., Liu, Y. L., Lo, Y. C., Yuh, C. H., Yu, G. Y., Lo, J. F., Luo, Y. P., Xiang, R., and Chuang, T. H. Toll-like receptor 9 and 21 have different ligand recognition profiles and cooperatively mediate activity of CpG-oligodeoxynucleotides in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America 110*, 51 (2013), 20711–20716.

[658] Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y. M., Gale, M., Akira, S., Yonehara, S., Kato, A., and Fujita, T. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol 175*, 5 (2005), 2851–2858.

[659] Yoon, S. I., Kurnasov, O., Natarajan, V., Hong, M. S., Gudkov, A. V., Osterman, A. L., and Wilson, I. A. Structural Basis of TLR5-Flagellin Recognition and Signaling. *Science 335*, 6070 (2012), 859–864.

[660] Younger, J., Emmerson, L., Southwell, C., Lelliott, P., and Miller, K. Proliferation of East Antarctic Adélie penguins in response to historical deglaciation. *BMC Evol Biol 15* (2015), 236.

[661] Younger, J. L., Clucas, G. V., Kooyman, G., Wienecke, B., Rogers, A. D., Trathan, P. N., Hart, T., and Miller, K. J. Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Global Change Biology 21*, 6 (2015), 2215–2226.

[662] Yuan, S., Huang, S., Zhang, W., Wu, T., Dong, M., Yu, Y., Liu, T., Wu, K., Liu, H., Yang, M., Zhang, H., and Xu, A. An amphioxus TLR with dynamic embryonic expression pattern responses to pathogens and activates NF-kappaB pathway via MyD88. *Mol Immunol 46*, 11-12 (2009), 2348–2356.

[663] Zamora-Vilchis, I., Williams, S. E., and Johnson, C. N. Environmental temperature affects prevalence of blood parasites of birds on an elevation gradient: implications for disease in a warming climate. *PLoS One 7*, 6 (2012), e39208.

[664] Zarei Ghobadi, M., Mozhgani, S. H., Farzanehpour, M., and Behzadian, F. Identifying novel biomarkers of the pediatric influenza infection by weighted co-expression network analysis. *Virol J 16*, 1 (2019), 124.

[665] Zdanowski, M. K., Weglenski, P., Golik, P., Sasin, J. M., Borsuk, P., Zmuda, M. J., and Stankovic, A. Bacterial diversity in Adelie penguin, Pygoscelis adeliae, guano: molecular and morpho-physiological approaches. *Fems Microbiology Ecology 50*, 3 (2004), 163–173.

[666] Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xiong, Z., Liu, S.,

Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xiong, Y., Wang, G., Wang, B., Fan, Y., da Fonseca, R. R., Alfaro-Núñez, A., Schubert, M., Orlando, L., Mourier, T., Howard, J. T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M. V., Hara, E., Smith, J., Farré, M., Narayan, J., Slavov, G., Romanov, M. N., Borges, R., Machado, J. P., Khan, I., Springer, M. S., Gatesy, J., Hoffmann, F. G., Opazo, J. C., Håstad, O., Sawyer, R. H., Kim, H., Kim, K. W., Kim, H. J., Cho, S., Li, N., Huang, Y., Bruford, M. W., Zhan, X., Dixon, A., Bertelsen, M. F., Derryberry, E., Warren, W., Wilson, R. K., Li, S., Ray, D. A., Green, R. E., O'Brien, S. J., Griffin, D., Johnson, W. E., Haussler, D., Ryder, O. A., Willerslev, E., Graves, G. R., Alström, P., Fjeldså, J., Mindell, D. P., Edwards, S. V., Braun, E. L., Rahbek, C., Burt, D. W., Houde, P., Zhang, Y., Yang, H., et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science 346*, 6215 (2014), 1311–1320.

[667] Zhang, G., Rahbek, C., Graves, G. R., Lei, F., Jarvis, E. D., and Gilbert, M. T. Genomics: Bird sequencing project takes off. *Nature 522*, 7554 (2015), 34.

[668] Zhang, J., Randall, G., Higginbottom, A., Monk, P., Rice, C. M., and McKeating, J. A. CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J Virol 78*, 3 (2004), 1448–55.

[669] Zhang, J. Z. Evolution by gene duplication: an update. *Trends in Ecology and Evolution 18*, 6 (2003), 292–298.

[670] Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics 9* (2008), 40.

[671] Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J., and Gerstein, M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol 11*, 3 (2010), R26.

[672] Zhang, Z. K., Ohto, U., Shibata, T., Krayukhina, E., Taoka, M., Yamauchi, Y., Tanji, H., Isobe, T., Uchiyama, S., Miyake, K., and Shimizu, T. Structural Analysis Reveals that Toll-like Receptor 7 Is a Dual Receptor for Guanosine and Single-Stranded RNA. *Immunity 45*, 4 (2016), 737–748.

[673] Zhong, Y., Kinio, A., and Saleh, M. Functions of NOD-Like Receptors in Human Diseases. *Front Immunol 4* (2013), 333.

[674] Zhu, J., Ghosh, A., and Sarkar, S. N. OASL-a new player in controlling antiviral innate immunity. *Curr Opin Virol 12* (2015), 15–19.