Co-evolution of bats and viruses and how it has shaped host biology

Emilia Cecylia Skirmuntt

A thesis presented for the degree of

Doctor of Philosophy



St. Hilda's College

University of Oxford

Trinity 2021

Abstract

Co-evolution of bats and viruses and how it has shaped host biology

Emilia C Skirmuntt

Trinity 2021

Although the general link between bats and viral diseases is clear, tracing outbreaks to their bat origins can be difficult. It is believed that the lack of symptoms in bats may be caused by their long co-evolution with viruses, their immunological traits, or evolutionary adjustments arising from endogenized viral genes.

Our work confirmed that, despite a small genome size, bats have a comparable diversity of endogenous retroviruses (ERVs) to other mammals. Some of these ERVs are sufficiently preserved to be expressed, with potential effects for host biology. We detected an unusual diversity of complex retroviruses, which are generally rare in endogenous form, and a novel class of recombinant retrovirus with an alpharetrovirus-like *envelope* which has not previously been seen in mammalian genomes.

In parallel we studied Toll-like receptors (TLRs) in bats, a class of innate immunity receptors. Although highly conserved in vertebrates, the viral and bacterial ligand recognition properties of TLRs can have subtle variations between species, which may have driven the evolutionary pathways of interactions with infectious diseases. We found that bat species differ between each other with respect to the array of TLR genes and discuss these in the context of a broader mammalian profile.

In addition, we studied the mammalian ACE2 receptor, which is an entry receptor for SARS viruses. With coronaviruses which can change tropism quickly, the host range can at least partially be predicted with knowledge of the receptor specificity. We analysed

the ACE2 SARS binding region of several vertebrates and discovered that it can vary even between related species e.g. bats.

Collectively these results strengthen the unique position the *Chiroptera* order has in regards to its relationship with viruses, both exogenous and endogenous and confirm the special role of bats among other mammals as virus reservoirs.

Acknowledgments

First, I would like to thank my supervisors, Aris Katzourakis and Adrian Smith, for their advice and support over the last five years. I am deeply thankful to Aris for believing in my project since the beginning, while Adrian was kind enough to accept me into his group for the second part of my DPhil. Both have guided me throughout the course of my degree, have encouraged to engage with new projects, and were supportive during very difficult periods.

I am also very grateful to the Zoology Department for their generosity and financial support for the lab-centric aspect of this project, as well as St. Hilda's College for their financial help when I needed it.

I want to say thank you to all of my friends and colleagues from the Zoology Department and the Peter Medawar Building for Pathogen Research. It has been a pleasure to work on my DPhil here, and I have been very lucky to be a part of such dynamic and friendly groups.

Particular thanks to the Zoology Graduate Office, and especially Heather Green who always was there were needed, relentlessly answering all my questions and emails. Without her all of the non-academic parts of pursuing a DPhil would be much more difficult.

Many thanks to the members of Aris' and Adrian's groups, both current and past students and post-docs. Thanks for amazing discussions, attending conferences together, and helpful comments about my work and projects. I need to say particular thanks to Stephen Preston and Steven Fiddaman for their help with the lab work, and Jose Nino Barreat for his assistance with coding.

I was very lucky to be a part of amazing projects with some brilliant academics. I want to say thank you to Emma Teeling for inviting me to join the Bat1K Consortium and take part in their project, as well as being very welcoming during my visits to University College Dublin. I also want say thank you to Marina Escalera-Zamudio for her critical support and sharing her expertise on viruses and their evolution. Many thanks to Stathis Giotis for providing me with cell lines, and Mads Bertelsen from the Copenhagen Zoo, and Stephen Rossiter for all the samples.

I also would like to thank my former MSc supervisor and director Atholl Johnston who always believed in me and supported me in my pursuit of applying for this DPhil.

Thanks to all my friends who supported me through the course of my DPhil. In particular I wanted to say thank you to Stephanie Holt, Ashmita Randhawa, and Jonathan Green for their helpful comments on my writing, and Urszula Michalska on her help with some of the graphical side aspects of my project.

I would like to thank Rob Williams who was always there when I needed him and who has helped and supported me all the time. I also wanted to thank him for his support and expertise in coding and all of the evenings we spent looking at these parts of my work. Without him the last years of my DPhil would have been much harder and far more lonely.

Finally, I would like to say thank you to my Grandmother Stanisława Irena Skirmuntt. She never had the opportunity to see me as a DPhil student but her ceaseless love and support were one of the major drivers which convinced me to apply for a DPhil degree.

Statement of Contribution and Associated Publications

The work presented in this thesis includes research that has been published in peer-reviewed journals, as well as papers that are currently in preparation. Below I present a summary of my contribution each of the works referenced, ordered by chapter.

Chapter 1

***Skirmuntt, Emilia Cecylia**, Marina Escalera-Zamudio, Emma C. Teeling, Adrian Smith, and Aris Katzourakis. "The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses." Annual Review of Virology 7 (2020): 103-119.*

I wrote the manuscript and designed all of the figures apart from the Figure 1.1 which was provided by ECT. MEZ and ECT provided editorial assistance. AS and AK supervised the writing and provided editorial assistance.

Chapter 2

***Skirmuntt, Emilia Cecylia**, and Aris Katzourakis. "The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology." Virus research 270 (2019): 197645.*

I wrote the manuscript, prepared all the figures, and performed all analyses. AK supervised the project and provided editorial assistance.

Chapter 3

Chapter based on an extract from the published article: *Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A. and Burkitt-Gray, L., 2020. Six reference-quality genomes reveal evolution of bat adaptations. Nature, 583(7817), pp.578-584.*

I wrote the referenced chapter and performed all of the analyses. The pipeline that used (ERVin) to detect the retroviral reads in bat genome assemblies was designed by me and the code was written by Robert Williams. Figures 3.2 and 3.3 were prepared by Liliana Davalos and based on results obtained by me. Figure 3.2a is prepared and based on the analysis performed by LBG. AK supervised the project and provided editorial assistance.

Chapter 4

I wrote the manuscript and performed all analyses. AK and AS supervised the project and provided editorial assistance.

Chapter 5

I wrote the manuscript and performed all the analyses. AK and AS supervised the project and provided editorial assistance.

Contents

List of Figures 6

List of Tables 26

Glossary 42

1 Introduction 46

1.1 Genome evolution, evolutionary virology and paleovirology 46

1.2 EDIs, syncytin-like genes, and viral gene co-optation 48

1.3 The Potential Role of Endogenous Viral Elements in the Evolution of Bats as Reservoirs for Zoonotic Viruses 51

1.3.1 Endogenous retroviruses 53

1.3.2 Non-retroviral endogenous viral elements 56

1.3.3 Other transposable elements 63

1.3.4 Bat immune system 64

1.3.5 The role of endogenous retroviruses and endogenous viral elements in shaping bat immunity 68

1.4 Conclusions 72

2 The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology 76

2.1 Introduction . . . 76

2.2 Materials and methods . . . 79

2.2.1 Database search and sequence analysis . . . 79

2.2.2 Sequence alignments and phylogenetic tree reconstruction . . . 80

2.2.3 Orthologous sequence search and annotation . . . 82

2.2.4 PAML analysis of orthologous sequences . . . 82

2.3 Results . . . 83

2.3.1 Search for gammaretroviral *envs* . . . 83

2.3.2 Retroviral envelope phylogenetic tree reconstruction . . . 84

2.3.3 Orthologous sequence identification . . . 89

2.3.4 dN/dS analysis for orthologous sequences . . . 92

2.3.5 Selected ortholog sequence analysis . . . 94

2.4 Discussion . . . 99

2.5 Conclusions . . . 103

2.6 Laboratory work design: Bat *syncytin*-like gene cell-cell fusion assay and tissue expression . . . 105

2.6.1 Materials . . . 106

2.6.2 Synthetic gene construction and Gibson assembly . . . 107

2.6.3 Cell-cell fusion assay . . . 111

2.6.4 Real Time qRT-PCR . . . 111

3 Integrated retroviruses in six new bat genomes 113

3.1 Introduction . . . 113

3.2 Materials and Methods . . . 120

3.3 Results . . . 122

3.4 Discussion . . . 134
3.5 Conclusions . . . 136

4 Toll-like receptors in bat species 138
4.1 Introduction . . . 138
4.2 Materials and Methods . . . 143
4.2.1 Database search, sequence alignments and phylogenetic tree reconstruction . . . 143
4.2.2 Domain architecture prediction . . . 144
4.2.3 Selection analysis . . . 144
4.3 Results . . . 146
4.4 Discussion . . . 166
4.5 Conclusions . . . 173

5 Change of receptor affinity and emergence of the new zoonotic pathogens: ACE2 receptor genetic differences amongst animals 189
5.1 Introduction . . . 189
5.2 Materials and methods . . . 193
5.2.1 Database search and alignments . . . 193
5.2.2 Phylogenetic tree reconstructions . . . 194
5.2.3 Selection and statistical analysis . . . 194
5.3 Results . . . 196
5.4 Discussion . . . 208
5.5 Conclusions . . . 215

6 Discussion and Conclusions 217

6.1 Key results, implications, and future prospects of this study 217

6.1.1 Chapter 2: The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology 219

6.1.2 Chapter 3: Integrated retroviruses in six new bat genomes 220

6.1.3 Chapter 4: Toll-like receptors in bat species 222

6.1.4 Chapter 5: The change of receptor affinity and emergence of the new zoonotic pathogens: ACE2 receptor genetic differences amongst animals 224

6.1.5 General implications and future prospects for the bat-virus interactions and co-evolution 225

6.2 General future prospects and directions for the host-virus co-evolution studies and implications for the science and global infectious diseases threat 228

7 Appendix to Chapter 2 233

8 Appendix to Chapter 3 252

9 Appendix to Chapter 4 315

10 Appendix to Chapter 5 348

Bibliography 361

List of Figures

1.1 Overall consensus bat family–level phylogeny based on [1] 52

1.2 Genome architecture of exogenous retroviruses and endogenous retroviruses. (a) Exogenous retroviral proteins: group-specific antigen (Gag), polymerase (Pol), and envelope (Env). (b) Exogenous retroviruses: long terminal repeats (LTRs), group-specific antigen (Gag-like), polymerase (Pol-like), and envelope (Env-like). 54

1.3 Genome architecture of exogenous bornaviruses and endogenous bornavirus elements. (a) Exogenous bornavirus proteins: nucleoprotein (NP), phosphoprotein (P), accessory p-10 protein (X), matrix protein (M), glycoprotein (GP), and RNA-dependent RNA polymerase (L). (b) Endogenous bornavirus-like elements (EBLs): polymerase (EBL-L), nucleoprotein (EBL-NP), glycoprotein (EBL-GP), and matrix (EBL-M). 57

1.4 Genome architecture of exogenous filoviruses and endogenous filovirus elements. (a) Exogenous filovirus proteins: nucleoprotein (NP), nonstructural proteins (VP35, VP40, VP30, and VP24), envelope glycoprotein (GP), and polymerase (L). (b) Endogenous filovirus-like elements (EFLs): nucleoprotein (EFL-NP) and nonstructural protein (EFL-VP35). 59

1.5 Genome architecture of exogenous parvoviruses and endogenous parvovirus elements. (a) Exogenous parvovirus proteins: replicase (Rep or NS) and capsid protein (Cap or VP). (b) Endogenous parvovirus-like elements (EPLs): replicase (EPL-NS) and capsid protein (EPL-VP). 62

1.6 Tolerance to infection. Infection can lead to a strong immune response or a lack of it and a range of responses in-between. A strong response can lead to the death of the host or clearance of infection and recovery. If the host gets infected again with the same virus, it might develop a strong immune response again or tolerance to the infection. 65

1.7 Immunity and infection. Exogenous virus infection can lead to co-option of the viral elements resulting in the presence of beneficial endogenous viral elements (EVEs) in the genome. Those elements can sometimes be used by the host in a host-virus arms race to evade infection and are called endogenous viral element-derived immunity (EDI). EDIs, environment, and ecological traits of the host all influence the host's immunity and its immune response to the infection with exogenous viruses, its clearance, or further co-option. 70

2.1 Structure of the retroviral genome and TM domain. 77

2.2 Number of envelope sequences in each bat species with *env* of length ≥400 aa 84

2.3 Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats (marked with different colours). Branches containing single species or large clusters of single family sequences collapsed for clarity. Continues on next two pages (list of all the sequences in **Table A 7.3**). 86

2.4 Simplified cladogram of the phylogeny of analysed bat species, with estimated dates (numbers at nodes) of family splits [2] and types of placentation in each of the species. Dates are all in millions of years ago (Mya). All analysed bats have discoidal type placentation. Bats marked with dark blue dots have mesometrial haemochorial type placentation, light blue dots indicate antimesometrial haemochorial type placentation, and pink dots indicate mesometrial endotheliochorial type placentation. . 90

2.5 Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats and reference sequences. *Syncytin* sequences and other reference sequences are marked with black colour. Orthologous sequences are marked with one of 31 colours – numbers in the legend correspond to the positions on the tree. 91

2.6 ORF length distribution amongst bats 99

3.1 The ERVin software pipeline. Chosen viral sequences are run in a tBLASTn mode against genome assemblies. Target genome areas with hits are extended 3kB each way and run in nBLAST mode against each other to detect retroviral LTRs. Reciprocal BLAST is run on hits flanked by the LTRs, and single hits giving a list of matching viral proteins. Additional separate step is comparing libraries of hits assigned to each viral protein detected in the target genome to find neighbouring sequences and full viral integrations. 121

3.2 a) Viral families identified in more than one genus mapped to phylogenetic tree of six bat species and seven additional mammals. Using reciprocal BLAST searches and a custom comprehensive library of viral protein sequences, six bat genomes and seven mammalian outgroups were screened for the presence of non-retroviral EVEs. Endogenous sequences identified as *Adenoviridae, Parvoviridae, Filoviridae* and *Bornaviridae* were represented across several mammalian genera. b) Bar plots show numbers of viral proteins of all seven *Retroviridae* genera detected in the genomes of our six bats. Beta-like integrations are most common for Pol and Gag proteins and gamma-like integrations are most common for Env proteins. Overall, the highest number of integrations was observed in *Myotis* (n = 630), followed by *Rousettus* (n = 334) with *Phyllostomus* containing the lowest. c) Alignment exemplifies that an ERV found in *Rhinolophus* (scaffold_m29_p_13:24821733-24822323) best matches the Env protein of an avian endogenous virus, which belongs to the alpharetrovirus group. d) UCSC genome browser screenshot (https://genome-public.pks.mpg.de/) of a 104 Mb scaffold (scaffold_m29_p_5) of the *Rhinolophus* assembly shows detected ERVs as an annotation track. 123

3.3 Mosaic plot showing the relative numbers of three retroviral gene sequences found in six bat genomes compared to seven mammalian reference genomes. 124

3.4 Full deltaretrovirus-like integration in *Molossus molossus* genome. a) Simplified figure showing reverse oriented sequence architecture with the detected genes and sequence features: LTRs (orange), Gag (dark pink), Pol (teal), Env (light pink), and pX ORF (green). b) Placement of each viral-like sequence in the *Molossus* genomic region together with all the ORFs (blue) in the right frame and direction. c) First reciprocal tBLAStn match for the Env protein 126

3.5 Gamma-alpha-like integration in *Phyllostomus discolor* genome. a) Simplified figure showing forward oriented sequence architecture with all detected genes and sequence features: LTRs (orange), Gag (dark pink), Pol (teal), Env (light pink), and stop codons (black bars). All genes in full ORFs marked with arrows. Genes containing stop codons marked with squares b) Placement of each viral-like sequence in the *Phyllostomus* genomic region together with all the ORFs (blue) in the right frame and direction. C) First reciprocal tBLAStn match for the Env protein 127

3.6 Lentiretrovirus-like integration in *Pipistrellus kuhlii* genome. a) Simplified figure showing reverse oriented sequence architecture with all detected genes and sequence features: Gag (dark pink), Pol (teal), LINE-1 (red), and stop codons (black bars). All genes in full ORFs marked with arrows. Genes containing stop codons marked with squares b) Placement of each viral-like sequence in the *Pipistrellus* genomic region together with all the ORFs (blue) in the right frame and direction. 128

3.7 Reconstructed phylogenetic tree of viral *pol* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink), and reference sequences (black). Bootstrap values are shown where the values are ≥70%. The tips of the phylogeny are labelled with the species name, position in the reference genomes and the direction of the sequences (N- negative/P- positive). Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. Continued on the next page. 130

3.7 Continued. Reconstructed phylogenetic tree of viral *pol* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink), and reference sequences (black). Bootstrap values are shown where the values are ≥70%. The tips of the phylogeny are labelled with the species name, position in the reference genomes and the direction of the sequences (N- negative/P- positive). Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. 131

3.8 Reconstructed phylogenetic tree of viral *env* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink) and reference sequences (black). Bootstrap values are shown where the values are $\geq 70\%$. The tips of the phylogeny are labelled with the species name, and the position in the reference genome. Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. Continued on the next page. 132

3.8 Continued. Reconstructed phylogenetic tree of viral *env* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink) and reference sequences (black). Bootstrap values are shown where the values are $\geq 70\%$. The tips of the phylogeny are labelled with the species name, and the position in the reference genome.Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. . 133

4.1 Localization of Toll-like receptors 1-13 with examples of agonists and signalling pathways. TLRs 1, 6, and 10 form heterodimers with TLR2 which are localised on the cell membrane. TLRs 4 and 5 form homodimers on the cell membrane while TLRs 11 and 12 can form both homo- and heterodimers and are localised on the cell membrane, while TLR11 can be also present on the endosome. TLRs 3, 7, 8, 9, and 13 are localised in endosomes. 141

4.2 Schematic of intron-exon organisation of bat TLRs. Length of the introns based on the *Myotis myotis* sequences. TLR7 and TLR9 starts with the first coding exon containing only start codon (ATG) followed by an intron, which is similar to pattern present in other mammals. 149

4.3 Positively selected site alignment for the TLRs 1, 2, 4, 5, 6, and 10. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *R. ferrumequinum*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*. 151

4.4 Positively selected site alignment for the NA-TLRs 7, 8 and 9. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *R. ferrumequinum*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*. 152

4.5 Positively selected site alignment for the TLRs 11, 12 and 13. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *M.natalensis*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*. 153

4.6 Positively selected sites in all bat TLRs. Posterior probability determined using the Bayes Empirical Bayes (BEB) method [3]. Red dots show sites with posterior probability of $\geq$0.90, and black dots show posterior probability of $\geq$0.50. Red line indicate extracellular domain, blue shows transmembrane domain, and green TIR domain. 155

4.7 Phylogenetic tree reconstruction showing TLR1, TLR6, and TLR10 of bats and reference sequences. Tree is midpoint rooted. Mammalian reference sequences marked in orange, *Yinpterachiroptera* sequences marked in blue, *Yangochiroptera* sequences marked in pink. 157

4.8 Phylogenetic tree reconstruction showing TLR7 and TLR8 within bat and reference sequences. Tree is midpoint rooted. Mammalian reference sequences marked in orange, *Yinpterachiroptera* sequences marked in blue, *Yangochiroptera* sequences marked in pink. 158

4.9 TIR domain tree for bats and all reference sequences. Each TLR set marked with a separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%. Continued on next two pages. 160

4.9 Continued. TIR domain tree for bats and all reference sequences. Each TLR set marked with separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%. 161

4.9 Continued. TIR domain tree for bats and all reference sequences. Each TLR set marked with separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%. 162

4.10 Tree showing episodic diversifying selection between branches of the TLR8 phylogenetic tree of bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection. 164

4.11 Tree showing episodic diversifying selection between branches of the TLR1 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection. 164

4.12 Tree showing episodic diversifying selection between branches of the TLR2 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection. 165

4.13 Tree showing episodic diversifying selection between branches of the TLR4 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection. 165

4.14 Tree showing episodic diversifying selection between branches of the TLR7 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection. 166
4.15 Predicted protein domain architecture of TLR1 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 176
4.16 Predicted protein domain architecture of TLR2 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 177
4.17 Predicted protein domain architecture of TLR3 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 178
4.18 Predicted protein domain architecture of TLR4 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 179
4.19 Predicted protein domain architecture of TLR5 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 180
4.20 Predicted protein domain architecture of TLR6 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. Myotis davidii and Hipossiderios armiger sequences just partial due to the gaps in assembly scaffold 181
4.21 Predicted protein domain architecture of TLR7 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 182

4.22 Predicted protein domain architecture of TLR8 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 183

4.23 Predicted protein domain architecture of TLR9 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. Myotis davidii and Myotis brandtii sequences just partial due to the gaps in assembly scaffold 184

4.24 Predicted protein domain architecture of TLR10 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 185

4.25 Predicted protein domain architecture of TLR11and TLR11-pseudogenes in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. 186

4.26 Predicted protein domain architecture of TLR12 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 187

4.27 Predicted protein domain architecture of TLR13 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. . . 188

5.1 Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Continued on the next pages. 197

5.1 Continued. Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. 198

5.1 Continued. Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. 199

5.2 Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%. Continued on the next pages. 200

5.2 Continued. Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%. . . . 201

5.2 Continued. Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%. 202

5.3 Amino acid alignment of SAR-CoV 2 ACE2 binding sites with identity of ≥90% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters. 204

5.4 Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of ≥80% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters. 205

5.5 Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of ≥70% or higher in reference to the *Homo sapiens* sequence. Dots mark mismatched characters. 206

5.6 Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of ≥60% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters. 207

5.7 3D structure of ACE2 protein based on *R.ferrumequinum* sequence visualised through ChimeraX [7, 8]. SARS-CoVr receptor binding domains marked with red colour; sites under positive selection marked with green colour. 212

7.1 VesperOrtho1 alignment with annotations, consensus sequence and identity across the sequence. 245

7.2 VesperOrtho3 alignment with annotations, consensus sequence and identity across the sequence. 246

7.3 VesperOrtho6 alignment with annotations, consensus sequence and identity across the sequence. 247

7.4 VesperOrtho7 alignment with annotations, consensus sequence and identity across the sequence. 248

7.5 MegabatOrtho1 alignment with annotations, consensus sequence and identity across sequence. 249

7.6 MegabatOrtho6 alignment with annotations, consensus sequence and identity across sequence. 250

7.7 MegabatOrtho10 alignment with annotations, consensus sequence and identity across sequence. 251

9.1 TLR1 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 322

9.2 TLR2 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 323

9.3 TLR3 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 324

9.4 TLR4 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 325

9.5 TLR5 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 326

9.6 TLR6 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 327

9.7 TLR7 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 328

9.8 TLR8 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 329

9.9 TLR9 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive- between 99% and 30%, red - below 30% identity). 330

9.10 TLR10 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 331

9.11 TLR11 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 332
9.12 TLR12 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 333
9.13 TLR13 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity). 334
9.14 TLR1 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 335
9.15 TLR2 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 336
9.16 TLR3 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 337

9.17 TLR4 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 338
9.18 TLR5 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 339
9.19 TLR6 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 340
9.20 TLR7 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 341
9.21 TLR8 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 342
9.22 TLR9 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 343
9.23 TLR10 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 344
9.24 TLR11 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 345

9.25 TLR12 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 346

9.26 TLR13 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour. 347

List of Tables

2.1 Probes of exogenous and endogenous retroviral envelope glycoproteins . . 79

2.2 List of the reference sequences 81

2.3 Ortholog sets with dN/dS results of one-ratio model. * full envelope from start to stop codon, ** only TM domain analysed, *** stop codon inside TM domain, § dN/dS value significantly below 1.0, †significant result. . . 94

2.4 Data on selected *syncytin*-like sequences showing the number of species in which a sequence was detected, its dN/dS result, and the length of the sequence in aa. 105

2.5 List of available bat cell lines 107

2.6 List of cell lines 111

3.1 Gamma-alpha retrovirus recombinant data. The table shows the genomic location, length of each protein, best viral match generated with reciprocal BLAST, and which viral family it corresponds to. 128

4.1 Summary of all TLRs present for each bat species and reference sequence. TLR presence shown in green, pseudogenes shown in pink, and missing genes shown in yellow. Each bat species is also described by general classification name, suborder name, family name, and distribution. 148

4.2 dN/dS results for all TLR sets *codeml* one-ratio model. *dN/dS value significantly below 1.0 150

4.3 Summary of the data for positively selected sites (PSS) for each set of TLRs. First column summarizes the length of each TLR sequence as number of amino acids (aa), second column shows pairwise percent identity of the sequences as both aa and nucleotides (nuc), third column shows the number of PSSs detected in each TLR by at least one method, fourth column displays the number of PSSs which were detected by at least two or more methods implying a strong positive selection pressure, the fifth column shows the percentage of PSSs across the whole gene. 154

5.1 Sites under positive selection for the ACE2 binding sites alignment detected by each used Datamonkey model and *codeml* Bayes Empirical Bayes (BEB) procedure. Sites detected by more than one model in bold. 208

7.1 Bat genome assemblies data 233

7.2 dNdS data. * difference of likelihood ratio test, and degrees of freedom (df) with significance
level of 0.05 in brackets 234

7.2 Continued. dNdS data. * difference of likelihood ratio test, and degrees of freedom (df) with significance
level of 0.05 in brackets 235

7.3 Sequence positions on a phylogenetic tree 236

7.3 Continued. Sequence positions on a phylogenetic tree. 237

7.3 Continued. Sequence positions on a phylogenetic tree. 238

7.3 Continued. Sequence positions on a phylogenetic tree. 239

7.3 Continued. Sequence positions on a phylogenetic tree. 240

7.3 Continued. Sequence positions on a phylogenetic tree. 241

7.3 Continued. Sequence positions on a phylogenetic tree. 242
7.3 Continued. Sequence position on a phylogenetic tree. 243
7.3 Continued. Sequences positions on a phylogenetic tree. 244

8.1 Accession numbers for exogenous and endogenous viral genes used in the ERV BLAST search and analysis 253
8.2 List of all integrations found in *Mollosus mollosus* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 254
8.3 List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 255
8.4 Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 256
8.5 Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 257
8.6 Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 258
8.7 List of all integrations found in Phyllostomus discolor containing two or more genes with name of the viral family, scaffold, and position on the scaffold 259
8.8 List of all integrations found in *Pipistrellus kuhlii* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 260
8.9 List of all integrations found in *Rhinolophus ferrumequinum* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 261

8.10 List of all integrations found in *Roussetus aegyptiacus* containing two or more genes with name of the viral family, scaffold, and position on the scaffold 262

8.11 Continued. List of all integrations found in Roussetus aegyptiacus containing two or more genes with name of the viral family, scaffold, and position on the scaffold 263

8.12 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name 264

8.12 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 265

8.13 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 266

8.13 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 267

8.14 Reciprocal BLAST hits for the betaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 268

8.15 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 269

8.16 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 269

8.17 Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 270

8.17 Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name 271

8.17 Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 272

8.18 Reciprocal BLAST hits for the deltaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 272

8.19 Reciprocal BLAST hits for the deltaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 272

8.20 Reciprocal BLAST hits for the deltaretroviruses Env hits in Molossus molossus genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name 273

8.21 Reciprocal BLAST hits for the epsilonretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 273

8.22 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 274

8.22 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 275

8.22 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 276

8.22 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 277

8.22 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 278

8.23 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 279

8.23 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 280

8.23 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 281

8.23 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 282

8.24 Reciprocal BLAST hits for the betaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 283

8.25 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 283

8.26 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 284

8.27 Reciprocal BLAST hits for the gammaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 285

8.27 Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 286

8.28 Reciprocal BLAST hits for the epsilonretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 287

8.29 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 287

8.30 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 288

8.31 Reciprocal BLAST hits for the betaretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 288

8.32 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 289

8.33 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 289

8.34 Reciprocal BLAST hits for the gammaretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 290

8.35 Reciprocal BLAST hits for the alpharetrovirus Env hits in *Phyllostomus discolor* genome. The table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match the common name 290

8.36 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 291

8.37 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 292

8.38 Reciprocal BLAST hits for the betaretrovirus Env hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 293

8.39 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 293

8.40 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 294

8.41 Reciprocal BLAST hits for the gammaretrovirus Env hits in Pipistrellus kuhlii genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name 295

8.42 Reciprocal BLAST hits for the epsilonretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 295

8.43 Reciprocal BLAST hits for the epsilonretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 295

8.44 Reciprocal BLAST hits for the epsilonretrovirus Env hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 296

8.45 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 297

8.45 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. . . 298

8.45 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. . . 299
8.46 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 300
8.46 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. . . 301
8.47 Reciprocal BLAST hits for the betaretrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 302
8.48 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 302
8.49 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 303

8.50 Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 304

8.51 Reciprocal BLAST hits for the alpharetrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 305

8.52 Reciprocal BLAST hits for the spumaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 305

8.53 Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 306

8.53 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 307

8.53 Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 308

8.54 Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 309

8.54 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 310

8.54 Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 311

8.55 Reciprocal BLAST hits for the betaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 311

8.56 Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 312

8.57 Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 312

8.58 Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 313

8.58 Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 314

8.59 Reciprocal BLAST hits for the alpharetrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name. 314

9.1 Bat and reference genomes statistical data 316

9.2 The dN/dS ratio for each TLR set in all tested PAML *codeml* models . . 317

9.3 *Codeml* model comparison for each set of the TLRs 318

9.4 Number of sites under positive selection (FEL, SLAC, FUBAR, *codeml* M8, MEME) and number of branches under positive selection (bsREL) for each analysed model and each bat TLR set. 319

9.5 Sites under positive selection (FEL, SLAC, FUBAR, *codeml* M8, MEME) for each bat TLR set. Sites found in by other studies underlined in the first column, and sites found by more than one model in bold. 320

9.6 Number of sites under negative selection for each analysed Datamonkey model (FEL, SLAC, FUBAR) in each bat TLR set. 321

10.1 List of species used in the analysis with accession number, Latin name, class, order, family, and common English name 357

10.2 List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score. 358

10.2 Continued. List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score. 359

10.2 Continued. List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score. 360

10.3 LRT test for the branch-specific model (M1/M2), the site-specific model (M8a/M8), and the branch-site model (BSAn/BSA). p1 is the proportion of sites having $\omega > 1$ 360

Glossary

adaptive immunity Also known as acquired immunity. The adaptive immune response is specific to the pathogen presented. 66

agonist A chemical that binds to a receptor and activates the receptor to produce a biological response. 12, 66, 140, 141, 166, 170, 172

antagonist A substance which interferes with or inhibits the physiological action of another. 60

co-infection Also known as a superinfection. The simultaneous infection of a host by multiple pathogen species. 53

complex retroviruses Retroviruses which have a greater variety of gene products. 56, 129, 135, 218, 221

derepression The removal of a mechanism which decreased or inhibited the expression of a gene. 53

endogenization The acquisition of a new genetic sequence in the genome of the host population through a vertical transfer. 53, 61, 89, 119

exaptation Also known as co-option. The process by which genes acquire functions for which they were not originally adapted. 49, 136

germline The population of a multicellular organism's cells that pass on their genetic material to the offspring. 48, 52, 76, 114, 142

horizontal transmission The process in which e.g. viruses are transmitted among individuals of the same generation. 48

innate immunity A host's first line of defense against invading pathogens. 222, 228

ligand Any molecule or atom that irreversibly binds to a receiving protein molecule, known as a receptor. 66, 138, 140, 142, 154, 156, 169, 170, 172, 174

lineage A sequence of populations, organisms, cells, or genes each of which is considered to have evolved from its predecessor. 46, 77, 115, 118, 135, 146, 163, 171

model species A non-human species that is extensively studied to understand particular biological phenomena. 175, 215, 227

molecular clock A technique that uses the mutation rate of biomolecules to deduce the time when two or more life forms diverged. 117

nonhomologous recombination A major pathway for the repair of chromosomal double-strand breaks in the DNA of somatic cells. 52, 61

open reading frame The part of a gene that has the ability to be translated. 60, 61, 78, 80, 99

oxidative stress An imbalance between the production and accumulation of reactive oxygen species and a biological system's ability to detoxify or to repair the resulting damage. 103, 170, 171

phenotype The set of observable characteristics or traits of an organism. 64, 67, 72, 73

provirus A virus genome that is integrated into the DNA of a host cell. 117

pseudogene A DNA sequence that resembles a gene but has been mutated into an inactive form over the course of evolution. 67, 102, 140, 144, 146, 147, 163, 169, 171, 172

recombination An exchange of genetic material between different organisms or two pieces of DNA/chromosomes which leads to the production of a new combination of alleles. 53, 76, 102, 135, 221

reservoir A population which is chronically infested with the causative agent of a disease and can act as a source of further infections. 51, 56, 60, 64, 72, 78, 115, 117, 134, 135, 168, 171, 190–193, 208, 211, 217, 218, 221, 223–232

restriction factor An anti-viral protein of the host that counteracts or 'restricts' viral replication. 48, 71

retroelement An endogenous component of eukaryotic genomes that is able to amplify to new locations in the genome through an RNA intermediate. 52, 76

RNA interference A biological process in which RNA molecules are involved in sequence-specific suppression of gene expression. 58

seroprevalence The level of a pathogen in a population, as measured in for e.g. blood serum. 60

speciation A process occurring when a group within a species separates from other members of its species and develops its own unique characteristics. 100, 101

super-spreader An unusually contagious organism infected with a pathogen. 56

syncytia Cells containing several nuclei, formed by fusion of cells or by division of nuclei. 111

transposon Also known as a transposable element. A DNA sequence that can change its position within a genome, sometimes altering the cell's genetic identity and genome size. 63, 223

tropism The capability of a virus to infect a distinct group of cells in the host or different group of hosts than it's reservoir species. 190, 191, 215, 216, 231

wild type The phenotype of the typical form of a species as it occurs in nature. 50

1 Introduction

1.1 Genome evolution, evolutionary virology and paleovirology

The contemporary application of the word evolution to the process of development of species over time was popularised by proponents of Charles Darwin's theory of natural selection [9] following its first published use by Charles Lyell in 1832. The contemporary definition of evolution entered the vocabulary of biologists and from that point started to change and 'evolve' at a similar rate to that of developments in the biological sciences, and the growth of our knowledge about the natural environment and living organisms.

The simplest definition that applies to both living organisms and biological entities like viruses is a change of the characteristics of a species over time which relies on the process of natural selection. Additionally, Leigh van Valen in 1973 characterised evolution as 'the control of development by ecology' [10].

Genomes of living organisms and viruses evolve and change constantly. The rate of evolution is the speed of genetic change in a lineage over a period of time. Viruses display the highest rate of evolution of all biological entities [11], while multicellular organisms like humans or other mammals display some of the lowest rates of evolution seen in living organisms [12]. But without a doubt, even if slow and almost impossible

to observe over a relatively short period of time, evolution still is one of the major forces of nature.

Genomes can evolve and change through random point mutations: substitutions, insertions and deletions; gene or genomic region duplications and rearrangements; or the acquisition of genes from other species. All of these are governed by different pressures and processes and happen in various proportions depending on species, genome size, type of genetic material, or even genomic region [13]. Some of these changes might give an individual advantages or disadvantages in comparison to other individuals and be preserved or lost in the population.

The special case we see within the category of such influences is the constant arms-race between pathogens and their hosts. In general, pathogens evolve to both evade host immune responses and be more efficient at infection and replication, and hosts try to overcome these changes through alterations in their immune responses and physiological characteristics which might defeat the viral threat or dampen the pathological effects of the infection. Mutual evolution of viruses and their hosts and the subsequent impact on the virus-host arms-race is studied by a part of virological sciences called evolutionary virology.

Paleovirology is the name of a relatively young discipline of evolutionary virology, which has emerged within the last decade. It concentrates on the study of ancient viral sequences in host genomes within evolutionary timescales, and studies the influence of prehistoric gene transfers on species evolution [14]. This knowledge can help us in viral-disease prevention strategy development and monitoring methods, as well as granting us a deeper understanding of the evolutionary history of host-virus interactions.

Methods used in paleovirology can be divided into two approaches: direct and indirect

[14]. Viruses are not meaningfully 'fossilized' in the traditional mineralised sense of the word so we cannot study extinct ancient viruses via physical samples. Our only window is through their exogenous descendants and the fragile traces left in host genomes. Direct approaches are based on identification of such genomic 'fossil' sequences left in host genomes [14, 15, 16]. When viruses do not leave a trace in the host genome, we can use indirect paleovirological methods: looking at modern viral descendants of ancient viruses or investigating the effects of ancient viruses on their hosts' antiviral genes [14, 17]. The best examples of this are innate immune genes like TRIM5α [18], SAMHD1 [19], and the APOBEC gene family [20, 21, 22, 23, 24], which evolved during a virus-host arms race to play a role as restriction factors in antiviral immunity. To achieve the best results, we can combine direct and indirect approaches, or use both methods to validate one another.

1.2 EDIs, syncytin-like genes, and viral gene co-optation

Viruses can be a force driving genome evolution of their hosts, driving changes of their immune genes, but also being a source of novel genetic material. Many exogenous viruses enter the genomes of their hosts upon infection, and if this occurs in the germline of the host they are subsequently passed via vertical rather than horizontal transmission [14]. Some integrate into the host genome as an obligatory step during their replication (e.g. retroviruses), while others do so by chance or through dormant steps in their life cycle (e.g. filoviruses). These integrated vertically transmitted viral sequences start evolving at the host rate of mutation rather than that of their viral progenitors [25]. This leads to two destinies for such sequences - mutation, partial loss of the coding sequence and a slow 'decay' if its function is harmful or non-advantageous for the host's survival, or fixation and preservation in genomes of the host population if the sequence provides a selective

advantage for the host – so called exaptation [26]. Viral sequences integrated into a host genome are called endogenous viral elements (EVEs), or endogenous retroviruses (ERVs) in cases of sequences of retroviral origin [14]. In some cases, a host can use such sequences during subsequent viral infections as an EVE-derived immunity (EDI) [27].

ERVs are an integral part of animal genomes, constituting up to 8% of the human genome [28] and around 10% in mice [29]. While some endogenous sequences are just evolutionary 'left-overs', fragmented and mutated remnants of ancient infections affecting our ancestors, some were altered and used as genes in the viral-host arms-race and later acting as stand-alone genes in biological processes [26]. We have several examples of such genes discovered very recently, one of the most widely studied examples being the retroviral envelope glycoproteins known as *syncytins*[30].

The placenta is an autonomous and transient organ that appeared in mammalian ancestors around 150 Myr ago [30]. It facilitates feeding and oxygenating an embryo, and the secretion of hormones and growth factors. It helps isolate foetus from mother, and reduces mother-foetus conflict [31]. This invasive 'tumour-like' syncytial plate is unique to mammals but the method of its formation was largely unknown prior to identification of retroviral envelope proteins [32, 33, 34]. The term *syncytin*, refers to captured retroviral envelope (*env*) glycoprotein genes, which take part in one of the main processes of placenta morphogenesis during pregnancy in mammals, the formation of syncytiotrophoblast from mononuclear cytotrophoblast [35, 36, 37, 38]. Trophoblast invasion is a carefully controlled process. Its down regulation might lead to placental disorders like preeclampsia, or cancers such as choriocarcinoma [38, 39].

Retroviral envelope genes in the transmembrane (TM) region contain the immunosuppressive domain (ISD), which helps the virus to down regulate the immune system of the host during infection through cytokine pathways [40]. Functionality of this domain was

conserved in both human *syncytin-2* and mouse *syncytin-B*, but not all *syncytins* have immunosuppressive ability, which is determined only by two single amino acid changes in its *env* ISD [41]. Recently, it has been demonstrated experimentally that *syncytins* can play a much bigger role in an organism's biological functions and development than was previously suspected. *Syncytins* take part in myoblast fusion and muscle development in mice [42]. *Syncytin-B* knockout mice showed delay in muscle development and reduction in myofiber size, thus male mice were phenotypically similar to female mice, while female mice did not show differences in their size in comparison to the wild type. *Syncytins* were also found to be expressed in the muscles of primates, ruminants and carnivores [42].

In mammalian genomes, we can find several other *env* sequences with immunosuppressive properties [43, 44, 45, 46]. In some case it has not been determined whether or not these other *envs* play a role in biological processes, or even in placentation [43, 46], but it has been proposed that these sequences might be old 'degenerated *syncytins*' that started 'decaying' and losing their functionality over time [47, 48]. This function was picked up later by new, functional versions in other ERVs [47, 48, 49, 50].

In the genomes of animals we can also find a number of other endogenous viral sequences with proven EDI functions, of which the following are some examples: Friend virus susceptibility protein 1 and 4 (*Fv1* and *Fv4*) in mice that encode retroviral proteins, block infections by exogenous retroviruses that can lead to mammary tumour in lactating females [45, 51, 52]. Retroviral proteins have specialised in controlling biological processes like the expression of amylase in human salivary glands [53, 54], or having a major role in development of memory, learning and thought processes in the case of recently discovered *arc* gene, which encodes a retrovirus-like group-specific antigen (Gag) protein [55]. Such genes are relatively rare but there is a number of co-opted viral genes that were

preserved in mammalian genomes for millions of years but still did not have a confirmed function. One such example are NIRVs, which are filovirus-like integrations in bat and marsupial genomes [56, 57].

1.3 The Potential Role of Endogenous Viral Elements in the Evolution of Bats as Reservoirs for Zoonotic Viruses

All bats are placed into the order *Chiroptera*, which is the second most speciose order of mammals, representing ∼20% of all known mammals [58], united by their unique ability to fly [1]. Coupling flight with laryngeal echolocation, bats are successful nocturnal predators that are found throughout the globe, where they have evolved to exist in diverse ecological niches, feeding on insects, small mammals, fish, blood, nectar, fruit, and pollen [1]. Order *Chiroptera* is divided into 21 families, placed into two subordinal groups, *Yangochiroptera* and *Yinpterochiroptera* [1, 2, 59], estimated to have diverged from other Laurasiatherian orders ∼65 million years ago (Mya) [60] **(Figure 1.1)**.

It is evident that diminishing bat populations can have huge consequences on the equilibrium of ecosystems; hence, a substantial effort is being put into bat conservation [61]. Nonetheless, in recent times, bats have also been associated with an increasing number of zoonotic viral diseases (the diseases that can jump from non-human animals to humans), that could pose a threat to animal and human health [62, 63, 64]. Despite sampling biases, there is evidence that bats do harbour a significantly higher proportion of zoonotic viruses than other mammalian orders [63]. Yet, the question of why bats seem to be a reservoir for many different viruses remains unclear [64, 65, 66].

Retroviruses are RNA viruses that can insert copies of their genomes into the genome

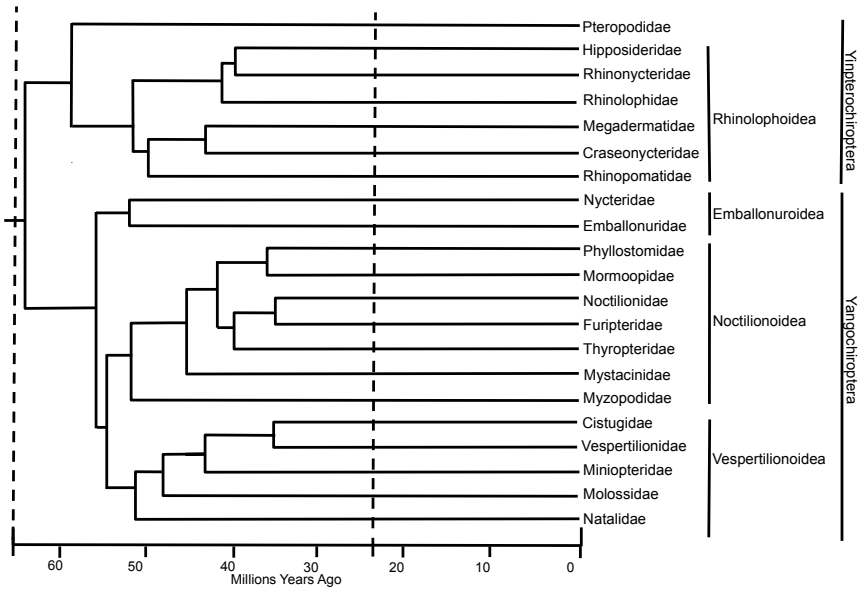


Figure 1.1: Overall consensus bat family–level phylogeny based on [1]

of their host [67]. These viruses encode for a reverse transcriptase, used for transcribing viral RNA into DNA during virus replication in a process called reverse transcription. Retroviruses are divided into those with simple (alpha-, beta-, gamma-, and epsilon-retroviruses) and complex (lenti-, delta-, and spumaviruses) genomes and exist in both exogenous and endogenous forms [67]. While exogenous retroviruses remain infectious and are transmitted horizontally, endogenous retroviruses (ERVs) result from those that are integrated into the genome of their hosts, have gained access to the germline, and subsequently inherited vertically and became fixed in the host population [67]. More rarely, non-retroviral RNA and DNA viruses can also be integrated within the host genome during infection, mediated by nonhomologous recombination or by interaction with retroelements [27]. Endogenous genomic elements homologous to non-retroviral viruses found so far in the genomes of different animal species include borna-, bunya-, filo-, orthomyxo-, reo-, and rhabdovirus sequences [11]. Throughout this review, we use the term EVEs (endogenous viral elements) to refer to all endogenous elements (retro-

viral or non-retroviral in origin), while we use the term ERVs for endogenous elements of retroviral origin. When EVEs become integrated in the genome of a host, they can eventually become inactive genetic fossils by accumulating mutations through time [25]. Some of these elements may still retain the ability to be actively transcribed or can even become reactivated through recombination or derepression (e.g. through co-infection with other virus), with potential implications for host biology [68, 69, 70, 71].

1.3.1 Endogenous retroviruses

Retroviruses are the most common representatives among all endogenous viral sequences found within animal genomes. ERVs are grouped into three classes: Class I ERVs related to the exogenous gammaretroviruses, Class II ERVs related to the betaretroviruses, and Class III ERVs distantly related to spumaviruses [67] **(Figure 1.2)**. ERVs that do not fit into these three classes such as lentiviral ERVs have also been described as integrated in animal genomes [72, 73], although these are relatively rare. ERVs have been identified in all vertebrate genomes studied so far, with a larger number and diversity biased toward the gamma- and betaretroviruses [74, 75]. For example, it has been shown that compared with other ERVs, many of the beta- and gammaretrovirus ERVs found in different animal genomes (including those of bats) have crossed species barriers and undergone recent integration events [76, 77, 78], while others are still undergoing endogenization, as is the case of the exogenous/endogenous koala gammaretrovirus [79].

Many unique ERVs have been described in different bat species [24, 80, 81, 82]. The nature of the ERV elements found in different bat genomes is indicative of a complex amplification history, characterized by a mixture of retrotransposition (amplification and reintegration into other genomic locations) and reinfection (characterized by the reactivation and production of viral particles, which in turn infect other cells of the

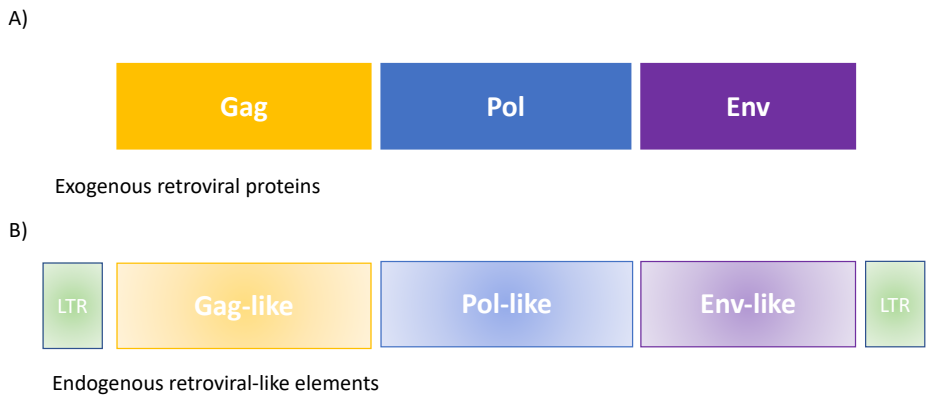


Figure 1.2: Genome architecture of exogenous retroviruses and endogenous retroviruses. (a) Exogenous retroviral proteins: group-specific antigen (Gag), polymerase (Pol), and envelope (Env). (b) Exogenous retroviruses: long terminal repeats (LTRs), group-specific antigen (Gag-like), polymerase (Pol-like), and envelope (Env-like).

host organism and then become endogenized) [81]. Multiple gammaretrovirus ERVs have been described in many different bat genomes, some of which are of ancient origin and are present in high copy numbers [24, 80, 82, 83]. Phylogenetic analysis has led to the hypothesis that gammaretrovirus ERVs might have originated in bats, as the bat gammaretrovirus ERVs are positioned basally to all other mammalian gammaretrovirus ERVs on the phylogenetic tree [83]. Within the genome of the little brown bat (*Myotis lucifugus*), over 360 complete gamma- and betaretrovirus ERVs were identified, approximately 25% of which were unclassified, while 64% were determined to be a result of recent proviral integrations [84]. Of interest, some of the identified gammaretrovirus ERVs are intact and potentially replication competent, suggesting that many might still have exogenous counterparts circulating in different hosts, as has been speculated for other ERVs found in other bat species [78, 84]. Some of the betaretrovirus ERVs are

actively transcribed and expressed, while further genomic analyses revealed intact and potentially functional copies in different bat genomes [76, 85]. However, the biological consequences of these elements being actively transcribed remain unknown.

A more recent study on six high-quality bat genome assemblies identified even higher numbers of integrations within greater mouse-eared bat (*Myotis myotis*) containing 630 *pol* sequences, followed by Egyptian fruit bat (*Roussetus aegyptiacus*) with 334 integrations [24]. Most of those sequences are Class I and II, but some of the integrations distantly resemble alpharetroviruses (in pale spear-nosed bat - *Phyllostomus discolor*; greater horseshoe bat - *Rhinolophus ferrumequinum*; and Egyptian fruit bat - *Rousettus aegyptiacus* genomes), which is the first observation of alpharetrovirus-like retroviruses in mammalian genomes [24]. The Kuhl's pipistrelle (*Pipistrellus kuhlii*) genome also contained *gag*-like lentiretroviral sequences that are very rarely observed in endogenous form [24], although these were short fragments and whether they truly result from ancient lentiviral integrations is yet to be determined. All those results show that bat genomes can contain a vast diversity of ERV sequences, among the highest observed across mammalian genomes, despite their small genome sizes.

Very few exogenous deltaretroviruses have been identified to date and even fewer deltaretrovirus ERVs [86]. Some of the few examples of deltaretrovirus ERV fragments were found in the miniopterid long-fingered bat (*Miniopterus natalensis*) genome [87] and horseshoe bat genome [88] and later with a full *pol* sequence in a velvety free-tailed bat (*Molossus molossus*) genome [24]. Similarly, the one example of a spumaretrovirus potential ERV was detected within a intermediate horseshoe bat (*Rhinolophus affinis*) genome [84]. Although it was unclear if the first described spumaretrovirus fragment was endogenous or exogenous, recent results describing the presence of five spumaretrovirus fragments in *Rhinolophus aegyptiacus* confirmed that the sequences are endogenous [24]. Most of

these spuma- and deltaretrovirus ERV sequences show a high degree of fragmentation, evidencing likely ancient integrations [24, 84, 86, 87, 88]. Those results show that some bat genomes appear to be prone to ERV colonization by complex retroviruses, although the reasons why remain elusive. It has been suggested based on the phylogenetic distribution of bat and non-bat ERV sequences that bats are more capable of receiving viruses from other non-bat reservoirs and then act as super-spreaders [85]. In agreement with this observation, other studies also support the hypothesis that bats might actually act as reservoirs of extant viruses to other animal species [89, 90, 91].

1.3.2 Non-retroviral endogenous viral elements

The most commonly detected non-retroviral EVEs in animal genomes include filovirus, bornavirus, and parvovirus sequences [27, 56, 92, 93, 94]. Although most of these elements are found to be inactive, a few are preserved and may be actively transcribed in tissues, suggesting that these viral elements could still have potential biological functions despite having been integrated millions of years ago [95]. Next, we describe the most commonly found EVEs within bat genomes.

Endogenous bornavirus-like elements

Borna disease viruses (BDVs) are single-stranded negative-sense RNA viruses with an ∼9 kb long genome that encodes for six proteins: nucleoprotein (NP), phosphoprotein (P), accessory p-10 protein (X), matrix protein (M), glycoprotein (GP), and RNA-dependent RNA polymerase (L) [96] **(Figure 1.3)**. BDVs were first described in horses and subsequently widely identified in mammals, birds, and reptiles. However, the implications of bornavirus infections remain unclear, and a natural reservoir has not yet been identified

[97]. Endogenous bornavirus-like elements (EBLs) have been detected in the genomes of many different animals [27, 92, 95, 98]. The integration of EBLs in animal genomes is potentially attributed to the fact that BDVs have a unique non-cytolytic replication strategy that takes place inside the nucleus; even if there is no direct reverse transcription stage, this nuclear replication makes these viruses more likely to be endogenized than viruses that primarily replicate in the cytoplasm [96].

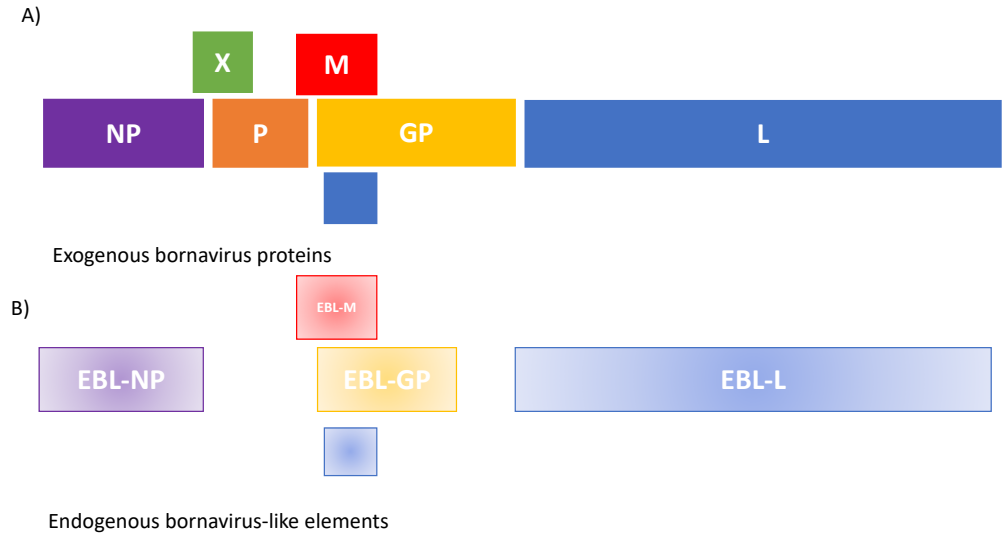


Figure 1.3: Genome architecture of exogenous bornaviruses and endogenous bornavirus elements. (a) Exogenous bornavirus proteins: nucleoprotein (NP), phosphoprotein (P), accessory p-10 protein (X), matrix protein (M), glycoprotein (GP), and RNA-dependent RNA polymerase (L). (b) Endogenous bornavirus-like elements (EBLs): polymerase (EBL-L), nucleoprotein (EBL-NP), glycoprotein (EBL-GP), and matrix (EBL-M).

Genomic analyses of over 10 different bat species belonging to both *Yangochiroptera* and the *Yinpterochiroptera* suborders **(Figure 1.1)** revealed the presence of different endogenous sequences homologous to four BDV genes: polymerase (EBL-L), nucleoprotein (EBL-NP), glycoprotein (EBL-GP), and matrix (EBL-M) [27, 92, 95, 98] **(Figure**

1.3). Many of these sequences, including full-length preserved copies, are located close to long interspersed nuclear genome elements (LINEs) that often flank transcription site duplications, suggesting that EBL integration is mediated through LINE element-facilitated integration [27, 92, 95, 98]. The presence of EBL sequences in distantly related bat species (David's myotis - *Myotis davidii* and Parnell's mustached bat - *Pteronotus parnellii*) that diverged at least ∼55 Mya has led to the conclusion that the integration of EBLs must have taken place at least 50 Mya, during an early evolutionary stage of the *Chiroptera* [98]. Nonetheless, EBLs seem to be much more common within the *Yangochiroptera*, as reflected by an overall higher copy number and diversity. It has been proposed that the smaller number of EBLs within the *Yinpterochiroptera* could be associated with a loss of LINE activity [98].

Research on other mammals suggests that the EBL-NP gene copies might inhibit the replication of exogenous bornaviruses in vitro [99]. Another proposed function for EBL genes is that the non-coding RNAs might help maintain an immune memory against modern bornavirus infections [100, 101, 102]. Other reports state that EBL genes may mimic an RNA-dependent RNA polymerase function within the RNA interference machinery [95, 103], playing a role in gene regulation and host immune defense against a range of different viruses with homologous proteins [92, 104]. Active EBL transcription has been described for some bat tissues, suggesting that some EBL elements might have maintained some of their biological properties and act similar to the other mammalian endogenous bornavirus-like sequences [95, 105].

Endogenous filovirus-like elements

Filoviruses are single-stranded negative-sense RNA viruses with a genome that is approximately 20 kb long, encoding for seven different proteins: the nucleoprotein (NP),

non-structural proteins (VP35, VP40, VP30, and VP24), envelope glycoprotein (GP), and polymerase (L) [106] **(Figure 1.4)**. The *Filoviridae* family has four recognized genera of extant viruses: *Marburgvirus* (MARV), *Ebolavirus* (EBOV), the *Cuevavirus* [with only one known representative species, *Lloviu virus* (LLOV)], and the *Dianlovirus* [again with only one known representative species, *Mengla virus* (MLAV)] [107, 108, 109]. Both LLOV and MLAV have been detected so far in only two bat species: European Schreiber's long-fingered bat (*Miniopterus schreibersii*) and Leschenault's rousette (*Rousettus leschenaultia*) from China [107, 109].

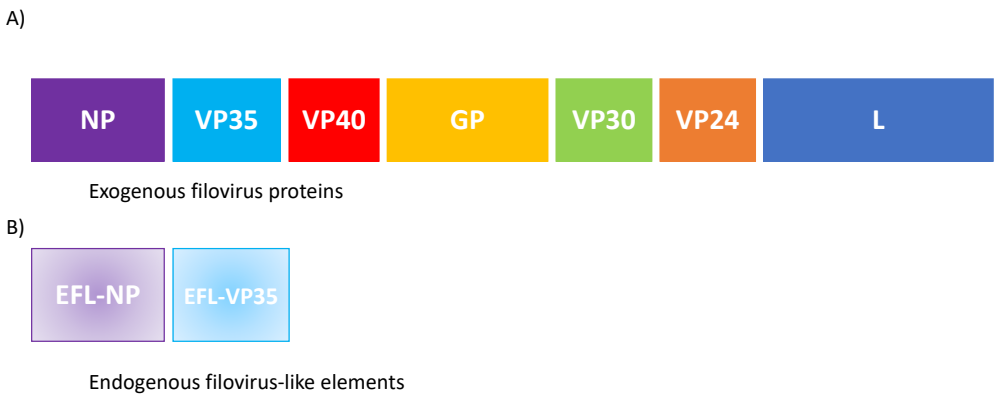


Figure 1.4: Genome architecture of exogenous filoviruses and endogenous filovirus elements. (a) Exogenous filovirus proteins: nucleoprotein (NP), nonstructural proteins (VP35, VP40, VP30, and VP24), envelope glycoprotein (GP), and polymerase (L). (b) Endogenous filovirus-like elements (EFLs): nucleoprotein (EFL-NP) and nonstructural protein (EFL-VP35).

Some filoviruses, such as EBOV and MARV, cause lethal haemorrhagic fever in humans and other mammalian species but are suspected to cause only a mild to asymptomatic infection in bats [107, 108, 109]. Thus, bats have been proposed to be the original

reservoir host for filoviruses [110]. However, EBOV and MARV seroprevalence in bat populations is below 5%, which is significantly lower than what would be expected for the seroprevalence of a virus within a reservoir host species (e.g., seroprevalence for lyssaviruses in bat colonies can be as high as ∼60–70%) [111, 112]. Thus, the hypothesis of bats being a reservoir for filoviruses remains speculative, as no EBOV and MARV viruses have yet been isolated from any bat species, and there is no clear evidence for zoonotic filoviral transmissions from bats to other animals.

Endogenous filovirus-like elements (EFLs) have been identified in several different mammalian genomes, including those of bats and rodents, with a higher copy number of intact sequences in the *M. lucifugus* bat genome [92, 56]. As with BDVs, filoviruses lack a reverse transcriptase. Furthermore, they do not replicate within the nucleus, which makes their integration into the host genome even more unlikely. Thus, the presence of filovirus sequences in vertebrate genomes has led to the hypothesis that EFL integration is associated with retrotransposon activity. As in the case of EBLs, transcription site duplications were often observed in close proximity to EFLs, suggestive also of LINE-mediated integration [27, 56, 92]. Endogenous sequences homologous to the EBOV/MARV N and VP35 genes were detected in the genomes of different bat species, displaying nearly complete open reading frames, including an interferon (IFN) inhibitory domain and other motifs needed for viral replication and transcription [27, 56, 92] **(Figure 1.4)**. An assay with a *M. lucifugus* VP35-like sequence found it to have potential as an IFN antagonist but not polymerase cofactor in human kidney cell lines, suggesting it could have been preserved in the bat genome with suppressive effects on host immunity while its functions acting in viral transcription and replication were discarded [113].

Endogenous parvovirus-like elements

The *Parvoviridae* (including autonomous parvoviruses and the dependoparvoviruses) are single stranded DNA viruses with a genome size ranging typically between 4 and 6 kb, encoding for only two open reading frames: a replicase (Rep or NS1) and a capsid protein (Cap or VP1) [114] **(Figure 1.5)**. Endogenous sequences related to both autonomous parvoviruses and dependoparvoviruses have been detected in different animal genomes, for example in those of rodents, rabbits, marsupials, and African elephants [115, 116, 117, 118]. They were also found in the genomes of bats, with nearly complete insertions homologous to the parvovirus NS gene [94, 119]. Again, LINEs have been observed close to the integration site of some endogenous parvovirus-like elements (EPLs), which is suggestive of LINE-mediated integration as a common mechanism for EVE endogenization. DNA virus integration can also occur by nonhomologous recombination [27], which could be more common in genomic regions that are highly unstable. As double-strand breaks are targets for integration, this could result in viral integrations being more frequently placed by chance within these regions [120].

The NS protein is critical in reducing IFN responses and maintaining viral replication. Thus, NS proteins tend to be more conserved than other proteins in extant parvoviruses [121]. Regarding the biological implications of EPLs, it has been shown that the NS protein is involved in selective cell apoptosis [122]. It has been hypothesized that the presence of EPLs might provide the host with antiviral immunity against extant parvovirus infections [116]. This is supported by the observation that some animal species with endogenous parvo-like sequences are less likely to be infected with related exogenous parvoviruses, suggesting that parvovirus sequences might be captured by the host during double-stranded DNA repair and further used as protection against other viruses [117].

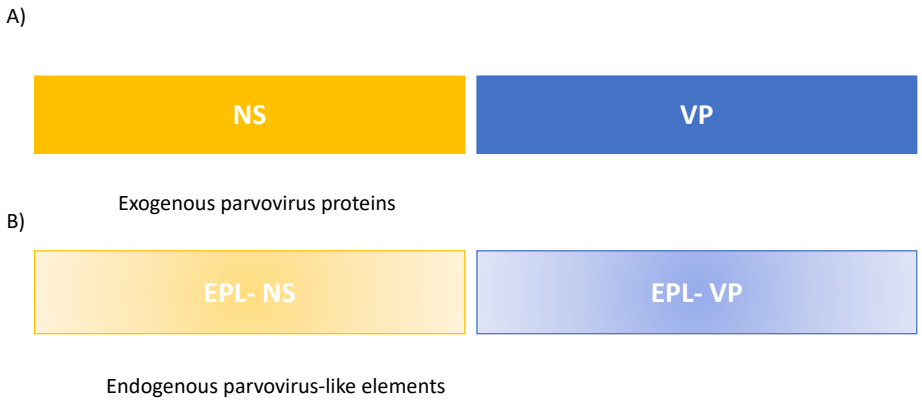


Figure 1.5: Genome architecture of exogenous parvoviruses and endogenous parvovirus elements. (a) Exogenous parvovirus proteins: replicase (Rep or NS) and capsid protein (Cap or VP). (b) Endogenous parvovirus-like elements (EPLs): replicase (EPL-NS) and capsid protein (EPL-VP).

1.3.3 Other transposable elements

Not only viruses can be a source of genetic material and changes in the genomes of animals. Class II transposable elements are DNA based transposons, previously observed in their active state mostly in plants and invertebrates and thought to be inactive in mammalian genomes [123, 124]. Recently it was found that this type of transposons was introduced into bat genomes (*Myotis spp.*) around 15 Mya, through horizontal transfer from an unknown source [125, 126]. Diversification of *Myotis* bats occurred around 12-13 Mya, which corresponds to the potential activity of DNA transposons in these species [127, 128]. The viral sequences might be the best candidates for horizontal introduction of DNA transposons into bat genomes [128], but also it might be the transposons facilitating viral integrations into the certain parts of the hosts genomes [129].

Myotis bats have relatively small genomes even in comparison to other bat species (around 2000 Mb) [130], 5% of which is made up of ERVs [84], and around 10% of class II transposons [128, 131]. Genetic tolerance of bats, and their ability to safely harbour large loads of various transposable elements in their genomes was observed and reported before [77, 130]. Some of these elements (helitrons, hAT, TcMariner, piggyback) are still active in several species of bats [123, 125], while being also associated with genome evolution of other species through rearrangements and restructuring of chromosomes, reshuffling and duplication of host genes as well as spontaneous excisions while capturing host genes in the process [123, 125, 129, 132, 133]. We know that endogenous viruses in the *Vespertilionidae* bat family were acquired in very complex ways through waves of retransposition and reinfection events [81]. It was also suggested that pressure on bat genomes given by co-evolution with many viral species might have had a huge impact on bat genome evolution including their immune system [24, 77].

1.3.4 Bat immune system

Just as with other mammals, bats are susceptible to infection with a range of pathogens, including viruses that can have a significant effect on their populations [134, 135]. However, there is evidence that some bat species can be persistently, yet asymptomatically, infected with viruses that may be lethal to other mammals [111, 136]. Thus, bats are hypothesized to display a virus-tolerant phenotype as an immune strategy to control viral induced pathology [137, 138]. Despite the importance of understanding the complex relationship between *Chiropteran* immunology and the disease ecology of the viruses that they host, bat immunology still remains a relatively under explored field [137, 138, 139, 140]. However, several studies have highlighted the importance of studying bat immunology in the context of their potential role as reservoirs for viruses.

Basic immunological processes in bats appear to be very similar to those in other mammals, with comparable overall cell populations such as macrophages, B and T cells, and signalling cascades [141, 142, 143]. One of the examples depicting host-pathogen co-evolution shaping the bat immune system is the presence of positively selected mutations within the cell receptor associated with filovirus entry [Niemann-Pick C1 protein (NPC1)] for some bat species, making bat cells less susceptible to infection [144]. This suggests that some bats might show resistance to infection by filoviruses, resulting from a long-standing co-evolution shaping the bat NPC1 receptor and immune evasion mechanisms of the viruses [145, 146, 147]. However, despite bats being able to mount effective immune responses to various pathogens, such responses can sometimes be atypical regarding both innate and adaptive immune systems [148] **(Figure 1.6)**. Few studies have examined adaptive immune systems of bats, although there is evidence that the naive immunoglobulin repertoire may be larger in bats compared with other mammals [149].

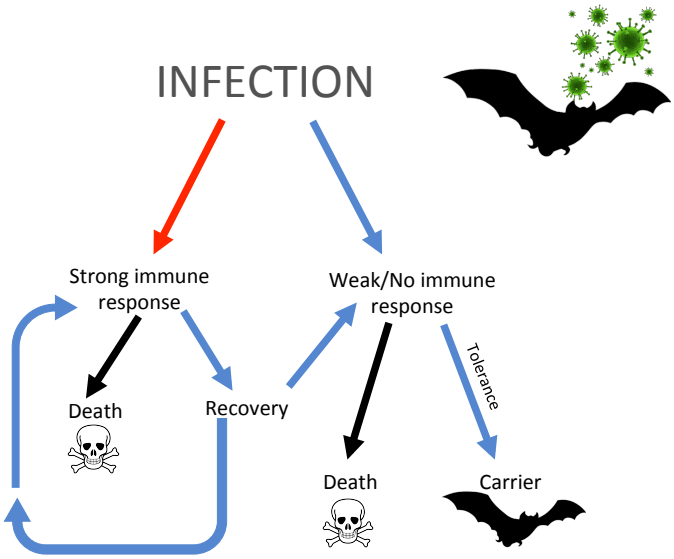


Figure 1.6: Tolerance to infection. Infection can lead to a strong immune response or a lack of it and a range of responses in-between. A strong response can lead to the death of the host or clearance of infection and recovery. If the host gets infected again with the same virus, it might develop a strong immune response again or tolerance to the infection.

The Toll-like receptors (TLRs) are a class of innate immune receptors that recognize a wide variety of pathogen-associated molecular patterns. The mammalian TLRs 3, 7, and 8 are activated by nucleic acid ligands and thus are often referred to as the viral-sensing TLRs, while TLR9 is not virus specific but still can detect viral pathogens [150]. A study comparing TLRs in eight different bat species with other mammalian TLRs revealed that despite being highly conserved [151, 152], bat TLRs 3, 7, 8, and 9 show unique changes within agonist binding sites, some of which involve strong positive selection predicted to alter nucleic acid binding motifs [153]. Missing components of the bat innate immune system include the PYHIN locus, which corresponds to a family of proteins recently characterized as innate immune sensors for microbial DNA that activate the type I IFN signalling cascade [154], or the loss of the killer cell immunoglobulin-like receptor and of the killer cell lectin-like receptor loci in some bats, suggesting that bat natural killer (NK) cells might use alternative receptors to recognize major histocompatibility complex (MHC) class I–presented peptides [77]. But it is still not known if the NK cells have the same effector functions in bats as in other mammals.

Similarly, the constitutive phosphorylation of the STAT1 gene (*Signal transducer and activator of transcription 1*) in some bat species supports the so-called always-on hypothesis, which states that the bat immune system is always turned on [140]. As STAT1 mediates cellular responses to IFN signalling pathways, it has been speculated that the constitutive activation state of this factor could limit viral replication [155].

Comparative studies between the bat and other mammalian innate immune systems reveal evidence that some bats may rely more on adaptive immunity in order to compensate for other atypical responses [137, 138, 140]. For example, a single point mutation in the STING gene (*Stimulator of Interferon response CGAMP interactor 1*) in some bat genomes results in a dampened production of IFN type III [156]. Other studies

have shown a stronger type II and type III IFN response in bat cell lines compared with human cells, suggesting a species-specific secretion profile of antiviral cytokines in the response against specific viruses [157]. Thus, bats may have evolved unique nucleic acid sensing and IFN response mechanisms [140].

Nonetheless, it is important to highlight that the differences observed in the immune response of individual bat species compared with other bats and other mammals do not necessarily reflect a hindered/altered immune response. A good example for this is the missing TLR8, which is disrupted by a retrotransposon insertion and preserved as a pseudogene in birds [158], or the TLR11 expressed in mice but not in humans [159]. In those examples, different organisms have developed compensatory immune responses that still render their immune systems fully functioning. Moreover, when studying bat immunology, the huge diversity and wide range of unique adaptations in *Chiropterans* should be considered. Bats are not a single entity; there are many different ecological and physiological factors (such as feeding ecology, generation time, roosting behaviour, and inter and intraspecies ecological interactions) that are likely to have differentially contributed to particular disease dynamics and a virus-tolerant phenotype observed for some bat species [90]. As the virus-tolerant phenotype is likely to result from intricate evolutionary dynamics, genome-wide comparative studies encompassing more bat species and contrasting other mammals are necessary for elucidating the complexity of a general bat immune system. Initiatives such as the Bat1K project [160], which aims to sequence the genomes of all extant bat species (∼1,300 species) to chromosome-level assembly, are expected to facilitate large-scale immune profiling comparative approaches. Finally, as *Chiropteran* immunology studies have focused on only a few bat species and/or use *in vitro* systems that cannot replicate the complexity of the immunological context, the development of new bat cell lines from more bat species and the use of established bat colonies is expected to greatly assist in the study of disease ecology, bat behaviour, and

bat immunology [161].

1.3.5 The role of endogenous retroviruses and endogenous viral elements in shaping bat immunity

While most EVEs have typically been considered genetic fossils with no apparent function, an increasing number of studies have shown that many of these may play a key role in some biological processes [68]. For example, the integration of EVEs is proposed to provide a source of structural variation for host genomes and their presence can act as foci for genomic rearrangements, which can have dramatic effects on the regulation of gene expression [26]. The integration of EVE sequences has been associated with the emergence of new host genes, used either to combat exogenous viruses or for the evolution of certain physiological processes, such as the development of the placenta in mammals [26, 162]. EVEs found in several species contain repetitive sequence clusters called Piwi-interacting RNA (piRNA), which can act as traps for sequences that will integrate into them, such as transposable elements and viruses [163]. Such clusters of piRNAs were found to be generated by EBL-NPs, which are also present in bats [105]. A functionally analogous system recognizing self-sequences from non self viruses is very well understood in prokaryotes (CRISPR-Cas9), and there are studies of EVE-derived piRNAs associated with viral recognition in insects [164, 165], although it is still poorly understood in mammals.

The long evolutionary relationship between bats, ancient viral infections, and EVEs might hold the secret (at least partially) to understanding the complex host-virus dynamics with some bat species. The bats have a relatively small genome in comparison with other mammals, but high viral activity and a vast number of ancient viral genomic

insertions observed in bats [24] may have shaped the differential capacity for endogenous viruses to colonize bat genomes [166]. Some of those integrations are likely to have resulted in highly active transposable elements with functional properties that are yet unknown [126, 167]. It is also plausible that some of the very ancient viral integrations predating the divergence of the *Yangochiroptera* and *Yinpterochiroptera* may have even shaped the early evolution of the bat immune system, but their function might have been subsequently lost and their mechanism of action may no longer be recognizable. However, a link between the number and diversity of extinct and extant viral infections has not yet been established.

It has been shown that ERV sequences (particularly the *gag* and *env* genes) can be part of an inheritable immunity in animals [44, 68], and ERV-mediated or ERV-targeted immunity is important to fight exogenous retroviruses in some mammals, as in the example of ERVs and immunity to the exogenous Jaagsiekte sheep retrovirus in sheep [44, 68]. Very few exogenous retroviruses have been described in bats so far [168, 169], thus it could be speculated that superinfection by exogenous retroviruses and consequent integration of multiple ERVs may have led to resistance to infection by retroviruses. Moreover, some full-length ERVs found in mammalian genomes are located close to immune genes or are found to be duplicated in close proximity to DNA transposable elements, thus possibly altering the expression of immune-related and other genes [170].

The process of co-option of EVEs for important roles in the evolution of immune systems is referred to as EVE-derived immunity (EDI) [171] **(Figure 1.7)**, and we hypothesize that EDI may have also played an important role in the evolution of the bat immune system. Some experimental models have implied a role for EVE-derived immunity in antibody development in animals necessary for maintaining an immune memory against active exogenous viruses [68, 102]. Examples for potential EVE-derived immunity in bats

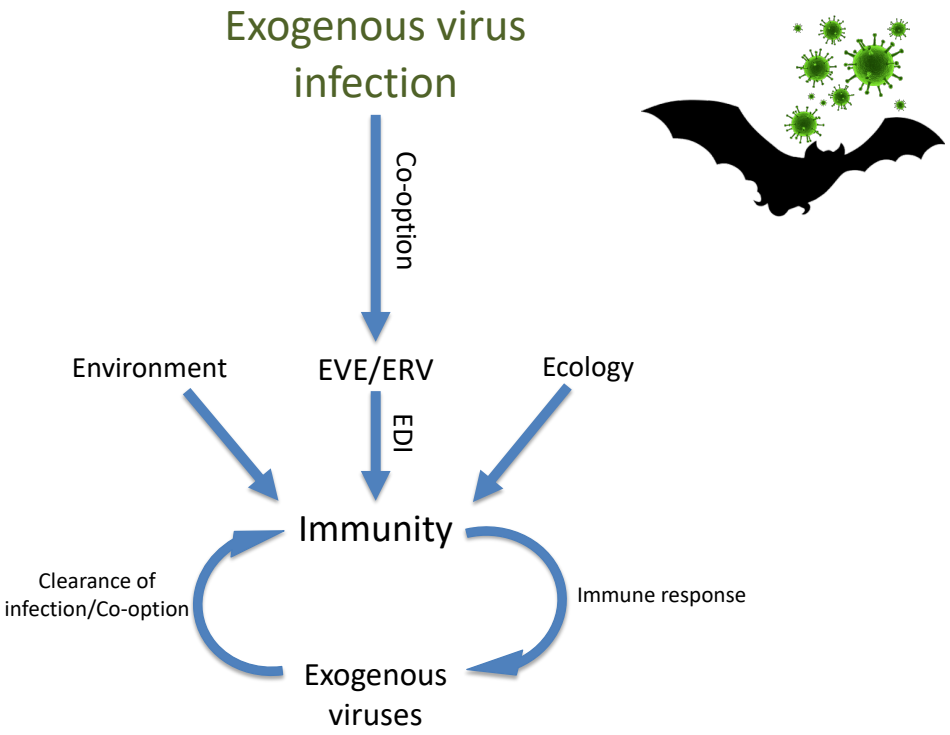


Figure 1.7: Immunity and infection. Exogenous virus infection can lead to co-option of the viral elements resulting in the presence of beneficial endogenous viral elements (EVEs) in the genome. Those elements can sometimes be used by the host in a host-virus arms race to evade infection and are called endogenous viral element-derived immunity (EDI). EDIs, environment, and ecological traits of the host all influence the host's immunity and its immune response to the infection with exogenous viruses, its clearance, or further co-option.

are the EBOV/MARV-like endogenous sequences homologous to the NP and VP35 genes [56]. For extant EBOV/MARV viruses, the co-expression of NP and VP35 is known to inhibit type I IFN production [172], suggesting that conserved NP and VP35-like sequences found in different bat genomes might have retained some functional properties [56], thus resulting in asymptomatic infections of filoviruses in bats. As is the case for EFLs, resistance to extant viral infections associated with EVE-derived immunity has also been hypothesized for EBLs and EPLs [68]. EVE-derived immunity to extant viral infections in bats is an appealing hypothesis, especially when considering the large diversity of EVEs found within bat genomes compared with other mammals, thus highlighting the importance of further studying bat EVEs within an immunological context. Bat genomes have been shown to contain up to four different viral families, which is among the most diverse complements of EVEs found in mammals to date, despite their small genome sizes [24, 94]. ERVs from all three classes were found in bats in numbers higher than in other mammals with the exception of the mouse (*Mus musculus*) genome. Moreover, bats have an unusual diversity of ERV sequences, with some resembling retroviral families previously not found in mammalian genomes [24].

Ancient viral infections may have also shaped the evolution of the host's antiviral genes through the virus-host arms race [14]. The best examples of this are the TRIM (Tripartite motif containing protein), the SAMHD1 (Sterile alpha-motif and histidine-aspartate domain-containing protein 1), and the APOBEC (Apolipoprotein B messenger RNA editing enzyme catalytic polypeptide-like) genes, which are thought to have evolved as restriction factors against retroviral infections, driven by a long-standing co-evolution [19, 21, 173]. It has been hypothesized that the low retroviral content found within the genome of the common vampire bat, *Desmodus rotundus*, (as compared with other bat species) is associated with the expansion of the TRIM5 antiretroviral gene family through natural selection [94]. The adaptation of the common vampire bat to sanguiv-

ory (and thus to a greater exposure to blood-borne viruses) might have played a role in the evolution of resistance to retroviral infection, limiting ERV genome colonization [94]. However, in contrast to blood-feeding bats, it has been shown that some fruit-feeding flying foxes (*Pteropodidae*) possess the largest and most diverse array of APOBEC3 genes identified to date, while several of these showed antiretroviral activity in vitro [174].

Expressed endogenous elements may also have an effect on the T- and B-cell repertoires with expression of EVE-encoded antigens leading to central or peripheral immunological tolerance [175, 176]. The consequence of such tolerance events could mean that either or both T- and B-cell responses against particular antigens of incoming viral infections (with epitopes that relate to the sequences in endogenous elements) may be muted or even absent. In terms of the incoming viral infection, this may lead to reduced clearance or increased persistence.

1.4 Conclusions

As suspected reservoirs for emerging viral diseases, and as a highly diverse order with unique adaptive traits contributing to ecosystem balance, bats should be considered as a research priority. Particular effort should be made in studying their immune system and empirically testing predictions such as those developed from their comparative biology, linking the role of bats as hosts for viruses through the genetic adaptations to their immune responses [177]. Moreover, exploring the possible effects of viral infections and EVE integration associated with a virus-tolerant phenotype in bats could shed some light on the disease ecology of viruses with zoonotic potential that bats carry. Studying bat ecology and evolution will also provide a better understanding of zoonotic virus-host dynamics, as well as the long-term co-evolutionary process between hosts and some

viruses.

The virus-tolerant phenotype observed in bat species is likely to result from a combination of order- and even species-specific ecological and behavioural factors, such as longevity, metabolic rate, and social interactions among communities [166, 178]. Therefore, the large diversity within the *Chiroptera* and thus their differential immune responses should be integrated through large-scale comparative studies in order to be able to model a generalized bat immune response. Further developments of *in vitro* and *in vivo* systems, such as new bat cell lines and the long-term non-lethal study of established bat colonies using new molecular techniques [179], will also aid in uncovering bat immune responses, hopefully limiting the use of extensive and invasive sampling.

EVEs have been shown to have important roles in host biology, where they can benefit their hosts in battling exogenous viral infections [171], and are essential factors acting in reproduction as demonstrated by the *syncytin* genes [38]. Novel roles of EVEs continue to be discovered, such as evidence that EVEs can act as a system analogous to CRISPR-Cas in prokaryotes, becoming templates to piRNAs, which can act in silencing of transposable elements [163]. We still may not know many of the functions that EVEs have in their host organisms, but the nature of bats being such potent carriers of diverse endogenous sequences might shed some light on the evolutionary relationships between mammals and their constituent EVEs.

In the following chapters we described our work exploring different aspects of millions of years of virus-bat co-evolution. In this chapter we discussed what we know so far about bat immune systems and immune responses, especially in the context of ancient viral integrations (**Chapter 1**). We followed that with a study identifying endogenous retroviruses with potential biological functions in a number of species (**Chapter 2**), and extending it further to other retroviral genera and bat representatives (**Chapter**

3). We also searched a number of bat genomes for the sequences of innate immune receptors which are the first line of defence against pathogens (**Chapter 4**). We ended this work with an *in-silico* study of animal SARS-CoV and SARS-CoV 2 entry receptors in a broader vertebrate context which also highlighted special place of bats among other animals (**Chapter 5**).

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

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

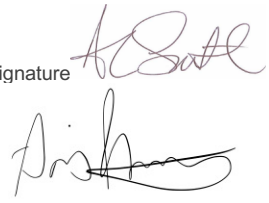
Title of Paper	The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses.
Publication Status	Published
Publication Details	Skirmuntt, Emilia Cecylia, Marina Escalera-Zamudio, Emma C. Teeling, Adrian Smith, and Aris Katzourakis. "The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses." Annual Review of Virology 7 (2020): 103-119

Student Confirmation

Student Name:	Emilia Cecylia Skirmuntt		
Contribution to the Paper	Wrote the manuscript and designed all of the figures apart from the Figure 1.1 which was provided by ECT.		
Signature Emilia Skirmuntt		Date	16.06.2021

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: Adrian Smith and Aris Katzourakis		
Supervisor comments		
Signature 	Date	17.06.2021

2 The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology

2.1 Introduction

Viral genes in the host genome arise when viruses become permanently integrated into a host cell's genome, either through an obligatory step in their replication process (as with retroviruses) or chance events such as interaction with cellular retroelements or illegitimate recombination with host genomic DNA [27]. If this integration occurs in germline cells, the virus can be transmitted vertically between generations in the form of an endogenous viral element (EVE) or in the case of retroviruses, endogenous retrovirus (ERV) [17, 27]. ERVs arising from ancient viral infections can persist and accumulate in the genomes of their hosts, constituting, for example, ∼8% of the genome in humans [28], around 10% in mice [180], and around 5% in bats [84].

The retroviral envelope (*env*) plays a crucial role in viral infection, controlling attachment to the receptors and cell entry [181], as well as host immune evasion [182, 183, 184].

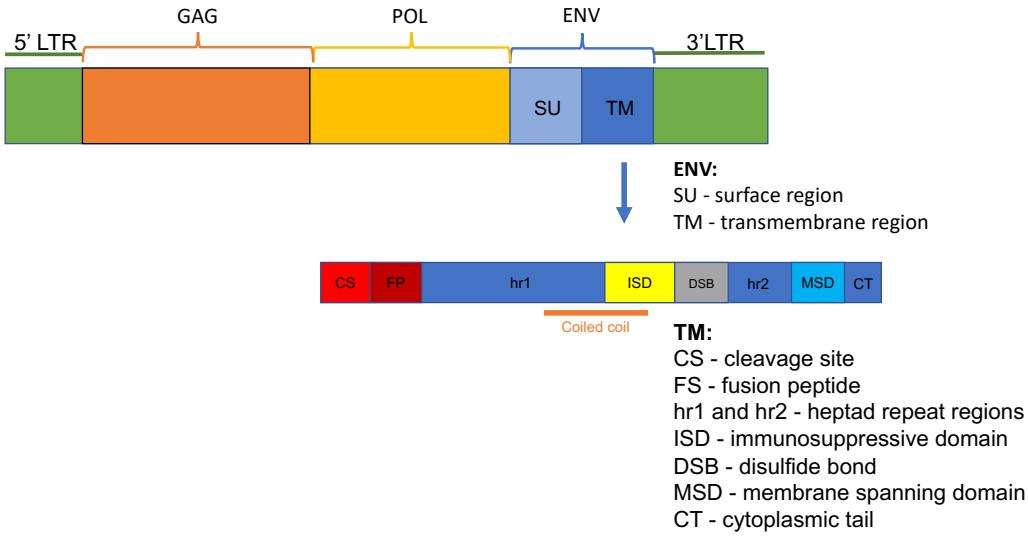


Figure 2.1: Structure of the retroviral genome and TM domain.

The protein it encodes contains two main domains: surface (SU) and transmembrane (TM) - the SU interacts with different receptors on the cell surface during the first part of the infection process. The TM domain regulates fusogenicity - fusion of viral and cellular membranes during viral entry [41, 185], and also contains an immunosuppressive region (ISD): 17–20 amino-acids which help retroviruses in evasion of the host immune system [186, 187, 188, 189, 190, 191, 192, 193] **(Figure 2.1)**.

Multiple retroviral *env* genes exist in vertebrate genomes in the form of ERVs, some of which have a demonstrable function in the biology of the host, notably in the context of EVE-derived immunity [171]. For example, sheep endogenous betaretroviruses both protect sheep against the Jaagsiekte sheep retrovirus through receptor interference, and are required in sheep conception and placental development [37, 44]. One of the most well studied co-opted ERVs is an *env* with a TM domain and an ISD region - *syncytin*. *Syncytins* are retroviral *env* glycoprotein genes independently captured and endogenized by different mammalian lineages [30, 49, 194, 195, 196]. They can influence myoblast

fusion and sexual dimorphism in some species [42], and there is a possibility that some of them might be immunosuppressive and prevent mother-foetus conflict, but formal evidence is still missing [31, 197, 198]. *Syncytins* have not been found yet in some of the orders of mammals like whales and dolphins (odd-toed ungulates), pigs (even-toed ungulates), elephants (*Proboscidean*), or bats (*Chiroptera*). Bats can be an interesting model for placentation and *syncytin* studies, because they have two types of placenta unevenly divided between both *Chiroptera* suborders. *Syncytins* are homologous to each other, but not necessarily orthologous [30], and types of placentation might differ according to the type of ancestral viral sequence.

Bats are the second largest order of mammals, representing about 20% of all classified species. *Chiroptera* can be divided into two sub-orders: *Yinpterachiroptera* and *Yangochiroptera*. Bats are also major reservoir for ∼100 different viruses, including viruses able to induce zoonotic diseases – infectious diseases of animals that can be transmitted to humans [62, 66, 199, 200, 201]. The mechanism for how bats can be asymptomatic hosts for diseases deadly to other species (e.g. Ebola and rabies) is however not yet fully understood. It has been suggested that the lack of symptoms of viral infections and low mortality rate in bats might be due to immune adaptations that result from their long-term co-evolution with viruses [56, 81, 98, 110, 202, 203, 204, 205, 206, 207, 208, 209, 210].

In this article, we describe the results of an extensive search for ERV envelopes, run on all 14 publicly available bat genomes (*Myotis lucifugus, Myotis brandtii, Myotis davidii, Miniopterus natalensis, Eptesicus fuscus, Pteropus alecto, Pteropus vampyrus, Rousettus aegyptiacus, Eidolon helvum, Megaderma lyra, Pteronotus parnelli, Hipposiderios armiger, Rhinolopus sinicus*, and *Rhinolopus ferrumequinum*). We looked for *envs* inside open reading frames (ORFs), searched for orthology between identified sequences,

performed selection analysis and reconstructed a phylogenetic tree for all the detected *envs* and reference sequences. To the best of our knowledge, a single comprehensive analysis of all of these genomes has not been previously conducted.

2.2 Materials and methods

2.2.1 Database search and sequence analysis

Retroviral endogenous envelope-like sequences were searched for by local BLAST [211] in 14 bat genomes (*Myotis lucifugus* [GCA_000147115.1], *Myotis brandtii* [GCA_000412655.1], *Myotis davidii* [GCA_000327345.1], *Miniopterus natalensis* [GCA_001595765.1], *Eptesicus fuscus* [GCA_000308155.1], *Pteropus alecto* [GCA_000325575.1], *Pteropus vampyrus* [GCA_000151845.2], *Roussetus aegyptiacus* [GCA_001466805.2], *Eidolon helvum* [GCA_000465285.1], *Megaderma lyra* [GCA_000465345.1], *Pteronotus parnelli* [GCA_000465405.1], *Hipposiderios armiger* [GCA_001890085.1], *Rhinolopus sinicus* [GCA_001888835.1] and *Rhinolopus ferrumequinum* [GCA_000465495.1]) **(Table A 7.1)** with 9 probes of known exogenous and endogenous retroviral envelope glycoproteins (**Table 2.1**).

Abbreviation	Name	Accession number
MuLV	Murine leukemia virus	AAV68489.1
KoRV	Koala retrovirus	YP_008169844.1
GaLV	Gibbon ape leukemia virus	AAC96086.1
BaEV	Baboon endogenous virus	BAP91049.1
HERV-W	Human endogenous virus W	AAF74215.1
HERV-K	Human endogenous virus K	CAB56604.1
MDEV	Mus dunni endogenous virus	AAC31806.1
PERV	Porcine endogenus virus	ADG27336.1
FeLV	Feline leukemia virus	NP_047256.1

Table 2.1: Probes of exogenous and endogenous retroviral envelope glycoproteins

From local BLAST search results, using Python (ver.2.7) custom scripts we selected sequences of length ≥100 amino acids (aa) with an e-value ≤0.09. We extracted contigs containing selected sequences from assembled genomes and searched them for open reading frames (ORF) ≥200 aa in length (from one stop codon to the next stop codon). Use of the Bedtools package [212] allowed us to select 1447 unique sequences from all bat genomes complying with the above requirements. We ran reciprocal BLAST against the chosen sequences and discarded all which did not return results for gammaretroviruses with the first hits. We manually inspected sequences and discarded ones which did not contain the highly conserved C-X-6-C-C motif or were duplicates. We used NCBI's Conserved Domain search tool [213] to determine the exact placement of the *env* sequences and what other viral genes and domains can be found in the closest genomic neighbourhood.

2.2.2 Sequence alignments and phylogenetic tree reconstruction

We manually aligned the 737 chosen *env* TM regions using the Aliview software [4]. Putative *env* sequences with accompanying genes were sorted into categories based on the presence of conserved TM domains such as a cleavage site, hydrophobic domain, and C-X-6-C-C motif. We aligned sequences according to the generated consensus sequence consisting of the most conserved motifs (78 aa) –R-X-K-R- [X_{-46-64}] -S-Q-T-L-V-Q-D-Q-V-D-S-L-S-E-V-V-L-Q-N-R-R-G-L-D-L-L-T-A-E-K-G-G-L-C-A-A-L-G-E-E-C-C-F-Y-V-N-Q-S-G-L-V-R-D-[X_{-27-46}]-I-L-P-F-L-G-P-L-I-V-L-L-M-I-L-L-F-G-P-C-I-L-[X_{-2}]-L-L-L-X-L-X-F-G-P-C-X-[X_{-17}]. In the final alignment, we included only those sequences, which had conserved motifs present at least in 50% (≥ 38 aa).

Abbreviation	Name	Accession number
BaEV	Baboon endogenous virus	NC_022517.1
BRTV	Bat gammaretrovirus	KF547869.1
envH3	Chimpanzee H3 envelope	AJ862649.1
HEMO	Human Endogenous MER34 ORF	XM019829500
HTLV	Human T-lymphotropic virus 1	AB600229
BLV	Bovine leukemia virus	M35242
FeLV	Feline leukemia virus	NC_001940.1
FERV	Feline endogenous virus	LC005748.1
GaLV	Gibbon ape leukemia virus	AF055063.1
HERV-H	Human endogenous virus H	AH007767.2
KoRV	Koala retrovirus	NC_021704.1
MDEV	Mus dunni endogenous virus	AF053745.1
MuLV	Murine Leukemia virus	AY277737.1
PercomORF	Stickleback endogenous envelope	AANH01003690.1 [37666-38871]
PERV	Porcine endogenus virus	GU980187.1
SRV-Y	Simian retrovirus Y	AB758125.1
Syn Mab-1	*Syncytin Mab-1*	MG254888
Syn Car-1	*Syncytin Car-1*	JN587095.1
Syn Opo-1	*Syncytin Opo-1*	NM_001308377.1
Syn Ory-1	*Syncytin Ory-1*	NM_001305593.1
Syn Rum-1	*Syncytin Rum-1*	JX412964.1
Syn Ten-1	*Syncytin Ten-1*	NM_001305586.1
Syn-1	*Syncytin 1*	DQ256474.1
Syn-2	*Syncytin 2*	KC010510.1
Syn-A	*Syncytin A*	NM_001013751.2
Syn-B	*Syncytin B*	AY849978.1

Table 2.2: List of the reference sequences

To reconstruct a phylogenetic tree of retroviral *env* sequences in 14 bat genomes, we added to the alignment sequences of the known mammalian *syncytins*, and a selection of the closest viral matches for both endogenous and exogenous viruses **(Table 2.2)**. We first ran Prottest [214] to determine which model to use with the RAxML (ver.8) package [6], with the best model being the JTT + G model under AICc scoring criteria. We used MrBayes software [215] with the SRD06 substitution model [216] for reconstructing a phylogenetic tree of all identified orthologous sets with reference sequences.

2.2.3 Orthologous sequence search and annotation

We selected all sequences ≥400 aa in length. Sequences long enough and containing the smallest number of gaps were further extended to generate flanking regions, which we subsequently ran in BLAST searches against the Refseq and WGS *Chiroptera* databases. We then compared the results to the accession numbers on the phylogenetic tree. From the bat genome databases, we extracted all the sequences marked as orthologous and manually annotated them in Geneious R10 software (ver. 10.1.3; http://www.geneious.com, ; [5]).

2.2.4 PAML analysis of orthologous sequences

We tested 31 sequence alignments with long envelope sequences using the *codeml* algorithm from the PAML: Phylogenetic Analysis by Maximum Likelihood package [217]. We ran several different models for each set of orthologs. The one-ratio test (M_0 – single ω ratio for whole tree) was run for all sequence sets. If all sequences from the set contained stop codons or frame shifts we analysed only part of the sequence starting from the 5′ end stop codon or frame shift to the next stop codon/frame shift, or the end of the sequence. In the case of sets that contained only two orthologs we ran M_0 in pairwise mode. To check for the possibility of neutrality of the whole sequence we ran M_0 with ω fixed to 1 and compared log-likelihood ratios (LRTs) for M_0 ω= 1 vs. M_0. In sets with three or more sequences we ran additional analyses for M_1 (separate dN/dS ratios for each branch of the tree) and M_2 (giving two or more dN/dS ratios to appointed branches). LRTs calculated for M_0 vs. M_1 and M_0 vs. M_2 showed that the one-ratio model is the best fit for all the sequence sets. For sets of three or more sequences we ran NSsite models - M_1: neutral (two-state, $\omega > 1$ disallowed), M_2: positive selection (two

states, $\omega > 1$ allowed), M_3: discrete, M_7: beta (beta distribution, $\omega > 1$ disallowed), M_8: beta& $\omega > 1$ (beta distribution, $\omega > 1$ allowed). We compared log-likelihood ratios for our data: M_1 vs. M_2, M_0 vs. M_3 and M_7 vs. M_8. The level of significance for LRTs was calculated using a $\sim \chi^2$ table, with twice the difference of log likelihood between the models (2 Δ lnL) compared to a $\sim \chi^2$ distribution. The number of degrees of freedom (df) corresponds to the difference in number of parameters between the models **(Table A 7.2)**.

2.3 Results

2.3.1 Search for gammaretroviral *envs*

Through local BLAST searches with 9 retroviral envelope glycoprotein probes (**Table 2.1**), we initially identified 51,245 potential envelope sequences in the 14 bat genomes. We selected sequences with an *env* length of ≥400 aa, which were in ORFs. We manually checked sequences for the presence of an ISD and ran reciprocal BLAST on them to attest if the sequence is of gammaretroviral origin, choosing all the sequences with a gammaretrovirus as first hit in the BLAST search. Sequences with long flanking regions (∼500 base pairs on both ends) were extended and run in BLAST against the Refseq and WGS *Chiroptera* database to find orthologs. The quality of the data in some contigs prevented us from using some portions of the chosen sequences because of the presence of large gaps. Low quality data with too many gaps or missing flanking regions means that a number of orthologous sequences likely remain undetected. In general, species with the highest number of identified sequences were also the genomes with the highest data quality and the lowest number of gaps (*P. vamprus, M. lucifugus, H. armiger, R. aegyptiacus*), species with the fewest number of sequences in the global alignment were

from the assemblies with the lowest data quality (*P. parnelli, E. helvum, M. natalensis*) (**Figure 2.2** and **Table A 7.1**).

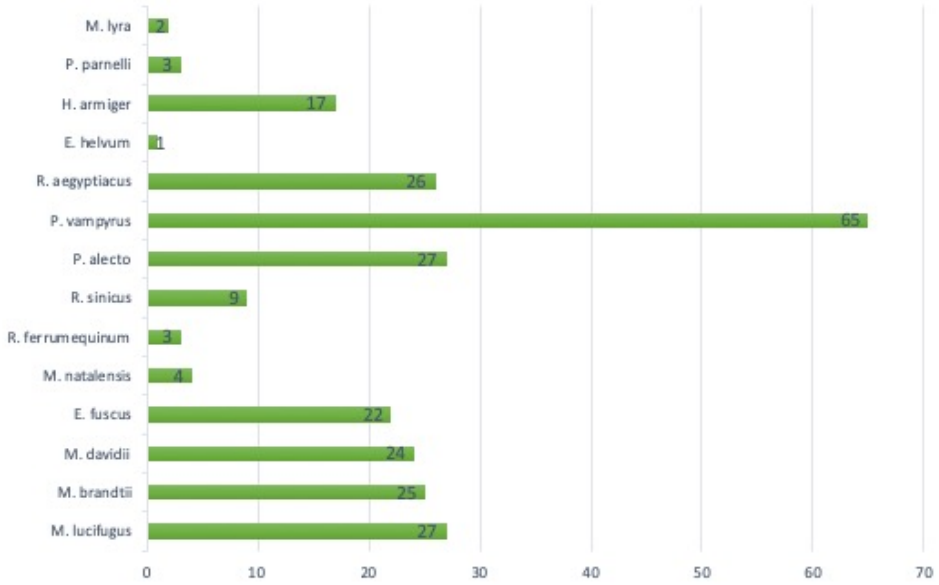


Figure 2.2: Number of envelope sequences in each bat species with *env* of length ≥400 aa

2.3.2 Retroviral envelope phylogenetic tree reconstruction

All the sequences found by us shared several features of gammaretroviral *env*, such as putative furin/PACE cleavage sites between SU and TM domains, hydrophobic domains (involved in fusogenic activity of *env* –anchoring it to the plasma cell membrane), and the highly conserved disulphide bond (C-X-6 -C-C) characteristic of gammaretroviruses [186, 218].

Despite poor data quality in some genomes, we reconstructed a phylogenetic tree of bat gammaretroviral envelope sequences, and based on that we searched and found

several examples of orthologous sequences (**Figure 2.3**). On the reconstruction of the phylogenetic tree most of the branches contained clusters of sequences from the same bat species or same families.

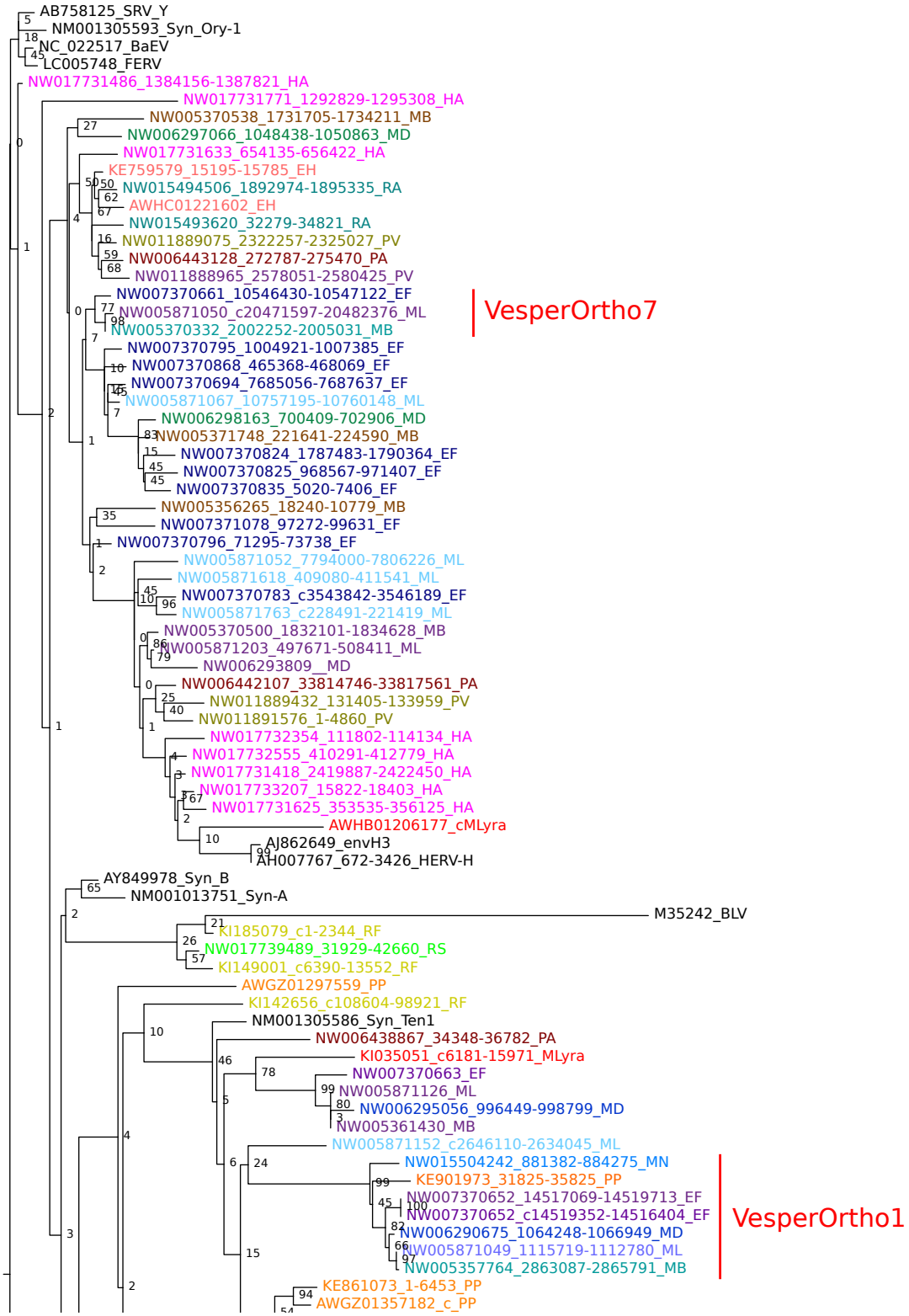
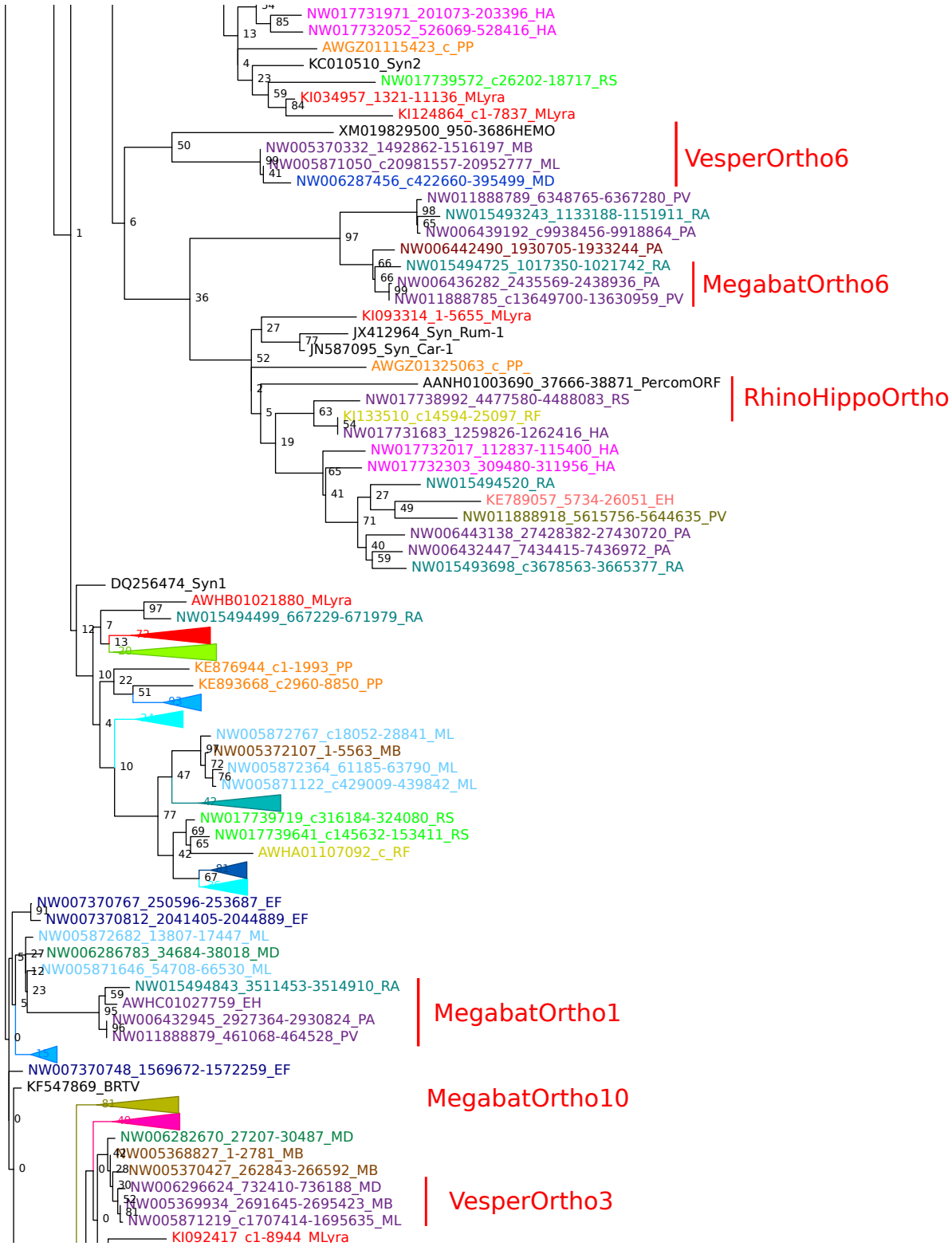
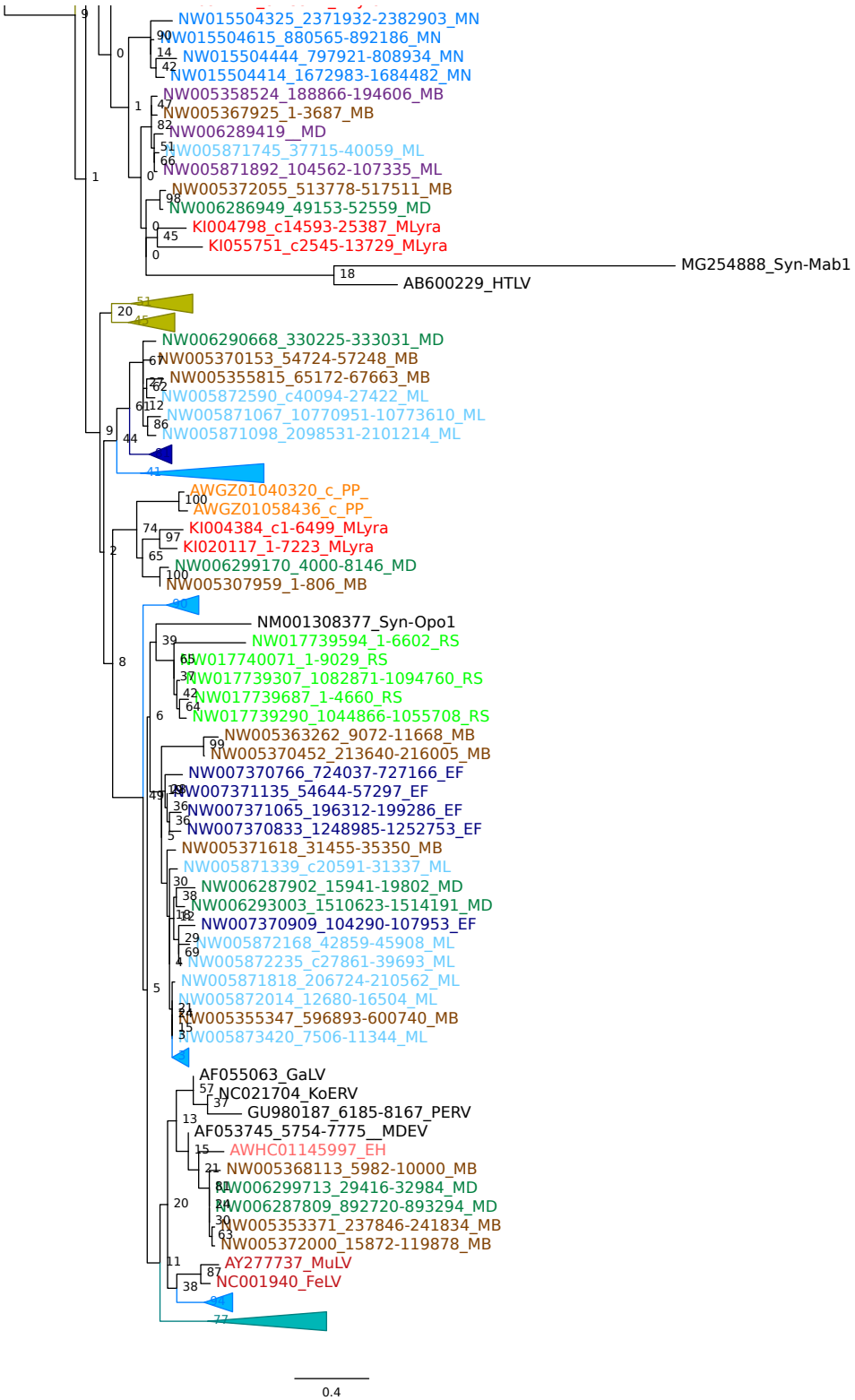


Figure 2.3: Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats (marked with different colours). Branches containing single species or large clusters of single family sequences collapsed for clarity. Continues on next two pages (list of all the sequences in **Table A 7.3**).


NW017731971_201073-203396_HA
NW017732052_526069-528416_HA
AWGZ01115423_c_PP
KC010510_Syn2
NW017739572_c26202-18717_RS
KI034957_1321-11136_MLyra
KI124864_c1-7837_MLyra
XM019829500_950-3686HEMO
NW005370332_1492862-1516197_MB
NW005871050_c20981557-20952777_ML
NW006287456_c422660-395499_MD
VesperOrtho6
NW011888789_6348765-6367280_PV
NW015493243_1133188-1151911_RA
NW006439192_c9938456-9918864_PA
NW006442490_1930705-1933244_PA
NW015494725_1017350-1021742_RA
NW006436282_2435569-2438936_PA
NW011888785_c13649700-13630959_PV
MegabatOrtho6
KI093314_1-5655_MLyra
JX412964_Syn_Rum-1
JN587095_Syn_Car-1
AWGZ01325063_c_PP_
AANH01003690_37666-38871_PercomORF
NW017738992_4477580-4488083_RS
KI133510_c14594-25097_RF
NW017731683_1259826-1262416_HA
NW017732017_112837-115400_HA
NW017732303_309480-311956_HA
NW015494520_RA
KE789057_5734-26051_EH
NW011888918_5615756-5644635_PV
NW006443138_27428382-27430720_PA
NW006432447_7434415-7436972_PA
NW015493698_c3678563-3665377_RA
RhinoHippoOrtho
DQ256474_Syn1
AWHB01021880_MLyra
NW015494499_667229-671979_RA
KE876944_c1-1993_PP
KE893668_c2960-8850_PP
NW005872767_c18052-28841_ML
NW005372107_1-5563_MB
NW005872364_61185-63790_ML
NW005871122_c429009-439842_ML
NW017739719_c316184-324080_RS
NW017739641_c145632-153411_RS
AWHA01107092_c_RF
NW007370767_250596-253687_EF
NW007370812_2041405-2044889_EF
NW005872682_13807-17447_ML
NW006286783_34684-38018_MD
NW005871646_54708-66530_ML
NW015494843_3511453-3514910_RA
AWHC01027759_EH
NW006432945_2927364-2930824_PA
NW011888879_461068-464528_PV
MegabatOrtho1
NW007370748_1569672-1572259_EF
KF547869_BRTV
MegabatOrtho10
NW006282670_27207-30487_MD
NW005368827_1-2781_MB
NW005370427_262843-266592_MB
NW006296624_732410-736188_MD
NW005369934_2691645-2695423_MB
NW005871219_c1707414-1695635_ML
KI092417_c1-8944_MLyra
VesperOrtho3


NW015504325_2371932-2382903_MN
NW015504615_880565-892186_MN
NW015504444_797921-808934_MN
NW015504414_1672983-1684482_MN
NW005358524_188866-194606_MB
NW005367925_1-3687_MB
NW006289419__MD
NW005871745_37715-40059_ML
NW005871892_104562-107335_ML
NW005372055_513778-517511_MB
NW006286949_49153-52559_MD
KI004798_c14593-25387_MLyra
KI055751_c2545-13729_MLyra
MG254888_Syn-Mab1
AB600229_HTLV
NW006290668_330225-333031_MD
NW005370153_54724-57248_MB
NW005355815_65172-67663_MB
NW005872590_c40094-27422_ML
NW005871067_10770951-10773610_ML
NW005871098_2098531-2101214_ML
AWGZ01040320_c_PP_
AWGZ01058436_c_PP_
KI004384_c1-6499_MLyra
KI020117_1-7223_MLyra
NW006299170_4000-8146_MD
NW005307959_1-806_MB
NM001308377_Syn-Opo1
NW017739594_1-6602_RS
NW017740071_1-9029_RS
NW017739307_1082871-1094760_RS
NW017739687_1-4660_RS
NW017739290_1044866-1055708_RS
NW005363262_9072-11668_MB
NW005370452_213640-216005_MB
NW007370766_724037-727166_EF
NW007371135_54644-57297_EF
NW007371065_196312-199286_EF
NW007370833_1248985-1252753_EF
NW005371618_31455-35350_MB
NW005871339_c20591-31337_ML
NW006287902_15941-19802_MD
NW006293003_1510623-1514191_MD
NW007370909_104290-107953_EF
NW005872168_42859-45908_ML
NW005872235_c27861-39693_ML
NW005871818_206724-210562_ML
NW005872014_12680-16504_ML
NW005355347_596893-600740_MB
NW005873420_7506-11344_ML
AF055063_GaLV
NC021704_KoERV
GU980187_6185-8167_PERV
AF053745_5754-7775__MDEV
AWHC01145997_EH
NW005368113_5982-10000_MB
NW006299713_29416-32984_MD
NW006287809_892720-893294_MD
NW005353371_237846-241834_MB
NW005372000_15872-119878_MB
AY277737_MuLV
NC001940_FeLV
0.4

2.3.3 Orthologous sequence identification

From all of the identified envelope sequences containing ISDs, we were able to identify 34 orthologous sets, although with 3 sets containing TM too short for further analysis. We did not find any orthologous sequences shared between *Yinpterachiroptera* and *Yangochiroptera*, but that may be due to the poor data quality. The split between *Yinpterachiroptera* and *Yangochiroptera* is dated to ∼58.9 Mya, and divisions between families within the two suborders occurred around 54 and 53.8 Mya respectively. Based on our results, we assume none of the orthologs found by us are older than the splits between suborder families. Most identified sequences were in the vesper bat species or the *Pteropodidae* species and were not coding competent with a no significant evidence of selection. These sequences date back no later than to the vesper family split (∼15.5 Mya), and the split between *Pteropodidae* species (∼22.6 Mya). We detected two pairs of sets that are shared between families, one with *Rhinolophidae* - *Hipposideridae* sequences (∼36.9 Mya), and one which was found in representatives of three *Yangochiroptera* families- *Vespertilionidae, Mormoopidae* and *Miniopteridae*, which sets the suspected date of endogenization no later than the suborder family division (∼53.8 Mya) (**Figure 2.4**) [2].

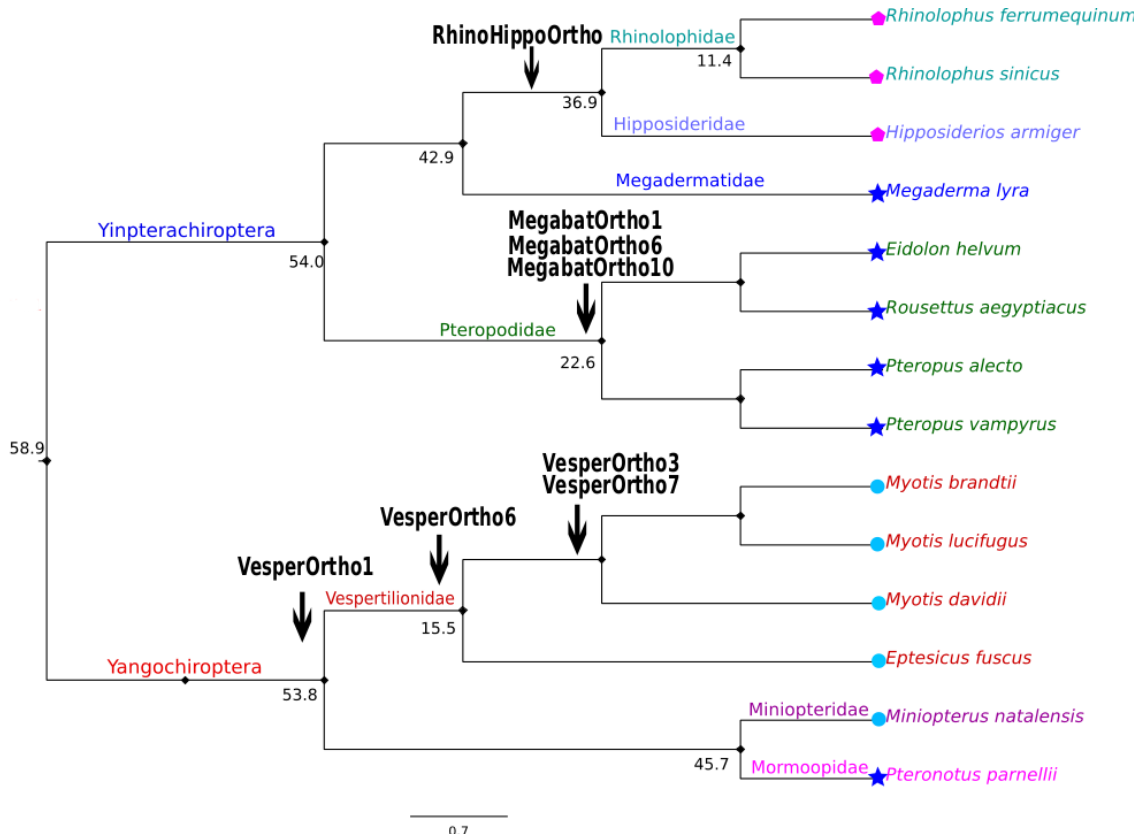


Figure 2.4: Simplified cladogram of the phylogeny of analysed bat species, with estimated dates (numbers at nodes) of family splits [2] and types of placentation in each of the species. Dates are all in millions of years ago (Mya). All analysed bats have discoidal type placentation. Bats marked with dark blue dots have mesometrial haemochorial type placentation, light blue dots indicate antimesometrial haemochorial type placentation, and pink dots indicate mesometrial endotheliochorial type placentation.

We also reconstructed a phylogenetic tree that shows all identified true orthologs along with reference sequences (**Figure 2.5**). All sets of orthologs were marked with different colours to distinguish them from each other. Some of the sequences do not cluster with their orthologs due to too short sequences in the alignment (seen in sets MegabatOrtho7, 11, 15, 17, 19) but further analysis on the rest of the coding sequences and flanking regions confirmed that those are still orthologous. The overall architecture of both trees and placement of orthologous sequences in relation to reference sequences correspond

broadly with each other.

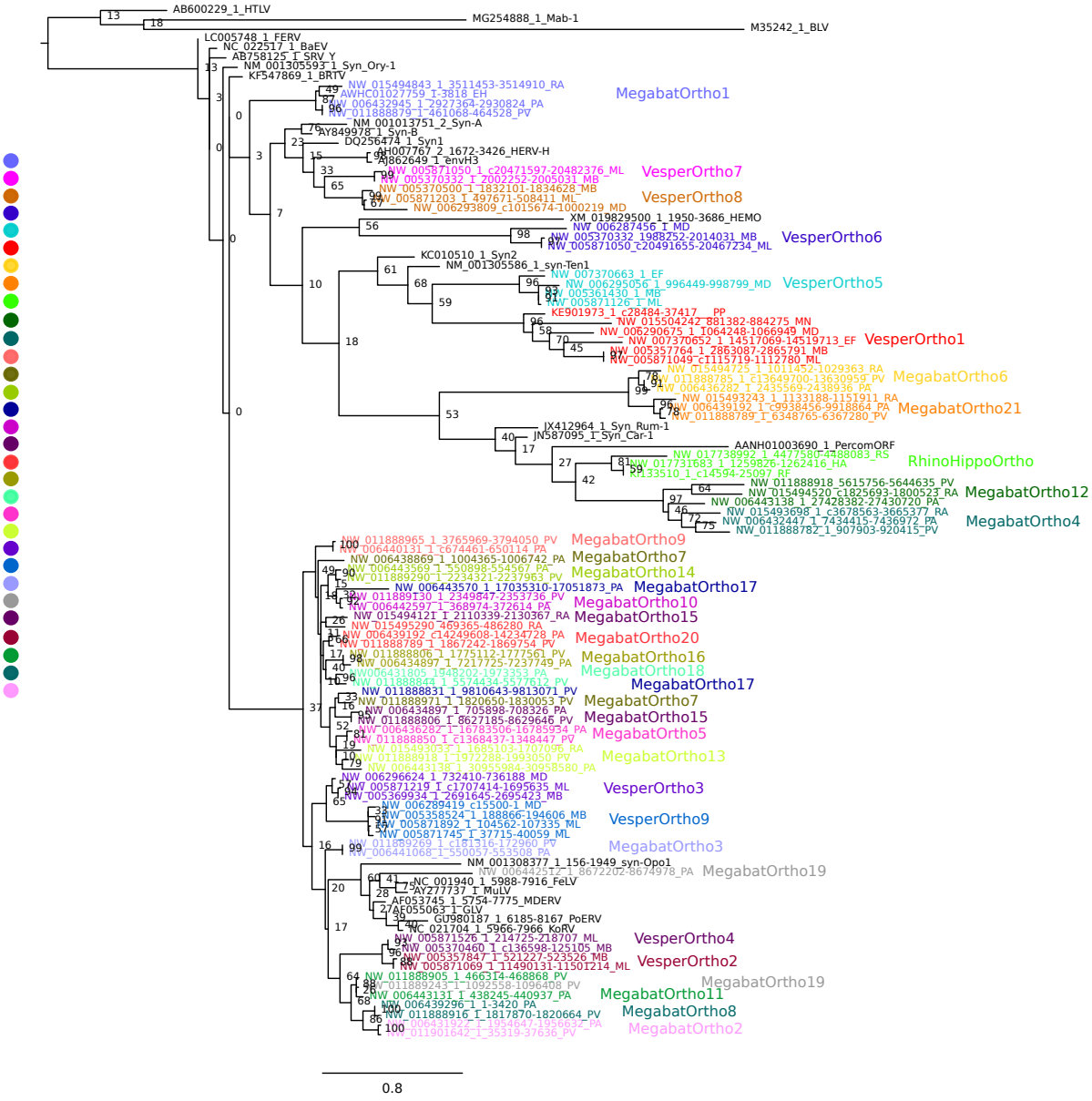


Figure 2.5: Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats and reference sequences. *Syncytin* sequences and other reference sequences are marked with black colour. Orthologous sequences are marked with one of 31 colours – numbers in the legend correspond to the positions on the tree.

RhinoHippoOrtho along with MegabatOrtho4 and 12 all contain three sequences and cluster together with ruminant and carnivore *syncytins* as well as *syncytin*–like sequences

found in spiny-rayed fish. The only other *Yinpterachiroptera* sequence clustering together with reference sequences is MegabatOrtho19 where *P. alecto* branch with reference sequences of exogenous viruses. This is probably caused by the one of orthologs being too short within the fragment used for an alignment to cluster together with its true orthologous sequence. VesperOrtho1 and VesperOrtho5 are the largest sets of orthologs we found, with VesperOrtho1 containing sequences from three different bat families: *Vespertilionidae, Miniopteridae* and *Mormoopidae*. They both cluster together but also with tenrec *syncytin* and primate *syncytin 2*. VesperOrtho6 contains 3 vesper bat sequences and lays on the tree close to newly found *syncytin*-like sequence expressed in human placenta - HEMO. VesperOrtho7 (2 sequences) and VesperOrtho8 (3 sequences) branch together with endogenous viral sequences HERV-H and envH3.

2.3.4 dN/dS analysis for orthologous sequences

dN/dS analysis (ω ratio) is a calculation of the rate of non-synonymous substitutions (alteration of the amino acid sequence) to synonymous substitution (changes in genetic code which do not alter the translated amino acid) in a sequence. It gives an indication of how strong the effect of natural selection is on a given protein. A higher number of non-synonymous mutations than synonymous ones suggests that the sequence is under positive selection ($\omega > 1$), while a higher number of synonymous mutations than non-synonymous ones indicates negative purifying selection ($\omega < 1$), while $\omega = 1$ is consistent with neutral evolution. Non-synonymous changes can alter the function of the gene, and are more likely to be deleterious, hence in coding sequences their number tends to be lower than synonymous ones. Functionally conserved sequences usually will have a higher number of synonymous mutations than non-synonymous ones. In our orthologs, we found representatives for all three types of ω ratio (**Table 2.3**). We also determined

statistically whether the whole sequence is consistent with neutrality. To do that we calculated the likelihood ratio test (LRT), a statistical test which compares the fit of two statistical models, for the one-ratio model (null model) and the one-ratio model with ω fixed to 1 (alternative model). That allowed us to reject neutrality in some of our sequences, with the degrees of freedom equal to 1 for all the tests.

We tested 31 ortholog sequence alignments using the PAML: Phylogenetic Analysis by Maximum Likelihood package [217]. PAML analyses are highly accurate in detecting adaptive molecular evolution, as well as positive selection acting on specific sequence sites [219, 220]. We ran a one-ratio model (M_0) test for all ortholog sets – a strict model that allows a single dN/dS ratio for all branches. According to this model, the majority of the tested sequences showed evidence of purifying selection. Purifying selection works by maintaining the function of the gene, meaning that most coding genes will have a dN/dS below 1. The rest of the tested sets show evidence of positive selection – indicating continuous adaptive evolution, and some show evidence of neutral evolution.

The likelihood-ratio test for neutrality of the whole sequence allowed us to reject neutrality in 6 out of 31 sequence sets (VesperOrtho1 and 9, MegabatOrtho3, 6, 19, and Rhino&HippoOrtho). All these alignments show evidence of purifying selection.

For all ortholog sets with the number of sequences greater than 2 (15 sets) we ran additional site models (M_0, M_1 and M_7) and compared neutral models to the selection ones (M_2 and M_8). We also ran M3 - a discrete model, which allows two classes of sites to have dN/dS > 1. From these we compared one-ratio model (M_0) with the discrete model (M_3), neutral model (M_1) with positive selection model (M_2), and beta model (M_7) with beta&$\omega > 1$ (M_8). All values and results are listed in the Appendix A - **Table A 7.2**.

Ortholog set	n = #	dN/dS (one-ratio)	Sequence size (aa)	2Δ lnL (for df=1 and p-value=0.05)
MegabatOrtho1	3	0.7799	550*	2.23
MegabatOrtho2	2	0.8264	107	0.040764
MegabatOrtho3	2	0.4561§	561	7.23193†
MegabatOrtho4	3	1.1498	186	0.278162
MegabatOrtho5	2	0.3708§	102	1.326184
MegabatOrtho6	3	0.5345§	626*	10.974878†
MegabatOrtho7	2	1.0502	125	0.011084
MegabatOrtho8	2	2.4412	231	0.374048
MegabatOrtho9	2	2.7831	185**	1.135522
MegabatOrtho10	2	4.5747	517*	3.27657
MegabatOrtho11	2	0.6593	109**	0.302754
MegabatOrtho12	3	1.0074	161	0.000512
MegabatOrtho13	3	1.2891	202	0.53744
MegabatOrtho14	2	0.9991	476***	0.000008
MegabatOrtho15	3	0.6208	170	2.04
MegabatOrtho16	2	1.7135	156	0.706314
MegabatOrtho17	2	0.5471§	190**	2.84
MegabatOrtho18	2	1.9117	113	0.389932
MegabatOrtho19	2	0.3652§	434	8.32724†
MegabatOrtho20	3	0.8718	183**	0.179014
MegabatOrtho21	3	2.6365	117	2.289654
VesperOrtho1	6	0.4790§	480*	61.389886†
VesperOrtho2	2	2.1077	82	0.49425
VesperOrtho3	3	0.9698	594*	0.015048
VesperOrtho4	2	0.7137	452***	0.845898
VesperOrtho5	4	0.7001	415***	2.04
VesperOrtho6	3	0.9430	138	0.11324
VesperOrtho7	3	0.4374§	538*	3.22
VesperOrtho8	3	1.2085	246	0.23843
VesperOrtho9	3	0.4714§	182	4.93258†
Rhino&HippoOrtho	3	0.5920§	405	11.009006†

Table 2.3: Ortholog sets with dN/dS results of one-ratio model. * full envelope from start to stop codon, ** only TM domain analysed, *** stop codon inside TM domain, § dN/dS value significantly below 1.0, †significant result.

2.3.5 Selected ortholog sequence analysis

We selected ortholog sets which contained full *env* (from start to stop codon) or were branching closely to reference sequences. This gave us six sets: VesperOrtho1, 3, 7 and MegabatOrtho1, 6 and 10. On further analysis, we also selected VesperOrtho6 which branches closely to the HEMO reference sequence and RhinoHippoOrtho which contains representatives of the two families. Six alignments were annotated using Geneious (ver. 10.1.3) [5]. All of these sequences contain all canonical motifs found in gammaretroviral *env* proteins: a hydrophobic signal peptide in SU region, furin cleavage site between SU and TM domains (R/K-X-R/K-R), fusion peptide, immunosuppressive domain, C-X-6-C-C motif, and transmembrane domain.

VesperOrtho1 is a set of six orthologs from *M. lucifugus, M. davidii, M. brandtii, E. fuscus, M. natalensis* and *P. parnelli* bats, spanning three *Yangochiroptera* families (*Vespertilionidae, Miniopteridae, Mormoopidae*). The *env* alignment is 480 aa in length and 57.2% of the positions in these sequences are identical to each other (**Figure A 7.1**). We observed the highest similarities between vesper bat species, with most notable aa changes in *P. parnelli* and *M. natalensis* in both of the canonical motifs and the rest of the sequence. *M. natalensis* has a 5 aa gap (position 269–274) before the fusion peptide and lacking threonine/methionine on position 409, *M. davidii* lacks glutamine on position 364. There are no frameshifts or stop codons in the middle of the *env* sequence, but *P. parnelli* does not contain a stop codon at the end of the coding sequence, instead having a lysine. In close genomic proximity to the sequence (∼300 aa) we found a fragmented *pol* region - with reverse transcriptase, RNase and integrase in *M. brandtii*, reverse transcriptase and RNase in *M. natalensis*, and RNase in *M. lucifugus, M. davidii* and *E. fuscus*. In *P. parnelli* we detected only partial RNase but that is probably due to the genomic scaffold being truncated on one end. We could not detect LTRs on both ends of the sequences although this might be caused by poor data quality in some of the sequences. This *env* is under purifying selection (**Table A 7.2**).

VesperOrtho3 is an ortholog containing 3 sequences from the *Myotis* genus: *M. brandtii, M. davidii* and *M. lucifugus* (**Figure A 7.2**). The alignment contains 90.5% aa that are identical and contains all canonical *env* features with no frame shifts or stop codons and only single amino acid changes throughout the 593 aa length of the sequences. We detected the presence of Gag and Pol proteins as well as a zinc finger in close proximity to *env*. The dN/dS for this gene is 0.9698 which suggests neutral evolution, and we could not reject neutrality statistically (**Table A 7.2**). All three sequences are flanked by LTRs. The dN/dS score and the presence of complete retroviral proteins between LTRs is consistent with neutrality.

VesperOrtho7 is a set of two *Myotis* bat sequences (*M. brandtii* and *M.lucifugus*), which are almost identical to each other (98.1%) with only 10 single aa substitutions and no frame shifts or stop codons (**Figure A 7.4**). The full *env* (537 aa) contains all canonical motifs and in its closest proximity we detected fragmented Pol and Gag proteins, containing frame shifts and stop codons. All glycoproteins are flanked by LTRs. The dN/dS for those sequences is 0.4374, which implies purifying selection. However the neutrality test for the whole gene was inconclusive with a borderline result (**Table A 7.2**). On the 3′ flanking region, ∼1000 aa from the *env* sequences, we detected the presence of the transposable element.

MegabatOrtho1 contains full *env* sequences (550 aa) from 4 megabats: *P. alecto, P. vampyrus*, and *R. aegyptiacus* and *E. helvum* (**Figure A 7.5**). These sequences are 81.6% identical to each other. All *env* canonical motifs are present, and there are no frameshifts or stop codons in the alignment. We observed a gap in position 192 instead of arginine in *E. helvum* and *R. aegyptiacus*, and the biggest differences in aa changes in these two bats. All bats have Gag and Pol sequences present and LTRs on both ends of the viral protein sequences, apart from *E. helvum* due to a truncated contig at the 5′ end, which contains only a partial *pol*. According to dN/dS analysis the sequence seems to be under purifying selection, as we could reject the possibility of neutrality for the whole gene (**Table A 7.2**).

MegabatOrtho6 is an alignment containing full *env* sequences of three *Yinpterachiroptera* bats: *P. alecto, P. vampyrus*, and *R. aegyptiacus*, from the family *Pteropodidae* (**Figure A 7.6**). The *env* sequences are 625 aa long, contain all canonical *env* regions, and show 89.1% identity to each other. We observed just minor single aa changes across the whole sequence, including in the region of fusion peptide, ISD and TM region. These were mostly seen in *R. aegyptiacus*. There are no frame shifts or stop codons across the whole

coding *env*. Apart from an envelope length in the range of ∼500 aa, we found fragmented Gag and Pol glycoproteins, as well as short LTR on both ends of all viral glycoproteins. dN/dS shows purifying selection with neutrality being rejected (**Table A 7.2**).

MegabatOrtho10 is another ortholog set containing only 2 sequences (*P. alecto* and *P. vampyrus*) (**Figure A 7.7**). A possible insertion between the same flanking regions was detected also in *R. aegyptiacus* but containing only the Pol glycoprotein. In both sequences containing a full *env* sequence there is also fragmented and frame shifted *gag* and *pol*, along with a zinc finger domain. All viral proteins are between long but fragmented LTRs. *Env* is 516 aa long, 97.9% identical for both sequences with only 11 single aa substitutions. One of the substitutions is present in the C-X-6-C-C motif of ISD, where in *P. vampyrus* one of the cysteines is substituted by tyrosine. The sequence is under positive selection according to dN/dS result, although the neutrality test for the whole sequence is inconclusive with a borderline value (**Table A 7.2**).

We identified vesper ortholog sequences branching closely with the HEMO reference sequence, which is an endogenous retroviral envelope protein seen in the blood of pregnant women, pluripotent stem cells, and tumours, but with unusual immunosuppressive domain not seen in other retroviruses [221]. Assigned as ortholog set VesperOrtho6, it contains 3 sequences from all 3 available *Myotis* bat genomes. We found a partial sequence in *E. fuscus*, but it is too fragmented to align with the others, containing large strings of undefined nucleotides. Close to the *env* sequence there is also a fragmented partial *pol* gene. The *env* sequence itself can be identified in *M. brandtii* and *M. davidii*, but it is very short (138 aa) and it contains only the 3′ end part of the TM domain with an ISD. The *M. lucifugus* genomic region is too fragmented to align with other *env* sequences but it has a *pol* region. All sequences are flanked by LTRs. As a reference in our alignment we used cat HEMO sequence ISD (LQN-HHLDLSAAQQGRT), as their

HEMO is most closely related to chiropteran HEMO [221]. The bat sequence branching with the reference showed similarities both in aa comparison (**Figure A 7.3**) with 8 aa same as in the 17 aa in cat HEMO ISD, and they grouped together in a phylogenetic tree reconstruction from the nucleotide alignment.

RhinoHippoOrtho is one of two ortholog sets containing sequences from two different bat families (*Hipposideridae* and *Rhinolophidae*). *Env* sequences from three bats are 405 aa long but end prematurely on the 5′ end. It contains a stop codon in *H.armiger* on the 5′end and 3′ end in *R. sinicus* and *R. ferrumequinum*. Fusion peptide, ISD and transmembrane region are present in all three sequences. Despite a short sequence dN/dS is low suggesting purifying selection, with neutrality rejected for the whole sequence (**Table A 7.2**). Close to the Env we also detected a partial Pol glycoprotein, with both being flanked by LTRs. Low dN/dS with only partial *env* sequence recovered from the data might be a result of this sequence having previously been functional, with the function lost over time, and remnants of the sequence left as a viral fossil.

The chosen ortholog sets present two different forms of immunosuppressive domains (**Figures A 7.1** to **7.7**). MegabatOrtho1, VesperOrtho 3 and 7 show variants with high levels of similarity to the 17 amino acids of the consensus immunosuppressive domain (e.g. FELV, PERV, MULV) with all of the aforementioned ISDs similar to the consensus in 12–13 aa [40]. VesperOrtho1 and MegabatOrtho6 contain immunosuppressive domains that are more similar to those seen in squirrels and woodchucks (Env-Mar1) [222] or HERV4-1 or ERV-3 [40, 41, 223] with levels of similarity going up to 16/17 aa when comparing MegabatOrtho6/Env-Mar1 and to 11/17 aa when comparing VesperOrtho1/Env-Mar1 ISD. Both kinds of ISD can be seen in species with endotheliochorial (e.g. cats, dogs) [194, 224] and haemochorial placentation (e.g. humans, mice, rabbits, rats) [35, 47, 225, 226, 227].

2.4 Discussion

Despite poor data quality in the case of some of the analysed genomes our study shows that bat genomes contain at least 32 sets of viral derived sequences of unknown function that could potentially be *syncytins* and/or EDIs. Large, intact *envs* in open reading frames were found clustering closely on a phylogenetic tree with *syncytin* sequences of other species, which might indicate that such sequences are good candidates for further bat *syncytin*/EDI research.

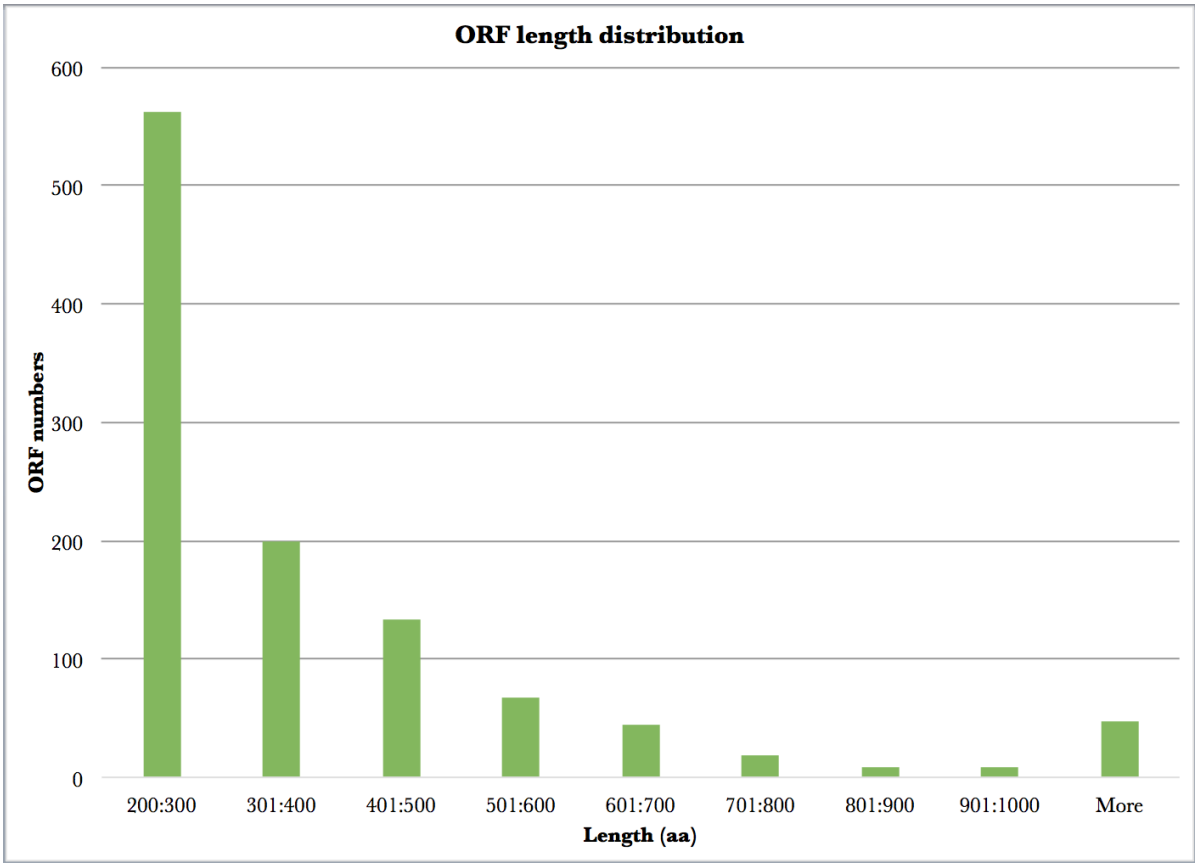


Figure 2.6: ORF length distribution amongst bats

We detected 1447 long ORFs overlapping BLAST matches to *env* sequences (**Figure 2.2**). Most of these are probably a result of relatively recent retroviral integrations,

but could also represent preserved ancient coding sequences. After entering the host genome through vertical transfer, endogenous viruses start evolving at the host rate of mutation rather than that of their viral progenitors [25]. This leads to two destinies for such sequences - accumulation of mutations, with partial loss of the coding sequence and a slow 'decay' if its function is harmful or non-advantageous for the host's survival; or fixation and preservation in genomes of the host population if the sequence provides a selective advantage for the host [26]. It is unclear what proportion of ERVs seen today might have gone through either of these routes. Seeing large numbers of viral sequences in ORFs most probably means that these are recent integrations in the genome that have not started or have just begun deactivating and decaying. However, since the probability of maintaining an intact ORF for an extended period of time without purifying selection is low [27], several of the identified *envs* in ORFs could be under purifying selection (**Figure 2.6**). ORFs of length higher than 1k aa are most probably due to sequencing mistakes in a genome and large gaps of unidentified bases in the middle of the sequence, as this is longer than any known retroviral *env* [15, 40]. Of the all identified *envs* in ORFs, 737 were suitable for further analysis, containing almost full envelope sequences which allows them to be included in the final alignment and phylogenetic analysis.

On our reconstructed phylogenetic tree, we observed many branches with the sequences restricted to a single bat species (**Figure 2.3**). This suggests that the given sequences came from a recent integration event that occurred after speciation. The same is true for clusters containing sequences of only *Pteropodid* bats (*P. alecto, P. vampyrus, R. aegyptiacus, E. helvum* - which split from other bat families around 22.6 Mya) or only vesper bats (*M. lucifugus, M. davidii, M. brandtii, E. fuscus* - which split from other *Yangochiroptera* around 53.8 Mya) [2]. These types of subtrees have been collapsed for clarity in the tree. Branches containing different bat species and reference retroviral or *syncytin* sequences are of particular interest; they might contain sequences of similar

architecture to known ERVs and EDIs dispersed amongst different families, indicating integrations predating speciation events and preserved in a potentially coding state for a long period of time, which might mean that they have a recent function in the host.

Of all the identified *envs* in ORFs, 84 candidates in 32 sets were considered further as potentially functional host genes on the basis of presence of all gammaretroviral canonical motfis and the immunosuppressive domain, as well as being present in more than one analysed bat genome (**Figure 2.5**). From all of these we want to distinguish two specific sets which display the most promising features for further analysis. VesperOrtho1 is a set of 6 sequences from 3 different *Yangochiroptera* families, containing a full envelope sequence, and displaying full canonical features of gammaretroviral *env*. dN/dS analysis showed that the sequences are under purifying selection. The other candidate is the MegabatOrtho6 set of 3 *Yinpterachiroptera* full envelope sequences showing all the gammaretroviral features and evidence of purifying selection. In close proximity to both of these sets we found fragmented *gag* and *pol* genes, a feature that has also seen before in case of many of the known *syncytin* genes. Because orthologs from *Yinpterachiroptera* and *Yangochiroptera* cluster with reference *syncytins* from different types of mammals, we speculate that those potential candidates might be responsible for the presence of different types of placentation seen in bat species. We found several orthologous sequences in different bat families, suggesting some of them predate family splits and remain conserved, which may indicate function in a genome.

Interestingly, MegabatOrtho6 sequences contain a canonical receptor binding motif in their SU region, but do not cluster closely with BaEV/SRV-Y viruses, which also contain that motif. In our analysis we did not use the SU region in the alignment, since it is very poorly conserved, thus unsuitable for phylogenetic analysis. We suspect that it is possible that the surprising positioning of those orthologs on the tree might be caused by

recombination or deletion in those sequences, which made them cluster with candidates other than the expected one.

Yangochiroptera have haemochorial placentation, while different *Yinpterachiroptera* utilise either endoteliochorial or haemochorial placentation (**Figure 2.4**) [59, 228]. Because of this, bats are a good model to show how different viral progenitors of *syncytin* are responsible for different placenta types in mammals. In contrast to most other mammals, *Yangochiroptera* females are larger than males both in weight, size, and wingspan [229, 230, 231, 232]. *Syncytins* also seem to be involved in muscle development as they have been found expressed in the muscles of rodents, primates, carnivores and ruminants [42]. Levels of expression of potential bat *syncytins* in muscles, as well as correlations of expression with sex, could shed additional light on the evolution of placental mammals.

The clustering of the VesperOrtho6 sequences with HEMO might mean that this is an old and unused ancestor of a sequence similar to HEMO, which could have shared a similar function. The HEMO gene was introduced into the genome of a mammalian ancestor around 100-120 Mya (before the split of Laurasiatherians and Euarchontoglires) and it can be detected in conserved full-length form in simians and cats [221]. The presence of VesperOrtho6 in all analysed *Vespertilionidae* bats might suggest that this sequence is specific for that family, which might suggest that other families could have HEMO homologs placed in different genomic regions compared to vesper bats. We ran BLAST analysis comparing sequences found in bats to other mammalian sequences and found similarities in 5′ end flanking region which further confirms that sequences in bats might be orthologs of HEMO, present in their genomes as pseudogenes. From our analysis, as expected, bat HEMO-like sequences are not under selection.

2.5 Conclusions

Long term co-evolution between bats and viruses might confer the ability to survive viral infections, and to safely harbour viruses and large numbers of highly active transposable elements in their genomes [126, 128]. It is reasonable to expect that this co-evolution at some point resulted in co-option of viral genes that may have played a role in antiviral immunity [170]. It is possible that bats might have a different set-point for triggering the immune response and harbour pathogens controlling their replication in tissues rather than clearing them and setting up immune-pathological responses in infected tissues [156]. As the only mammals capable of active flight, bats have cells that are under high oxidative stress and limiting processes, which might be additionally damaging, and might be a viable explanation for their immunological differences in susceptibility to pathogens [77, 154]. We hypothesise that bat resilience may also stem from captured ERV genes, which help bats to overcome viral infections and suppress symptoms. There is a need for more analysis and laboratory tests to mark out which of these sequences might have immunosuppressive or antiviral potential, and if any are substantially expressed in tissues, particularly the placenta. To make such analyses more thorough, more bat genome assemblies from different families with good data quality will be needed.

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

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

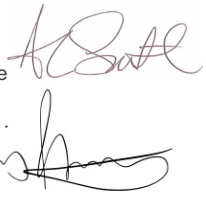
Title of Paper	The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology.
Publication Status	Published
Publication Details	Skirmuntt, Emilia Cecylia, and Aris Katzourakis. "The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology." Virus research 270 (2019): 197645.

Student Confirmation

Student Name:	Emilia Cecylia Skirmuntt		
Contribution to the Paper	Wrote the manuscript, prepared all the figures, and performed all analyses.		
Signature Emilia Skirmuntt		Date	16.06.2021

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: Aris Katzourakis and Adrian Smith		
Supervisor comments		
Signature 	Date	17.06.2021

2.6 Laboratory work design: Bat *syncytin*-like gene cell-cell fusion assay and tissue expression

In order to check whether sequences detected can be categorised as *syncytins* we selected three sequences based on their dN/dS score, their presence in multiple bat species, and full-length status: MegabatOrtho1 (MO1), MegabatOrtho6 (MO6), and VesperOrtho1 (VO1) (**Table 2.4**).

In our *in-silico* analysis (**Chapter 2**) we also detected two other possible candidates: MegabatOrtho3 and MegabatOrtho19. MegabatOrtho3 contains a full *env* sequence with evidence of purifying selection, but it was detected only in 2 *Yinpterachiroptera* species (*P. alecto* and *P. vampyrus*). MegabatOrtho19 also showed evidence of purifying selection. This sequence is present close to *glial cell missing-1* gene found neighbouring to *syncytin* in mice and humans. The limiting factors are its presence in only 2 species, and that we detected only a partial *env* sequence.

In mammalian genomes, we can find several *env* sequences with immunosuppressive properties. But in order to be adjoined to the *syncytin* group the gene needs to exhibit all three canonical characteristics of *syncytins*: (A) be fusogenic in an *ex-vivo* cell-cell fusion assay, (B) be expressed in placental tissue at the level of the foeto-maternal interface, (C) and be conserved in a functional state in a number of species, while subjected to selective pressure with low rates of dN/dS [30].

	Number of species present	dN/dS	Length
MegabatOrtho1	3	0.7799	550 aa
MegabatOrtho6	3	0.5345	626 aa
VesperOtho1	6	0.479	480 aa

Table 2.4: Data on selected *syncytin*-like sequences showing the number of species in which a sequence was detected, its dN/dS result, and the length of the sequence in aa.

The cell-cell fusion assay was used previously in studies identifying *syncytin* genes in other species and taxa [49, 194, 195, 226, 227, 233, 234] using an array of different methods and vectors. Since the bat-*syncytin* receptor is still unknown this study will use a co-culture assay with a panel of different cell lines from humans, bats, and other mammals.

Due to the lack of fresh samples and the unknown status of the pregnancy for the two obtained placental samples available to us, we cannot perform an *in-situ* hybridisation assay to inspect the level of the expression of the *syncytin* candidates at the foeto-maternal interface.

The expression of the *syncytins* candidates should be at least 10 times higher in the placenta than in other tissues. Due to the fact that both placental samples and the other obtained tissue samples are from *R. aegyptiacus*, we will perform a qRT-PCR on these samples using primers for MO1 and MO6 sequences which were detected in this species.

2.6.1 Materials

Samples of large intestine, small intestine, spleen, kidney, muscle, and liver in RNALater (Sigma Aldrich) from 25 *R. aegyptiacus* bats were received from Copenhagen Zoo, Denmark. From these 20 (10 adult, 10 juvenile) were female and 5 were male (1 adult, 4 juvenile). Upon arrival samples were transferred to new empty cryotubes and stored in the -80°C freezer.

Two samples of *R. aegyptiacus* placenta received from Queen Mary's University, UK (S. Rossiter lab and Copenhagen Zoo) were received in RNALater, transferred to empty tubes and stored in -80°C.

Cell line	Species	Tissue
	Carollia prescipillata	lung
	Carollia prescipillata	kidney
	Carollia prescipillata	liver
PaBr	Pteropus alecto	brain
Tb1Lu	Tadarida brasiliensis	lung
AIK	Atribeus lituratus	kidney
Sliki	Sturnira lilium	kidney
	Sturnira lilium	lung
	Sturnira lilium	liver

Table 2.5: List of available bat cell lines

Total RNA and DNA will be extracted from the frozen samples from all six tissues from 5 juvenile and 5 adult females, and 1 adult and 3 juvenile males, as well as from two placental samples using DNA/RNA Mini Kit (Qiagen). 1mm zirconia/silica beads will be used for lysis and homogenisation of tissues. Genomic DNAs will be purified by phenol-chloroform extraction.

Bat cell lines frozen in FCS and 10% DMSO were received from Imperial College, UK (S. Giotis lab) and stored in liquid nitrogen (**Table 2.5**).

2.6.2 Synthetic gene construction and Gibson assembly

Gblocks are synthetically assembled double-stranded DNA fragments. We designed Gblocks containing full sequences of MO1 *R.aegyptiacus*, MO6 *R.aegyptiacus*, and VO1 *M.davidii* (IDT).

MO1 sequence: `ACAAGGATGACGATGACAAGATGGCTCCAATCAAGGCGATCATCATTCTAGGACTCATA`
`CGGACTGTTTTCTCTGCTTCGGGACATTGGGAATTCCAGGTTATTCGGACATGGGGGGATGAAGGGATAGTT`

AATGCCACTCGGTGGGATGGGCACCGCCACCCTGTTACCCTCCAATTTGATGCTTGTGGGGTAATGGAACCC
AAATCATGGACAAATTGTGGTGGGCTAGGATGGGAGCGGTATTATGCTACACAACATAAATACCTCTGTTTT
AAGTCAAGGCAGAAGGGTGACTGCGATAATATAGGAGAATTCTACTGTCCCTACTGGAGTTGCCCCCACGCC
CATTCGAGGGCTCTAGGCAGCGGAGTAAGGCTCCTTAGGCCCCCGGCAAACTGCTCCCTTGGTCACTGCAAC
CCTATTGCCTATGACATCAGGGACCCAGGGGACCCCAAGTGGAGCCAAGAGGTGATAGCTGGGGTCCGGATA
GATGGGAAAGGAATAGATCCAGGAGCTATCATTAAGATAGTCAAGTCCTGGGTAGTCAATCCTACCTCTTGC
CTGAGGCTCAAACCCCACAGTGCCGCAGGAACCCCTCCCACAGTTCGTCTCCACCCTTCTTGCCCCCTCCAA
ACATGGGGCCCTCTTGAATTCCCACTCTGGGAAATTGTCAATGCTACCCATCAACTTCAGCTAGTAGCCAAC
CCCTCCCTAGGGTCCGACTGCTGGCTTTGCACCTCTGTGACTACCTGGACCTATATAGGCCTTGGAGTAAAT
GGTATCGTCACCCCCGAGCATTCTGGCACTGCCTGGCCTGTCTGGCCCTCAGTAACGTCGGTTGGTAATGCC
AGCTGTGTACGAGGTACAGGAACCCAGCCTAATGTAACTGCTAATTGCACCCAAATCTGGCAGCTAAATTCC
GATACCTTGTACCGGCCCCCAAATGGCACGTATTTCTCTTGCTCAGATGGAGTGTACGCCCGACTCAACGGC
TCACGGTGCAGCTTATGTGCTATAATTATCTTACTGCCTAACCTCTCAGTTGTCCCCAGATCTGATGTAACC
TACCTGGCGGAGCCAGTCGGGCCCCAAATTCGCCAAGCACCCATATTAATATTGCTGCTGGCAGGAATCGGG
GCAATAGGGGGAATTGCCTCTGGAGAAGCCGCAAAAGGCGCAGCTAAACCTCAATTCTTTCAACTGTCCACT
CAGATAGATACAGACCTCAGTGCCTTAGAGAGTGCTCTAAAGACACTCCAAGATCAGGTAGACTCACTGGCA
GAGGCTGCTCTGCAGAATAGAAGGGCCCTCGACCTGCTTACCACCAGCCAAGCGGGCACAAGTCTTTACCTT
AAAGATCAATGCTGTTTCTATGCCAATAAGTCTGGAGTCGTACGAGAAAACCTAAGCCAGTTGTGTGCTCGA
ATCCAGAGGCGTTTAGCAGAAAGGGAGCAGCAAGGAACATGGTGGAACCCTTCACGATCCTGGGGTTCCTGG
CCGGTACTGCTCCTTGGGCCCCTGGTTGGACTTATCTTTGTCCTCATACTGGCACCCTGTGTTCTGAGCTGC
CTGTCAAATTTAATCTCAGAATGGGTAGAGGCCATCAAGCTTCACGTGGTCTGCCGCCACGCATATGAACCC
GTATCCCAACAATATGATGTGTGATCTGATATCGGTACCAGTCG

MO6 sequence: ACAAGGATGACGATGACAAGATGAGGAAAATCTTTCTTTTGTTAATTCAAGTAGGATTT
TGTATTCAGGTAATACAAGCAATGGTGTCTCCCAACCCACCCTGCAAGGGAATTATGGATACTGGGCACATA
TACAGGAGTCTATTAATACTTCTCATAAGCTCACCCTTGAGGGCAGCCACTGAGCCTCAATGCGATAAATGC
GTAGAGGCCATCCAAATAGGTGGTCAGGTCCAGCAGGCCTTTACCTCATATACCCATCAACCAGCCGCTTGT

TACGAATCGGCTACCATTTGCACTTATGCTGGCAGAAGGTACTGGATGGGATGGATAAAAACCAAATATGTA
GCAGGATACATTTGCCAAGGGCCACGCCCAGGGGCAGCAGGAAAGTGGGCCTGTTGGACATACTATGGGCAG
ACAGGTATCTCAGATGGAGGGGGTGTCCAAGACCAGGCTCGAGACCAGCTAGTCAAGGAGACTGTACAGCGA
GTAATAGAACAAACAAAACCTAAACCACTTAAAAAGACCTTAGGTCAAGACCTAGACCTCCCAACCCTGGGA
AAGAACCTCTTTGTTGACCTAGCTGAGAGGATAGCCCATGCAGTTGGTGTTACTAAATGCTGGGTTTGTGGA
GGAGCACTCATGGGTGAAGATTGGCCTTGGAAGGGAACTAGCCTAGATGCTTTTTACCTCCTTCAGTGGAAT
CACACGGTGGCTCAACATGAAGAGGGTCGACCCAGGATCTGGGTCCTCTCCTCAGAAGTAGTAGGAGAAGAA
TGTCTTAGCCGGATAGGTCCAAGATATACTCAGTGGGTTGGAGAGACTCCATGTAGAAGAATTCTATTTTTA
AACCACACAGACCTAAACTGGTGGCCCAAAGCTCCTTCTTGGTATTGGGCCAACAGTACCATTGCTAACTCC
TCCTTGTGCCACTCTCTTGGCCAAAGCTCAACTGTCACAAACTGTACTATAAAGGAAGGCACGGTCCCTAAC
CCCTTCCGTGCCATCCCAGCTATTGGCACATTCTGGGAAACTCTCAATTCCACTAGCCCCGGGAGTTGGAAA
GCACCAGATGGGCTCTTCTGGATTTGTGGGAAGAGGGCTTACACTCAACTCCCAACAAACTGGAAGGGAAGT
TGCACCATCGGGATCATTCAGCCTGGCTTCTTCCTCTTACCGGCCCAAGGAGGAAATGAGCTAGGTGTTCCC
CTCTATCAAGACTTAAAGGCCAGGACTAAGAAGTCAGTCGAAGTAAGTGGACAACAAACAAGGGAGGAGGGT
GAATGGCCATCTCACCGTATTATCCAATACTATGGCCCCGCAACTTGGGCCCGAAATGGGAGGTGGGGTTAC
CGGACCCCCATGTACATGCTCAATCGGATCCTTAGGTTACAGGCCACACTAGAGATTATTACCAACCAGATG
GCCTCGGCCCTAGAACTCCTGGCCACGCAACAGTCCCAGACGCGTGCGGCAATGTATCAGAACCGTTTAGCT
TTAGATTACTTGCTAGCAGAAGAAGGGAGAGTATGCAGGAAGTTTACCCCATATGATTGTTGTCTCCAAATA
GATGACAATGGCAAAGCTGTAACTGACACAGCTACTAATATCAGAAAAATGGCTCATGTACCCACCCAGAAT
TGGAATGGGTGGGGCCCTAGTAACCTCTTTGGAGGTTGTTGGTTCTCTGCTTTCAGGGGCTTTAAGATCTTA
GCTGGGATAGTACTCCCCCTCCTCGGGGCCTGCCTCCTTCTCCCATGCCTATTGCCTCTTGTTATCCGTTCA
ACTACTTCCCTCATAGAGGCTATAGTAGAAAGAAAAATTGCAGCAAAAGTCTTCATGCTGTGGAGGTACTCG
CCCCTGCCCCAGGAAAATCAAGAGATCAATGCCCTGTAATCTGATATCGGTACCAGTCG

VO1 sequence: ACAAGGATGACGATGACAAGATGAAGCTCTGGGCTGCGTTGTGTGTCCTGTCCCTCTGC
TGGAGCTGCACCCACGCGTGGAAGGATAATTCCTTCGTTCGAATCTCTCAATCCATAGCCTCCTCCGCTAAC
TTATCCCATTACTGGCTCTGTCATGTCAAACCTTCCTCCGGGCTTGATTCTGCTGACCCTCTGGTCATGCCC

GTTTTAAATTTCTCTCAGATCCCGAGTTTCCAAGTCACTTATGACATGAAATTCCCTAATGCTGTTTACCGA
GTGAGGCTTGTTAGTTATGAATTTCCTGTTCCTTGTTTCTTCCTCACCCCACCTGAGAGCTTGACAACCATG
GTTTCAGGTAAATACTGTAAAAAGACTGAATCCTCGGGGCACAAGAATTTGACCACGCCCCATGGAAGAACT
ACTGTTGCTTGTCAGTATAATAAGACATATTTCACTGACCCAGTTTCTCTCCCTTTGATTCGACAACGATGT
GCAAATCTGTCCTCTGCCTATGCTTCTACTGAACTATGCAGAACCTGGAGGGATCAAAACGCGACATACCTG
AAGGGCACCCCCACCCCCTTACAGAGTTGCCATATGACTCCATCATGGACCCTGCCTCGCACTGACCCCTGG
TTAGAAAACATCTTGAGTCACCTTCCCCAAGGGGACAGCCTTAATTTTCAGAATATGACTGGAATCTTATGT
GCTCCCATGGGGTACGTTTTTGTTTGTGGATCCCAAAATGAGGCCTGGGCCTATAAATGCCTTGACAGTTGG
CATGTTGGTGGAACCTGCCTATTGGGCTACTTGGTAACCCCATTCTCTGTCCAGAATGTTCTAGAAACTGTC
CATTGGACCAGTCCCCTCACTCTACTCACTAGGATTGCTCGGGACCTCCCTGAGGGAGACAATTATTTCTTT
GGCTACACCTTCCTTCCTTGGTGGAATGTGGCAGCTCATCAACATGTACTGAGGAACCTATCAAAAACCTTA
GATATTATTGCAAATGAAATGGCTACCACCCTTACTGGGCTGCAAACGTCCTTGGATTCTCTTGCTAAAGTT
GTACTTGATAATCGTATAGCTTTAGATGACATGCTAGCTGAACAAGGTGGAGTCTGTATTGCTGCTAACGCT
ACTTGCTGTACATCTATCAATTCCTCTTCTGAGGTCGAGACTCACATTGCTAAAATCGGACAACAGGCTACT
CGACGCCAACAAACCTTAGGACATCAGCACAAAACAAAGTTAGAGAGAATGCGGAACTGGTTCACAGACCTT
TTCTCGAGGAAACTGAATGGGTTCCAGTCGAACGTAGGTAGTATTCTCAAATTATGTCTCATTCTTCTTCTT
ATAGTTATGGTATTCATAATAATTATGTACTGTATAATTAAGAGCTGTGCTAAGAAGGCTAAGACTCCTAAG
ACAAATAGGGAAATTCCCATGGTAAATCTTTAATCTGATATCGGTACCAGTCG

GBlocks will be cloned into the p3xFLAG-CMV-9 vector (Sigma) using a Gibson assembly approach [235].

p3xFLAG-CMV-9 vector primers:
Forward TCT GAT ATC GGT ACC AGT CG
Reverse CTT GTC ATC GTC ATC CTT GT

Cell line	Species	Tissue
293T	Homo sapiens	kidney
HeLa	Homo sapiens	uterus; cervix
MCA205	Mus musculus	fibrosarcoma
208F	Rattus norvegicus	fibroblast
MDCK	Canis familliaris	kidney
A23	Cricetulus griseus	fibroblasts
G355.5	Felis catus	brain

Table 2.6: List of cell lines

2.6.3 Cell-cell fusion assay

The expression vector for *env* (p3xFLAG-CMV-9-env) and an empty vector will be introduced together with nlsLacZ vector into a panel of bat, human, and mammalian cell lines (**Tables 2.5 and 2.6**). Transfected cells will be co-cultured with the target cells from different species. After 48h of co-culture, cells will be fixed and stained with X-Gal. The fusion index will be calculated as [(N-S)/T]x100, with N standing for the number of nuclei in the syncytia, S the number of syncytia, and T the total number of nuclei. Human and mammalian cells will be grown in DMEM supplemented with 10% (vol/vol) FCS, 100 mg/mL streptomycin, and 100 U/mL penicillin. Bat cells will be grown in Glutamax DMEM supplemented with 15% (vol/vol) FCS, 100 mg/mL streptomycin, and 100 U/mL penicillin.

2.6.4 Real Time qRT-PCR

MO1 and MO6 transcript levels in *R.aegyptiacus* tissues will be investigated via real-time quantitative RT-PCR, by using primer pairs designed to be specific for these genes.

As a positive control we will use primers designed to be specific for two housekeeping genes: ActinB and GAPDH.

Primers for cDNA for MO1, MO6 and housekeeping genes:

-MO1

Forward GCC ATG CCC ACA ATC AGT TC

Reverse GGA TAC GGG TTC ATA TGC GTG

-MO6

Forward GCC TCA ATG CGA TAA ATG CGT

Reverse CTA TGG GTT GCA AGG AGT TAC AAA G

-ActinB

Forward CTG AAG TGA ACC CCC GTT GG

Reverse ATT CAA AGT CCT CGG CCA CAC TA

-GAPDH

Forward TGT GCT TCA AGG GAG TTG CTA A

Reverse GCC ATG CCC ACA ATC AGT TC

In silico analysis combined with qRT-PCR assays should identify a putative *syncytin* gene determining whether the selected sequences have all three canonical *syncytin* characteristics.

3 Integrated retroviruses in six new bat genomes

3.1 Introduction

Bats are one of the major sources of zoonotic viral diseases, harbouring and spreading pathogens like rabies, coronaviruses and Ebola [62]. However, there is mounting evidence suggesting that bats can tolerate and survive viral infections, paving the way for a long history of co-evolution [148]. Indeed, selection tests indicated ancestral bat adaptations in immune pathways (e.g. Innate immunity activator protein -INAVA, or STING1 which stimulates interferon genes) and also showed expansion of the viral restricting APOBEC3 gene cluster [24]. We screened high quality genome assemblies for inserted endogenous retroviral elements (ERVs) to ascertain the number and diversity of endogenous viral elements in bat genomes, which are traces of previous infections. The retroviral life cycle makes ERVs the largest group found among all EVEs (endogenous viral elements) in vertebrate genomes [27, 171]. ERV integrations make up to 10% of the mouse genome [236], 8% of the human genome [28], and according to previous analyses up to 5% of the bat genomes of selected species (*Myotis spp.*) [77]. Non-retroviral EVEs by contrast are far less numerous in animal genomes [27].

There are seven known genera of exogenous retroviruses: alpharetroviruses (e.g. Rous sarcoma virus), betaretroviruses (e.g. Mouse mammary tumour virus), gammaretroviruses (e.g. Feline leukemia virus), deltaretroviruses (e.g. Bovine leukemia virus), epsilonretroviruses (e.g. Wallaeye dermal sarcoma virus), lentiviruses (e.g. Human immunodeficiency virus 1) and spumaviruses, also known as 'foamy' viruses (e.g. Simian foamy virus). All of these can be grouped into three classes comprising exogenous viruses and their endogenous retrovirus counterparts: class I consists of the gamma- and epsilon viruses, class II of the alpha-, beta-, delta- and lentiviruses, and class III of the spumaviruses [237]. Most exogenous retroviruses and ERVs share the same three primary structural building blocks:

1. Group-specific antigen (*gag*) : ∼2000 nucleotides long, containing matrix (MA), capsid (CA), and nuclear binding protein (NBP)
2. Polymerase (*pol*) : ∼2900 nucleotides long with protease (PR), reverse transcriptase (RT), and integrase (IN)
3. Envelope (*env*) : ∼1800 nucleotides long, containing surface (SU) and transmembrane (TM).

ERVs from all three classes were identified in mammalian genomes, although the distribution of their sequences is very uneven and biased toward the gamma- and betaretroviruses, which might suggest different efficacy of entering germline cells or some existing barriers to replication after integration for some genera [75]. Multiple examples of ERVs were found in bat genomes, however the exogenous retroviruses were identified in bat only recently [168, 169]. A few studies reported full length ERVs which have potential to be replication competent or to be very recent integrations [24, 76, 80, 82, 83, 84].

Gammaretroviruses are common pathogens of many different orders of animals. They can cause sarcomas and leukaemias. They are also one of the most widespread class of

endogenous retroviruses. However, while being very commonly seen both in infection screening and co-opted in genomes, their evolutionary origin still is not well understood [83].

There have been numerous studies looking at bat co-evolution with viruses and their ERVs. Some have identified gammaretrovirus derived ERVs in *Tadarida* bats [238], as well as in *Rhinolophus ferrumequinum* bat, which contain numerous copies of gammaretroviral genes [80]. Additional mining of 10 other species of bats confirmed that 6 species (*Rhinolophus pusillus, Rhinolophus pearsoni, Rhinolophus megaphyllus, Rhinolophus affinis, Myotis ricketti, Pteropus alecto*) also contain gammaretroviral genes. Further analysis of the phylogenetic relationship between bat and non-bat mammalian gammaretroviruses shows a cluster of mammalian viruses within bat virus lineages, which may suggest that other mammalian gammaretroviruses could originate from bat viruses [80]. Gammaretroviral genes have also been identified in megabats (*Rousettus leschenaultii, Pteropus vampyrus*) and two more microbats (*Megaderma lyra, Myotis lucifugus*), further supporting the hypothesis that bats are important reservoir hosts of viruses and that certain mammalian gammaretroviruses might have originated in bats [83]. However, another study showed that rodents are the more probable source of retroviruses previously seen in bats, and that bats may actually be more capable of receiving retroviruses from non-bat reservoirs rather than serving as their origin [85]. This cross-species receptiveness could explain how bats acquired such a large number of retroviruses. This study also further confirmed the hypothesis that bats, along with rodents, are major carriers and spreaders of viruses among humans and other animals [85].

Even when looking at ERVs specific to bats, the situation is often quite complicated. Elements in *Vespertilionidae* bat genomes are indicative of a complex amplification history

characterized by a mixture of retrotransposition (reamplifying themselves into another location in a genome), and reinfection (characterized by elements becoming infectious again and producing viral particles which infect other cells of the host organism) [81]. Scans of the genome of *Myotis lucifugus* identified 362 complete proviral sequences from all three ERV classes (gamma-, beta-, and spumaretroviruses), 25% of which were from newly discovered subfamilies [84]. LTR divergence analysis show that 64% of the ERVs were recent proviral integrations no older than 10 million years (My) old. Cross species genomic comparison estimated the age of the oldest integration to be around 27 My old based on LTR divergence and orthologous site analysis between *Myotis lucifugus* and two other bat species, *Eptesicus fuscus* and *Myotis davidii*. One of the ERV copies was fully intact and potentially replication competent [84]. This analysis suggests that vast numbers of retroviruses might still be present in natural bat populations [84].

Betaretroviruses cause infections in a wide range of animal species. Exogenous and endogenous betaretroviruses are associated with several cancers and autoimmune diseases in sheep caused by the Jaagsiekte sheep retrovirus (JSRV) or immunosuppression in rhesus monkeys caused by Mason-Pfizer monkey virus (MPMV) [193, 239, 240, 241]. Betaretroviruses show a higher capacity to cross the species barrier in comparison to other types of retroviruses, which might be of additional concern in the case of zoonotic diseases [76]. Endogenous betaretroviruses have been found in more species than their exogenous counterparts including also in more species of bat, where they have shown evidence of cross-species transmissions [76, 77, 78]. This might be caused by different integration mechanisms for this class of the viruses, or better adaptation to replication after integration [75]. The betaretroviruses genus is divided into types D, B, and HERV-K (containing only human ERVs). Type B and D differ in morphology, length of LTRs, ways of RNA nuclear export and complement of accessory factors.

Various betaretroviruses of both types D and B were identified using transcriptomes and genome analyses of *P. alecto, P. vampyrus, M. lucifugus, R. megaphyllus* and *R. ferrumequinum* [76]. Study assigned them to 8 distinctive groups according to their genomic organizations. In all bats, betaretroviral genes were transcribed into messenger RNA (mRNA) and a full-length transcript was found to be expressed in *P. alecto*, while analysis of the genome revealed intact and potentially functional betaretrovirus in *M. lucifugus*. Molecular clock testing on LTRs suggested that integration of betaretroviral particles might predate the bat suborder split and occurred at approximately 36 Mya. This strongly suggests that bats were betaretroviral reservoirs for most of their evolutionary history and still could be [76, 85].

Novel endogenous type D betaretroviruses were also found in the genome of the common vampire bat (*Desmodus rotundus*)[78]. It is present in other bats from this family, but not all of them, suggesting earlier integration rather than a family split. On a phylogenetic tree, it branches closer with rodent and New World monkey betaretroviruses, which is indicative of cross-species transmission, while the age of the provirus infers that the exogenous virus might still be in circulation in bat populations. The *D.rotundus* betaretroviral sequence corresponds to sequences found in another phyllostomid bat - the Seba's short-tailed bat (*Carollia prescipillata*), which seems to be most closely related to the squirrel monkey retrovirus, but might also have been transferred to both hosts through an unknown, possibly murid, host [242]. Additional class II sequences were also found in *M. lucifugus* with 157 elements divided into 6 families, which exceeds the amount of sequences found by the same experiment in the class I group, but the total genomic DNA from class II (9.2Mb) was much smaller than the one derived from class I (31.5Mb) [84].

Alpharetroviruses are simple retroviruses which so far has been observed to replicate

exclusively in birds, hence sometimes they are referred to as avian retroviruses [86], but some laboratory experiments shown that they might also affect rats [243, 244], although no alpha-like viruses, both endogenous and exogenous, were found in non-avian species to this date. Exogenous alpharetroviruses can cause leukaemias and sarcomas in birds. Analysis shown that alpha-like ERVs are more heterogeneous than mammalian ERVs [245]. Apart from that alpha ERVs show patterns which suggests gradual evolution from beta, to alphabeta and finally alpharetroviruses [245], which also suggests that avian retroviruses evolved in birds interdependently from the rest of retroviruses over the last 150 million years [245]. There is also a unique ERV group found in birds (TgERV-F), which is the only gammaretroviral recombinant in the avian genome and possesses alpha-, beta- and gammaretroviral like features [16]. This ERV gained mammalian like *env* at some point on the course of its evolution which might have help it gain access to the new hosts emphasising role of *env* in generating novel viral lineages [16].

Deltaretroviruses are another example of class II ERVs. It is interesting that no endogenous representatives of this family had been found until recently, when a sequence was found in a genome of long fingered bat (*Miniopterus natalensis*) [87]. Single and heavily fragmented copies of this ERV were found in this bat, but not in the families closest to *Miniopteridae*, suggesting that the integration took place 20 to 45 Mya [87]. Exogenous deltaretroviruses were identified in only a very small number of species, but they are important human and veterinary pathogens (e.g. bovine leukaemia virus - BLV, human T-lymphotropic virus type 1 - HTLV-1). Recently, exogenous deltaretroviruses were also found in bat species making them the only class II retroviruses found in bats in exogenous form [168].

Lentiviruses are causing a wide range of chronic diseases in humans and animals like immunodeficiencies and neurological disorders. They are an object of intensive studies

since the HIV emergence. Although the first genomic analyses of exogenous lentiviruses suggested modern origin, further discovery of the new endogenous subgroup in European rabbits shown an ancient origin of this genera [72]. However, lentivirus endogenization seems to be a very rare event. To this date we known only about few examples: rabbit endogenous lentivirus K (RELIK) in rabbits and hares [72, 246], *Mustelidae* endogenous lentivirus (MELV) in weasels [247, 248], prosimian immunodeficiency virus (PSIV) in lemurs [249, 250], sloth endogenous foamy virus (SloEFV) in sloths [251], coelacanth endogenous foamy-like virus (CoeEFV) in coelacanths[252], and *Galeopterus variegatus* endogenous lentivirus (GvaELV) in colugos [73]. All of these ERVs form distant clades branching far from exogenous lentiviral phylogeny, with CoeEFV placed in the basal branches suggesting ancient marine origin of the foamy viruses [252].

Within class III of retroviruses, spumaviruses, known also as foamy viruses, infect a wide array of hosts, although a human spumavirus has yet to be found. It contains Gag, Pol and Env proteins plus two non-structural proteins, Tas (transcriptional activator) and Bet (with unclear function) [253]. Virions outside of the host cells are observed to contain DNA rather than RNA, because reverse transcription takes place during budding and assembly on the exit pathway and not entry and integration like with the other retroviruses [254]. Endogenous foamy viruses were seen in bats, but in much smaller numbers than the other two classes. Two full proviral sequences were identified in a genome of *M. lucifugus* (likely to have invaded the host genome ∼4 Mya). In addition, a large number of non-autonomous and decayed retroviral copies were identified in that bat. Their fragmentation and mutation state suggest that they probably predate the radiation of eutherian mammals [84].

Bats possess extraordinary adaptations, including flight, echolocation, extreme longevity and unique immunity. High-quality genomes are crucial for understanding the molecular

basis and evolution of these traits, but to our knowledge to this date such bat genome assemblies did not exist. To understand the evolution of bats and the molecular basis of their special relationship to viruses, we analysed reference-quality genomes of six bat species as part of the Bat1K global genome consortium (http://bat1k.com) [1], which were generated in coordination with the Vertebrate Genome Project (https://vertebrategenomesproject.org). These six bat genera span both major suborders *Yinpterochiroptera* (*Rhinolophus ferrumequinum* and *Rousettus aegyptiacus*) and *Yangochiroptera* (*Phyllostomus discolor*, *Myotis myotis, Pipistrellus kuhlii, Molossus molossus*)[59], represent extremes in bat longevity [255], possess major adaptations in bat sensory perception [1] and can better survive viral infections as compared with other mammals [138].

3.2 Materials and Methods

The six bat genomes and seven additional mammalian genomes were inspected for the presence of ERVs. Potential integrations were identified using local BLAST [211] with 14 probes for the viral proteins Gag, Pol, and Env from each genus of *Retroviridae*: alpha-, beta-, delta-, epsilon-, gamma-, lenti-, and spumaretroviruses (**Table A 8.1**). Using Python (ver.3.6+) custom software (ERVin – https://github.com/strongles/ervin; **Figure 3.1**) we extracted sequences with an e-value $\leq$0.009 for all of the proteins and of length $\geq$400 aa for Pol regions and $\geq$200 aa for both Gag and Env regions. We grouped found sequences according to their taxonomic relation to the first returned hit given by reciprocal BLAST (**Tables A 8.12** to **A 8.59**). The Pol and Env regions for each viral protein family passing quality thresholds (highly conserved 200 aa region ending with 'Y[M/V]DD' motif for Pol, and highly conserved 68 aa region containing immunosuppressive domain (ISD) ending at a short disulfide bond loop 'C-X6-C-C' or

'C-Xn-C' for Env) were aligned using MUSCLE within the Aliview software [4]. We manually inspected sequences and corrected the alignments. We discarded sequences which were shorter than 50 aa.

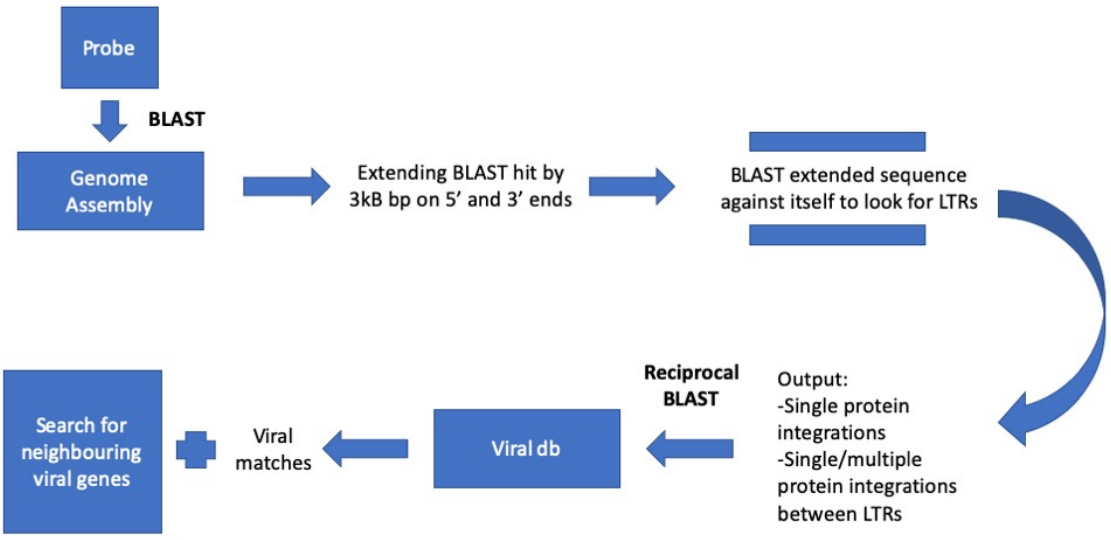


Figure 3.1: The ERVin software pipeline. Chosen viral sequences are run in a tBLASTn mode against genome assemblies. Target genome areas with hits are extended 3kB each way and run in nBLAST mode against each other to detect retroviral LTRs. Reciprocal BLAST is run on hits flanked by the LTRs, and single hits giving a list of matching viral proteins. Additional separate step is comparing libraries of hits assigned to each viral protein detected in the target genome to find neighbouring sequences and full viral integrations.

With Python (ver.3.6+) custom pipeline we distinguished all the integrations which contained combinations of two or three of the found sequences (*gag-pol-env, gag-pol, gag-env, env-pol*) from all of the ERV classes (**Tables A 8.2** to **A 8.11**).

To reconstruct the phylogenetic trees of the retroviral Pol and Env protein sequences for all 6 bat genomes and the viral probes used in the BLAST searches, we first ran Prottest (ver.3.4.2)[214] to determine which model to use with RAxML (ver.8) package

[6]. The best model was the VT+G model for Pol (**Figure 3.7**), and JTT+G for Env according to the AICc scoring criteria (**Figure 3.8**).

Finally, for the ERVs from the least abundant types in our analysis (delta-, lenti-, and alpha-like) which contained at least two canonical genes we performed additional analysis of neighbouring genomic architecture. We extended genomic regions for those co-opted genes for 3 000 base pairs (bp) in each direction and analysed these regions with tBLASTn against the viruses database and with NCBI's Conserved Domains Database (CDD) [256]. Then we annotated all found sequences and ORFs in the Geneious R10 software (ver.10.2.3; https://www.geneious.com) (**Figures 3.4 - 3.6**).

3.3 Results

The retroviral life cycle makes endogenous retroviruses the largest group found amongst all of the endogenous viral elements (EVEs) in vertebrate genomes [27, 171]. Our analysis confirmed the presence of retroviral proteins from all of the retroviral classes in the analysed genomes. As expected with respect to findings in other mammalian and bat genomes the highest number of integrations come from beta- and gamma-like retroviruses [76, 82], with a higher number of beta-like integrations in the cases of Pol and Gag proteins and gamma-like integrations for Env proteins in most of the bats. The only species which differs from the given pattern is *M. mollosus* which displayed a higher number of gamma-like sequences for all 3 viral proteins. Apart from that, we also detected integrations for the following retroviral families: delta (*M. molossus*), epsilon (*P. kuhlii*), and spuma (*R. ferrumequinum*) for the Pol region; lenti and epsilon for the Gag region (both in *P. kuhlii*); and alpha for the Env region (*P. discolor, R. ferrumequinum, R. aegyptiacus*) (**Figure 3.2b**).

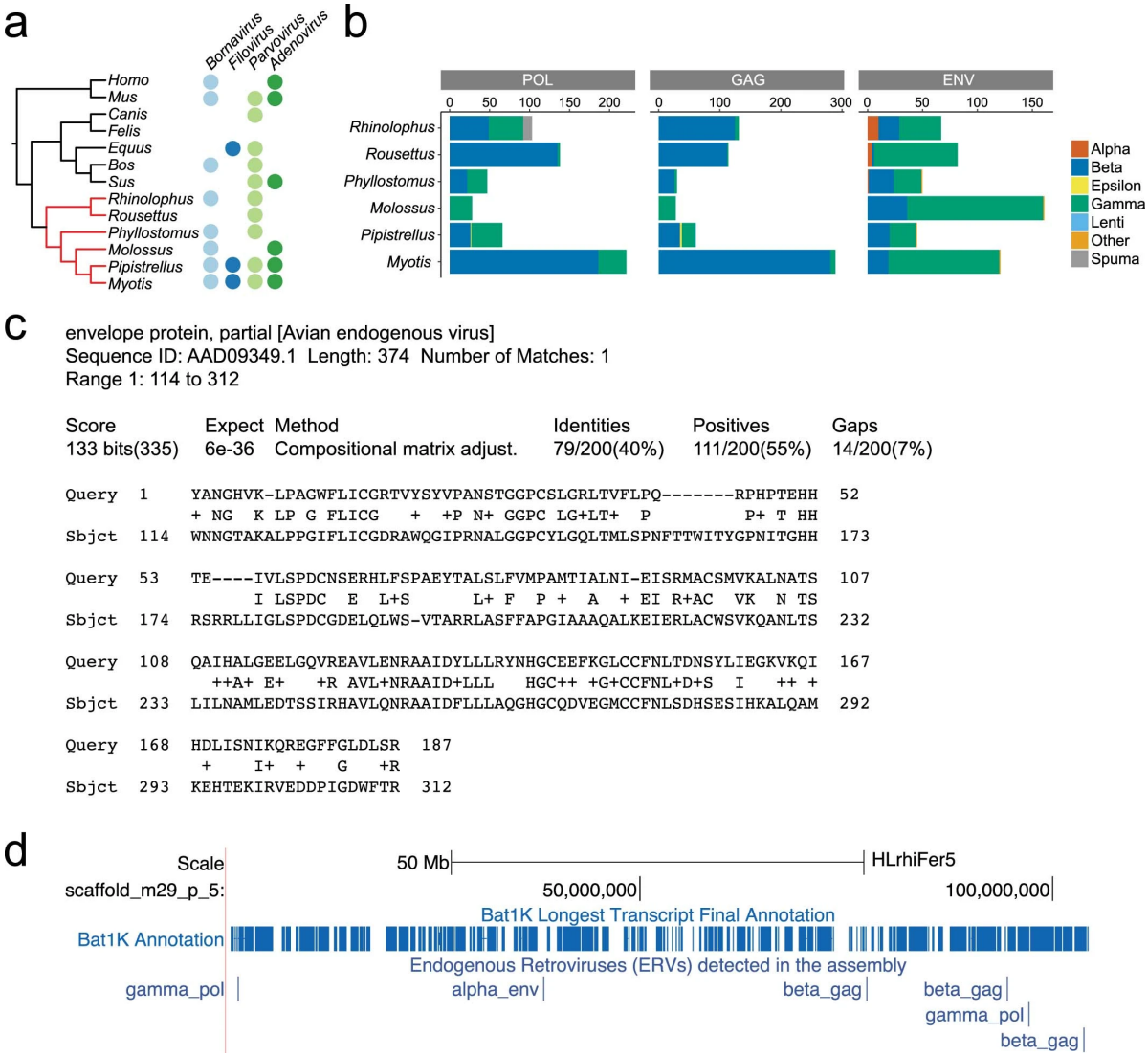


Figure 3.2: a) Viral families identified in more than one genus mapped to phylogenetic tree of six bat species and seven additional mammals. Using reciprocal BLAST searches and a custom comprehensive library of viral protein sequences, six bat genomes and seven mammalian outgroups were screened for the presence of non-retroviral EVEs. Endogenous sequences identified as *Adenoviridae, Parvoviridae, Filoviridae* and *Bornaviridae* were represented across several mammalian genera. b) Bar plots show numbers of viral proteins of all seven *Retroviridae* genera detected in the genomes of our six bats. Beta-like integrations are most common for Pol and Gag proteins and gamma-like integrations are most common for Env proteins. Overall, the highest number of integrations was observed in *Myotis* (n = 630), followed by *Rousettus* (n = 334) with *Phyllostomus* containing the lowest. c) Alignment exemplifies that an ERV found in *Rhinolophus* (scaffold_m29_p_13:24821733-24822323) best matches the Env protein of an avian endogenous virus, which belongs to the alpharetrovirus group. d) UCSC genome browser screenshot (https://genome-public.pks.mpg.de/) of a 104 Mb scaffold (scaffold_m29_p_5) of the *Rhinolophus* assembly shows detected ERVs as an annotation track.

The highest number of integrations was observed in *M. myotis* (with the exception of the Env region which was observed most frequently in the *M. molossus* genome), while the greatest variety of genera was observed in *P. kuhlii*. We also compared the numbers of Pol, Env, and Gag regions found in bats to the 7 mammalian reference genomes (**Figure 3.3**). Of all analysed genomes, *M. musculus* displayed the highest number of integrations of viral protein sequences. The numbers of Pol sequences found in the non-bat mammalian genomes and 5 of our 6 bat genomes were comparable to each other, with *M. myotis*, whose genome contained twice more Pol and Gag integrations, being an exception. Except for the mouse, all of the analysed bats exhibited more Env and Gag integrations in comparison to the other mammalian genomes.

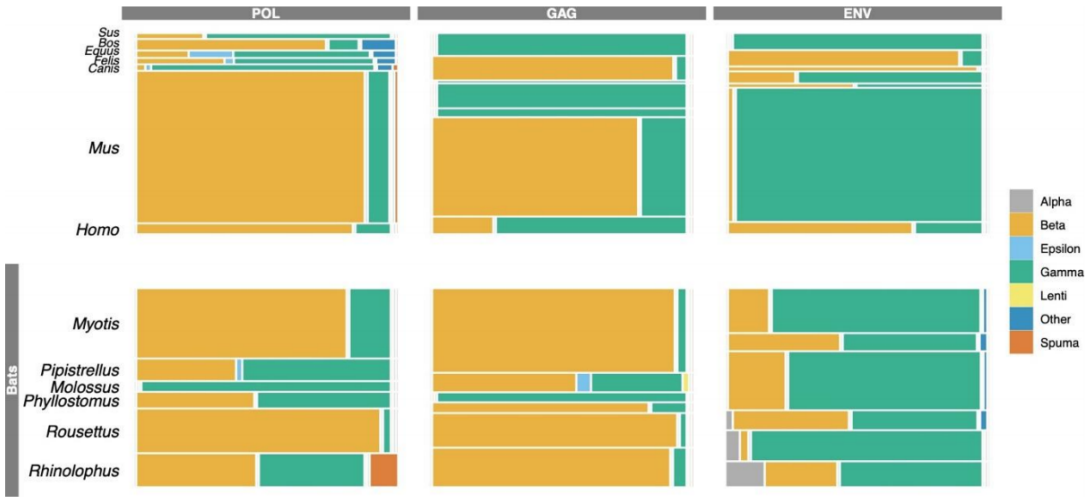


Figure 3.3: Mosaic plot showing the relative numbers of three retroviral gene sequences found in six bat genomes compared to seven mammalian reference genomes.

Overall, the highest number of integrations was observed in *Myotis* (n=630), followed by *Rousettus* (n=334) with *Phyllostomus* containing the lowest (n=126) (**Figure 3.2b** and **3.3**), and the highest number of integrations containing two or more genes was observed in *Myotis* (n=147), and the lowest in *Phyllostomus* (n=22), while *Rousettus* genome con-

tains the highest number of integrations of all three *gag-pol-env* genes (n=12) (**Tables A 8.2** to **A 8.11**). Additionally, we detected ERV sequences with hits for alpha- and lentiretroviruses in reciprocal BLAST searches. Until now, alpharetroviruses were considered as exclusively endogenous avian viruses [257]. Thus, our discovery of endogenous alpharetroviral-like elements in bats is the first record of these sequences in mammalian genomes, widening the known biodiversity of potential hosts for retrovirus transmission. We detected several alpha-like Env regions in *Phyllostomus, Rhinolophus*, and *Rousettus* (**Figure 3.2b, c, d** and **3.5**), showing that multiple and diverse bat species have been and possibly are being infected by alpharetroviruses. We also detected duplicated lentivirus *gag*-like fragments in *Pipistrellus*, which are rarely observed in endogenized form (**Figure 3.6**) [72], and were detected close to the LINE-1 transposable element. Transposable elements are an important driver of the genome evolution and might facilitate integrations of these elements into the host genome, and were found before close to endogenous viruses and retroviruses [27]. We also identified a full deltaretrovirus-like integration in the *Molossus* genome (**Figure 3.4**). The integrated ERVs are available as an annotation track in the Bat1K genome browser (https://genome-public.pks.mpg.de) (**Figure 3.2d**).

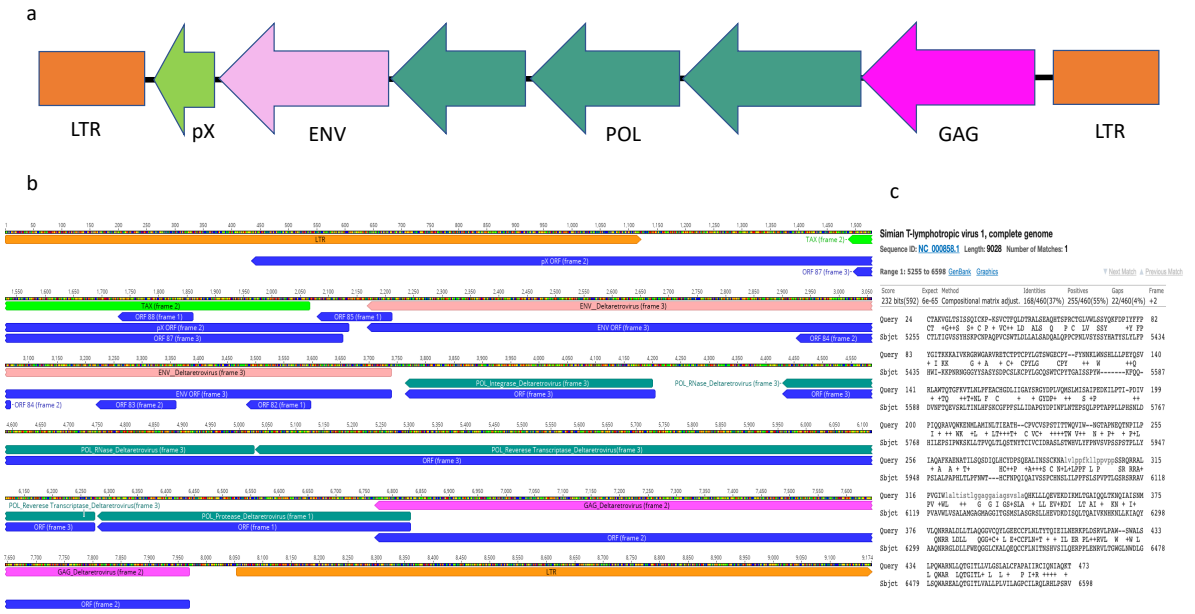


Figure 3.4: Full deltaretrovirus-like integration in *Molossus molossus* genome. a) Simplified figure showing reverse oriented sequence architecture with the detected genes and sequence features: LTRs (orange), Gag (dark pink), Pol (teal), Env (light pink), and pX ORF (green). b) Placement of each viral-like sequence in the *Molossus* genomic region together with all the ORFs (blue) in the right frame and direction. c) First reciprocal tBLAStn match for the Env protein

Further analysis of the deltaretrovirus-like integration in the *Molossus* genome (Molossus_9 8674906 - 8676474) detected all the canonical genes (*gag, pol, env* and pX) in ORFs with no stop codons or frameshifts, which are placed between two full LTRs (**Figure 3.4**). All the detected proteins returned hits to HTLV in reciprocal BLAST searches, while *env* and *pol* sequences both branch with deltaretroviral references on the *env* and *pol* phylogenetic tree reconstructions (**Figures 3.7** and **3.8**).

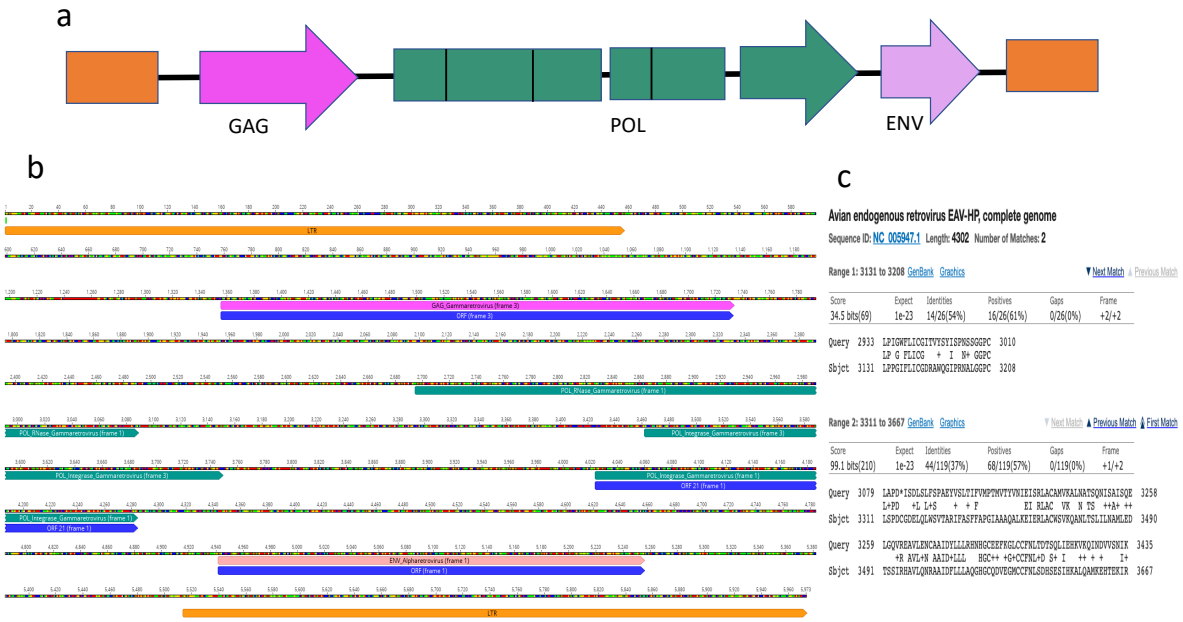


Figure 3.5: Gamma-alpha-like integration in *Phyllostomus discolor* genome. a) Simplified figure showing forward oriented sequence architecture with all detected genes and sequence features: LTRs (orange), Gag (dark pink), Pol (teal), Env (light pink), and stop codons (black bars). All genes in full ORFs marked with arrows. Genes containing stop codons marked with squares b) Placement of each viral-like sequence in the *Phyllostomus* genomic region together with all the ORFs (blue) in the right frame and direction. C) First reciprocal tBLAStn match for the Env protein

Extending the detected region of the alpharetrovirus-like *env* sequence in *Phyllostomus* genome (Phyllostomus_5 92481255 - 92481824) resulted in detection of full Gag, and fragmented Pol protein sequences. All proteins are placed between two full LTRs (**Figure 3.5**). Both *pol* and *gag* regions returned matches to gammaretroviruses in reciprocal BLAST, while the *env* sequence returned alpharetrovirus hit and branches closely with alpharetroviruses on the envelope phylogenetic tree reconstruction (**Figure 3.8** and **Table 3.1**). This might suggest detection of the first recombinant gamma-alpha sequence found in mammals or a new retroviral class. Previous reports identified gamma-alpha like sequences in birds, with an acquisition of the gamma-like *env* [16].

Recombinant gamma-alpha retrovirus					
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**	**Viral family**
ENV	scaffold_m19_p_5 92481255 92481824	191	NC_005947.1	Avian endogenous retrovirus, EAV-HP	Alpharetrovirus
POL	scaffold_m19_p_5 92483093 92484293	1210	NC_014474.1	Feline Leukemia virus	Gammaretrovirus
GAG	scaffold_m19_p_5 92487469 92487769	299	NP_955585.1	Moloney Murine Leukemia Virus	Gammaretrovirus

Table 3.1: Gamma-alpha retrovirus recombinant data. The table shows the genomic location, length of each protein, best viral match generated with reciprocal BLAST, and which viral family it corresponds to.

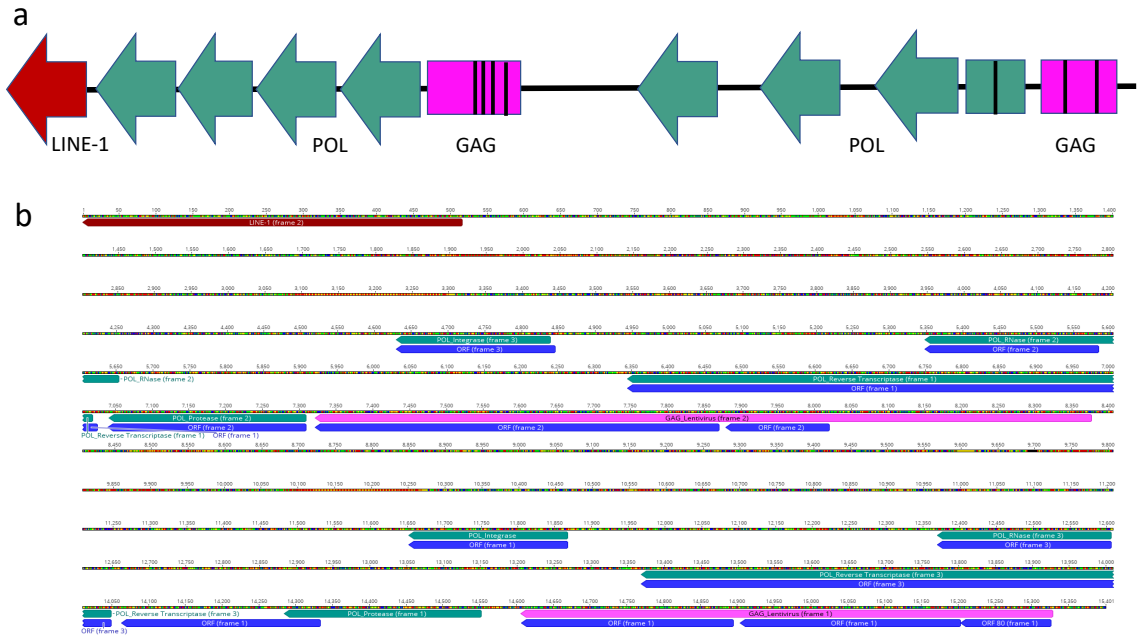


Figure 3.6: Lentiretrovirus-like integration in *Pipistrellus kuhlii* genome. a) Simplified figure showing reverse oriented sequence architecture with all detected genes and sequence features: Gag (dark pink), Pol (teal), LINE-1 (red), and stop codons (black bars). All genes in full ORFs marked with arrows. Genes containing stop codons marked with squares b) Placement of each viral-like sequence in the *Pipistrellus* genomic region together with all the ORFs (blue) in the right frame and direction.

Lentivirus-like integration found in the *Pipistrellus* genome showed duplications of *gag* and *pol* regions in a close proximity to the LINE-1 transposable element (**Figure 3.6**).

In two of the analysed genomes we also found epsilonretrovirus-like integrations. One of the integrations in *Pipistrellus* genome contained fragmented *gag* and *pol* sequences.

While integration in the *Molossus* genome contained fragment of reverse transcriptase Pol gene.

To identify historical ancestral transmission events, we reconstructed a phylogenetic tree from our recovered ERVs with the known viral protein 'probe' sequences (**Table A 2.1**) for all six bat genomes and seven mammalian outgroups (**Figure 3.7** and **3.8**). The majority of sequences group as single bat species clusters, suggesting that relatively recent integration events, more than ancestral transmission (**Figure 3.7** and **3.8**) govern the ERV diversity. While most ERVs are simple retroviruses, consisting of *gag, pol* and *env* genes, we found an unusual diversity of complex retroviruses in bats, which are generally rare in endogenous form [27, 72, 87] (**Figure 3.7** and **3.8**). We detected a clade of 5 *Rhinolophus pol* sequences clustered together with reference foamy retroviruses – Feline Foamy Virus (FFV) and Bovine Foamy Virus (BFV). Foamy retroviruses in bats were detected before in metagenomic data from *Rhinolophus affinis* [258], however, until now it was unclear whether these sequences represented exogenous or endogenous viruses [259]. With the detection of these sequences, we can now confirm the presence of endogenous spumaretroviruses in the *Rhinolophus ferrumequinum* genome, which furthers our understanding of the historical transmission dynamics of this pathogen. We also detected *pol* and *env* sequences in the *Molossus* genome clustering closely with reference delta sequences (Bovine Leukemia virus – BLV, Human T-lymphotropic Virus – HTLV). *Pol* and *env* regions for delta retroviruses in bats have not been detected before, with only partial *gag* and a single LTR identified previously in *Miniopterus* and *Rhinolophus* species [87, 88]. On the *env* phylogenetic tree we detected a cluster of sequences close to alpharetrvirus reference sequences from three different species of bats - *Rhinolophus, Rousettus* and *Phyllostomus*, which might suggest wide distribution of alpharetroviruses in bat species.

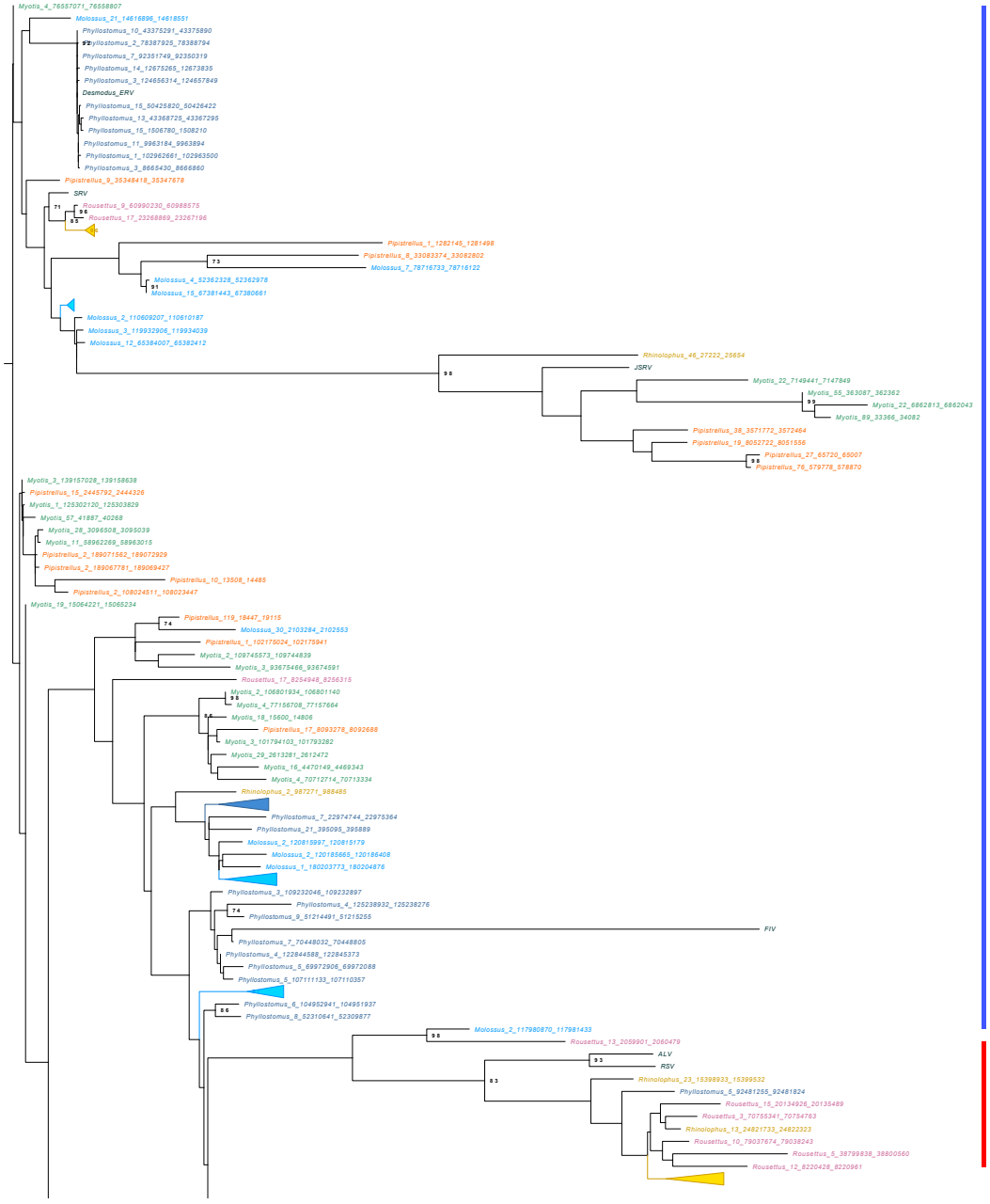


Figure 3.7: Reconstructed phylogenetic tree of viral *pol* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink), and reference sequences (black). Bootstrap values are shown where the values are ≥70%. The tips of the phylogeny are labelled with the species name, position in the reference genomes and the direction of the sequences (N- negative/P- positive). Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. Continued on the next page.



Figure 3.7: Continued. Reconstructed phylogenetic tree of viral *pol* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink), and reference sequences (black). Bootstrap values are shown where the values are ≥70%. The tips of the phylogeny are labelled with the species name, position in the reference genomes and the direction of the sequences (N- negative/P-positive). Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity.

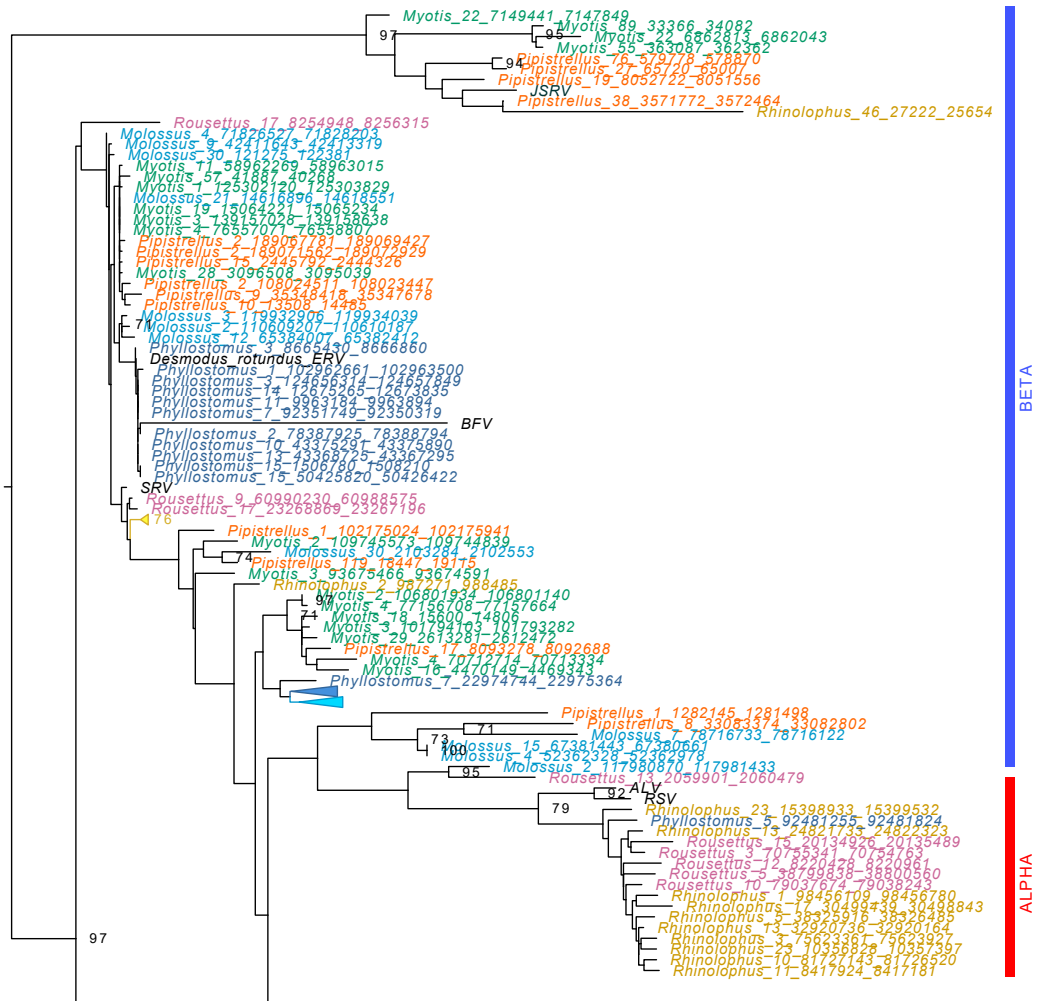


Figure 3.8: Reconstructed phylogenetic tree of viral *env* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink) and reference sequences (black). Bootstrap values are shown where the values are $\geq 70\%$. The tips of the phylogeny are labelled with the species name, and the position in the reference genome. Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity. Continued on the next page.

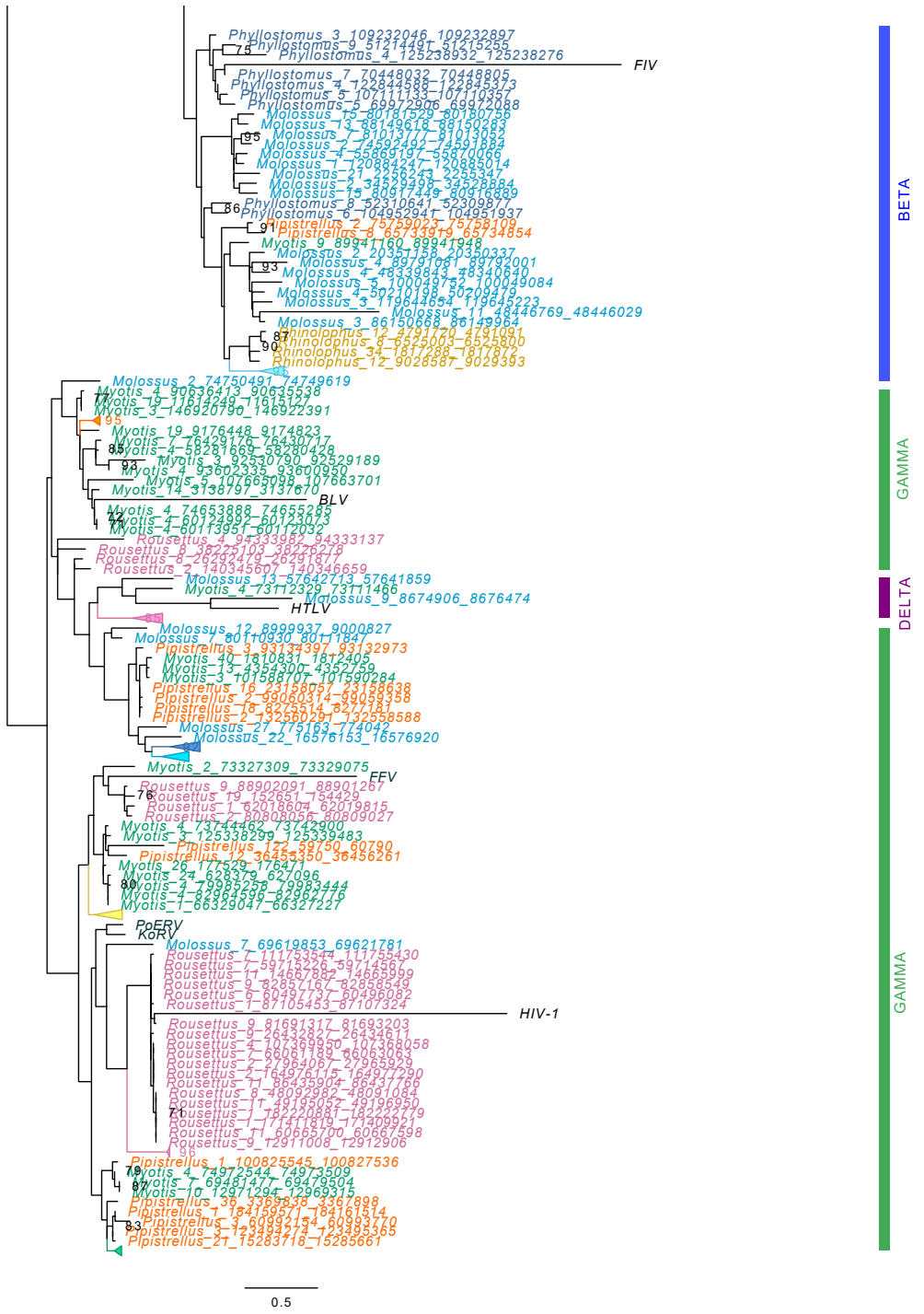


Figure 3.8: Continued. Reconstructed phylogenetic tree of viral *env* sequences found in six species of bats. Each species used in the ERV search is marked with a colour: *Phyllostomus* (navy blue), *Myotis* (green), *Pipistrellus* (orange), *Rhinolophus* (yellow), *Molossus* (light blue), *Rousettus* (pink) and reference sequences (black). Bootstrap values are shown where the values are $\geq 70\%$. The tips of the phylogeny are labelled with the species name, and the position in the reference genome.Coloured bars on the right hand side mark branches belonging to the specific retroviral families. Clades containing sequences from the same species collapsed for clarity.

3.4 Discussion

Although recognised as viral reservoirs [62] little is known about how bats survive a viral infection, nor the diversity of viruses present within individuals as they age [260]. Despite mounting evidence, some researchers have questioned if bats are really 'special' viral reservoirs and tolerant to disease or rather host many viruses due to large population sizes and their high mobility [261].

Bats can be persistently infected with different viruses while not showing any symptoms of disease, despite having high virus titres in their tissues, and can shed virions for months or years after primary infection [200]. Whether bats evolved to tolerate viral infections or to limit viral replication without developing symptoms of the disease is still a matter of debate. In fact it seems as though bats can intertwine these two strategies quite successfully [138, 262, 263]. Bat's have evolved several mechanisms to limit excessive inflammation, a process which can be involved in the response to disease and can also damage their DNA during flight, a process which generates high metabolic pressures [77, 140, 154, 264]. For example, bats lost the PYHIN gene family which is involved in inflammasome sensing [154]. The NLPRP3 inflammasome sensor which is involved in pathogen recognition has also been dampened in bats [264]. These two adaptations suggest that bat's have evolved to tolerate viral infections and to not develop symptoms of the disease. On the other hand, bats constitutively express three interferon genes, even in unstimulated bat tissues [155]. Interferons (IFNs) are antiviral defensive proteins that limit viral replication. In contrast, constant expression of IFNs might be a driver behind the high pathogenicity of bat viruses when they jump to other species [140]. Bat-borne viruses might have adapted to high IFN environments present in bat cells. Upon infecting another host which does not display such conditions, these viruses can quickly overwhelm the new host's antiviral response which can lead to elevated replication and

much higher pathogenicity [140].

Our results additionally confirm the special role of bats as virus reservoirs among other mammals. Apart from detecting a vast array of different classes of retroviruses and viruses from other families, we detected a novel type of retroviral recombination with gamma-like *pol* and *gag*, but alpha-like *env*. Alpha-like sequences were not found previously in mammals and gamma-alpha recombinations were reported only in birds [16]. *Env* genes are a key factor which determines the host range of the retrovirus, and *env* gene recombination might be an important factor for the virus in gaining new hosts. While cross-species transmissions between vertebrates are believed to be an extremely rare occurrence [16, 76], our finding shows that there might be a novel retroviral lineage present in bats which at some point in its evolutionary history gained an alpharetroviral *env*. Alpharetroviruses have to date been seen exclusively in birds, but gammaretroviruses have been seen previously in both birds and mammals. We hypothesize that at some point in its evolution a mammalian retrovirus gained alpharetrovirus *env* with the ability to jump between bat and bird taxa. Other recombinant retroviruses have been seen in bats before [76], so a discovery of a completely new class of recombinants only strengthens the unique position the *Chiroptera* order has in regards to its relationship with viruses, both exogenous and endogenous.

Overall these results show that bat genomes contain a surprising diversity of ERVs, with some sequences never previously recorded in mammalian genomes, confirming interactions between bats and complex retroviruses, which endogenize exceptionally rarely. This suggests that our high-quality genomes can enable the discovery and classification of novel viruses, not possible before. These integrations are indicative of past viral infections, highlighting which viruses bat species have co-evolved with and tolerated, and thus, can help us better predict potential zoonotic spillover events and direct routine

viral monitoring in key species and populations. In addition, bats, as one of the largest orders of mammals, are an excellent model to observe how co-evolution with viruses can shape the mammalian genome over evolutionary timescales. For example, recent analysis looking at APOBEC3 hypermutation in mammals proved that these class of genes evolved in response to conflict with retroviruses [23]. The expansion of the APOBEC3 genes in bats reported in several studies could well be a result of a co-evolutionary arms race shaped by ancient retroviral invasions, and could contribute to the restriction in copy number of endogenous viruses in some bat species, or advanced decay of ERV fragments in their genomes [22, 24, 174].

3.5 Conclusions

All six analyzed bat genomes contain endogenous viral sequences not found previously in other mammalian species. Bats, as one of the oldest and largest orders of mammals, can be a very good model to observe how the long term co-evolution with viruses can re-shape a mammalian genome over a long period of time. All bat species evolved independently for a very long time so the results regarding ERVs and EVEs might differ in each one of them giving a possibility to perform cross-species analysis and examine the changes in number and possible exaptation of those sequences over time and in different environments. Given that our findings were generated from only six bat genomes we can be confident that further cross-species comparison with similar quality bat genomes will bring even greater insight.

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

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Six reference-quality genomes reveal evolution of bat adaptations.
Publication Status	Published
Publication Details	Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A. and Burkitt-Gray, L., 2020. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*, *583*(7817), pp.578-584.

Student Confirmation

Student Name:	Emilia C Skirmuntt		
Contribution to the Paper	Wrote the endogenous retroviruses section and performed all of the analyses on the endogenous retroviruses.		
Signature Emilia Skirmuntt		Date	16.06.2021

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: Aris Katzourakis		
Supervisor comments I note that the version of the chapter in the thesis has been rewritten substantially by the student from the perspective of the section of a large multi authored consortium that they were part of.		
Signature 	Date	17.06.2021

4 Toll-like receptors in bat species

4.1 Introduction

In order to mount an effective immune response against a pathogen, it first needs to be detected by the immune system of the organism. The immune system can be divided broadly into two subsystems – innate and adaptive. The innate immune system provides a preconfigured response driven by a smaller number of germ-line encoded receptors. The adaptive immune system provides a more tailored response driven by millions of clonally constructed specific receptors. But, similarly to the adaptive immune system, the innate immune system also possesses the ability to differentiate non-self structures from those of the host. Toll-like receptors (TLRs) are a class of pattern recognition receptors (PRRs) conserved across the animal kingdom. They are located at the cell surface or in endosomal compartments (TLRs 3, 7, 8, 9) which makes them the first line of defence against several different kinds of pathogens, such as viruses, bacteria or fungi. TLRs recognise molecules called damage-associated molecular patterns (DAMPS), which are released from damaged or dying cells, and motifs called pathogen-associated molecular patterns (PAMPS), such as nucleic acid ligands (e.g. unmethylated DNA, double stranded RNA - dsRNA, single stranded RNA - ssRNA) present in invading pathogens.

Following PAMP detection, PRRs directly activate dendritic cells (DC), which are sentinel cells of the immune system. This activation serves as a bridge between the innate and adaptive immune systems, since DC activation is crucial in inducing primary T-cell responses [265, 266, 267, 268, 269].

There are several genes in the TLR family but all are structurally similar [270, 271, 272] comprising three regions:
- an ectodomain formed from 19-26 leucine-rich repeats (LRR) capped with N- and C-terminal ends (LRR-NT and LRR-CT) which interacts with microbial components
- a transmembrane domain (TM)
- a TIR domain, the TLRs cytoplasmic tail responsible for activating intracellular pathways

The TIR domain is a conserved domain shared by TLRs and Interleukin-1 receptor. When PAMPs are recognised by the PRR-like TLRs it results in the activation of a range of different signalling pathways alternating gene expression and initiating the production of pro-inflammatory cytokines, chemokines, and interferons (IFNs) [150, 273, 274, 275]. In regards to the majority of TLRs, when the receptor recognises the pathogen it recruits cytoplasmic TIR-domain adaptor proteins [Myeloid differentiation primary response protein (MyD88) and TIR-domain-containing adapter-inducing interferon β (TRIF)] to activate Nuclear factor κB (NF-κB) signalling (**Figure 4.1**).TLR3 does not use the MyD88 dependant pathway, but instead by using the TRIF-dependant pathway it can elicit anti-pathogenic and antiviral responses activating transcription factors like NF-κB, Activator protein 1 (Ap-1), and IFNs [150, 267, 273, 274, 275] (**Figure 4.1**). The majority of cells in the body express TLRs, but the largest repertoire of them is displayed by phagocytic cells (e.g. dendritic cells, neutrophils, and macrophages), the first immune cells which encounter pathogens [267, 269]. NF-κB in bats was recently

found to be altered with several components and genes missing or under strong positive selection [24] which may contribute to immune adaptations in the *Chiroptera* order [24].

In general animals have between 10 to 13 TLRs, with 11 identified in humans and 12 in mice. There are additional types of TLRs found in other orders of animals with TLR15 and 21 found in birds; 14, 16, 21, 22 in amphibians; and TLRs 14, 21, 22, 23 in fish. Some of them can also be missing, for example TLR8 and 9 are not observed in birds [276], and TLR4 is absent in several fish and amphibians [277]. TLRs present in their full length in certain species can be represented just as a pseudogenes in others, for example TLR11 and 12 pseudogenes in humans [278]. TLRs can be grouped according to the agonists they can recognise: TLRs 3, 7, 8, 9, 21 recognise nucleic acids, with 3, 7, and 8 targeting RNA (typically virally derived but in the case of TLR 7 also RNA of bacterial origin [279]) (**Figure 4.1**). TLR9 in mammals and 21 in non-mammals respond to viral, bacterial or protozoan DNA [273]. TLRs 1, 2, 4, 5, 6, 10, and 11 are expressed on the surface of cells and respond to protein, lipid and carbohydrate agonists in bacteria, fungi and protozoa [273], while 2 and 4 can also bind viral proteins [280, 281]. TLR13 senses ribosomal RNA [282, 283] and its interactions with molecules involved in protein transport suggest its ligands might have not yet been fully identified, and it may have a role in viral recognition [275, 284, 285].

Exogenous pathogens are not the only things which can be recognised and controlled by TLRs. Substantial portions of vertebrate genomes are made up of ERVs which are 'fossilized' remnants of past retroviral infections co-opted and preserved through generations [27]. Most often those genomic left-overs are non-functional and silenced, but occasionally they can preserve their pathogenic functions and be re-activated through different mechanisms, for example in extracellular re-infection of germ-line cells or retro-

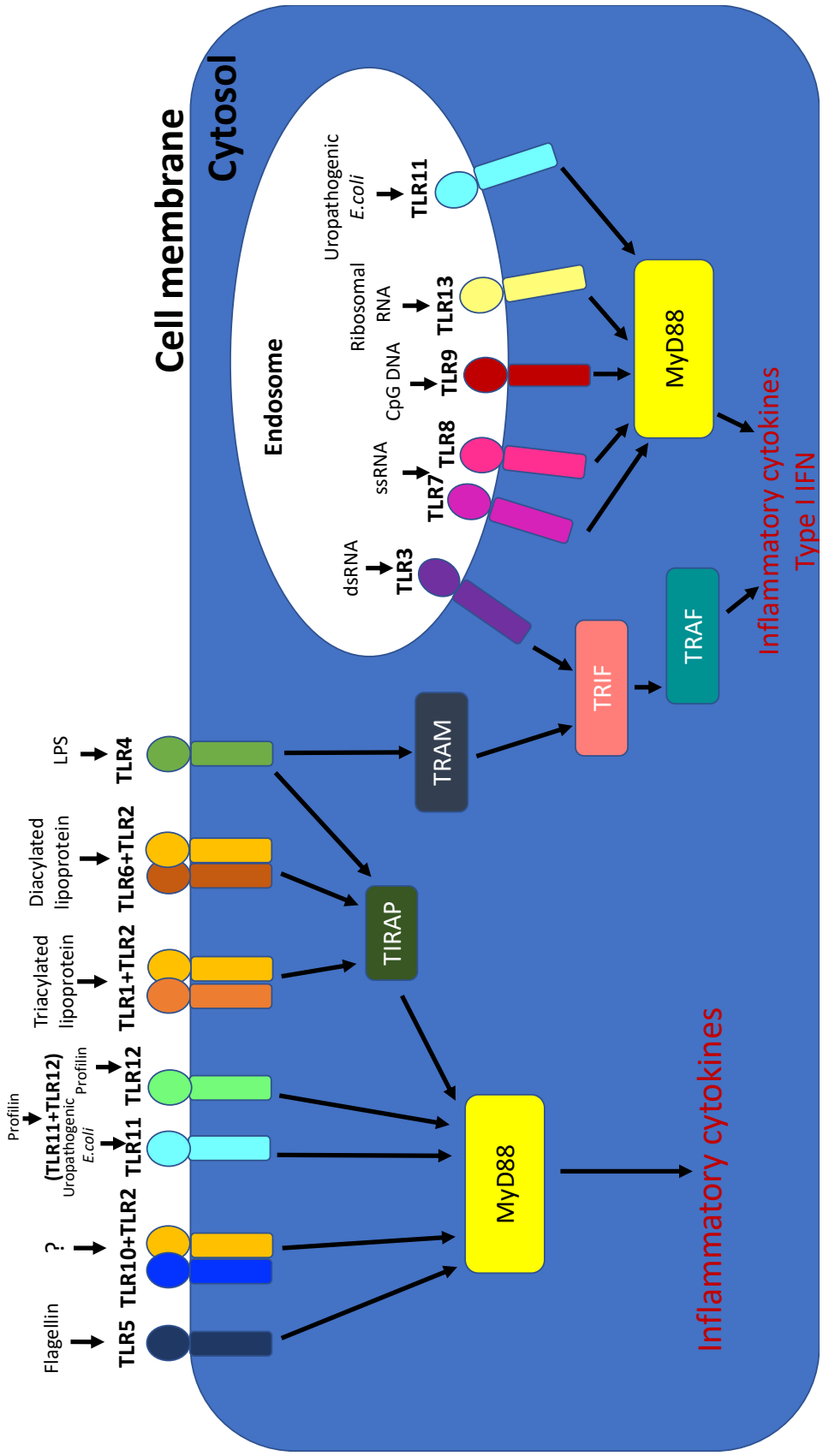


Figure 4.1: Localization of Toll-like receptors 1-13 with examples of agonists and signalling pathways. TLRs 1, 6, and 10 form heterodimers with TLR2 which are localised on the cell membrane. TLRs 4 and 5 form homodimers on the cell membrane while TLRs 11 and 12 can form both homo- and heterodimers and are localised on the cell membrane, while TLR11 can be also present on the endosome. TLRs 3, 7, 8, 9, and 13 are localised in endosomes.

transposition [274, 286, 287]. Mice deficient in nucleic acid sensing TLRs (NA-TLRs) showed enhanced murine endogenous retrovirus (MuERV) mRNA expression and viral particle formation [288]. The absence of TLR7 resulted in increased MuERV *gag* mRNA levels and gp70 Env protein expression [288]. A deficiency in IFN regulatory factor 5, which is a signalling molecule for TLRs 7 and 9, also leads to increased gp70 *gag* and *env* mRNA levels [288]. NA-TLR deficient mice developed acute T-cell lymphoblastic leukaemia, suggesting that the combined effects of these TLRs can provide protection against ERV-induced tumours [288]. TLR7 is also involved in the germline centre reaction which facilitates antibody production and class switching in acute response to infection [289, 290], as well as upregulation of type I IFNs [291]. Mice lacking this TLR were unable to produce antibodies to ERVs. ERVs can be activated sporadically and TLR7 seems to be able to induce an antibody response and control ERV-induced viraemia. NA-TLRs also promote autoimmune responses activating autoreactive B cells which could include ERV antigens and other auto-antigens [292].

Although highly conserved the ligand recognition properties of TLRs can have subtle variations in different species [293], which in turn might have driven the evolutionary pathways of infectious diseases [278]. Recent developments in whole genome sequencing showed that bats can have very unique immune responses against viruses and that they lack some gene families involved in inflammatory pathways implicated in antiviral responses [24, 77, 154], as well as constitutive upregulation of antiviral cytokines like IFNs [138, 262]. Since bats are one of the largest and oldest orders of mammals, with representatives living in a vast array of different environments and ecological niches, we expect that there may be some characters around the PRRs - in particular the TLRs - that might also differ from other mammalian orders as well as between *Chiropteran* species. In the search for bat TLRs we analysed six reference-quality bat genomes as well as less robust older genome assemblies of another nine bat species, and compared the

found sequences to reference sequences of other mammals. We found that bat species and families differ between each other when it goes to the numbers and types of TLRs present.

4.2 Materials and Methods

4.2.1 Database search, sequence alignments and phylogenetic tree reconstruction

With BLAST [211] we searched Refseq bat genome assemblies with a set of TLR reference sequences from other mammals (**Table A 9.1**). We analysed fifteen bat genomes (*Myotis myotis* [GCA_014108235.1], *Molossus molossus* [GCA_014108415.1], *Pipistrellus kuhlii* [GCA_014108245.1], *Phylostomus discolor* [GCA_004126475.3], *Rhinolophus ferrumequinum* [GCA_004115265.2], *Rousettus aegyptiacus* [GCA_014176215.1], *Desmodus rotundus* [GCA_002940915.2], *Pteropus alecto* [GCA_000325575.1], *Pteropus vampyrus* [GCA_000151845.2], *Hipposideros armiger* [GCA_001890085.1], *Eptesicus fuscus* [GCA_000308155.1], *Myotis lucifugus* [GCA_000147115.1], *Myotis brandtii* [GCA_000412655.1], *Myotis davidii* [GCA_000327345.1], *Miniopterus natalensis* [GCA_001595765.1]), and reference sequences of other mammals (*Homo sapiens* [GCA_000001405.28], *Eqqus caballus* [GCA_002863925.1], *Rattus norvegicus* [GCA_015227675.2], *Mus musculus* [GCA_000001635.9], *Felis catus* [GCA_000181335.4], *Canis lupus familiaris* [GCA_014441545.1], *Sus scrofa* [GCA_000003025.6], and *Bos turus* [GCA_002263795.2]).

In Aliview software [4] we manually aligned bat sequences together with the reference sequences for each of the separate TLR sets. In addition to this, we made alignments

of TLR1/6/10, and TLR7/8, as well as an alignment of all TLRs's TIR domains. Prepared alignments were run with Prottest (ver.3.4.2) [214] and the best models for each set (HIVb+G for TLR1, 2, 3, 6, 10, 12, 13; HIVb+I+G for TLR4, 7, 11, and combined alignment of TLR1/6/10; JTT+I+G for TLR5, 8, 9, and combined alignments of TLR7/8, and TIR domains) were later used to generate a phylogenetic tree reconstruction using the maximum likelihood method in RAxML software (ver.8.2.12) [6]. RAxML trees were visualised with FigTree (ver.1.4.4; http://tree.bio.ed.ac.uk/software/figtree). Generated trees were used later for the PAML selection analysis [217].

4.2.2 Domain architecture prediction

We used SMART (Simple Modular Architecture Research Tool) [294, 295] as a guide to determine the domain architecture of all found TLRs and pseudogenes from both the bat and the reference sequences (**Figures 4.15** to **4.27**). SMART identifies putative protein domains through homology search and predictive algorithms. All SMART predictions were then checked manually. We used TMHMM Server (ver.2.0; http://www.cbs.dtu.dk/services/TMHMM/) to identify the ectodomains, transmembrane domains, and TIR domains of the bat sequences. Domains detected by SMART and TMHMM were annotated on the alignment of the consensus *Chiroptera* sequences generated by the Geneious R10 software (ver.10.1.3; http://www.geneious.com, [5]) (**Figure A 9.1** to **A 9.13**).

4.2.3 Selection analysis

We tested all TLR sequence sets to ascertain which type of selection they are under. Under neutral selection the dN/dS (ω) ratio (dN: number of non-synonymous substitutions

per non-synonymous site/dS: number of synonymous substitutions per synonymous site) should not deviate from 1. Results below 1 can be interpreted as purifying selection, while above 1 can be read as indicating a positive selection [296]. The rates of dN/dS were estimated comparing several different models (**Table A 9.2**) using the *codeml* package from the PAML software (ver.4.9) [217] with the F3X4 codon frequency model, using phylogenetic trees previously reconstructed with RAxML software. First we tested whether a neutral evolution for the whole sequence can be rejected (M0 ω=1 vs. M0), then the M0 model (constant ω across all sites) vs. M3 (discreet model, which allows ω to vary among sites) to test if ω differs among sites. Further, we tested the M1 model (which assumes neutral selection) vs. the M2 model (assuming positive selection). In order to look for sites evolving under positive selection we used the M7 model (allowing codons evolving neutrally or under purifying selection) vs. the M8 model (assumes beta distribution for ω, which adds a class of sites under positive selection). All four nested model pairs were compared using a likelihood ratio test (LRT) with 1 degree of freedom (df) for the neutral evolution test, 14 df for the M0 vs. M3 test (apart from TLR6 set which uses 11 df), and 2 df for the M1 vs. M2 and M7 vs. M8 tests [3] (**Table A 9.3**).

The posterior probability of codons under selection in the M8 model was tested using the Bayes Empirical Bayes (BEB) algorithm [3]. In order to further validate persistent positive selection we ran several models from the Datamonkey server (www.datamonkey.org; [297, 298, 299]) on the bat TLR alignments and compared them to the PAML *codeml* M8 model (**Table A 9.5** and **Figures 4.3 - 4.5**). We implemented HyPhy models FEL (Fixed-Effect Likelihood) [299], SLAC (Single Likelihood Ancestor Counting) [299], and FUBAR (Fast, Unconstrained Bayesian AppRoximation) [300] to detect candidate codon numbers under positive selection. For the SLAC and FEL tests we considered codons with p-values <0.1, and for M8 and FUBAR codons with the posterior probabilities >0.9. Finally, to check for episodic positive selection we ran the bsREL model (Adap-

tive branch-site REL test for episodic diversification) [301] for bat and reference sequence alignments, and MEME model (Mixed Effects Model of Evolution) [302] for bat alignments (p-value threshold <0.05). The bsREL model estimates the proportion of sites which are under selection along phylogenetic tree branches at the same time allowing the evolutionary rate to vary across sites and branches [301], while MEME identifies lineage specific positive-selection events on specific sites, allowing the selection type to be purifying or neutral at the same site in other lineages [302].

The pairwise percent identity for each TLR set was calculated using the Geneious R10 software (ver.10.1.3; http://www.geneious.com; [5]) (**Table 4.3**).

4.3 Results

The complement of bat TLRs varies across species. All studied bat genomes have between 10 (*P. alecto* and *P.vampyrus*) and 13 TLRs (*M. natalensis, P.discolor*) (**Table 4.1**). TLRs 1, 2, 3, 4, 6, 7, 8, and 9 were detected in all fifteen bat species. TLR5 was present in all bats apart from *P.kuhlii* in which we detected a sequence corresponding to TLR5 but containing a stop codon at the position 532 aa. However, in this case we cannot rule out the possibility of a sequencing error. TLR10 is missing in the vesper bat genomes, but is still present in *M.natalensis* and *M.molossus*, which suggests that this gene was lost at some point after the *Vespertilionidae* family split from the *Miniopteridae* family (∼35-45 Mya [1]). TLR11 is missing in the three genomes belonging to the *Pteropodidae* family. TLR11 is present as a pseudogene in the six bat species from both *Yinpterachiroptera* (*H.armiger, R.ferrumequinum*), and *Yangochiroptera* (*M.davidii, M.lucifugus, D.rotundus, M.mollosus*) suborders. This pattern suggests that this gene might have been pseudogenised by an ancestor of *Rhinolophoidea*

(∼55-60 Mya [1]), and subsequently lost by the *Pteropodidae* family. The pseudogenization of the TLR11 sequence in some representatives of *Yangochiroptera* (*Vespertilionidae* and *Phyllostomidae*) families seem to be random and not connected with the deep phylogenetic relationships which exist between the species, and might suggest the impact of other environmental factors. Some of the representatives from these two families still display a functional copy of this gene (**Table 4.1**). We did not detect TLR12 as a functional copy or as a pseudogene in any of the *Yinpterachiroptera* genomes, which displayed an empty site between two flanking genes seen in other mammals. This suggests an ancient event which led to the loss of this gene after *Chiroptera* split into two suborders (∼55-65 Mya [1]). TLR12 was detected as a pseudogene in *D.rotundus* and *M.molossus*, which might suggest divergent loss of the TLR12 gene at the species level, since another *Phyllostomidae* representative (*P.discolor*), and the rest of tested *Yangochiroptera* bats still display functional copies of this receptor. TLR13, which is seen mostly in rodents, was detected in all bats apart from two of the tested *Pteropus* bat species in which it was detected as a pseudogene.

TLRs 3, 4, 7 and 9 contained introns of different lengths in all studied bat genomes (**Figure 4.2**). We have noticed differences between species in the length of the introns present, but not in the number of introns existing for each TLR, which also corresponded to the intron number seen in other mammalian species.

BATS	TLR1	TLR2	TLR3	TLR4	TLR5	TLR6	TLR7	TLR8	TLR9	TLR10	TLR11	TLR12	TLR13	Classification	Suborder	Family	Distribution
P.alecto											M	M	P	Megabat	Yinpterochiroptera	Pteropodidae	Australia
P.vampyrus											M	M	P	Megabat	Yinpterochiroptera	Pteropodidae	Malaysia
R.aegyptiacus											M	M		Megabat	Yinpterochiroptera	Pteropodidae	Africa
R.ferrumequinum											P	M		Megabat	Yinpterochiroptera	Rhinolophidae	Africa, Europe, Asia
H.armiger											P	M		Megabat	Yinpterochiroptera	Hipposideridae	Asia
E.fuscus										M				Microbat	Yangochiroptera	Vespertilionidae	America
M.lucifugus										M	P			Microbat	Yangochiroptera	Vespertilionidae	America
M.davidii										M	P			Microbat	Yangochiroptera	Vespertilionidae	China
M.myotis										M				Microbat	Yangochiroptera	Vespertilionidae	Europe
M.brandtii										M				Microbat	Yangochiroptera	Vespertilionidae	Eurasia
P.Kuhlii					P					M				Microbat	Yangochiroptera	Vespertilionidae	Africa, Europe, Asia
M.nattalensis														Microbat	Yangochiroptera	Miniopteridae	Africa
M.mollosus											P	P		Microbat	Yangochiroptera	Molossidae	America
P.discolor											P	P		Microbat	Yangochiroptera	Phyllostomidae	America
D.rotundus											P	P		Microbat	Yangochiroptera	Phyllostomidae	America
REFERENCES	TLR1	TLR2	TLR3	TLR4	TLR5	TLR6	TLR7	TLR8	TLR9	TLR10	TLR11	TLR12	TLR13				
H.sapiens											P	P	M				
M.musculus										P							
R.norvegicus																	
C.lupus											M	P	M				
F.catus											M	P	M				
S. scrofa												P	M				
B.taurus											M	P	P				
E.caballus													P				

Table 4.1: Summary of all TLRs present for each bat species and reference sequence. TLR presence shown in green, pseudogenes shown in pink, and missing genes shown in yellow. Each bat species is also described by general classification name, suborder name, family name, and distribution.

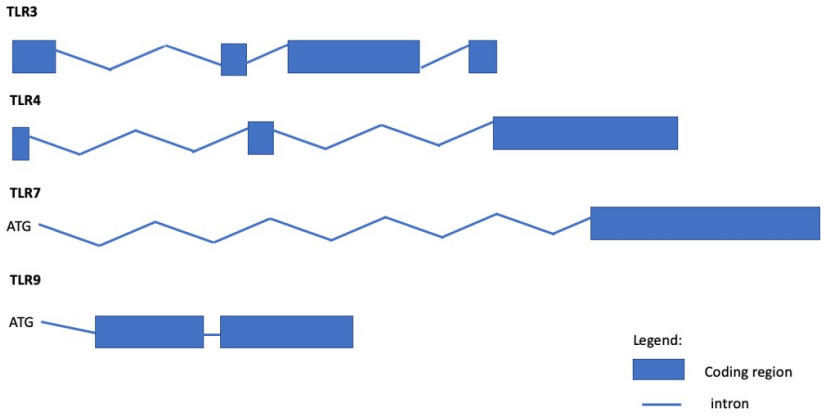


Figure 4.2: Schematic of intron-exon organisation of bat TLRs. Length of the introns based on the *Myotis myotis* sequences. TLR7 and TLR9 starts with the first coding exon containing only start codon (ATG) followed by an intron, which is similar to pattern present in other mammals.

All bat TLRs were under purifying selection, with the highest degree found in nucleic-acid sensing TLRs 3, 7, 9 as well as TLR5 which recognizes bacterial flagellin. dN/dS was below 1.0 for all of the TLRs, with results ranging between 0.1242 for TLR9 and 0.8425 for TLR11 (**Table 4.2**).

Posterior probability based on the *codeml* M8 BEB method showed the highest number of positively selected sites (PSS) (posterior probability of ≥0.90) in TLR11 (54 PSSs) and TLR8 (53 PSSs). The high number of PSSs was also detected in TLR4 (32 PSSs), TLR13 (27 PSSs) and TLR12 (21 PSSs) (**Figure 4.6**). TLR3 shows no PSSs with posterior probability of ≥0.90. Further comparison of the *codeml* M8 BEB method, with DataMonkey FEL, MEME, SLAC, and FUBAR models showed the highest number of sites under positive selection in TLR8 (79 in total, with 21 selected by more than two models) (**Table 4.3** and **Figure A 9.4**). This comparison also confirmed the lowest number of positively selected sites in TLR3 (20 sites in total, with 10 selected by two or more models). The highest percent of positively selected sites is present in TLR8

	dN/dS	2 Δ lnL (for df=1 and p-value=0.05)
TLR1	0.4464	3.0315
TLR2	0.4582	5.536632
TLR3	0.2764*	44.091464
TLR4	0.5423	31.616922
TLR5	0.284*	66.326224
TLR6	0.416	178.034216
TLR7	0.3342*	9.810122
TLR8	0.52491	35.53166
TLR9	0.1242*	597.908782
TLR10	0.465	4.634224
TLR11	0.8425	180.4276
TLR12	0.6862	78.211854
TLR13	0.5168	18.065392

Table 4.2: dN/dS results for all TLR sets *codeml* one-ratio model. *dN/dS value significantly below 1.0

(7.60%), and the lowest in TLR3 (2.21%) (**Table 4.3**). In summary all thirteen bat TLRs have 11628 aa with 454 PSSs, and of these 129 (28.41%) were detected by more than one method, which suggests strong positive selection pressures (**Table 4.3** and **Table A 9.4**). For all of the TLR genes the pairwise identity on nucleotide level is between 81.6-90.8% and on aa level between 77.8-88.9%, and the percentage of PSSs is 3.90% (**Table 4.3**). PSSs were mostly detected in the LRR binding regions (268 PSSs) and ectodomains (87 PSSs) (**Figure 4.3 and 4.6**). Most of the sites of negative selection were located in the TIR domains. The highest number of negatively selected sites was detected in TLR3, and TLR5 (**Table A 9.6**).

From all positively selected sites identified by the *codeml* M8 method thirty eight of them were detected previously in other mammalian species [303, 304, 305, 306], and an additional five in avian species [276, 307] (**Table A 9.5**). Two were detected in TLR1 (315, 317), seven in TLR2 (112, 114, 161, 322, 336, 454, 777), twenty three in TLR4 (3, 56, 239, 267, 269, 270, 320, 321, 323, 340, 342, 346, 365, 370, 383, 393, 459, 463, 473, 499, 543, 576, 632), three sites in TLR5 (312, 422, 530), and one in TLR10

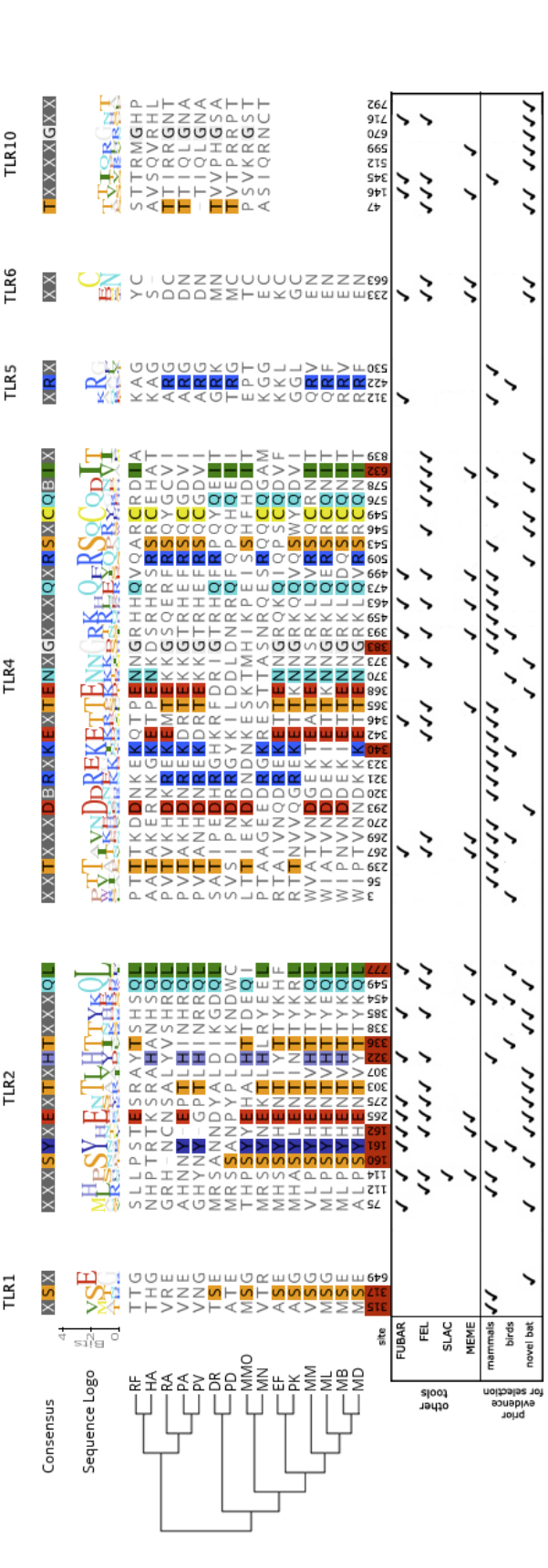


Figure 4.3: Positively selected site alignment for the TLRs 1, 2, 4, 5, 6, and 10. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *R. ferrumequinum*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*.

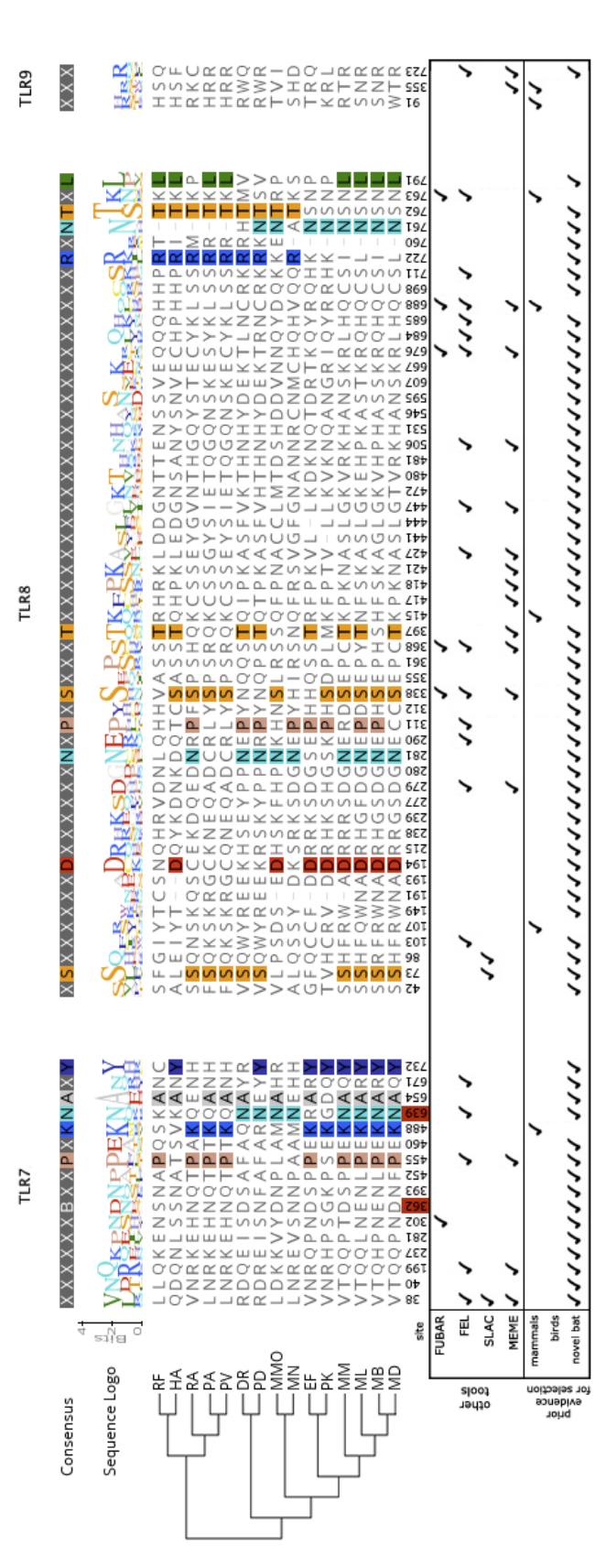


Figure 4.4: Positively selected site alignment for the NA-TLRs 7, 8 and 9. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *R. ferrumequinum*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*.

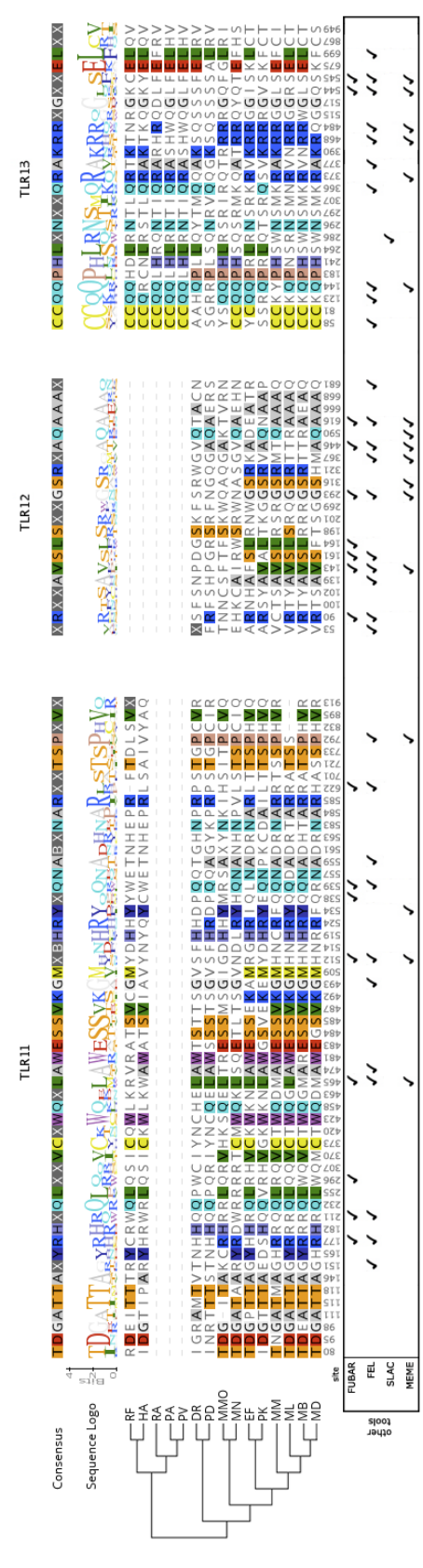


Figure 4.5: Positively selected site alignment for the TLRs 11, 12 and 13. Consensus aa and the aa coloured in the alignment are present in at least 50% of the sites displayed by each species. PSS position (according to the *M.natalensis*) given below the alignment; sites coloured in red has a known function. 'Other Tools' table indicate if the site was also found by a model alternative to *codeml* (FUBAR, FEL, SLAC, MEME). 'Prior evidence for selection' table shows if the site was reported previously as positively selected in either birds or mammals, or is it a novel site selected in bat species. Sequence was manually aligned in Aliview [4], and viewed in Geneious R10 [5]. The tree was generated in RAxML [6]. Abbreviations on the tree are for: RF, *Rhinolophus ferrumequinum*; HA, *Hipposiderios armiger*; RA, *Rousettus aegyptiacus*; PA, *Pteropus alecto*; PV, *Pteropus vampyrus*; DR, *Desmodus rotundus*; PD, *Phyllostomus discolor*; MMO, *Molossus molossus*; MN, *Miniopterus natalesis*; ED, *Eptesicus fuscus*; PK, *Pipistrellus kuhlii*; MM, *Myotis myotis*; ML, *Myotis lucifugus*; MB, *Myotis brandtii*; MD, *Myotis davidii*.

	Length of the seq (aa)	Pairwise % identity aa/nuc	Number of PSSs	Number of PSSs detected by more than one method	% PSS
TLR1	786	85.3/89.2	23	3	2.93%
TLR2	785	83.9/88.4	44	15	5.61%
TLR3	904	88.0/89.4	20	10	2.21%
TLR4	839	82.6/87.9	45	15	5.36%
TLR5	855	77.8/81.6	31	10	3.63%
TLR6	796	84.6/88.8	26	5	3.27%
TLR7	1045	88.9/90.8	26	4	2.49%
TLR8	1039	79.4/83.8	79	21	7.60%
TLR9	1005	79.3/85.0	23	3	2.28%
TLR10	811	84.6/89.2	20	4	2.47%
TLR11	917	80.5/87.5	54	13	5.78%
TLR12	896	81.6/88.5	31	16	3.45%
TLR13	950	87.0/90.5	33	10	3.47%
Sum	11628		455	129	3.90%

Table 4.3: Summary of the data for positively selected sites (PSS) for each set of TLRs. First column summarizes the length of each TLR sequence as number of amino acids (aa), second column shows pairwise percent identity of the sequences as both aa and nucleotides (nuc), third column shows the number of PSSs detected in each TLR by at least one method, fourth column displays the number of PSSs which were detected by at least two or more methods implying a strong positive selection pressure, the fifth column shows the percentage of PSSs across the whole gene.

(345) [276, 303, 304, 305, 306, 307] (**Figure 4.3**). From the NA-TLRs single site was detected in TLR7 (488), four in TLR8 (107, 415, 688, 763), and two in TLR9 (91, 355)[276, 304, 306] (**Figure 4.4**). Sites detected in TLR6, and TLRs 11-13 were not identified by previous studies in mammals and avian species (**Figure 4.5**). Some of the selected sites have proven function in the TLRs of other mammals and birds. Sites 315 and 317 in bats are corresponding to the sites responsible for TLR1 and TLR2 dimerization as well as lipopeptide binding [308]. TLR2 sites 160, 161, 162, 322, and 336 in other studies were found to be responsible for TLR2-(TLR1/TLR6) dimerization, lipopeptide binding, and dsRNA binding [308, 309]. TLR4 bat sites 340, 383 were found to act in LPS binding in both birds and mammals [310, 311, 312, 313], while site 632 is a part of the transmembrane domain for this receptor. Bat TLR7 site 362 is corresponding to ligand binding site [314], and site 639 is taking part in TLR7 dimerization [314].

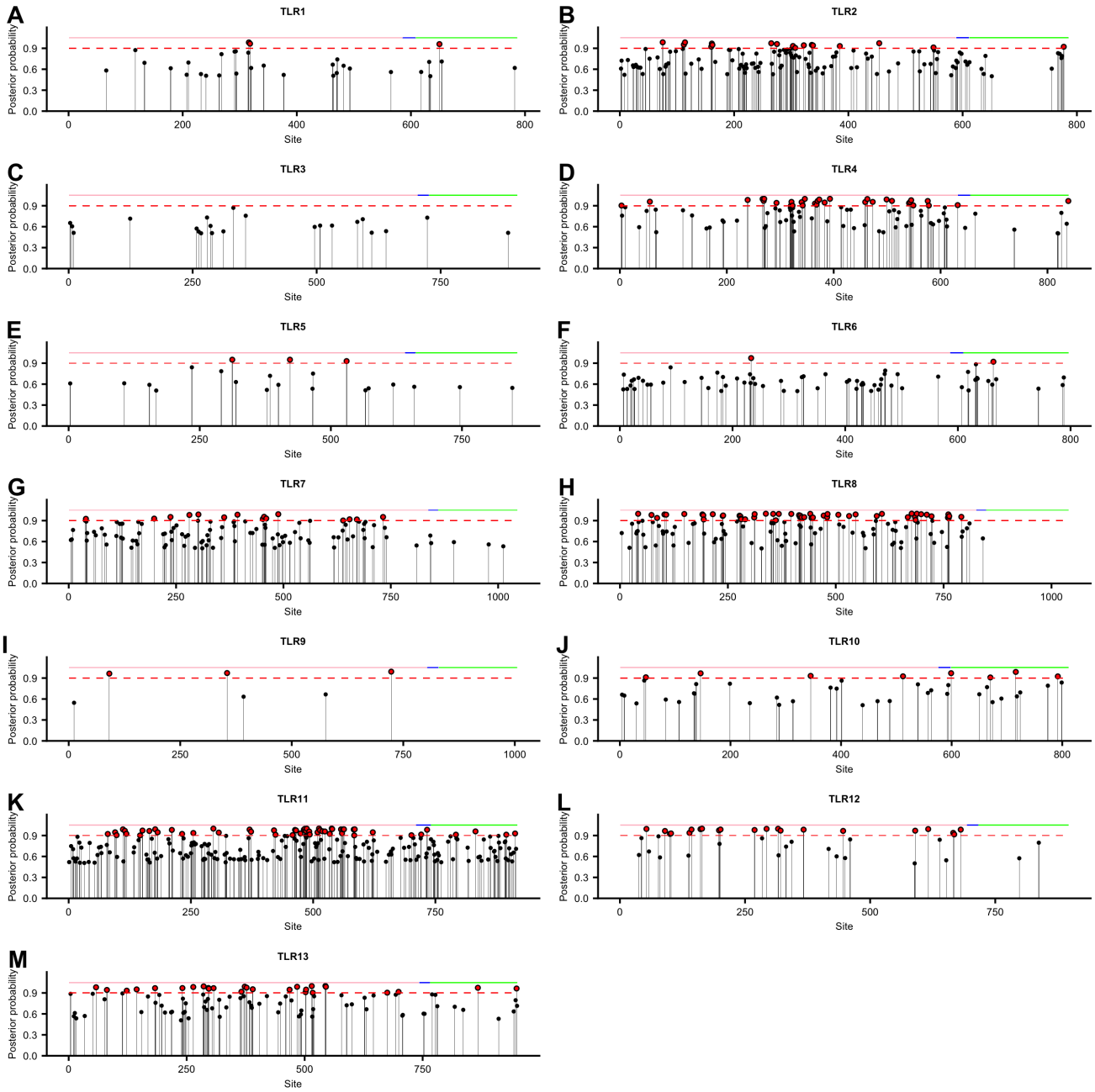


Figure 4.6: Positively selected sites in all bat TLRs. Posterior probability determined using the Bayes Empirical Bayes (BEB) method [3]. Red dots show sites with posterior probability of $\geq$0.90, and black dots show posterior probability of $\geq$0.50. Red line indicate extracellular domain, blue shows transmembrane domain, and green TIR domain.

SMART failed to identify all LRR domains in the analysed TLRs (**Figure 4.15 - 4.27**).

LRRs are diverse and each TLR is expected to contain one N-terminal and one C-terminal LRR flanking a variable number of LRR domains forming a ligand binding region of the receptor [271]. SMART detected N- and C- terminal LRRs only in a subset of all analysed sequences from bats and reference species, which was also observed previously in work identifying TLRs in *P.alecto* [152]. Missing LRR domains were manually corrected and annotated on the alignments (**Figure 9.1 - 9.13**), which resulted in a final TLR alignment containing between 16 (TLR10) to 25 LRR domains (TLR7, TLR8, TLR12).

Some groups of TLRs are highly similar across the whole sequence. One such group is that of the closely related TLRs 1, 6, and 10. Phylogenetic tree reconstruction of these three TLRs for both reference sequences and bats show three distinct clades for each distinct TLR. In each of these clades bats branch separately to the reference sequences, and within bats there is clear separation between *Yinpterachiroptera*, and *Yangochiroptera* (**Figure 4.7**).

Another closely related group of TLRs are nucleic acid sensing TLRs 7 and 8. The phylogenetic tree reconstruction for TLR7 and TLR8 shows a similar structure with separate clades for bats and reference sequences, and two separate branches for both bat suborders (**Figure 4.8**).

The ectodomain of different types of TLRs is usually highly divergent, but the TIR domain is conserved that makes it possible to align and compare all TLRs basing it on this part of the sequence. The phylogenetic tree reconstruction of TIR domains shows most of the separate TLR sequences branching together for both bats and other mammalian sequences (**Figure 4.9**). TLR 9 sequences branch together making three separate clades for reference sequences, *Yinptera*- and *Yangochiroptera*, with the exception of the *D.rotundus* sequence which branches separately from other bats. As expected TLR7

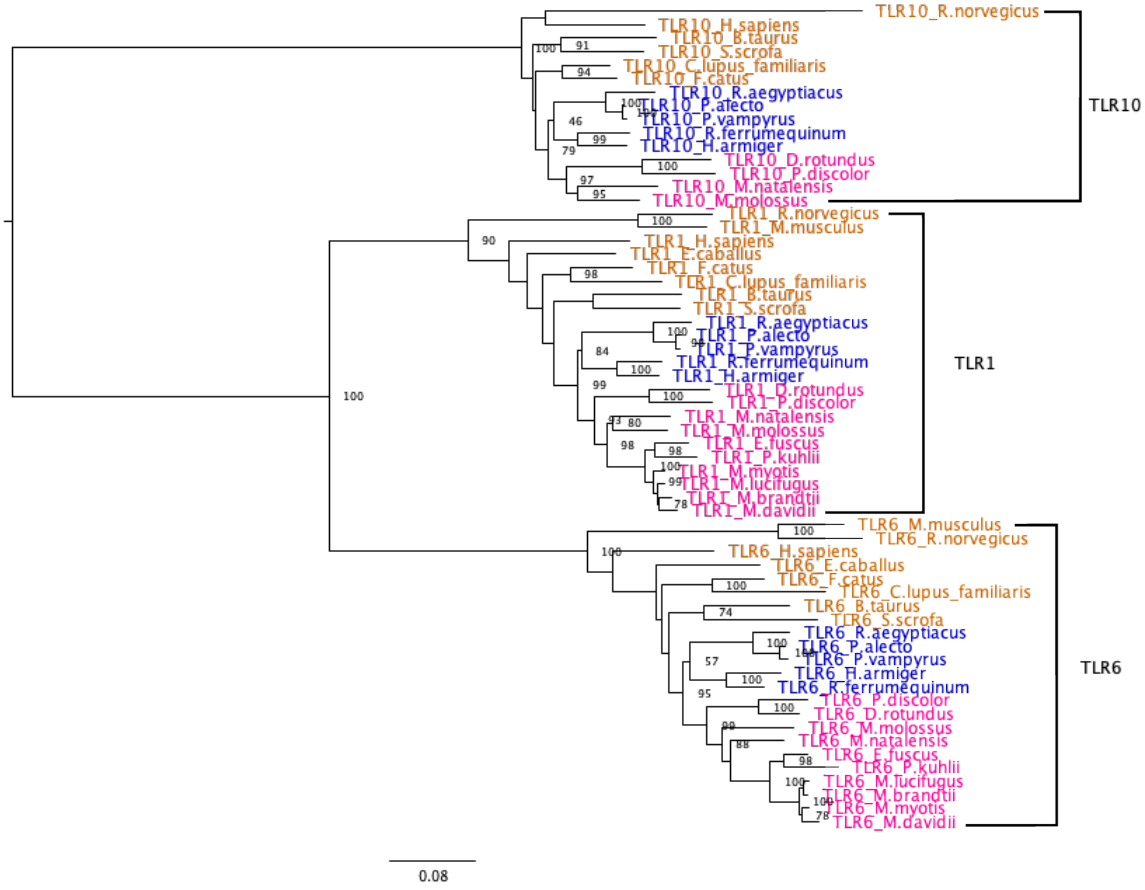


Figure 4.7: Phylogenetic tree reconstruction showing TLR1, TLR6, and TLR10 of bats and reference sequences. Tree is midpoint rooted. Mammalian reference sequences marked in orange, *Yinpterachiroptera* sequences marked in blue, *Yangochiroptera* sequences marked in pink.

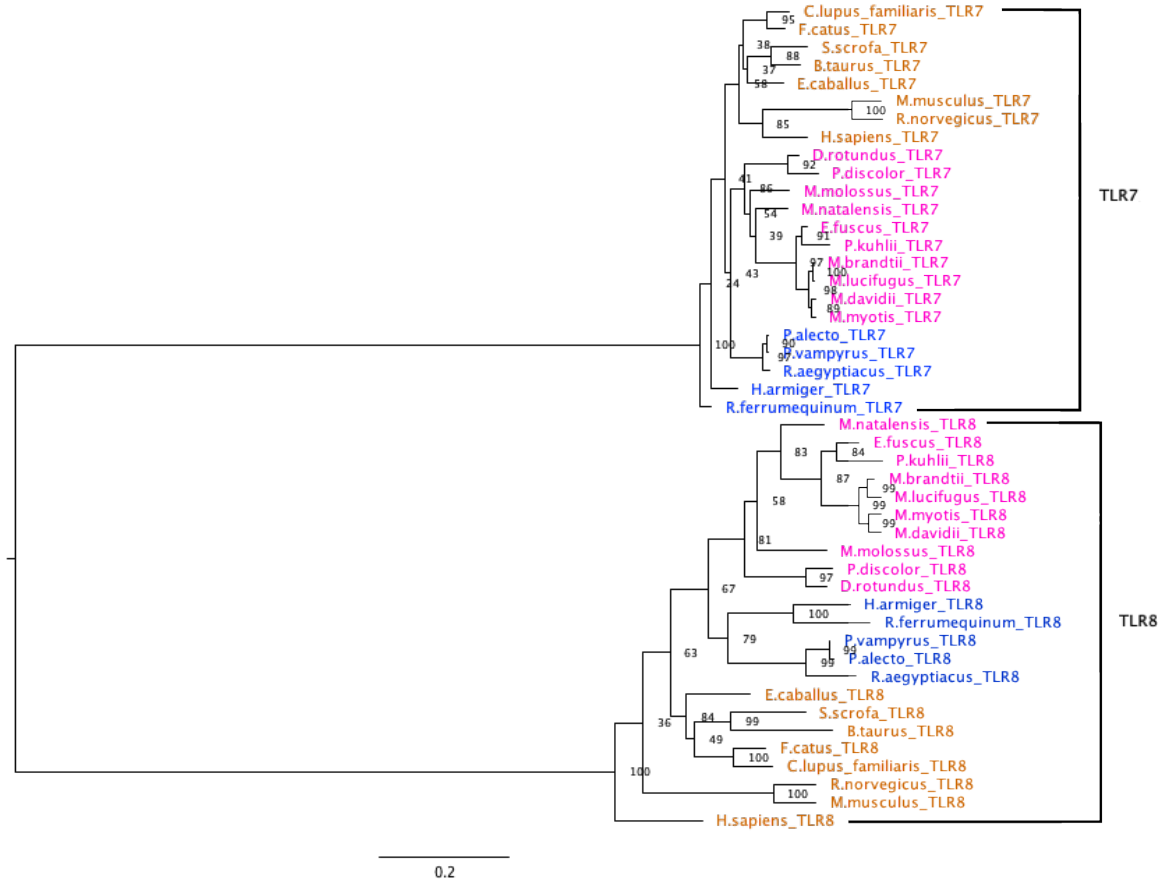


Figure 4.8: Phylogenetic tree reconstruction showing TLR7 and TLR8 within bat and reference sequences. Tree is midpoint rooted. Mammalian reference sequences marked in orange, *Yinpterachiroptera* sequences marked in blue, *Yangochiroptera* sequences marked in pink.

and TLR8 are placed on neighbouring branches also showing a clear distinction between reference sequences and bats. TLR7 bat sequences are highly similar, but reference sequences are placed on a very long branch far away from bats which might indicate distinctive changes in TIR domain of bats separating them from other mammals. TLR3 is placed on a separate branch from all of the other TLRs with *Yinpterachiroptera* sequences branching closer to other mammalian sequences than to *Yangochiroptera,* which might suggest changes in the TIR receptor after the separation of the two bat suborders. TLR2 is placed on a branch neighbouring that of TLRs 1, 6, and 10. It does not show clear separation between reference and bat sequences. Most of the reference sequences are placed between bat branches, apart from rodents which are placed separately. TLRs 1, 6 and 10 branch together. TLR10 shows separation between sequences of bats and other mammals. This is not reflected by TLR1 or TLR6 where mammalian sequences and bat sequences are assorted together. TLR1 and TLR6 rat and mouse sequences branch together which suggests that TLR6 rodent sequences are more closely related to TLR1 sequences than TLR6 sequences of other mammals. TLR4 and TLR5 are on neighbouring branches and display similar architecture as other sequences with separate groups of mammals and two bat suborders branching together. The same three distinct branch groups are visible in the case of two highly similar TLRs 11 and 12, as well as the more separate TLR13. As expected TLRs from the same families branch together, displaying separate groups for TLR 1/2/6/10, TLR 7/8, and TLR 11/12. TLR3 branches separately from other nucleic-acid sensing TLRs, and TLRs 4 and 5 which are representatives of separate families are placed on the neighbouring branches.

Additional phylogenetic tree reconstructions for each TLR separately showed strong support for all *Chiroptera* branches and other mammalian species (**Figures A 9.14-9.26**). All trees were rooted to the rodent outgroup. Relationships between reference species and bats, as well as relationships between bat species, were consistent in all

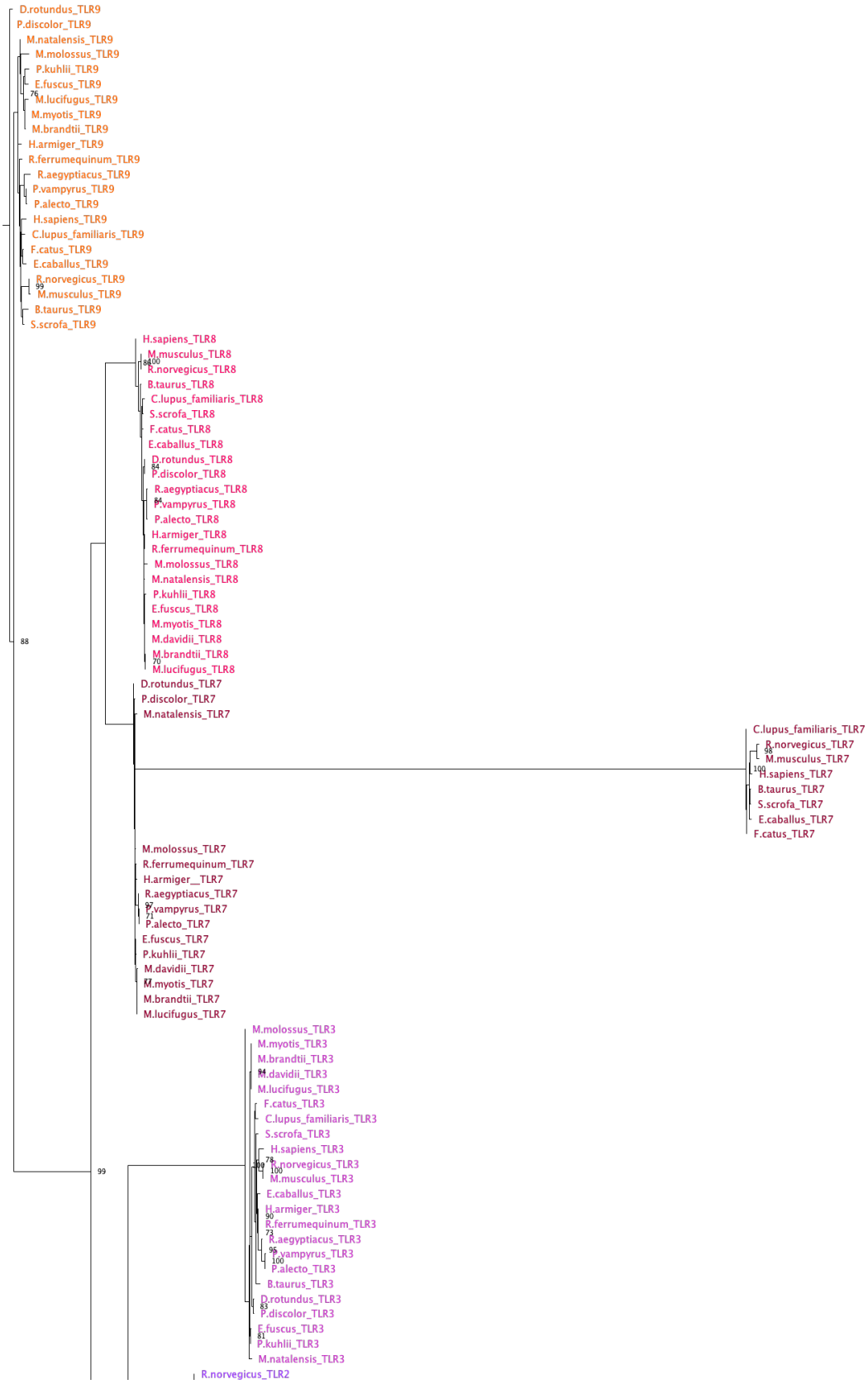


Figure 4.9: TIR domain tree for bats and all reference sequences. Each TLR set marked with a separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%. Continued on next two pages.

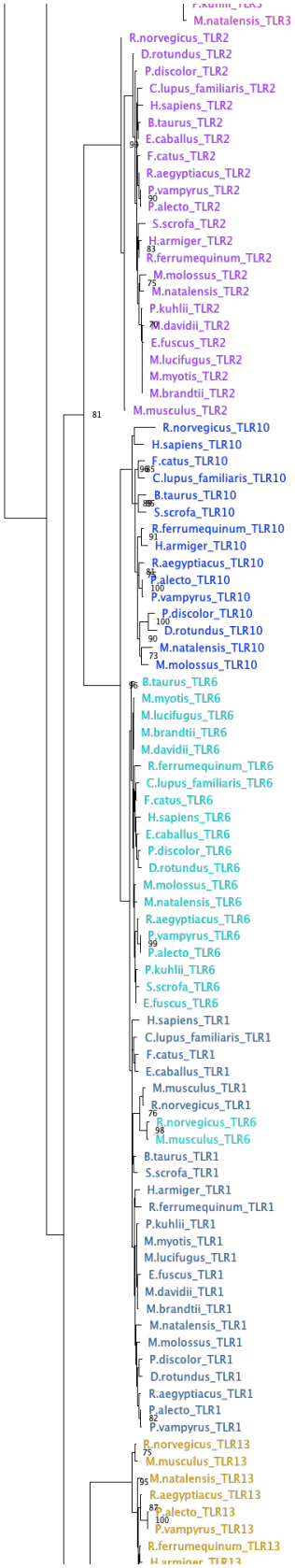


Figure 4.9: Continued. TIR domain tree for bats and all reference sequences. Each TLR set marked with separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%.

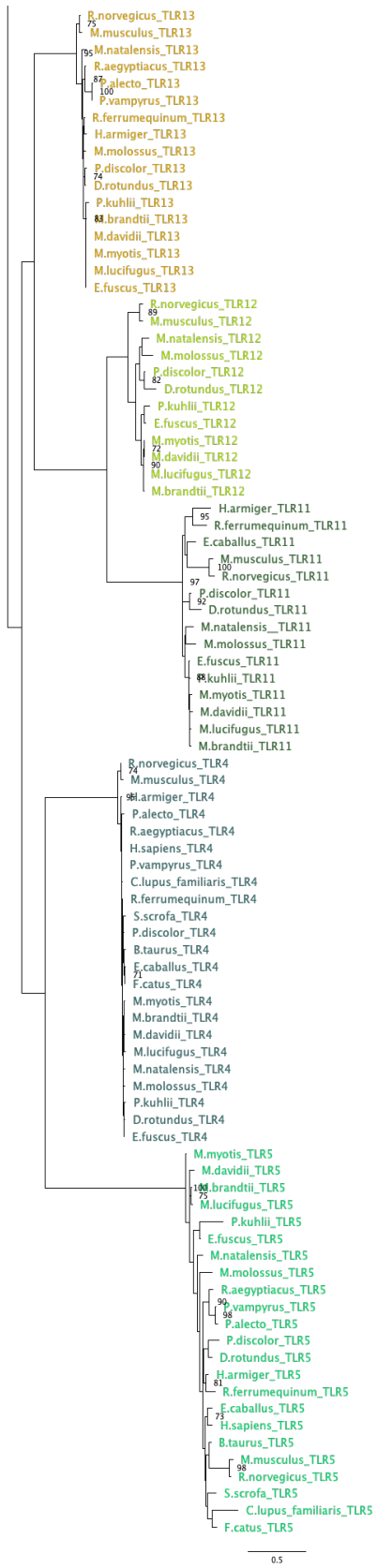


Figure 4.9: Continued. TIR domain tree for bats and all reference sequences. Each TLR set marked with separate colour. Tree is midpoint rooted. Bootstrap displayed for values ≥70%.

TLRs. Across almost all the TLRs all three groups - reference species, *Yinpterachiroptera*, and *Yangochiroptera* - branched together creating three separate clades. Other *Laurasiatheria* branched closest to both bat clades. In the case of TLR13 all *Pteropodidae* sequences (*P.alecto* and *P.vampyrus* pseudogenes, as well as *R.aegyptiacus* full sequence) were placed on the branch seen to be ancestral to both *Perissodactyla* and *Cetartiodactyla* reference species (**Figure A 9.26**). This might be caused by the long branch attraction event, with changes accumulated by the pseudogenised genes making these sequences seem more similar to other mammals rather than bats.

The analysis of the TLR phylogenetic tree reconstructions in search for positively selected branches with the Datamonkey bsREL tool (**Table A 9.4**) showed that the highest number of branches under positive selection was in TLR8 (10 out of 43) and that there was evidence of positive selection on the branch separating two bat suborders and the branches of individual species (**Figure 4.10**). TLR1 and TLR2 showed episodic diversification (i. e. affecting only a subset of lineages) on branches that separate bats from other species (**Figures 4.11** and **4.12**). On the TLR4 tree positive selection was observed on the two branches separating the *R. aegyptiacus* and *R. ferrumequinum* species from other bat species, while positive selection was also present on the rodent branch and on branches separating horses from carnivore species (**Figure 4.13**). TLR7 displayed positive selection on the branch separating bats from all the reference species, as well as a branch which separates *Yinpterachiroptera* and *Yangochiroptera*, and additionally the *R. ferrumequinum* branch (**Figure 4.14**). There was no diversifying selection detected between other non-bat species which suggests selection specific to bats only. For TLRs 3, 10, 11, and 12 no branches were identified as under positive selection.

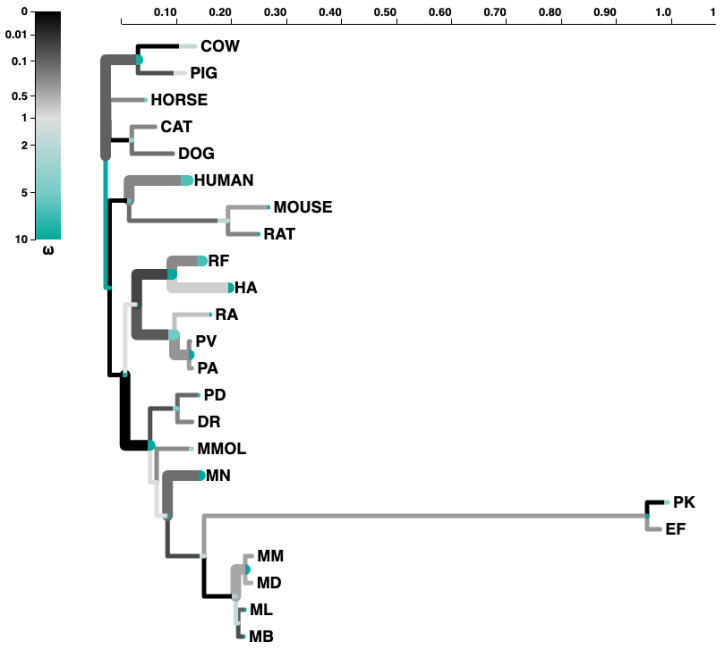


Figure 4.10: Tree showing episodic diversifying selection between branches of the TLR8 phylogenetic tree of bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection.

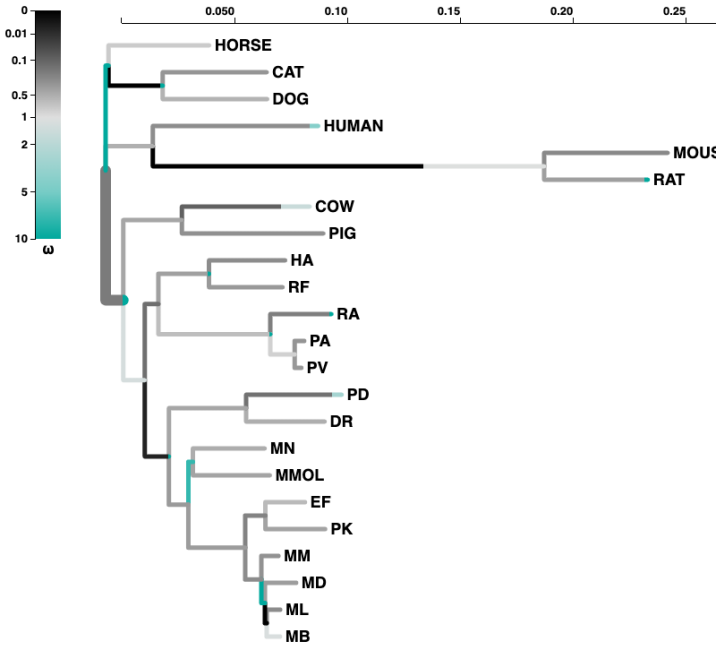


Figure 4.11: Tree showing episodic diversifying selection between branches of the TLR1 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection.

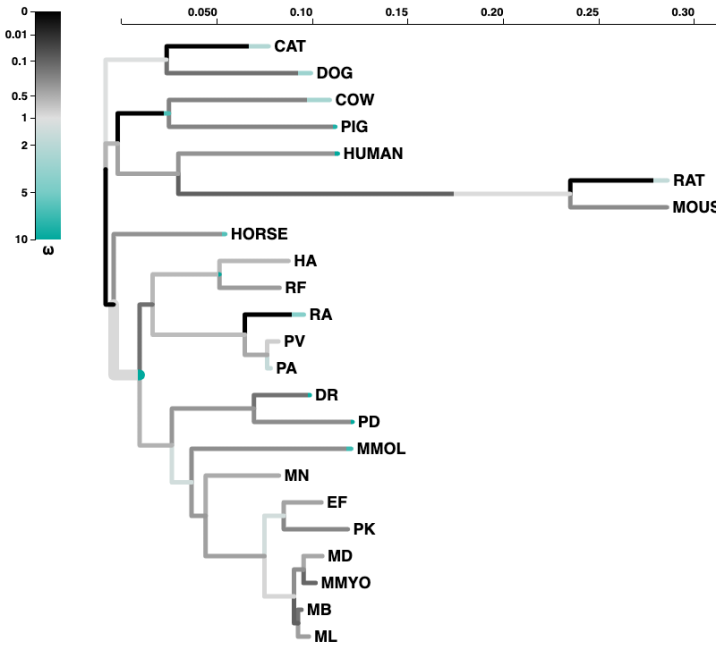


Figure 4.12: Tree showing episodic diversifying selection between branches of the TLR2 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection.

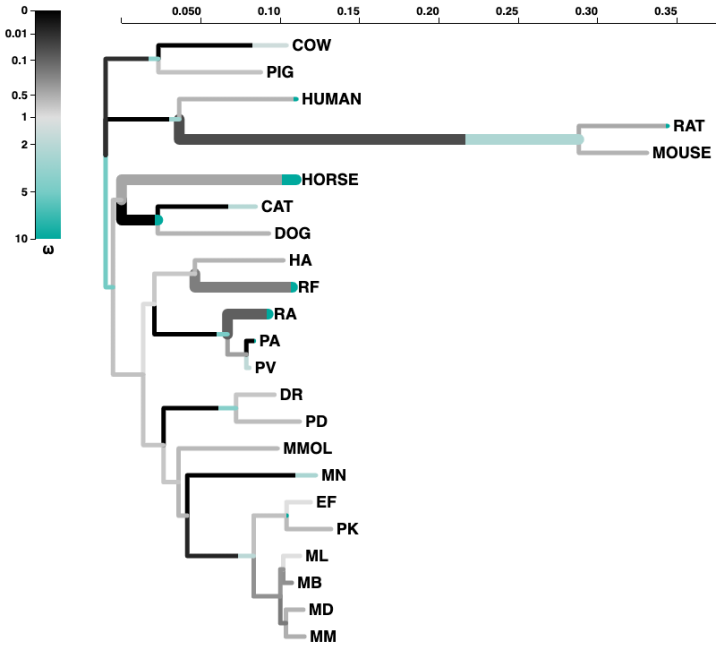


Figure 4.13: Tree showing episodic diversifying selection between branches of the TLR4 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection.

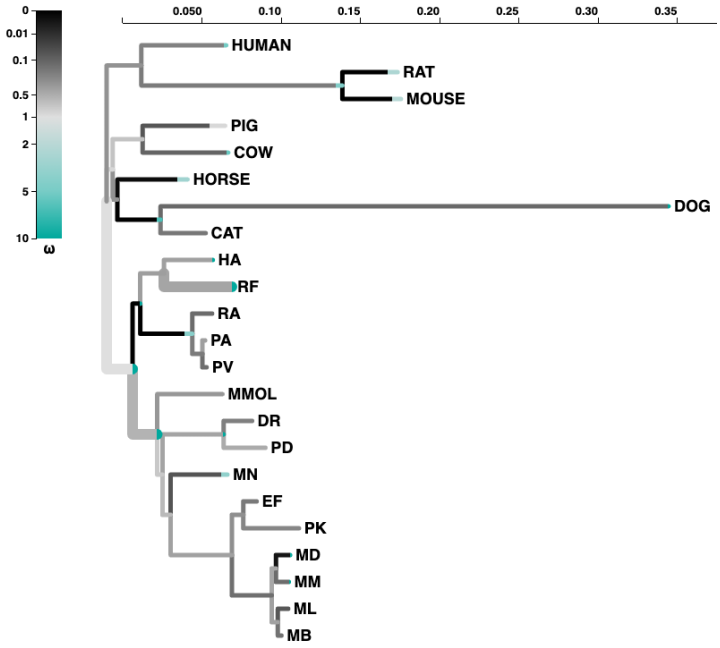


Figure 4.14: Tree showing episodic diversifying selection between branches of the TLR7 phylogenetic tree of the bat and reference sequences. Darker colour and thicker branches indicate stronger positive selection.

4.4 Discussion

TLRs are divided into six (TLR-1/2, TLR-3, TLR-4, TLR-5, TLR-7, TLR-11) [278] or eight (TLR-1/2, TLR-3, TLR-4, TLR-5, TLR-7, TLR-11, TLR-13, TLR-15) [277] major families. According to the first classification most vertebrates have at least one ortholog from each family [278]. We did not detect TLR5 in *P. kuhlii* which would counter this rule if our observation is not caused by a genome assembly artefact, the possibility of which must be additionally reconfirmed and tested.

TLRs can be classified into families, but also rather crudely into those involved in the recognition of viral or non-viral agonists. TLRs 3, 7, 8 and 9 are localised in endosomes and can recognise viral nucleic acids. TLRs 1, 2, 4, 5, 6, 10-13 are localised in cell membranes and recognise lipids, lipoproteins and proteins found in bacterial components and other non-viral varieties of pathogens, with the exception of TLR4 which can also

recognize some viral motifs [267, 277]. Natural selection is expected to have a strong effect on immune genes, especially on innate immune genes, which code for products that are in the first line of defence against invading pathogens. Studies on different species showed that selection patterns on different TLRs can be quite divergent, which may be caused by the different nature of the PAMPs or lower evolutionary dynamics in the case of viral TLRs which need to keep an established virus sensing function at a certain level at all times [277, 315]. It is worth mentioning that a similar range of dN/dS values can also be seen in reproduction-related genes in mammals, which are known to be a rapidly evolving group [316]. Our results show that some regions of bat TLRs are under purifying selection (**Table A 9.6**) which indicates functional constraints, driven by conserved functional parts of these genes, which are in contact with structurally conserved pathogen PAMPs [293]. In our results the highest number of sites under negative selection can be observed in the case of nucleic acid sensing TLRs (TLRs 3, 7, 8, 9). Similar patterns were observed in the case of TLRs in other mammals [293, 315], and also in fish, amphibians, and birds [277]. While in non-nucleic acid recognising TLRs selective constraints seem to be more relaxed, we detected an exception from this rule in the form of bat TLR5, which recognises bacterial flagellin but displays analogous low dN/dS values.

A much higher number of positively selected sites were detected amongst bats in comparison to most other vertebrate taxa. Our analysis considered outputs from a range of models (*codeml* M8, and Datamonkey FEL, MEME, SLAC, and FUBAR) but after scaling them down to just two (FUBAR and M8), which were used in other comprehensive study [277], the number of positively selected sites for the *Chiroptera* order was much higher (n=182) than that for other mammalian species (n=138) [277], and closer to the numbers reported for bird taxa (n=212), which have the highest numbers of positively selected sites among vertebrates, which can be partially caused by a high

divergence of this animals [276, 317, 318, 319]. Bats are not the only sources of zoonotic diseases. Indeed, birds spread the majority of zoonotic pathogens [320]. Avian species have large phylogenetically diverse populations around the world both as wild birds and as domesticated poultry. They gather in flocks and can migrate over long distances, giving them the potential to spread infectious diseases rapidly [321]. However, bats are the primary reservoirs and carriers of more than 200 viral diseases, which if transmitted to other species can cause enormous economic losses in agriculture, and mortality in humans [62, 199]. The old evolutionary relationship between birds and pathogens might be the factor driving selective pressures on their innate immune genes. Similar pressures can be at work in the case of bats, which are one of the oldest orders of mammals, co-evolving with viruses for millions of years, resulting in one of the longest evolutionary relationships with pathogens among mammals [59, 62, 77]. The proportion of positively selected sites in bats is also much higher than that observed in other mammals and also corresponds to the scores reported for birds [277], in most of the observed cases exceeding it (**Table 4.3**), which indicates a shift from purifying to diversifying selection in bats. Viruses and other pathogens are under high positive selection pressure to evade host immune responses, as are some host immune genes, which contributes to host defences in the host-pathogen arms-race. The fact that we detected positively selected sites mostly in LRR domains supports host-pathogen co-evolutionary interactions, since these sites are outward facing and responsible for molecular recognition of PAMPs (**Figure 4.6**). This pattern is consistent with other vertebrates [277, 293, 315].

The TLR-1 subfamily comprises TLRs 1, 2, 6 and 10 [277, 278]. TLR1 working as a heterodimer together with TLR2 detects lipopeptides of gram positive bacteria [322, 323, 324]. As a homodimer TLR2 can stimulate host responses to yeast through stimulation of NF-κB [325]. TLR2 also forms heterodimers with TLR6 and TLR10 [322, 323, 324], although the function of this last dimer is currently unclear [326]. TLR6 recognizes

tri- and dilipopeptides from gram positive bacteria, mycoplasma and fungal saccharides. TLR10 has not been as extensively studied and its ligands have not yet been fully catalogued [326, 327]. It is present as a pseudogene in mice which is disrupted by a retroviral insertion [327], although other mammalian genomes contain an intact TLR10 gene [278]. In all tested mammalian genomes TLRs 1, 6 and 10 are located in one genomic cluster. As those genes are close to each other their evolutionary relationship might not be independent, which might be corroborated by very similar dN/dS results (**Table 9.2**). Similar pattern can be observed in bats, although according to our data some bat species are missing TLR10 entirely (**Table 4.1**). That is especially interesting since TLR10 has been proposed to have a regulatory function that dampens TLR2 responses [327, 328]. Pathogen and danger-associated molecular patterns such as inflammation were confirmed to be dampened in bats [24, 138, 156, 264, 329]. Indeed bats are lacking several gene families involved in inflammasome response [24, 156, 264, 329]. Possession of TLR10, which is involved in inflammation dampening processes, should be expected widely in this order, but its lack may suggest that bats might have developed a more efficient way to conduct the same process or may use some other genes for the same purpose. It needs to be noted that we noticed a lack of this receptor only in the vesper bat species, which suggests that this kind of receptor was lost ∼35-45 Mya when the split between the *Vespertilionidae* and *Miniopteridae* families took place [1].

TLRs 3, 7, 8, and 9 are a group of nucleic acid sensing receptors, activated by ligands present in most pathogens including bacteria and viruses [275, 330]. TLR3 recognises double-stranded RNA [331], TLR7 and TLR8 are activated by single-stranded RNA [332, 333] and synthetic small molecules with structures resembling ribonucleotides [332, 333, 334, 335], and TLR9 ligands are unmethylated CpG sequences in DNA molecules [336, 337]. All four viral-sensing receptors are localised in cell endosomes [275]. In our analysis all bat genomes displayed working copies of all four genes. TLR7 and

TLR8 are in very close proximity to each other representing a tandem duplication in all species, together with birds in which TLR8 is a pseudoegene disrupted by chicken repeat 1 (CR1) retroviral-like insertion [158]. Similarly to the TLR-1 subfamily their evolutionary relationship might be influenced by their close genomic proximity, although this is not reflected by similar dN/dS scores, with the higher purifying selection observed for TLR7, and with TLR 8 displaying at least 3 times more sites under positive selection from the sum of all models (**Table 4.3**, **Figure 4.6** and **Figure 4.4**). TLR7 is expressed in haematopoietic cells such as plasmacytoid dendritic cells and B cells [66], while TLR8 is expressed mainly in macrophages and myeloid dendritic cells [337]. While, similarly to patterns observed in other animals, TLRs 3, 7, and 9 seem to be under strongest purifying selection from all bat TLRs, the exception of TLR8, which displays a highest number of PSSs might suggest that this gene is under strong diversifying selection or that some specific components of its signalling pathway might be driving the exceptionally high positive selection observed in this gene.

TLR4 is the sole member of the TLR-4 family. It is localised in the plasma membrane and endosomes and it activates pathways leading to inflammatory cytokine production similarly to TLR2 [280, 330]. It recognizes various ligands such as bacterial lipopolysaccharide (LPS) [338, 339], but can also sense other ligands , for example viral glycoproteins, mainly envelope proteins due to its cellular and endosomal location [275, 280]. An intact TLR4 gene was identified in all analysed genomes, but with a high proportion of positively selected sites (**Table 4.3**, **Figure 4.6** and **4.3**). A number of different viral glycoproteins have been shown to act as agonists for TLR4. A pretreatment with TLR4 ligands has been shown to protect mice against infections from the influenza virus [340], or SARS coronavirus [341]. On the other hand, other studies suggest that the TLR4-TRIF-TRAF6 axis might contribute to oxidative stress and lung inflammation leading to lung injury triggered by an influenza virus challenge [342, 343]. Thanks to their ability to

fly bats are exceptionally exposed to oxidative stress and inflammation, which might be responsible for changes in their immune pathways and in immune gene differences that dampen inflammation responses [138, 262]. Indeed, other animals with the capability of active flight are birds, although these two taxa evolved independently from each other for a very long time and might have evolved completely different strategies to deal with inflammation and oxidative stress generated by flight. In the case of bats the selected polymorphisms of TLR4 might have been present due to their ability to fly, especially in the light of bats being a potent reservoir for influenza A viruses [344, 345, 346, 347].

Another TLR which is a single representative of its family is TLR5, which is present as a pseudogene in some lineages of birds [276, 348], fish and reptiles [277] but in general is well conserved in mammals. TLR5 recognizes bacterial flagellin [349, 350] and is mainly expressed in epithelial cells and immune cells. A functional copy of the TLR5 gene was detected in all bats apart from one - *P. kuhlii* from the *Vespertilionidae* family. *P. kuhlii* has a stop codon in a part of its TLR5 copy, contrary to all other tested vesper bats (**Table 4.1**). This may suggest that the presence of the stop codon could have been caused by an assembly mistake, especially that in all other tested bat species this receptor shows evidence of the strong purifying selection (**Figure 4.6** and **Table 9.2**). This needs to be further investigated, possibly with an raw assembly data check, further *in-silico* studies on other closely related bats, or experimental laboratory tests on tissues or cell lines from this species. However, if the loss of TLR5 is confirmed to be correct for this species it might be an effect of random deletion or a selection by pathogenic effects linked to a flagellated bacterial pathogen.

TLR11 and TLR12 are highly similar to each other and can function either as homodimers or heterodimers [351, 352, 353]. Both recognise the profilin molecule present in *Toxoplasma gondii* and all other *Apicomplexa* [353, 354, 355], while TLR11, which is

located in the endolysosome, also recognises a component of uropathogenic *Esherichia colii*. TLR12 is located on myeloid cells and is unique to mammalian species. Both are present only in selected species, and in the case of mammals functional instances of TLRs 11 and 12 are seen predominantly in rodents, but are found as pseudogenes in other mammalian families. In some bat species both TLR11 and TLR12 were missing, but more often we observed this in the case of TLR11, with a fully functional TLR12. Apart from TLRs 11 and 12 having been observed to be missing in all analysed *Pteropodidae* representatives, in other families the presence or absence of these receptors seemed more species than family specific. Repeated gene loss has been seen before in the case of these TLRs in several mammals and birds [356, 357, 358]. Both of these TLRs displayed dN/dS values closest to 1 which might imply neutral selection, with TLR11 showing an accumulation of PSSs which might be a proof of strong diversifying selection acting on this gene (**Figure 4.6** and **Table 9.2**).

The functions and ligands of TLR13 remain poorly understood. It was found to recognise bacterial ribosomal S23 RNA [283, 359] and there is some evidence that it can recognise components associated with viral infections [360, 361], but the exact viral PAMP remains unknown [285]. According to one study this TLR is a member of the TLR-11 family [278], but according to a new more broad analysis TLR13 is a member of a separate TLR13 family together with TLRs 21, 22, and 23 which are present in taxas other than mammals [277]. TLR13 is present in mice but not in the human genome. Studies on TLR13 have found that its agonist is the same as the structure bound by some antibiotics, and that when bacterial antibiotic resistance has been acquired they were also able to evade detection by TLR13 [359]. Previous studies found that TLR13 in *P.alecto* and *P.vampyrus* recently lost its function from being transcribed, spliced and polyadenylated [152]. It was suggested that the lack of this receptor in humans and some other mammalian species might be a consequence of an ancient antimicrobial

resistance mechanism which rendered this receptor ineffective and subsequently lost in the evolution of these species [359]. The pseudogenization of the TLR13 gene in *Pteropus* bat species was reconfirmed by our analysis on high quality genomes. We also report that all of the other analysed bat species have an intact copy of this TLR suggesting that the function loss might have taken place during the split between *Rhinolophoidea* and *Pteropodidae* around 55 Mya. Such loss, observed only in one family, might suggest that this receptor might have also been rendered ineffective, either by evolution of pathogens evading TLR13 detection, or subsequent evolution of better recognition mechanisms by the host. It might have also been driven by the extinction of susceptible pathogens in the environment of the host resulting in the loss of the receptor which was no longer needed.

4.5 Conclusions

Previous comparative work on bat and mammalian TLRs showed differences in genomic architecture as well as selective pressures in these genes [151, 152]. However, we need to keep in mind that bats are one of the oldest order of mammals and are present in an array of different biological niches and environments, with species and whole families evolving separately for millions of years. Thus, we cannot treat bats as one single order with the same biological characteristics [262]. Large comprehensive analyses which have compared genomes of several different bats showed that their traits, gene functions and genomic architecture can vary from family to family and from species to species [24, 77, 78, 362, 363]. Also, in our results we detected that different species and families of bats have different numbers and types of TLRs. In general these differences were observed for whole families, but sometimes were seen only in particular species. This might be driven by the specific ecological traits of the given bats, by their habitat, or

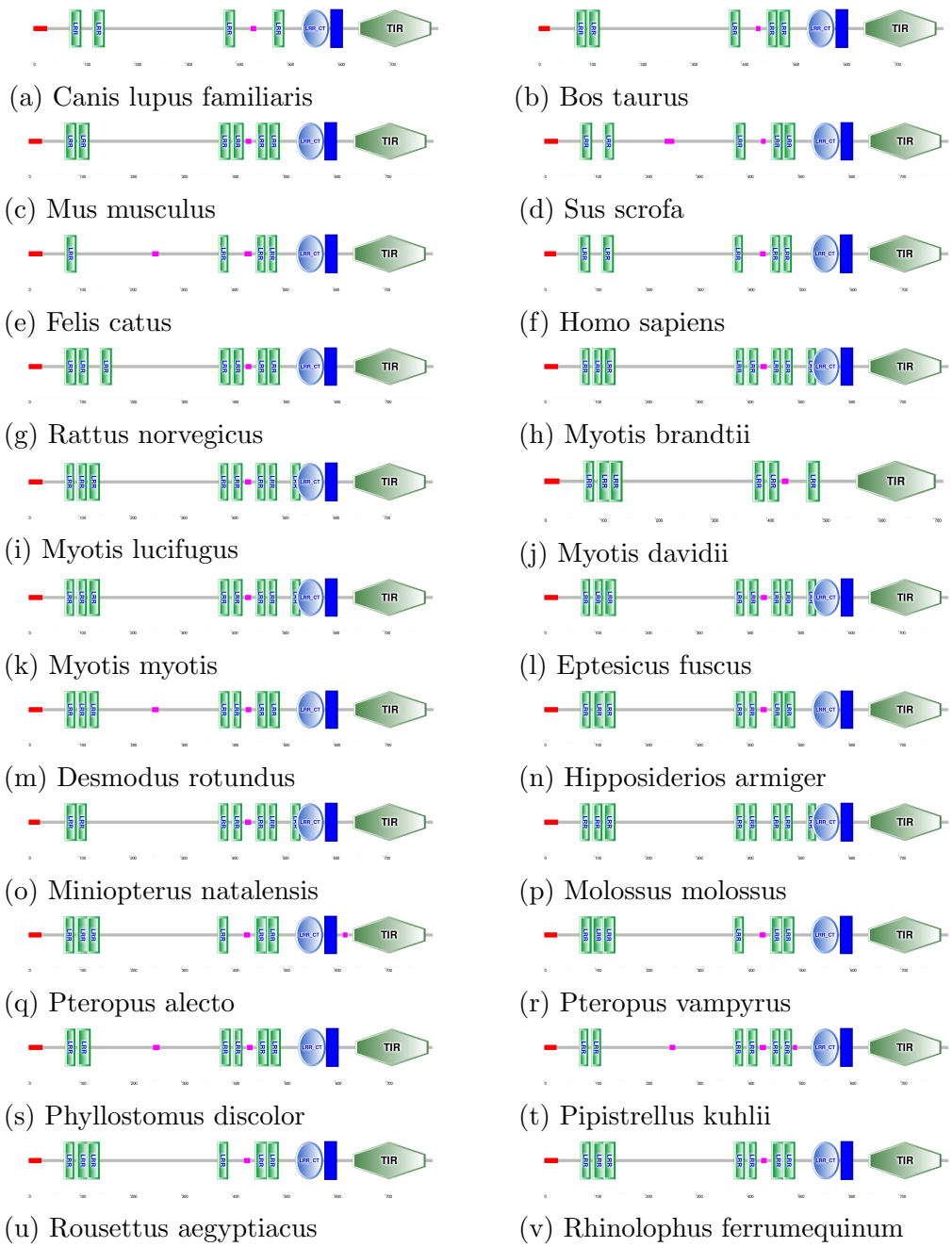
the presence or absence of certain types of pathogens in the environments in which these species are present.

As was suggested previously in the case of TLR13 [359], the lack of specific TLRs in different species might be driven by ancient co-evolution between pathogens and their hosts. We propose that this mechanism might have also taken place in the case of bats. Some pathogens acquired mechanisms to evade TLR ligand recognition. Immune receptors with no effective function in the host organism then were subsequently lost. This might have been followed or driven by developing other, more efficient ways of dealing with the pathogen invasion, or by the need to dampen inflammatory responses caused by TLRs as in the case of several other lost or altered inflammasome genes and cytokines observed in bats [24, 138, 156, 264, 329].

We also need to take into consideration a possibility of losing or changing TLRs to limit potential pathology driven by their over-activation. Some of these receptors might have been subsequently lost if their function was harmful to their hosts. Bats can be persistently infected with viruses at a much higher rate than other mammalian species [62, 138, 262]. This might be driven by the long co-evolution between bats and viruses which sometimes might lead to lower pathogenicity. The immune system is built from interconnected and intertwined pathways and imbalances can arise quickly if any of its fragments will lose parts which help to regulate or control other components of this system. Some of the genes or receptors which were lost in the bat-virus arms-race might have been involved in controlling some of the TLR immunomodulatory and inflammatory responses. That might have led to subsequent loss or change of the functions of the immune receptors to avoid damage caused by overactive responses leading to autoimmune and inflammatory diseases.

To fully understand the differences and special immune traits of the *Chiroptera* order

there is a need to sequence more high quality bat genomes and to conduct more analyses of TLRs across different bat families, followed with functional analysis to ascertain if bat TLRs have similar or diversed functions to these observed in other mammalian orders. Previous studies, together with our own, has laid groundwork to suggest that bat TLRs are under much higher positive selective pressures than these observed in other mammals, and that their TLR functions might significantly differ from TLRs as exhibited in model species.



(a) Canis lupus familiaris (b) Bos taurus

(c) Mus musculus (d) Sus scrofa

(e) Felis catus (f) Homo sapiens

(g) Rattus norvegicus (h) Myotis brandtii

(i) Myotis lucifugus (j) Myotis davidii

(k) Myotis myotis (l) Eptesicus fuscus

(m) Desmodus rotundus (n) Hipposiderios armiger

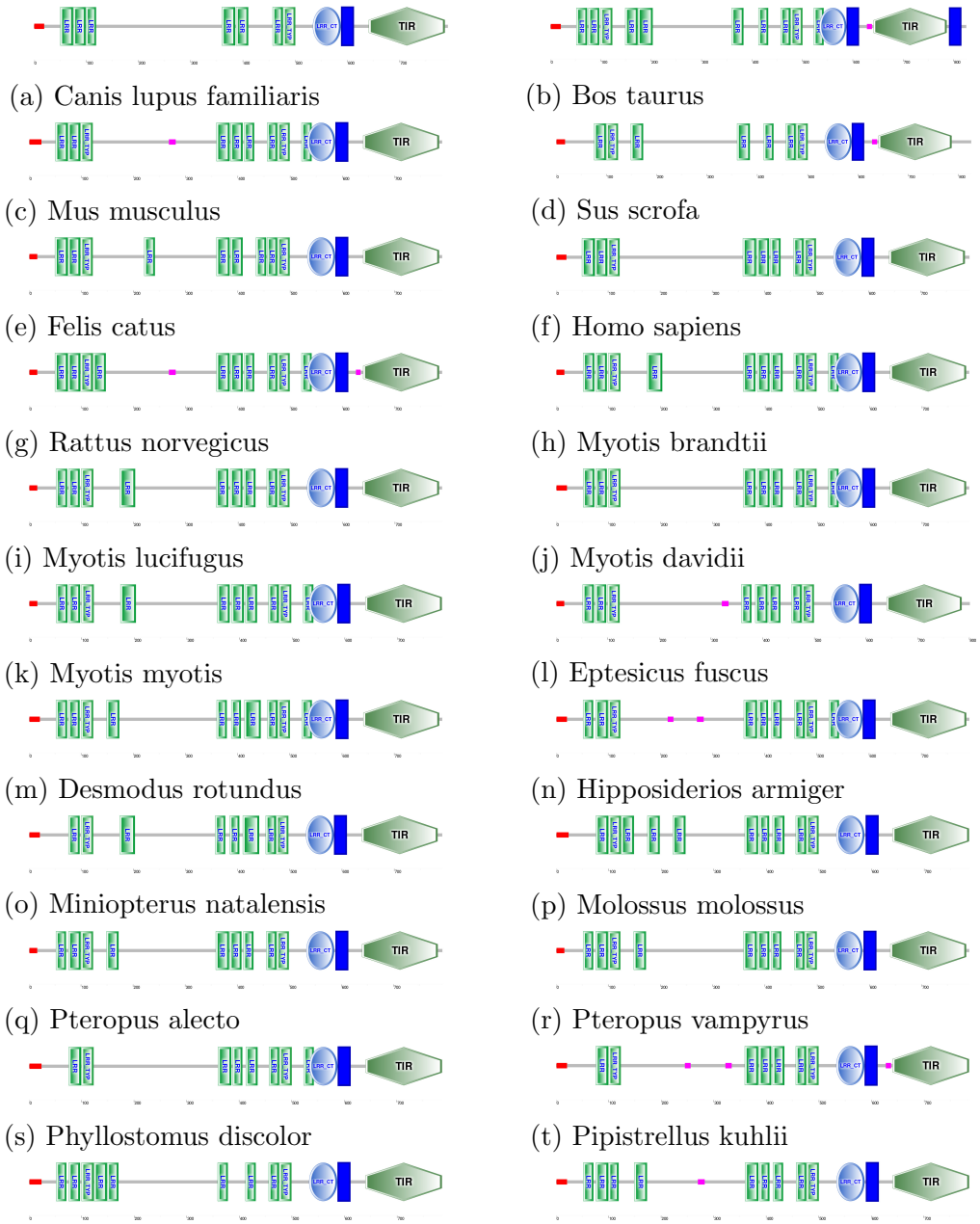
(o) Miniopterus natalensis (p) Molossus molossus

(q) Pteropus alecto (r) Pteropus vampyrus

(s) Phyllostomus discolor (t) Pipistrellus kuhlii

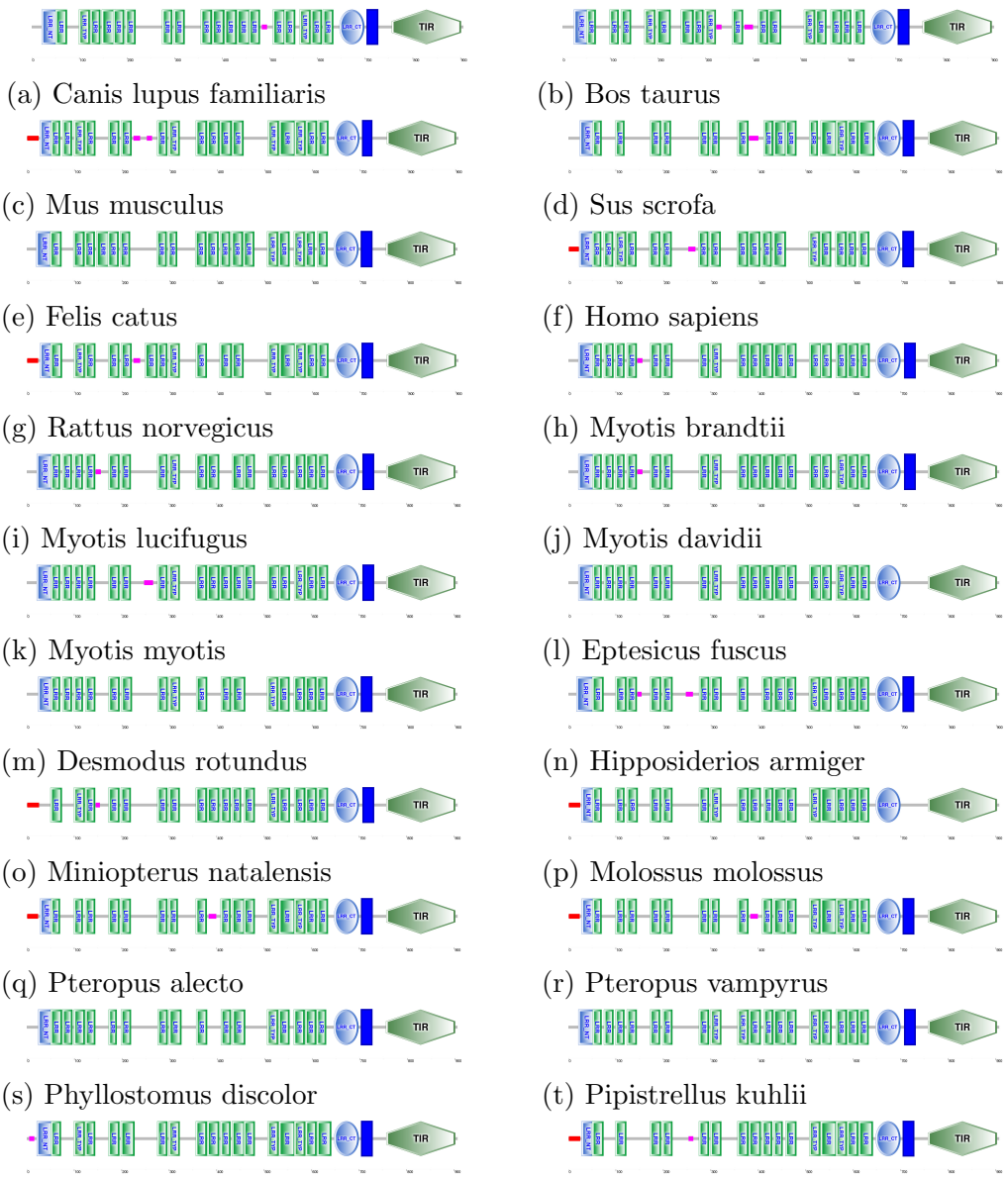
(u) Rousettus aegyptiacus (v) Rhinolophus ferrumequinum

Figure 4.15: Predicted protein domain architecture of TLR1 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.



(a) Canis lupus familiaris
(b) Bos taurus
(c) Mus musculus
(d) Sus scrofa
(e) Felis catus
(f) Homo sapiens
(g) Rattus norvegicus
(h) Myotis brandtii
(i) Myotis lucifugus
(j) Myotis davidii
(k) Myotis myotis
(l) Eptesicus fuscus
(m) Desmodus rotundus
(n) Hipposiderios armiger
(o) Miniopterus natalensis
(p) Molossus molossus
(q) Pteropus alecto
(r) Pteropus vampyrus
(s) Phyllostomus discolor
(t) Pipistrellus kuhlii
(u) Rousettus aegyptiacus
(v) Rhinolophus ferrumequinum

Figure 4.16: Predicted protein domain architecture of TLR2 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.



(a) Canis lupus familiaris (b) Bos taurus

(c) Mus musculus (d) Sus scrofa

(e) Felis catus (f) Homo sapiens

(g) Rattus norvegicus (h) Myotis brandtii

(i) Myotis lucifugus (j) Myotis davidii

(k) Myotis myotis (l) Eptesicus fuscus

(m) Desmodus rotundus (n) Hipposiderios armiger

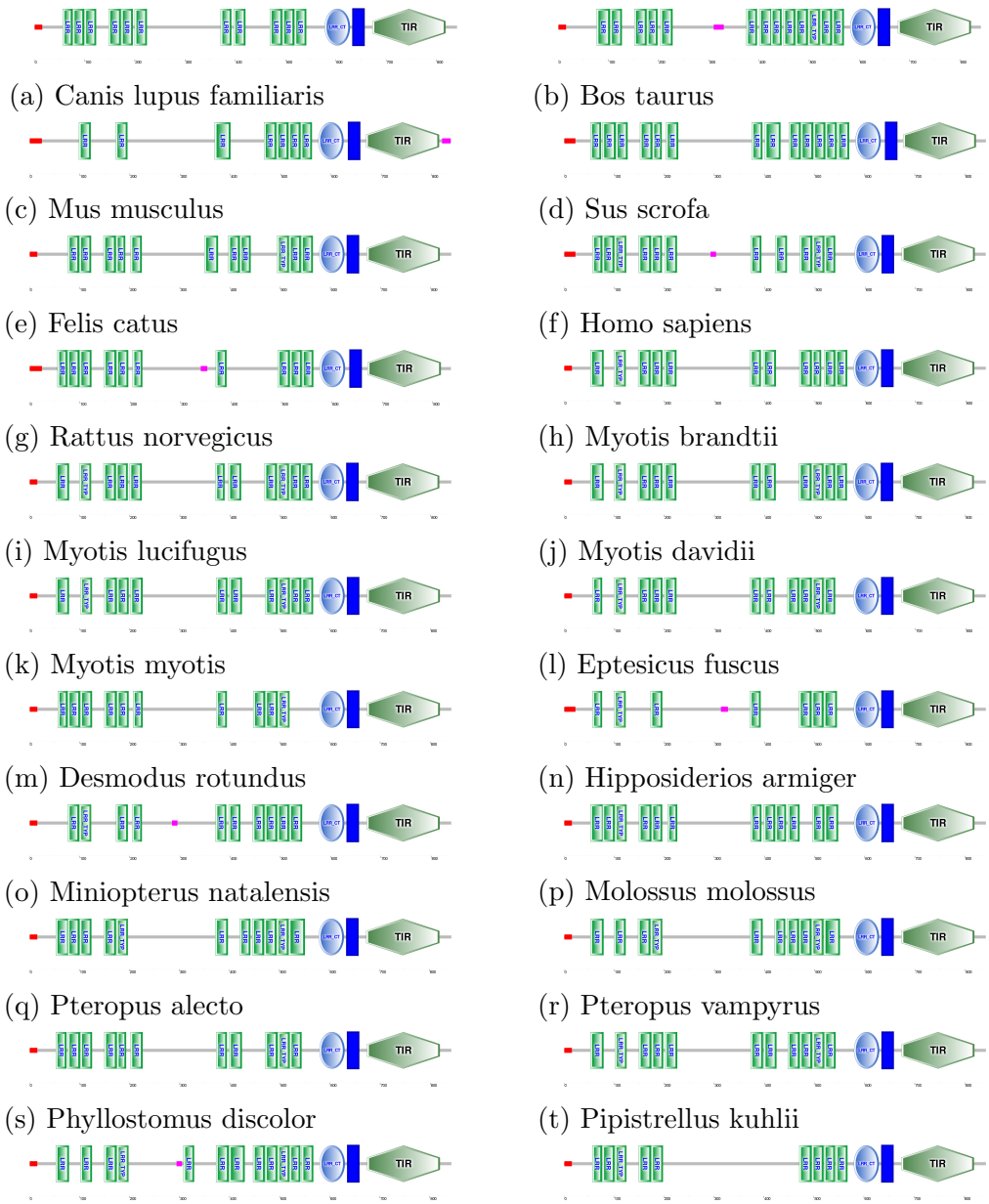
(o) Miniopterus natalensis (p) Molossus molossus

(q) Pteropus alecto (r) Pteropus vampyrus

(s) Phyllostomus discolor (t) Pipistrellus kuhlii

(u) Rousettus aegyptiacus (v) Rhinolophus ferrumequinum

Figure 4.17: Predicted protein domain architecture of TLR3 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.



(a) Canis lupus familiaris
(b) Bos taurus
(c) Mus musculus
(d) Sus scrofa
(e) Felis catus
(f) Homo sapiens
(g) Rattus norvegicus
(h) Myotis brandtii
(i) Myotis lucifugus
(j) Myotis davidii
(k) Myotis myotis
(l) Eptesicus fuscus
(m) Desmodus rotundus
(n) Hipposiderios armiger
(o) Miniopterus natalensis
(p) Molossus molossus
(q) Pteropus alecto
(r) Pteropus vampyrus
(s) Phyllostomus discolor
(t) Pipistrellus kuhlii
(u) Rousettus aegyptiacus
(v) Rhinolophus ferrumequinum

Figure 4.18: Predicted protein domain architecture of TLR4 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

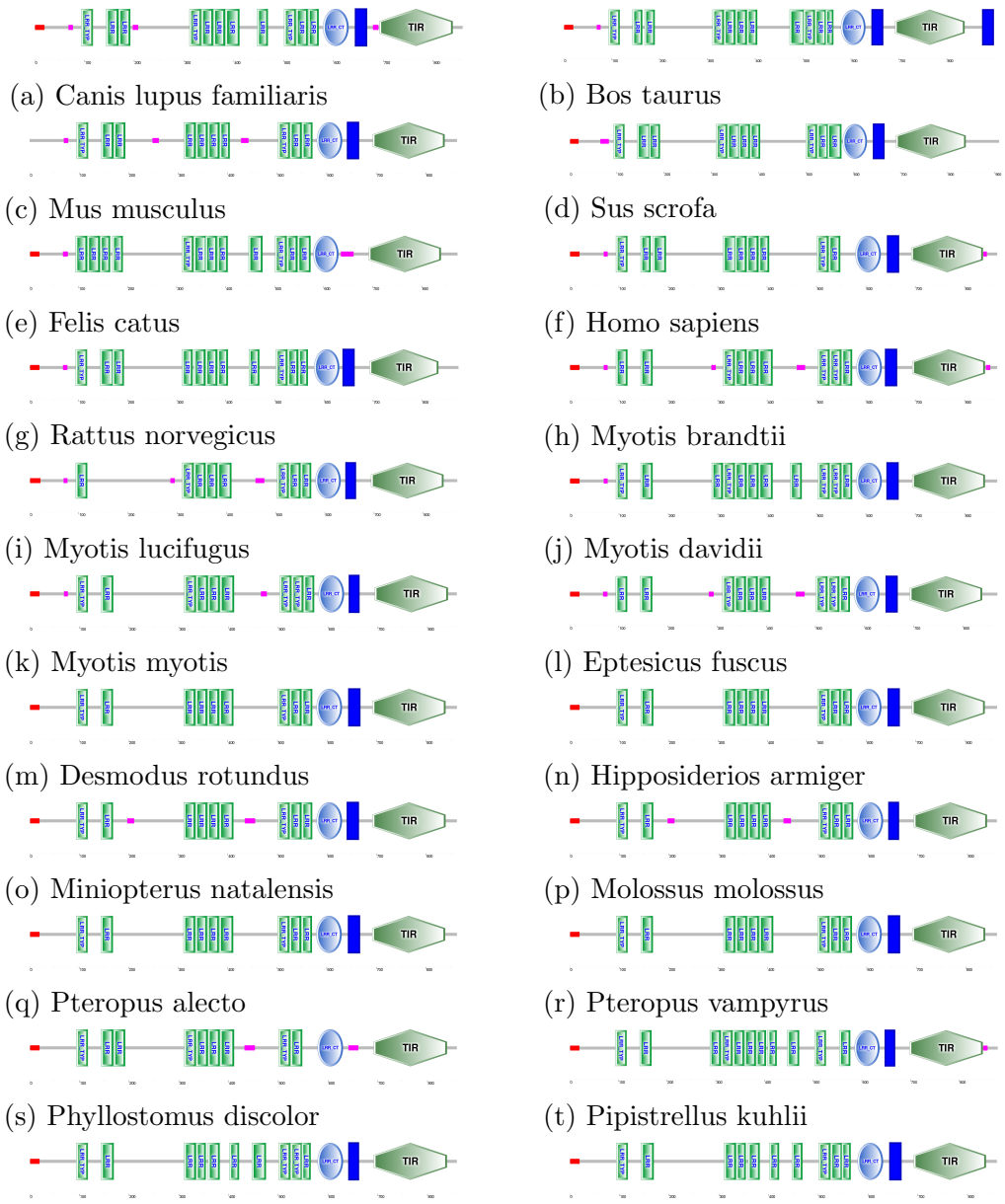


Figure 4.19: Predicted protein domain architecture of TLR5 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

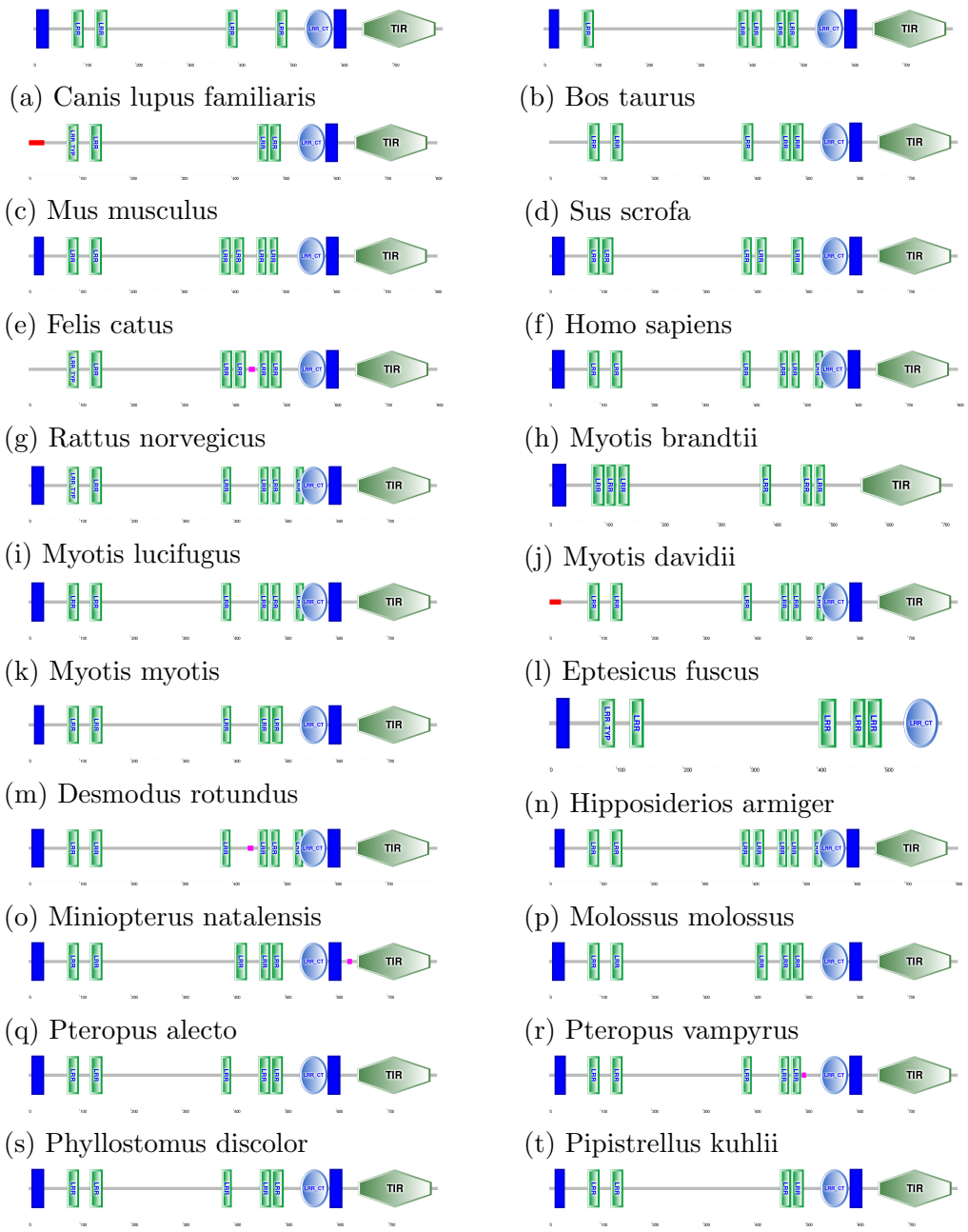


Figure 4.20: Predicted protein domain architecture of TLR6 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. Myotis davidii and Hipossiderios armiger sequences just partial due to the gaps in assembly scaffold

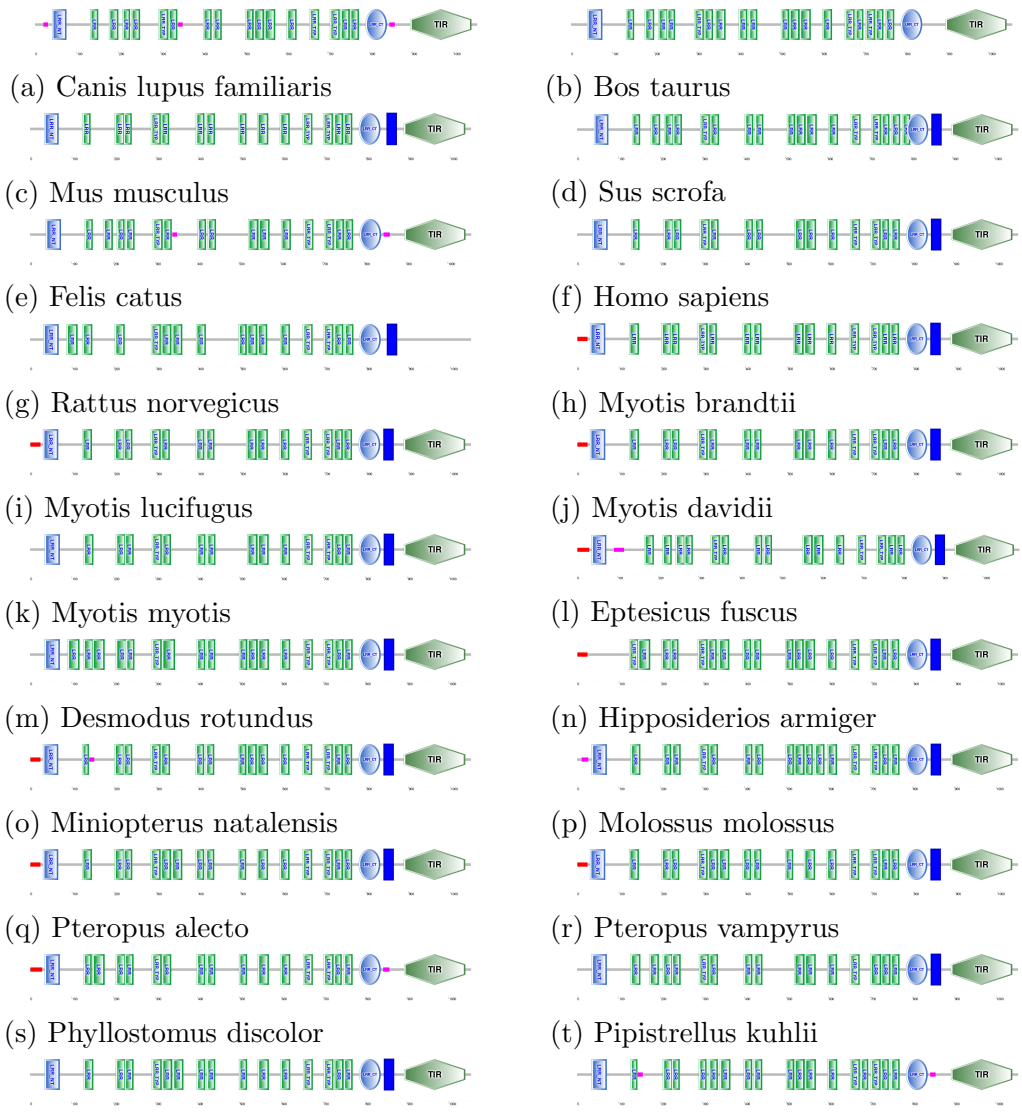


Figure 4.21: Predicted protein domain architecture of TLR7 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

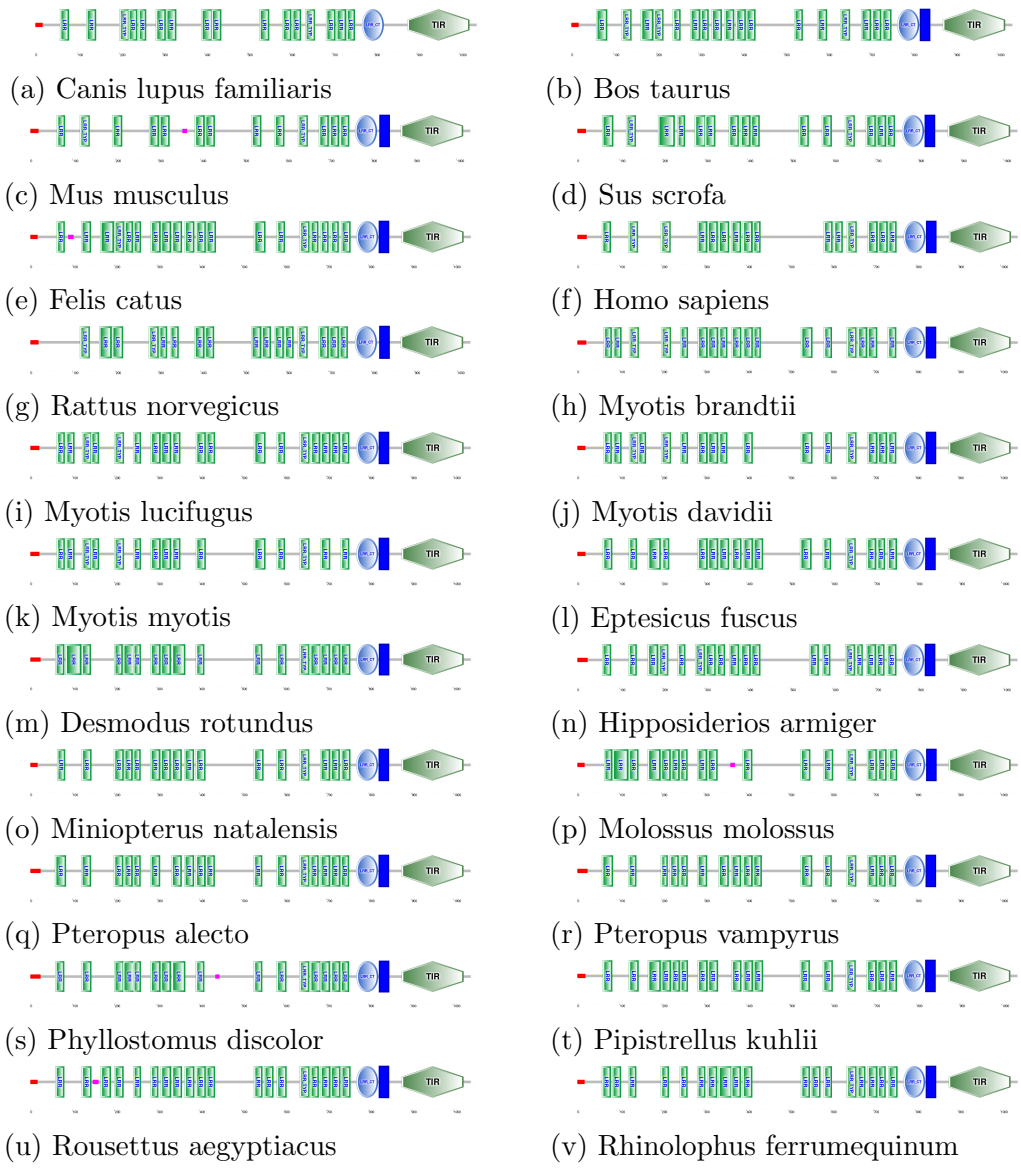


Figure 4.22: Predicted protein domain architecture of TLR8 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

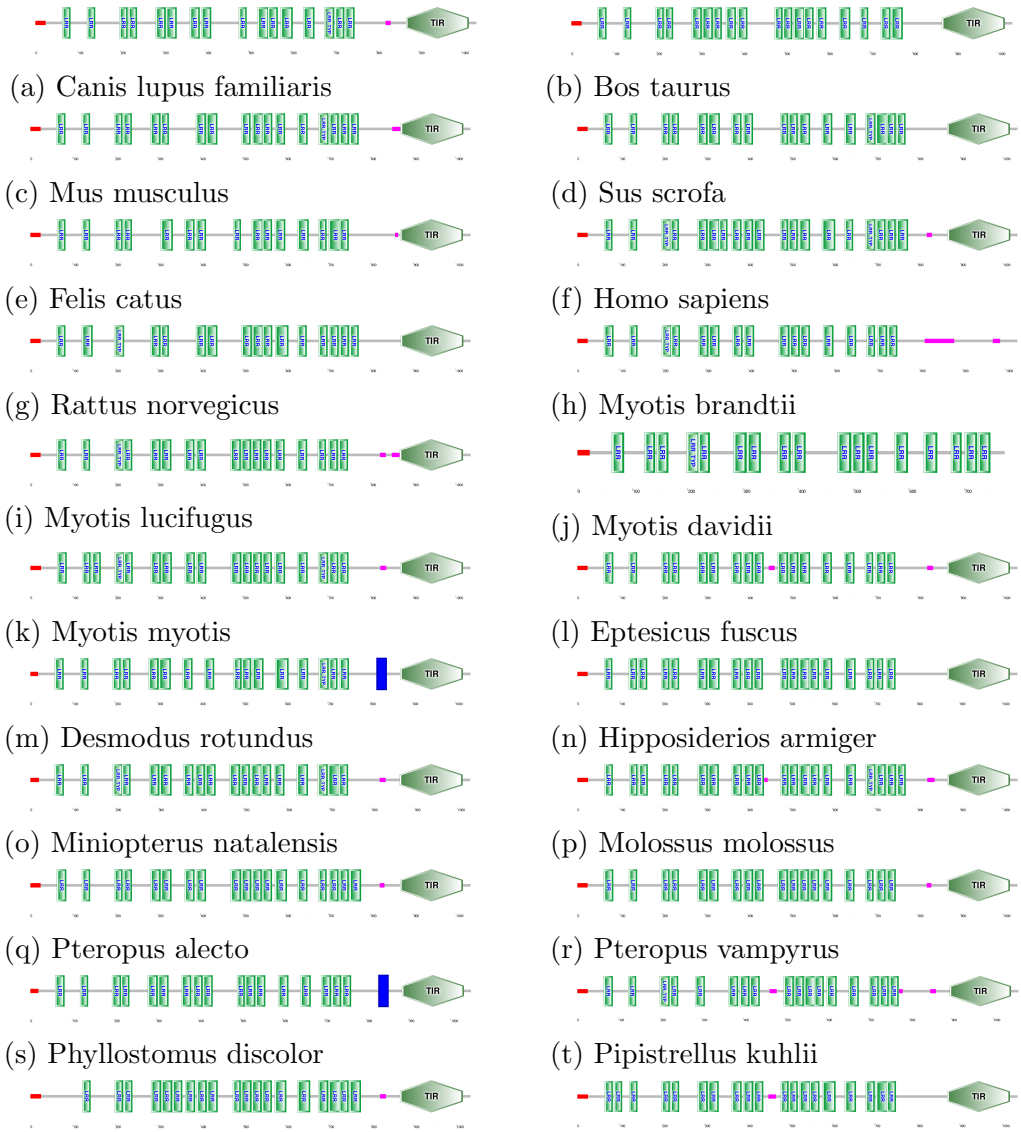
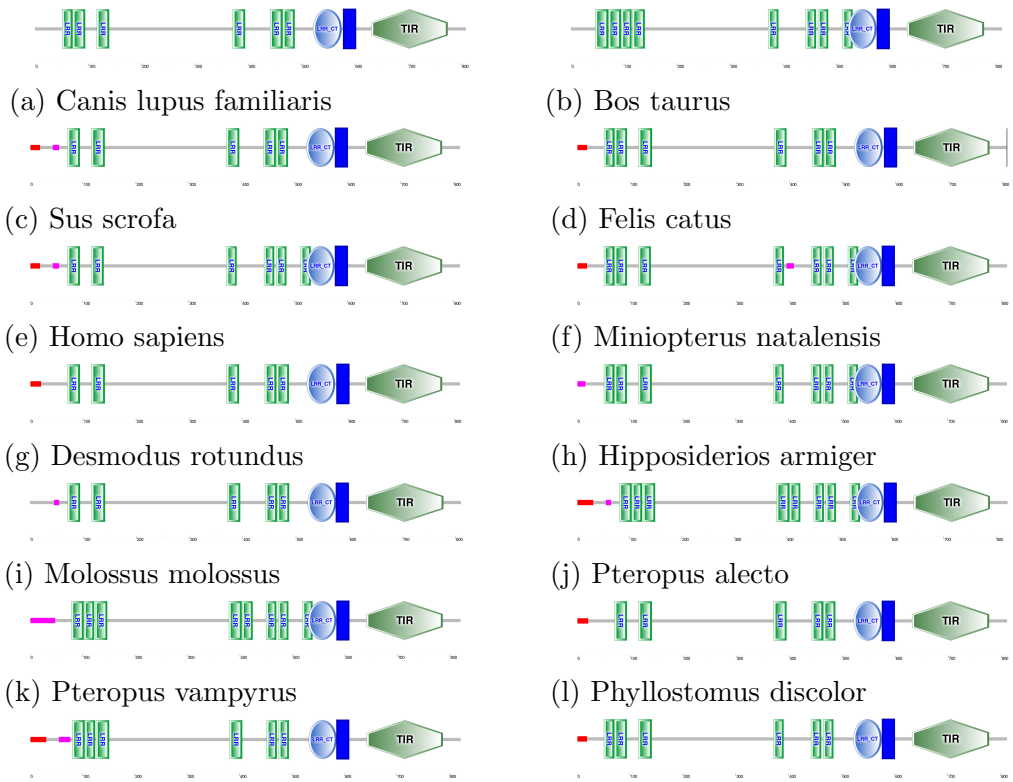
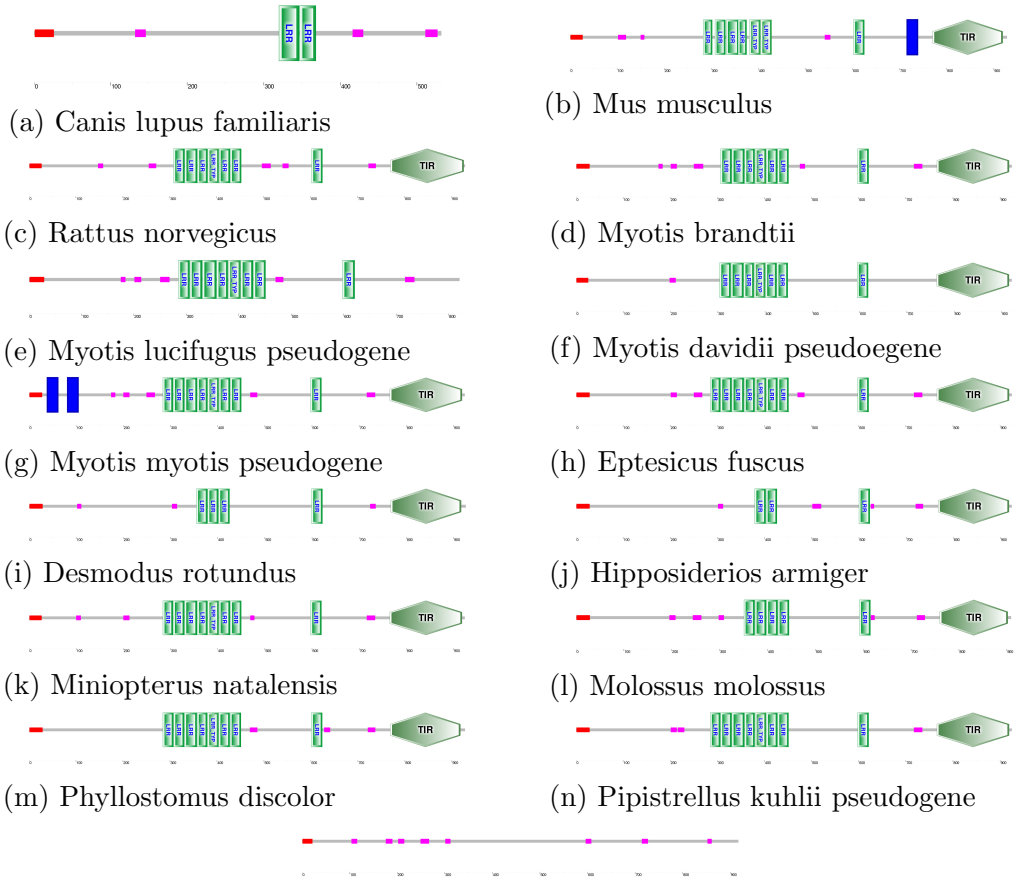


Figure 4.23: Predicted protein domain architecture of TLR9 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue. Myotis davidii and Myotis brandtii sequences just partial due to the gaps in assembly scaffold



(a) Canis lupus familiaris
(b) Bos taurus
(c) Sus scrofa
(d) Felis catus
(e) Homo sapiens
(f) Miniopterus natalensis
(g) Desmodus rotundus
(h) Hipposiderios armiger
(i) Molossus molossus
(j) Pteropus alecto
(k) Pteropus vampyrus
(l) Phyllostomus discolor
(m) Rousettus aegyptiacus
(n) Rhinolophus ferrumequinum

Figure 4.24: Predicted protein domain architecture of TLR10 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.



(a) Canis lupus familiaris

(b) Mus musculus

(c) Rattus norvegicus

(d) Myotis brandtii

(e) Myotis lucifugus pseudogene

(f) Myotis davidii pseudoegene

(g) Myotis myotis pseudogene

(h) Eptesicus fuscus

(i) Desmodus rotundus

(j) Hipposiderios armiger

(k) Miniopterus natalensis

(l) Molossus molossus

(m) Phyllostomus discolor

(n) Pipistrellus kuhlii pseudogene

(o) Rhinolophus ferrumequinum pseudoegene

Figure 4.25: Predicted protein domain architecture of TLR11and TLR11-pseudogenes in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

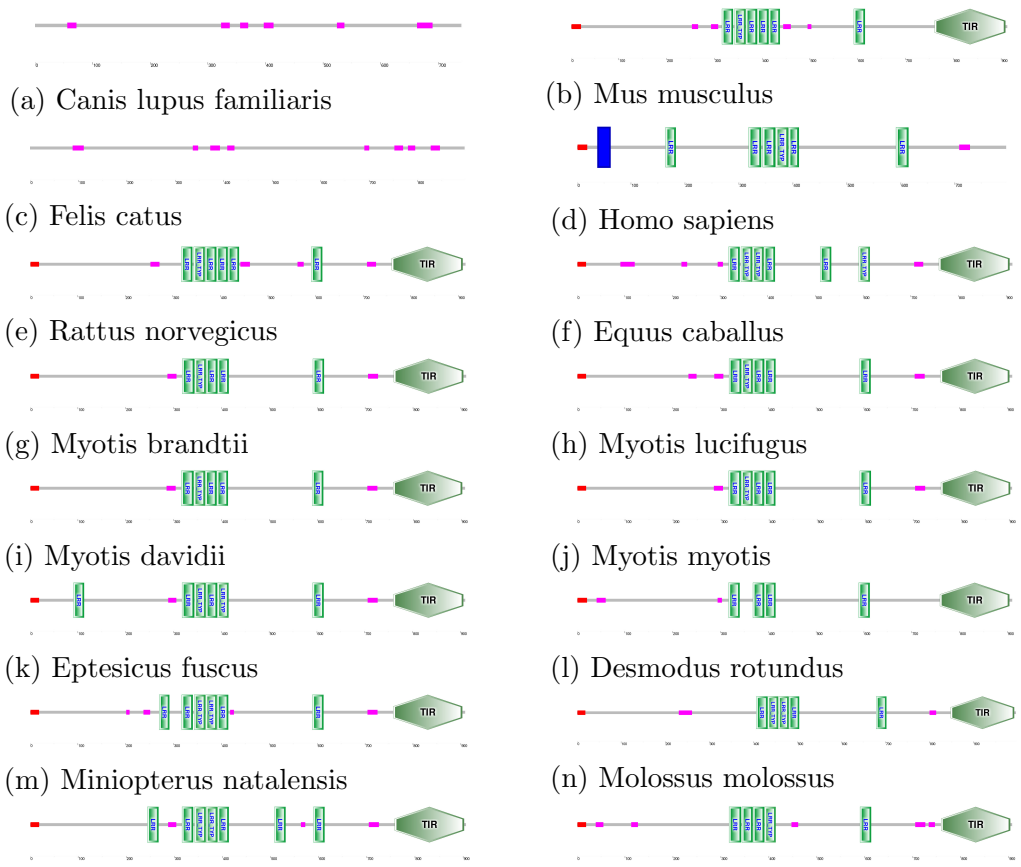
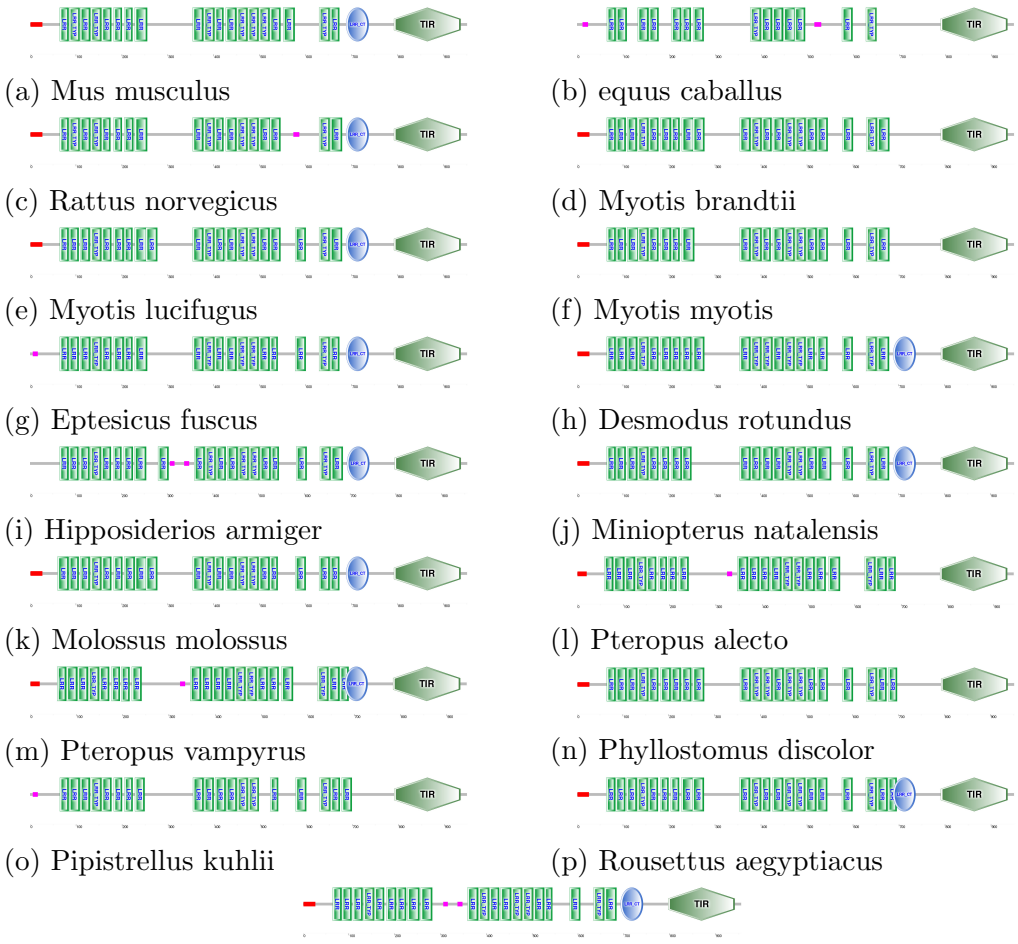


Figure 4.26: Predicted protein domain architecture of TLR12 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.



(a) Mus musculus

(b) equus caballus

(c) Rattus norvegicus

(d) Myotis brandtii

(e) Myotis lucifugus

(f) Myotis myotis

(g) Eptesicus fuscus

(h) Desmodus rotundus

(i) Hipposiderios armiger

(j) Miniopterus natalensis

(k) Molossus molossus

(l) Pteropus alecto

(m) Pteropus vampyrus

(n) Phyllostomus discolor

(o) Pipistrellus kuhlii

(p) Rousettus aegyptiacus

(q) Rhinolophus ferrumequinum

Figure 4.27: Predicted protein domain architecture of TLR13 in bats and reference genomes determined by SMART analysis. Signal peptides in red, low complexity regions in pink, and transmembrane domains in dark blue.

5 Change of receptor affinity and emergence of the new zoonotic pathogens: ACE2 receptor genetic differences amongst animals

5.1 Introduction

An epidemic is an interaction between a pathogen and a host population. The majority of infectious diseases that cause epidemics have a zoonotic origin [364]. Zoonotic spillover can happen in one of three ways: the introduction of the new pathogen into a susceptible host population; the pathogen gaining a new trait which allows it to infect a broader range of hosts; or the introduction of the pathogen into a new environment [365, 366, 367]. These three ways display different dynamics and might need to be dealt with in different ways. An understanding of the trigger of a zoonotic spillover event is crucial in epidemic control and future prevention [365], but can also be an important factor in protecting endangered taxa like primates or bats.

One of the major changes that drove the H5N1 avian influenza to infect mammalian hosts and start an epidemic was a change in receptor affinity [368, 369]. In order to control epidemic outbreaks we need to understand the interactions between pathogens and known receptors: to infect a host a virus first needs to bind to the receptor to establish an infection in the host cell. However, host cell receptors are still unknown for many zoonotic viruses. With pathogens that can change tropism quickly, for example coronaviruses [370], the host range can at least partially be predicted with knowledge of the receptor specificity [371].

By the end of 2019 an outbreak of an upper respiratory infection from an unknown source was reported in Wuhan, China. First taken for the influenza virus, it soon became clear that the pathogen responsible was a new zoonotic coronavirus. Coronaviruses are enveloped, non-segmented positive sense single-stranded viruses which can cause respiratory, gastrointestinal or central nervous system diseases [372]. They are widely distributed, can have broad receptor tropism, and have caused multiple epidemics over the years, including SARS-CoV and MERS-CoV.

Zoonotic pathogens can jump to new hosts in two ways: directly from their reservoir to the new host or through the use of a so-called intermediate host in which they can circulate for a long time with no visible symptoms. Circulation in such intermediate hosts can make pathogens better adapted to infect a human population through accumulation of mutations which might make them more adapted to new cell receptors [365]. Identifying reservoir species can be an important factor in future pandemic prevention. But finding an intermediate host or reservoir becomes harder with time and might ultimately prove impossible. We have previously seen this problem in the case of Ebola outbreaks in Africa [373], but also recently in case of the SARS-CoV 2 pandemic [374, 375]. For example, the SARS-CoV intermediate host is believed to be a mask palm

civet, while MERS-CoV can be transferred to humans from camels. In the case of SARS-CoV 2 both the reservoir and intermediate hosts are still unknown but coronaviruses of a very similar genomic profile have been found in a population of *Rhinolophus spp.* bats [376, 377, 378].

Virus host tropism can be partially predicted through the cell receptor comparison between different host species. Both SARS-CoV and SARS-CoV 2 have been found to interact through their spike glycoprotein (S) [379, 380] with angiotensin I converting enzyme 2 (ACE2) on the surface of the hosts cells. ACE2 is a dipeptidyl carboxypeptidase that regulates renin-angiotensin system, and is responsible for converting angiotensin II into angiotensin 1-7, and angiotensin I into angiotensin 1-9, with roles in regulation of blood pressure and inflammatory response [381, 382]. ACE2 is made up of a signal sequence at the N-terminus, a transmembrane sequence at the C-terminus, and an extracellular region, containing a zinc metallopeptidase domain and a collectrin homolog [383]. The ACE2 receptor is highly conserved in mammals which might result in a broad host range for coronaviruses [384, 379]. In humans ACE2 can be present in two forms: secreted or as a cell membrane protein [381]. The ACE2 protein is highly conserved in mammals. Mutations in human ACE2 tend to be low on a population level, with those affecting coronavirus infection susceptibility believed to be very rare [385, 386]. Residues on ACE2 tertiary structure are essential for binding with the SARS-CoV and SARS-CoV 2 receptor binding domain (RBD) on the virus S protein [376, 379, 387, 388]. Mutations in the S protein might change virus affinity towards infecting new hosts. For example, the mask palm civet SARS S gene acquired two mutations which increased its ability to bind to the human ACE2 receptor [384]. SARS-CoV and SARS-CoV2 S proteins both bind to the same mammalian receptor, but SARS-CoV 2 acquired mutations thanks to which it can bind to the human receptor with an affinity $\sim$10-15 times higher than that observed in SARS-CoV [376, 386, 389].

Most probably the SARS-CoV 2 ancestor originated in bats like *Rhinolophus spp.*, before jumping to an intermediate host and then finally to humans, following a similar path as other pandemic coronaviruses [334, 374, 384]. Bats are a major reservoir of betacoronaviruses, and bat betacoronavirus RaTG13 was found to be a very close relative of SARS-CoV 2 according to recent studies [376, 377, 378]. It was also found that SARS-CoV 2 is more closely related to RaTG13 (96,2%) than to SARS-CoV (79,5%), which suggests that SARS-CoV 2 may have originated in bats independently. The SARS-CoV 2 S protein is 98% identical to RaTG13 S protein apart from an insertion at the protease cleavage site which divides S1 and S2 regions [256, 390, 391]. This insertion was suggested as the cause of a higher SARS-CoV 2 transmissibility among humans, because virus is more efficient in binding to the human ACE2 receptor [392]. The outbreak of SARS-CoV 2 happened during the autumn and winter months when *Rhinolophus* bats were hibernating, which means that an intermediate source of spillover to humans is far more likely than direct transmission. Moreover, basing on the SARS-CoV 2 S protein residues it is unlikely that bat coronaviruses would be effective in infecting humans [393].

The intermediate host for this virus is still unknown with several candidates proposed over recent months, among them snakes, turtles, or pangolins [374, 394], although SARS-CoV 2 infections have also been detected in minks [395] and other carnivores [396, 397]. Most rodents display low susceptibility to SARS-CoV 2 infections, but some of them can be experimentally infected e.g. Golden Syrian hamsters [398]. Macaques [399, 400] and ferrets [401, 402] also displayed the potential for infection in experimental settings. Pangolin-CoV posses a high resemblance to the SARS-CoV 2 S protein at key residues similarly to RaTG13 [377, 393, 403]. Pangolin-CoV and civet-CoV share several key sites with bat coronaviruses, which might suggest that both ancestors of these viruses were circulating in bats before jumping to the intermediate species, like civets in case

of SARS-CoV and possibly pangolins in the case of SARS-CoV 2 [393]. Alternatively SARS-CoV 2 might have originated as a result of multiple jumps between bats and pangolins that could act as both reservoir and intermediate hosts allowing the virus to adjust better to infect new hosts like humans.

Comparative studies have been used before to investigate the possible SARS-CoV 2 host range basing it on its receptor affinity to bind with the S glycoprotein [277, 393, 404, 405, 406]. This research has attempted to analyse the ACE2 receptor from several different animal taxa to ascertain if their genetic make-up might help us predict species which have a potential to be SARS-CoV 2 reservoirs or intermediate hosts to add to the growing literature on SARS-CoV 2. The results that will be discussed in details in **Chapter V** not only confirm some of the previous findings and will add additional data to previously performed analyses.

5.2 Materials and methods

5.2.1 Database search and alignments

We retrieved ACE2 vertebrate orthologs from the NCBI Protein database and Refseq and WGS databases searched with BLAST [211]. All 301 hits (**Table A 10.1**) of the ACE2 gene sequence were aligned manually using Aliview software [4]. A separate alignment was made for the ACE2 SARS binding sites containing 81 amino acids (aa) in positions: 19-53/70-90/321-330/351-359 and 390-395 containing residues previously reported to interact with SARS-CoV 2 S glycoprotein: S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, N330, K353, G354, D355, R357, and R393 [376, 379, 405]; and human ACE2 RBD glycosylation sites which can affect

viral attachment N53, N90, and N322 [404, 405] (all aa according to *H.sapiens* sequence). The ACE2 SARS binding sites alignment was made using 285 sequences due to large gaps in the binding site region present in the sequences of a lower quality.

We compared percent identity matrix (% of identical bases in the alignment) of the ACE2 binding sites alignment to human sequence using Geneious R10 software (ver.10.1.3; http://www.geneious.com, [5]).

5.2.2 Phylogenetic tree reconstructions

Both alignments (ACE2 and ACE2 binding sites) were analysed with Prottest (ver.3.4.2.) [214] with the best models being JTT+I+G for the whole ACE2 sequence alignment and BLOSUM62+I+G for the binding site alignment. The maximum likelihood method in RAxML software (ver.8.2.12) [6] was used to generate the phylogenetic tree reconstructions. RAxML trees were visualised with FigTree (ver.1.4.4; http://tree.bio.ed.ac.uk/software/figtree). Generated trees were used later for the selection analysis. Both trees are rooted to the Australian ghostshark (*Callorhinchus milii*).

5.2.3 Selection and statistical analysis

We analysed the ACE2 binding site alignment in order to detect signatures of positive selection using the *codeml* package from Phylogenetic Analysis using Maximum Likelihood (PAML) (ver.4.9) [217] with the F3X4 codon frequency model and trees generated by RAxML software (ver.8.2.12) [6]. Under neutral evolution or purifying selection, the dN/dS ratio of the sequence should not deviate from 1 ($\omega \leq 1$). Scores deviating from that are evidence of positive selection ($\omega > 1$). The effect of a positive selection acting

on mammalian branches when held in comparison to the rest of the tree was tested by estimating ω values for the mammalian nodes and for the rest of the animal tree under the M1/M2 branch model [217]. The M1 model allows neutral evolution assuming an identical ω across all branches, and the M2 model constrains a specific node and allows positive selection ($\omega > 1$) to act on selected branches. For a conservative detection of sites evolving under ongoing positive selection within the ACE2 binding sites, the data was tested under the M8a/M8 site model. The M8 model allows all sites to evolve under positive selection ($\omega > 1$), and the M8a model uses ω fixed to 1 [217]. Models were evaluated under a likelihood ratio test (LRT) using a $\sim \chi^2$ distribution with the number of d.f. calculated from the number of parameters used and testing under a p-value <0.05. To attest whether mammalian-specific mutations had been under episodic positive selection we constrained the mammalian nodes from the rest of the tree under the branch-site model A (BSA). This model estimates ω values for the specific branches, classifying sites into four categories: 0 (purifying selection on all branches), 1 (neutral evolution on all branches), 2a (positive selection in selected branches and purifying selection for the rest of the tree), and 2b (purifying selection on the selected branch and neutral evolution for the rest of the tree) [217]. This model was evaluated with LRT against the null hypothesis (ω fixed to 1). The posterior probability of binding site codons under selection in BSA and M8/M8a models was inferred using the Bayes Empirical Bayes (BEB) procedure P $\geq$90% [3] (**Table 5.1**).

We also implemented several models from the Datamonkey server to further analyse positive selection of the ACE2 SARS binding residues: FEL (Fixed-Effect Likelihood) [299] and SLAC (Single Likelihood Ancestor Counting) [299] for codons with p-values ¡0.1; FUBAR (Fast, Unconstrained Bayesian AppRoximation) [300] for codons with the posterior probabilities >0.9; and finally, to check for episodic positive selection the MEME (Mixed Effects Model of Evolution) model [302] was used with a p-value threshold <

0.05 (**Table 5.1**).

5.3 Results

SARS-CoV 2 is a virus adapted to bind to a human ACE2 receptor [376, 379, 386, 389]. ACE2 orthologs should display residues similar to human ACE2 to confer efficient binding. In our analysis we look at SARS-CoV 2 binding sites together with neighbouring regions, due to insertions and duplications seen in some species and orders, which might result in shifts and changes to critical residues position. For example several fish species have an aa insertion at position N353 followed by K354 present in other animals. We removed this aa from our analysis to keep numbering of the positions consistent with an alignment. This change is present close to essential binding region residue that might be playing role in low susceptibility to SARS-CoV 2 and SARS-CoV infections in fish, since SARS-related (SARSr) viruses have never been found in fish.

We reconstructed phylogenetic trees for both full ACE2 sequence (**Figure 5.1**) and ACE2 SARSr virus binding sites in animals (**Figure 5.2**). On a full ACE2 tree reconstruction all animal classes and taxa branched together as expected from evolutionary relationships between taxa and families (**Figure 5.1**). There is clear distinction between separate classes, orders and families, with all amphibians (n=4), reptiles (n=17), birds (n=76), fish (n=68), and mammals (n=136) forming separate clades.

On the ACE2 SARS-CoV 2 binding site tree amphibian, reptile, fish and avian sequences still branch together forming four separate clades (**Figure 5.2**). Mammalian species form a separate clade but there is no clear distinction between families as seen on a full ACE2 tree, which might suggest diversifying positive selection pressures exhibited by mammalian species on these sites (**Figure 5.2**). All analysed bat species which branch

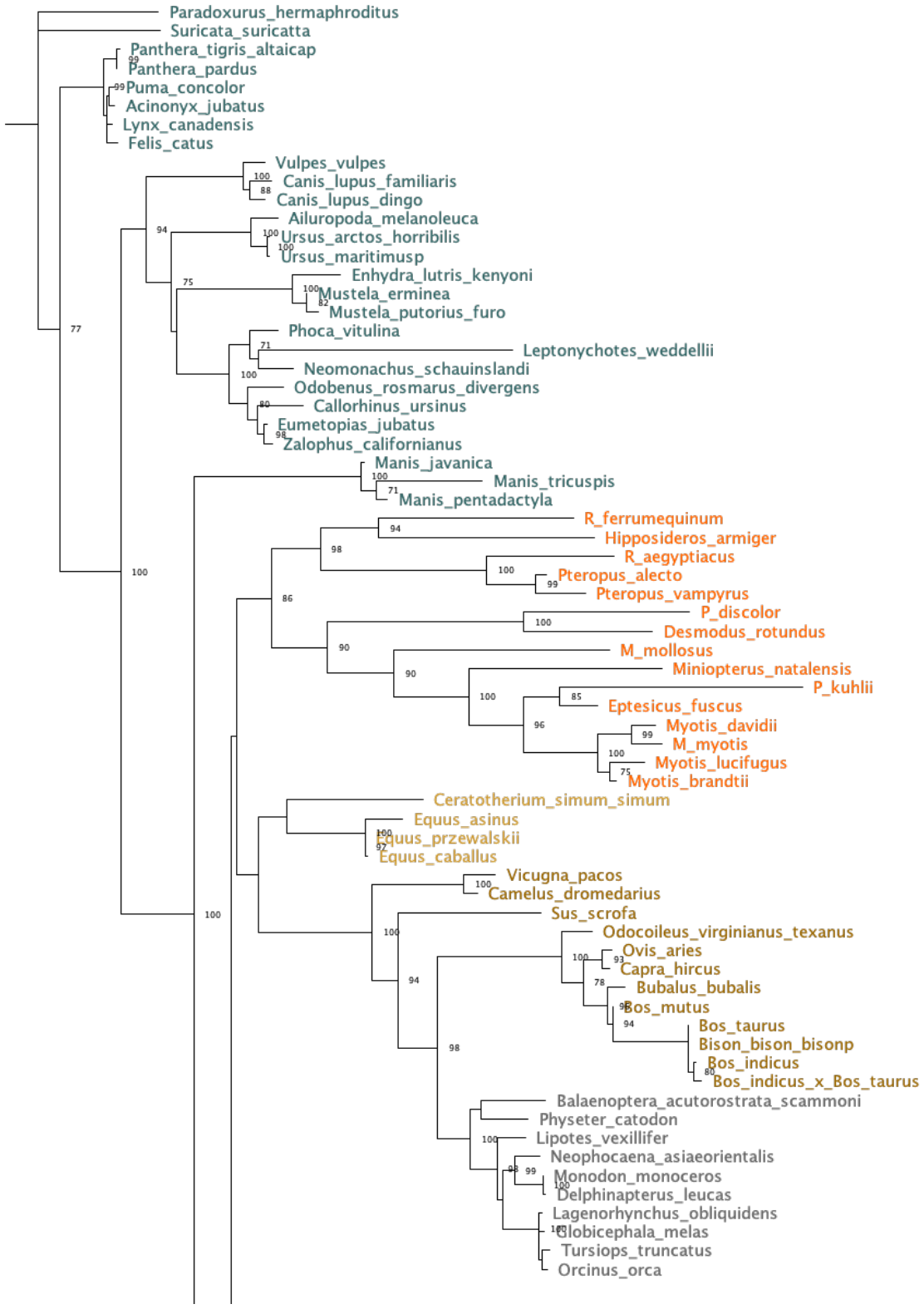


Figure 5.1: Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Continued on the next pages.

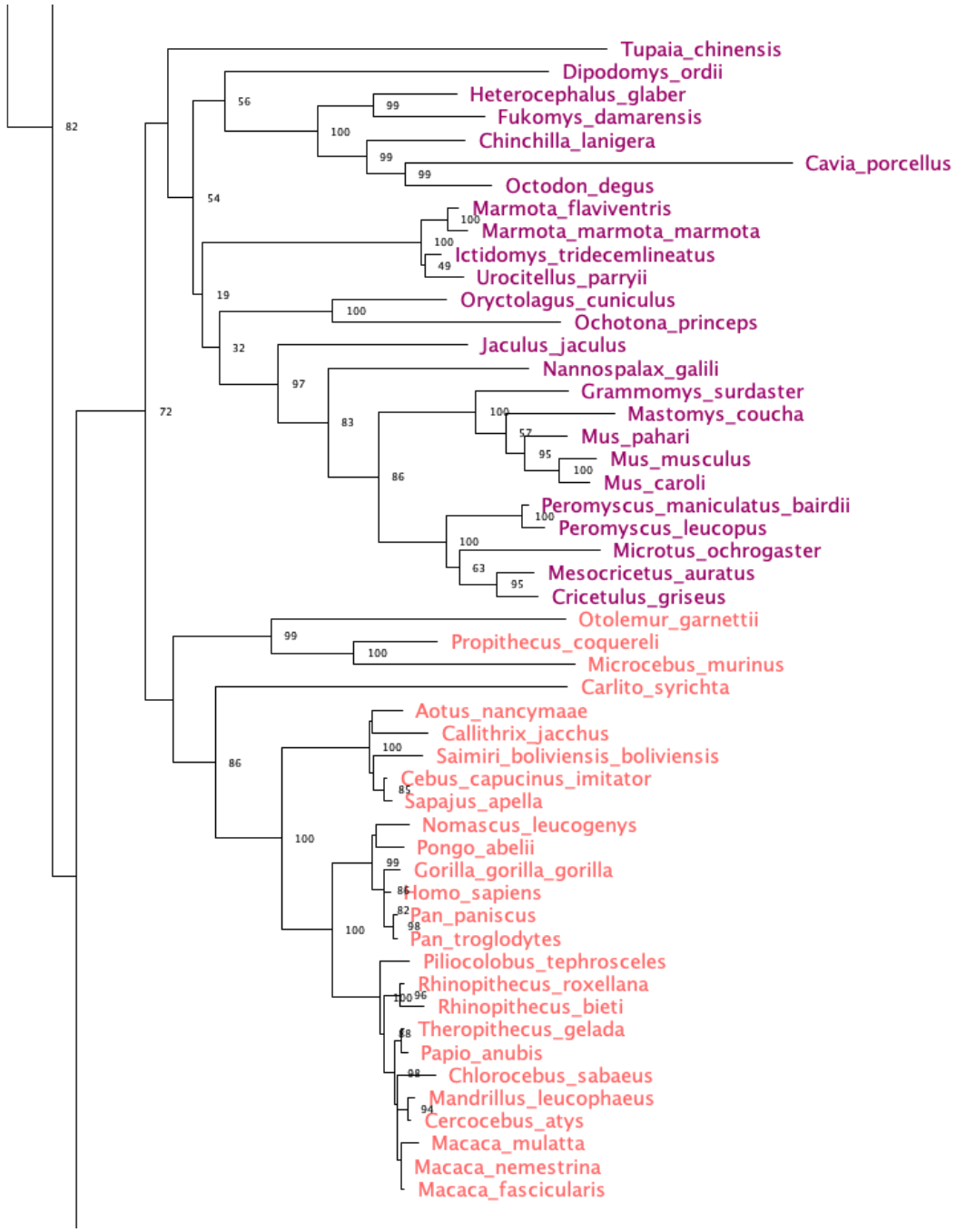


Figure 5.1: Continued. Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence.

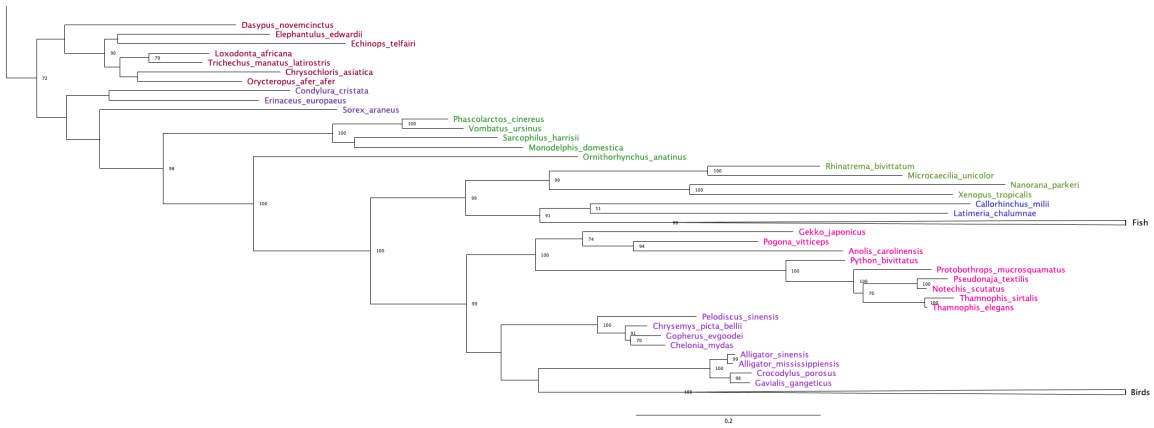


Figure 5.1: Continued. Phylogenetic tree reconstruction of full ACE2 sequence ortholog in animals. Separate classes marked with different colours. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence.

together on a full ACE2 tree are now branching with different taxa forming five different groups. The Egyptian Rousette bat (*Rousettus aegyptiacus*) is branching with the Cape Golden mole (*Chrysochloris asiatica*). All seven vesper bat species (*Myotis davidii, Myotis myotis, Myotis lucifugus, Eptesicus fuscus, Miniopterus natalensis, Pipistrellus kuhlii*) are on the same branch as representatives of four different families of mammals: three representatives of lemurs (*Microcebus murinus, Otolemur garnetti, Prooithecus coquereli*) that branch separately from other monkeys and primates, grey short-tailed oppossum (*Monodelphis domestica*), nine banded armadillo (*Dasypus novemcinctus*), and West European hedgehog (*Erinaceus europaeus*). The velvety free-tailed bat (*Mollosus molossus*) branches separately from other bat species together with Chinese tree shrew (*Tupaia Chinensis*) and Guinea pig (*Cavia porcellus*) on a rodent branch. Pale spear-nosed bat (*Phyllostomus discolor*) and common vampire bat (*Desmodus rotundus*) which both represent *Phyllostomidae* bat family branch with species like Arabian camel (*Camelus dromadarius*), alpaca (*Vicugna pacos*), and all three representatives of pangolin species (*Manis triscupis, Manis javanica, Manis pentadactyla*). Great roundleaf bat (*Hipposiderios armiger*) branches with horse representatives (*Equus spp.*). All three

representatives of *Pteropodidae* (*Pteropus alecto, Pteropus vampyrus, Rhinolophus ferrumequinum*) form a clade placed between carnivore branches formed by *Canidae* and *Felidae* species.

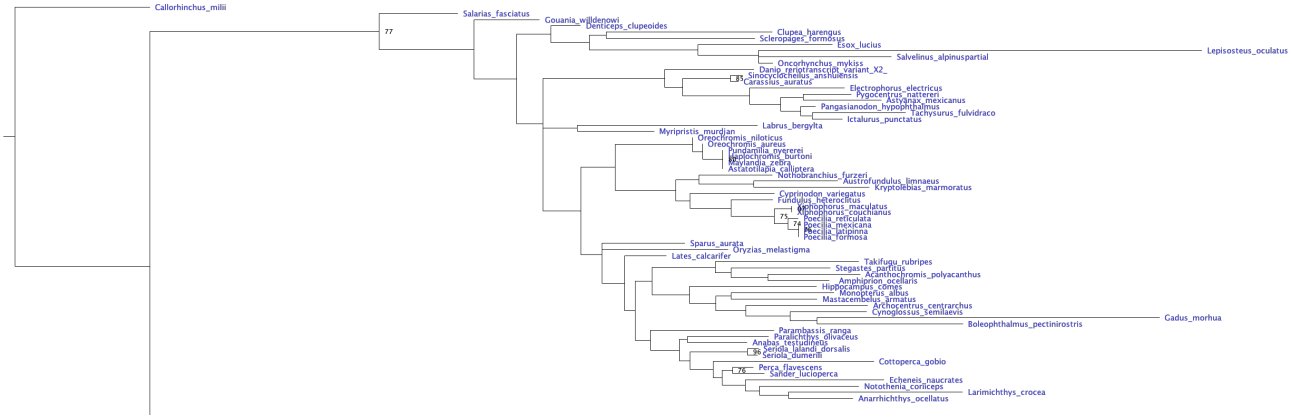


Figure 5.2: Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%. Continued on the next pages.

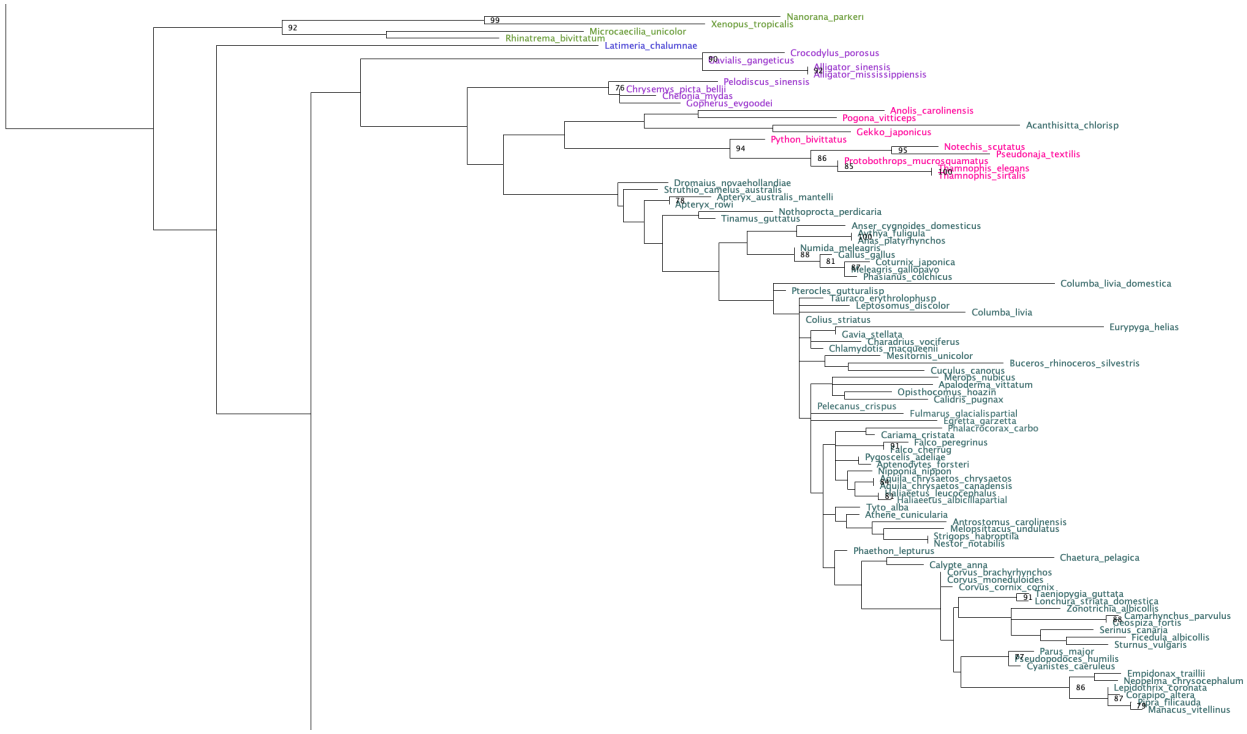


Figure 5.2: Continued. Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%.



Figure 5.2: Continued. Phylogenetic tree reconstruction of ACE 2 SARS-CoV binding regions in animals. Bird and fish clade collapsed for clarity. Tree is rooted to the Australian ghostshark (*Callorhinchus milii*) sequence. Boostraps displayed for the values ≥70%.

The percent identity matrix of ACE2 binding sites alignment showed the highest identity in primate species like macaques, gorillas and chimpanzees in comparison to the human reference sequence. All nineteen species which had an identity of 90% or higher in comparison to human ACE2 are representatives of primates (**Figure 5.3**). There are

49 species which scored 80 or higher (**Figure 5.4**). This group is more heterogeneous in terms of orders, and there are representatives of further related primates, carnivores like cats and dogs, rodents like squirrels and chinchillas, *Catacea* with representatives of whales, *Artiodactyla* like camels, and *Lagomorpha* (**Figure 5.4**). Species with a score between 70 and 79 are from orders of *Perissodactyla, Chiroptera*, further carnivores like palm civet and ermine, *Artiodactyla* like pig, *Catacea*, rodents like mouse, and all three species of pangolin (**Figure 5.5**). There are fifteen sequences with a score between 60-69, and among these are aardvark, marsupials like oppossum, tenrec, hedgehog, and three bats (*P. kuhlii, D. rotundus, P. discolor*) (**Figure 5.6**). All sequences with a score of 60 or higher are mammalian sequences apart from one sequence of an avian representative Chilean tinamou (*Nothoprocta perdicaria*) with a score of 61.25. All fish, reptiles, amphibians, and the remainder of bird sequences have scores lower than 60. Pairwise identity for all 285 species in comparison to the human sequence was 58.1% with pairwise positive similarity (BLSM62) 74.7%.

The LRT for the branch-specific model (M1/M2) was significant, meaning that positive selection affected a subset of sites during a specific evolutionary period (also called the episodic model of protein evolution) (**Table 10.3**). We ran a site specific model (M8a/M8) to detect if there were PSS within the binding region. LTR was also significant, detecting a few sites evolving under positive selection (**Table 10.3**). We also tested whether any of the sites on mammalian branches had been targets of positive selection at some point in their evolutionary history under the branch-site model A (BSA). The result again was significant with almost 5% of all tested sites found to be evolving under $\omega > 1$ in bats when compared to the rest of the tree at $P < 0.05$ (**Table 10.3**).

Selection tests run with Datamonkey and PAML *codeml* package identified several sites under positive selection in the binding sites. Some of them are selected by more than

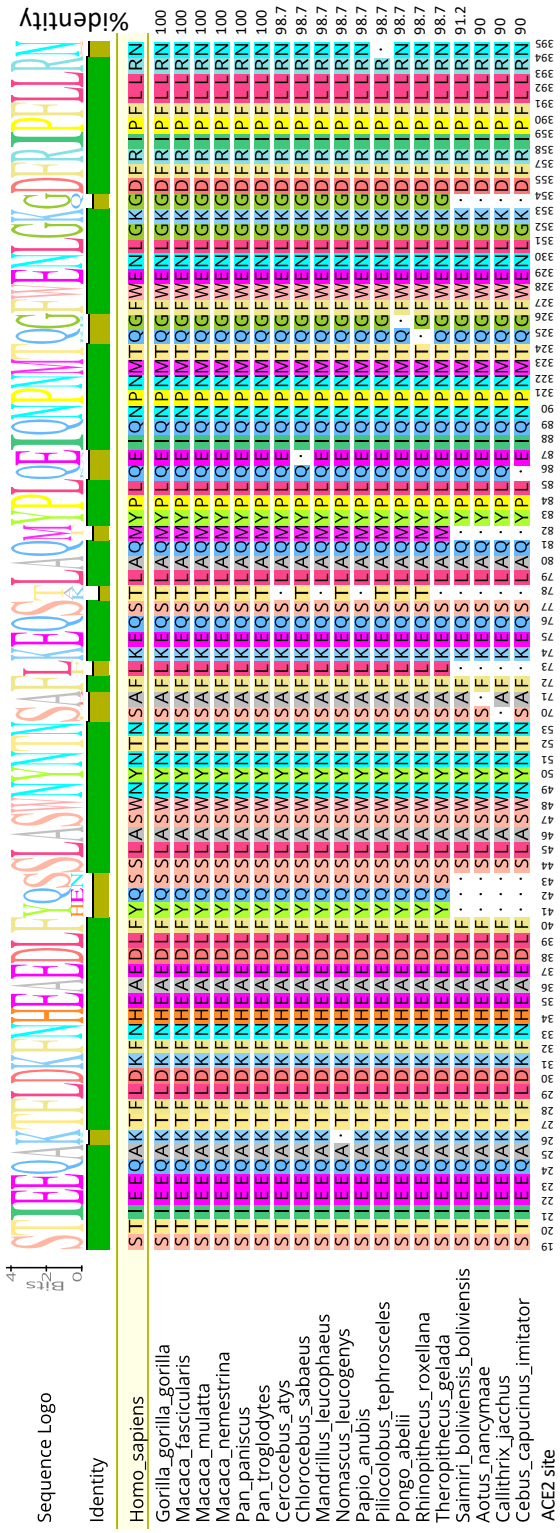


Figure 5.3: Amino acid alignment of SAR-CoV 2 ACE2 binding sites with identity of $\geq$90% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters.

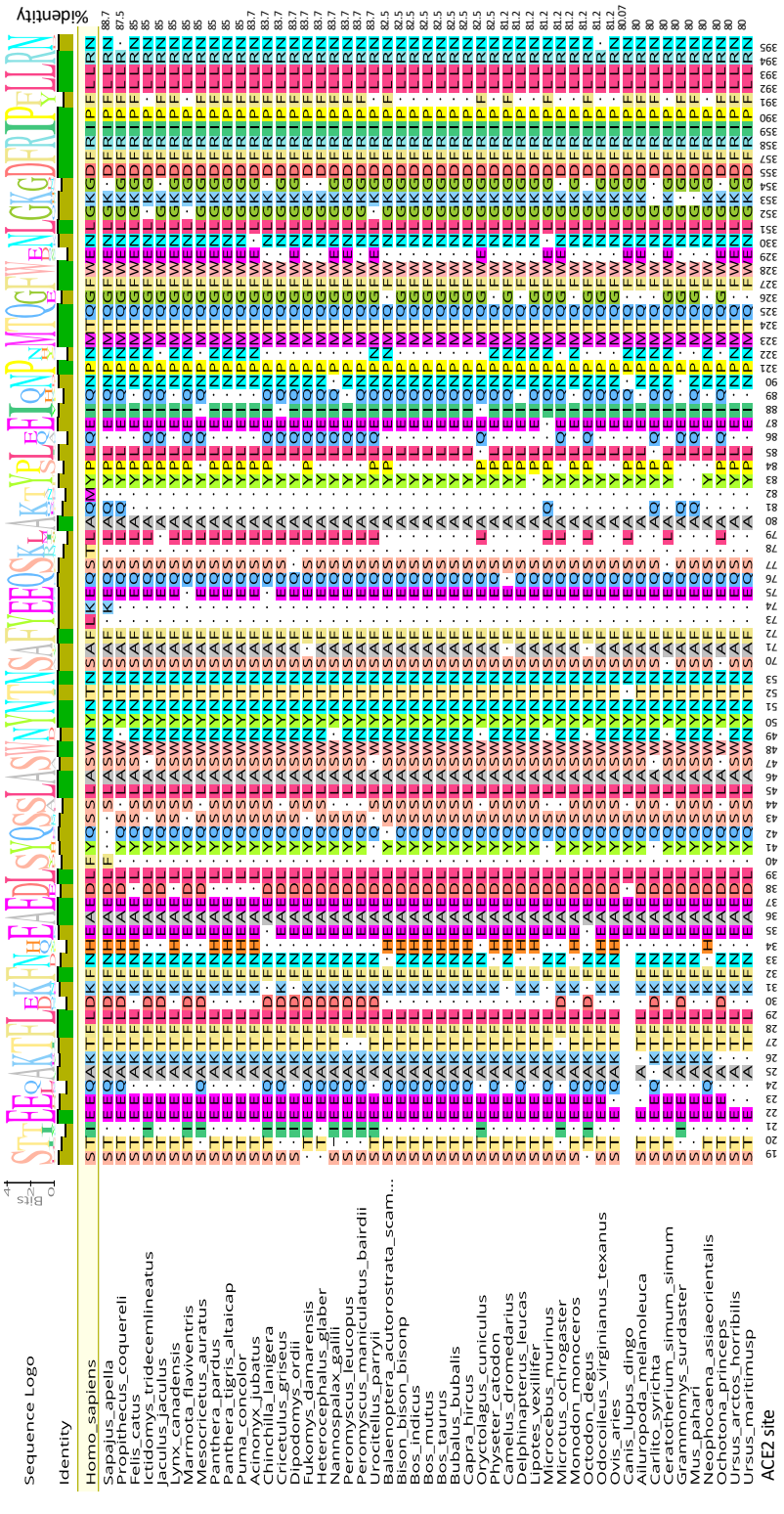


Figure 5.4: Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of $\geq$80% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters.

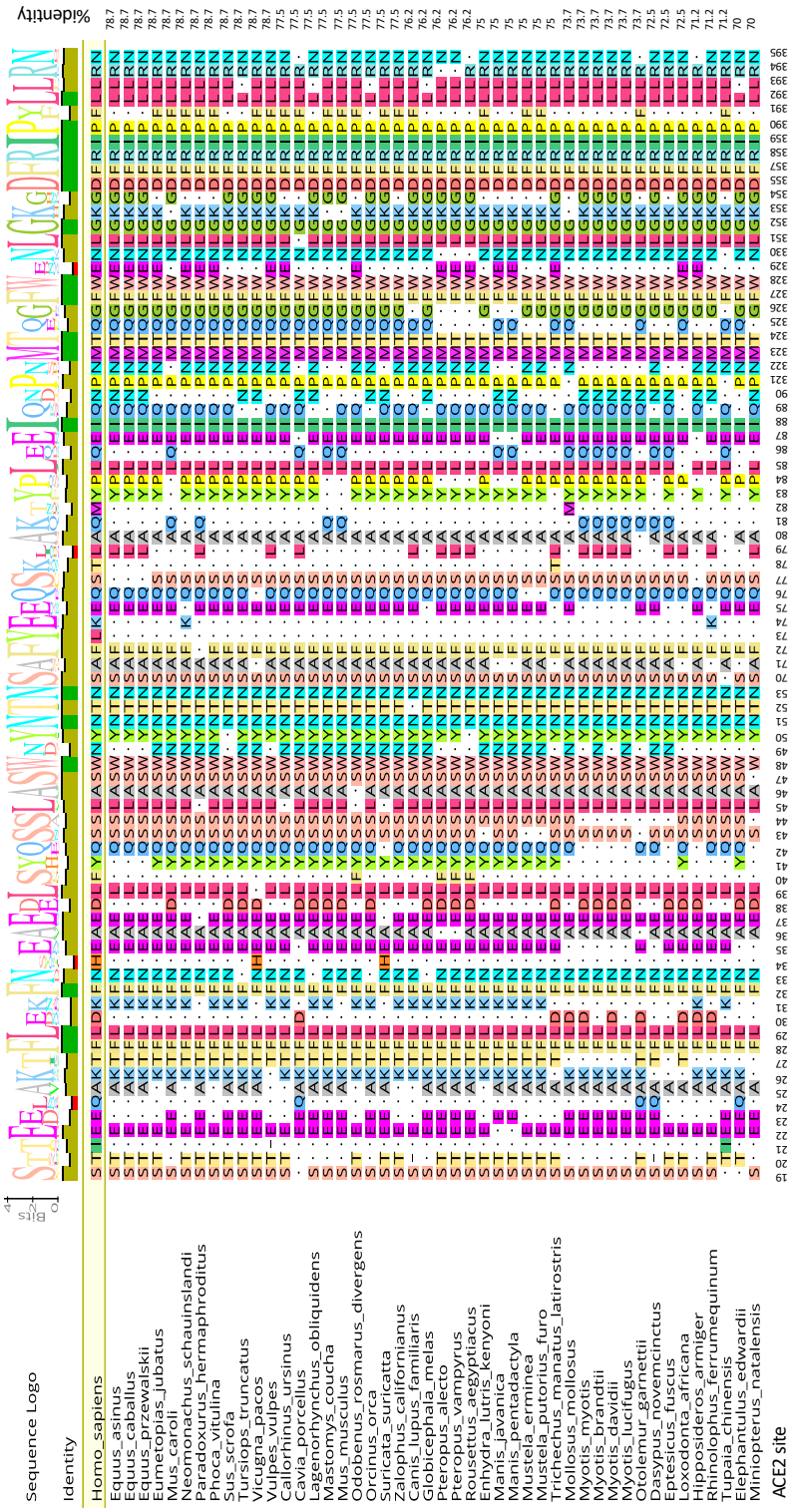


Figure 5.5: Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of $\geq$70% or higher in reference to the *Homo sapiens* sequence. Dots mark mismatched characters.

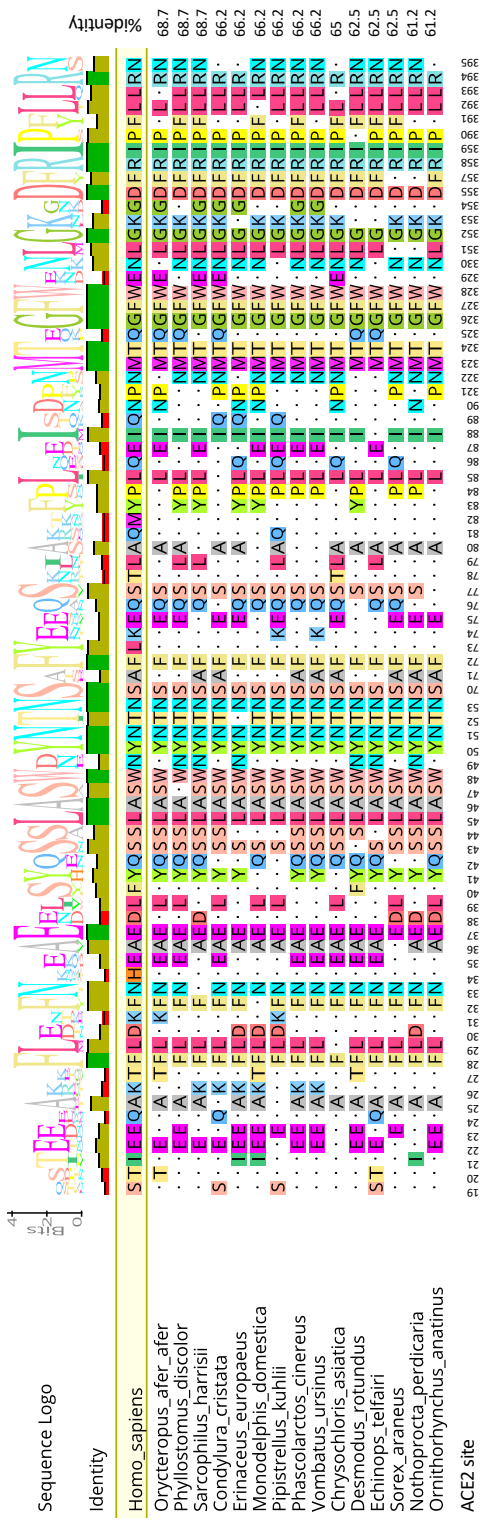


Figure 5.6: Amino acid alignment of SARS-CoV 2 ACE2 binding sites with identity of $\geq$60% in reference to the *Homo sapiens* sequence. Dots mark mismatched characters.

Positively Selected Sites					
	FUBAR	MEME	SLAC	FEL	PAML M2 (BEB)
Binding sites	27	22 **34** **41** 50 75 77 81 84 86 89 321 322 325 **326** 329 354 355	-	-	**34** **41** 82 **326**

Table 5.1: Sites under positive selection for the ACE2 binding sites alignment detected by each used Datamonkey model and *codeml* Bayes Empirical Bayes (BEB) procedure. Sites detected by more than one model in bold.

two tests which suggests strong positive selection. Positively selected sites of posterior probability higher than ≥0.9 was determined using BEB [3] procedure for M8 and BSA model which selected four residues for the ACE2 binding site alignment: 34, 41, 82, and 326 (**Table 5.1**). These sites were selected by more than one method which infers strong positive selection. Our analysis highlighted residues in which changes might disrupt virus binding affinity according to previous laboratory experiments [388, 407]: 84P, 354G, 355D (MEME), and 41Q (PAML, MEME), 82M (PAML).

The ACE2 protein is highly conserved in humans but some rare variants (S19P, I21V, E23K, K26R, T27A, N64K, T92I, Q102P and H378R) are predicted to increase susceptibility to the infection, or to show decreased binding to SARS-CoV 2 S glycoprotein (K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, M62V, K68E, F72V, Y83H, G326E, G352V, D355N, Q388L, and D509Y) [386]. After correcting binding sites alignment with the changed human variants we did not observe changes in the relative positions of species while comparing % identity matrix, and the only visible change was lower matrix scores for all tested species which might suggest that these aa differences are rare for the all classes of vertebrate tested.

5.4 Discussion

SARS-CoV and SARS-CoV 2 are both betacoronaviruses of the subgenus sarbecovirus, which are known to exclusively infect mammals with the natural reservoir believed to be

bats. Other genera of coronaviruses (alpha-, gamma-, delta-) were found to be able to infect both mammalian and avian species. This analysis confirmed that sarbecovirus's ACE2 binding sites, as well as the full ACE2 orthologs in fish, amphibians, reptiles, and birds are more distantly related to human ACE2 than ACE2 of other mammals (**Figures 5.1** and **5.2**). Binding sites of non-mammalian species have a very low % identity matrix which can be interpreted as a partial proof that SARSr viruses cannot infect other classes of vertebrates than mammals (**Figures 5.3** and **5.6**).

It is to be noted that the whole ACE2 gene alignment is a poor predictor of host affinity to SARSr viruses. ACE2 is a gene of a length of 805 aa and is very well conserved among vertebrates. Only ∼80 of these aa have an effect on the binding of the RBD S glycoprotein domain of SARSr viruses. Thus a more appropriate tool is the comparison of the selected binding regions which are in direct contact with the viral protein. The full ACE2 tree displayed conserved phylogenetic relationships between all the species used in the analysis (**Figure 5.1**). ACE2 restricted tree of the binding region showed a more mixed topology in the case of mammalian species (**Figure 5.2**). On the binding site tree *Phyllostomidae* bats species branched together with pangolins in which coronaviruses have displayed similar sequence patterns to SARS-CoV 2 [377, 393, 403] and camels which are an intermediate species for MERS-CoV. *Phyllostomidae* bats are living in the Americas while SARSr viruses have been mostly observed to originate in Asia (e.g. SARS-CoV, SARS-CoV 2), but since their ACE2 binding residues resembles pangolin sequence this might suggest that bats from this family may be a susceptible host for SARS-like viruses. The other branch is shared between *Rhinolophus ferrumequinum*, bat from a family which is known to be hosts for SARSr viruses, together with two Pteropus bats and carnivore clades with dogs, cats, and ferrets which have proven to be susceptible to infection with SARS-CoV 2, but also close to Asian palm civet which is a potential intermediate host for SARS-CoV. Primates have a high similarity between

their sequences and the human sequence, but lemurs branch further away from the other primates which suggests a longer evolutionary distance and might also mean lower susceptibility to SARSr infection.

Our binding site tree shows that *Chiroptera* ACE2 binding regions are exceptionally variable in comparison to other mammalian orders (**Figure 5.2**). This compliments another study which analysed ACE2 genes of primate species comparing them to *Rhinolophus* bats, showing that sequences from this genus contain all of the variations seen in non-primate mammalian species [408]. When comparing ACE2 binding residues of 46 bat species it was found that genetic variations were detected in all the analysed sites and from 25 residues 22 aa were under positive selection [406].

A study by Sun *et al.* [404] showed that in all tested species (n=24), except for the rat and mouse, two virus binding hotspots in positions 31 and 353 have same aa. That does not correlate with our data and both sites have much higher variability amongst mammals (**Figures 5.3 - 5.6**). It highlights how important it is to have a comprehensive analysis on a large number of species from different taxa [404]. Similarly to other studies we have also found sites which are positively selected or are adjacent to SARS-CoV 2 binding sites: 27T, 34H, 41Q, 81Q, 84P, 322N, 354G, and 355D (aa shown according to the *Homo sapiens* sequence; **Table 5.1**) [405, 406, 408, 409]. From these sites 41Q was picked up by more than two models in our analysis which suggests that it is under strong positive selection. Gene mutation experiments with SARS-CoV binding residues in human and rat ACE2 confirmed the high importance of residues 353-357 present in β-sheet 5, 31 and 41 present on α-helix 1, and 82-84 close to α helix 3, in which changes might disrupt SARS-CoV binding to the receptor [388, 407]. Based on human ACE2 aa positions, K353 residue forms a salt bridge with ACE2 D38, and K31 together with E35 forms a hydrogen bond with Q493 of the SARS-CoV 2 S protein [376, 379]. Replacement

of M82 residue and disruption of these bonds was proven to be major factor affecting the attachment of SARS-CoV [388]. Residues 41, 81, 84, 354, and 355 were highlighted by at least one of our models as being under positive selection (**Table 5.1** and **Figure 5.7**). The SARS-CoV 2 binding region is similar to that of SARS-CoV, with similarities probably driven by ACE2 conserved structure [379].

While bats are reservoir species for coronaviruses and a bat-borne virus is predicted to be ancestral to SARS-CoV and SARS-CoV 2, analyses on ACE2 binding sites of different bats show that bats are at low risk for SARS-CoV 2 infection and experiments on bat cells lines did not show infection when exposed to SARS-CoV 2 [406]. It also showed that even closely related species from the *Rhinolophus* genus differ when it comes to susceptibility to SARSr viruses [406]. Our phylogenetic tree reconstruction of the ACE2 binding sites also shows differences between bats, with some of them branching very distantly from the rest of *Chiroptera*, and others like *Phylostomidae* branching closely to species which are known to host SARSr viruses, like pangolins. But *Phyllostomidae* cell lines in a SARS-CoV and SARS-CoV 2 pseudovirus infection assay also showed that there are species specific differences in susceptibility to one or the other of these viruses [406].

Selection analyses showed that while the whole gene is highly conserved amongst vertebrates and in general under purifying selection, several residues in the SARS binding sites are under strong positive selection, including residues which can change binding affinity of the virus like for example site 41, which strong positive selection was indicated by two models (**Table 5.1** and **Figure 5.7**). Model detecting episodic diversifying selection flagged more critical sites which can impact virus binding to this receptor (82, 354, 355) that suggests these sites might have been under selection at some point in the mammalian evolution, which in turn might have impacted SARSr viruses distribution

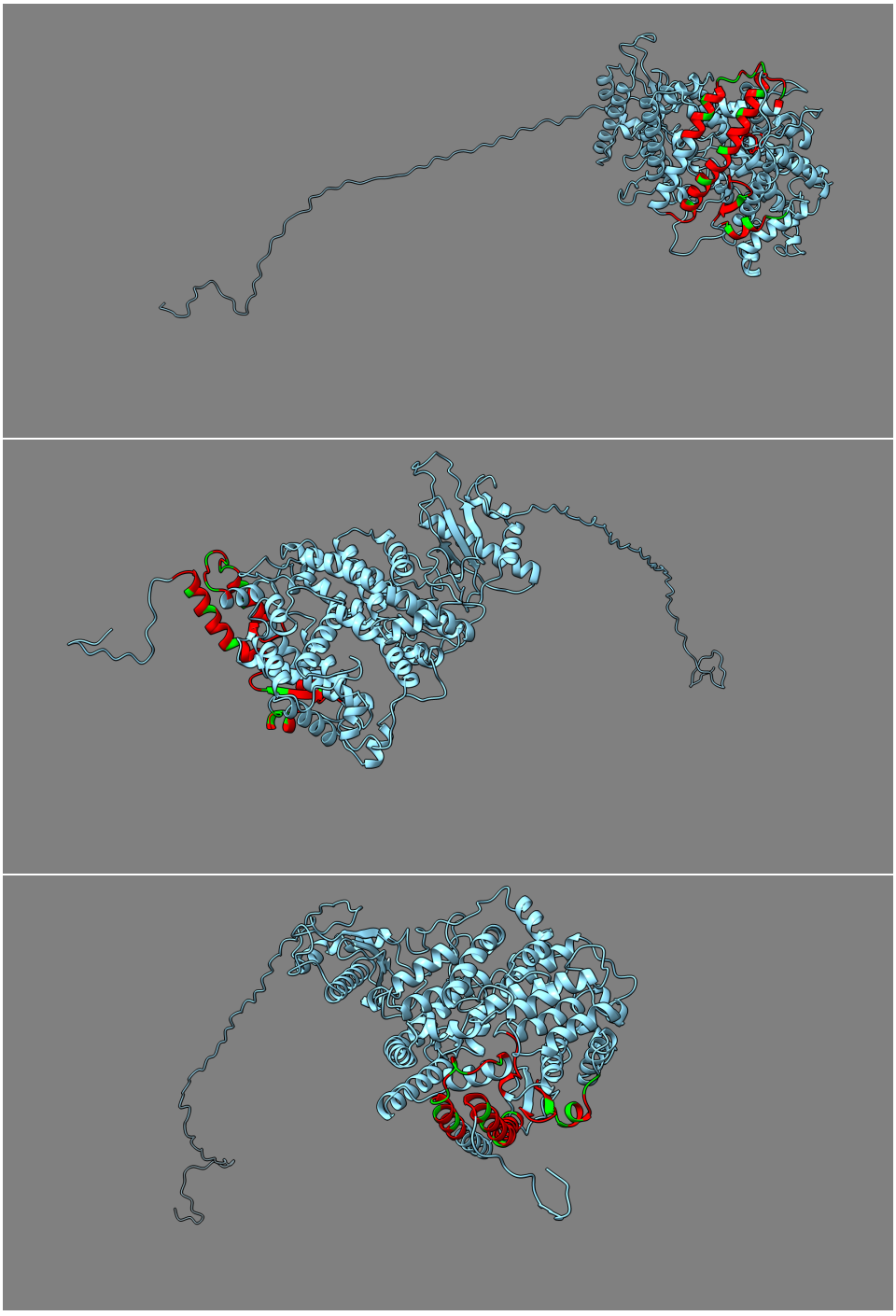


Figure 5.7: 3D structure of ACE2 protein based on *R.ferrumequinum* sequence visualised through ChimeraX [7, 8]. SARS-CoVr receptor binding domains marked with red colour; sites under positive selection marked with green colour.

(**Table 5.1** and **Figure 5.7**).

As expected, when comparing percent identity matrix of the ACE2 binding sites the most similar sites were from primates which are closely related to humans (**Figure 5.3**). The lower percent identities (80-60%) showed a mixture of different mammalian species with some of them already proved to be susceptible to SARSr infections like cats (85%), civets (78.75%), and pangolins (75%), while others proved their lack of susceptibility like mice (77.5%) but still displaying high identity scores (**Figures 5.4 - 5.6**). These results show that high aa identity scores are not the best predictor for verifying species susceptibility to the infection with SARSr viruses and need to be reconfirmed and followed by additional analysis, with biological characteristics of the species taken into consideration as well.

Receptor gene similarities are not a definitive predictor of host virus susceptibility. Additionally, cell line experiments cannot rule out that alternative receptors can be expressed in other cells or tissues. It is of note that the type of tissue and level of expression in the tissue of a receptor is also crucial when evaluating susceptibility of animals to the infection. Cats and mice both have high % identity of the binding sites (above 70%) and mouse branches closer on the phylogenetic tree reconstruction to human and primates than cat does. But cats proved to be more susceptible to natural infection [410] whereas infection experiments failed in mice. Data generated by the current studies suggests [411] that ACE2 expression patterns may be species specific, indicating that ACE2 is not expressed in a large number in mice, contrary to cats which express ACE2 orthologous on a high level in their skin and ears, making it easier for the virus to infect their cells [411, 412]. In humans ACE2 is expressed in tissues like testis, intestines, heart, kidney and pancreas [411], with the highest expression in the heart and testis, which follows clinical observations of possible cardiac disease and destruction of male germ cells

in testes in COVID-19, the disease caused by SARS-CoV 2 [277, 404, 413]. The same study found that ACE2 gene was highly expressed in kidney, liver, heart, muscle, and lung tissues in *Rhinolophus* bat; kidney, heart, large intestine, liver and tongue tissues in pangolins; and kidney, heart, liver, heart and lung tissues of ferrets. The fact that the tongue and lungs are exposed to air might make bats and pangolins more susceptible to infection with SARSr viruses, which are most often transmitted via respiratory fluids. In contrast to humans, kidneys were the tissues with the highest ACE2 expression in non-mammalian species [411].

Aa have different polarity and alternation of aa which changes the binding site polarity might potentially also prevent the virus from infecting the cell, which was suggested in the case of SARS-CoV [388]. It might be possible that complex ACE2 receptor interactions in different species can also influence virus binding affinity, as well as the possibility of alternative receptors existing in other species. Ferret scored 75 in % identity matrix (**Figure 5.5**), and proved to be susceptible to SARSr infections [401, 402]. But ferrets are also highly susceptible to other upper respiratory tract infections and serve as a model for respiratory disease. Because of the ferret's unique biology SARSr viruses might use highly efficient LSECTin (Liver sinusoidal endothelial cell lectin) receptors to enhance infectivity [405, 414, 415, 416].

Comparative analysis should be always performed on a large and diverse sample to detect all the patterns that may be present in the data. In the beginning of the SARS-CoV 2 pandemic initial analyses on the ACE2 receptor in mammalian and animal species compared only a very narrow number of species, a number which grew with further studies showing how the data can change with the choice of the right size of sample [277, 404, 405]. This analysis tried to follow that pattern and compare as large a number of species from multiple taxa as possible. It showed how significant differences there can

be not only between animal classes, but also orders, families or even species of the same family as can be seen in the example bats.

5.5 Conclusions

In this study we attempted to investigate if large-scale comparative analysis of host cell receptors could give us an insight into pandemic viruses binding affinity and receptor tropism. Our study showed that host receptors might substantially differ even between closely related species, patterns that would not be visible with the same methods used on much more limited data set, like e.g. just model species representatives. But the answer to the question on how broad certain virus host range can be is much more complex and cannot be based only on host receptor differences or similarities. The virus-host interactions are extremely complex and subject to many factors among the others: evolution of the host, and the pathogen, receptor number and distribution in tissues, or biophysical and biochemical properties of the cells, and virus-receptor binding mechanisms.

We recommend that further research concentrates on virus-receptor interactions and the types of receptors bound by pathogens as this might allow us to better understand the mechanisms relating to how spillovers occur, and predict possible sources and hotspots for pathogens with pandemic potential. More work on animal models could help establish target pathogens which have the potential to infect humans or jump between hosts.

In silico analysis of the host receptors can provide a useful guide but should be closely followed by laboratory work on respective receptors and species. As discussed above and proven by this and other studies, virus-receptor interactions are complex and genomic architecture does not always reflect real tissue and host tropism.

Apart from potential host susceptibility we should also look at the other part of the picture and investigate species not susceptible to infection. While highly specific virus receptor tropism is a barrier to infecting new hosts, we also need to learn about other intracellular barriers and host resistance factors which can prevent establishing an active infection. The differences in immune cells and receptors that can block infection at an early stage should be of particular interest [368]. Such mechanisms exist in a number of different animal species and might help answer some questions about infectious disease dynamics in the natural environment [368, 417, 418]. Additionally, large metagenomic analyses might be very useful in cases of various pathogens, although currently most analyses focus on bacteria rather than viruses [419]. Current experience with fast regular pandemic viruses shows that more work needs to be put into place to achieve a similar level of understanding in the case of viral pathogens.

6 Discussion and Conclusions

6.1 Key results, implications, and future prospects of this study

Driven by the recent SARS-CoV2 pandemic, the focus of the scientific community - as well as the public sector - has shifted to bats and bat-borne viruses. It is widely known that *Chiroptera* are the primary reservoirs and carriers of more than 200 viral diseases, which if transmitted to other species can cause enormous economic losses in agriculture, and mortality in humans and livestock [62, 199, 420, 421, 422]. Attempts to address this through culling have been unsuccessful. Decreases in bat population density does not have an impact on virus prevalence, and the increased levels of stress and higher birth rates in populations subjected to culling increase viral shedding [201]. However, endorsing such actions and portraying bats only in a negative light can have much more harmful consequences for the natural environment [423]. Bats have a large but overlooked ecological and economical importance, holding a special role in the ecosystem as pollinators, seed spreaders, and insectivores that control arthropod population [424, 425]. In light of this information, substantial effort has been put into bat preservation as it is evident that attempts to control their population can have catastrophic consequences within the ecosystem and in crop production [62, 426, 427]. In fact, the unique adaptations within the genomes of bats which allow them to tolerate viral infections without developing

symptoms might hold answers to multiple questions regarding infectious diseases, their co-evolution with animals, and zoonotic virus spillover [138, 177, 262]. In that regard bats should be seen as a research priority.

This thesis summarises work on the co-evolution of bats and viruses and our investigation on the implications of this long-standing relationship. First, we inspected bat genomes for the traces of ancient viral infections in search of co-opted viral genes which might have shaped bat immunity and in turn, their biology. Extending this search to reference quality genomes resulted in the discovery of a new retroviral class and several rare integrations with the complex retroviruses. This was followed by an examination of the bat's innate immune genes that highlighted bat-specific changes and adaptations that might have stemmed from host-pathogen arms-race. Following this, we also looked at sarbecovirus host's binding receptors in a broader mammalian context, knowing that bats are one of the major reservoirs of coronaviruses. This element of the study once again outlined divergent evolution for many *Chiroptera* displaying very different genomic profiles for this receptor across species and families.

All these findings help us to better understand bats' unique host-pathogen relationship, and add important data to what we already know about the immune systems of bats and what it is that makes them such a potent viral reservoir [263]. This is important in the light of the growing global pandemic risk and still existing gaps in our knowledge about host-virus co-evolution.

6.1.1 Chapter 2: The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology

First, we looked at gammaretroviral envelopes in several bat genomes (**Chapter 2**). Despite poor data quality we found a large number of co-opted viral sequences, some of them intact and with a potential to be expressed. Several of these sequences were detected in multiple species and were branching closely to *syncytins* - co-opted retroviral genes responsible for the development of placental reproduction in mammals, which have not been identified previously in *Chiroptera* [225, 428]. *Syncytins* found in other mammalian orders are homologous but not orthologous, which suggests that they were co-opted independently and have different retroviral ancestors [30]. Bats are a very divergent order of mammals displaying two types of placentation divided unevenly between *Yinpterachiroptera* and *Yangochiroptera* suborders. We have found multiple retroviral sequences under purifying selection that are present in bat representatives of one of the two suborders. If these candidate sequences prove that they can be categorised as *syncytins*, our findings could support the hypothesis that diverse types of placentation in mammals might be driven by different ERV ancestral sequences.

Even if sequences cannot be categorised as *syncytins*, several of these integrations which are under purifying selection can be a potential EDI. In general EDIs are very rare in vertebrate genomes [171], but we hypothesise that the resilience shown in bats in regard to displaying symptoms upon infection - or in some cases to clear infections more effectively than other mammals - may also stem from captured ERV genes as in the case of endogenous Jaagsiekte Sheep Retrovirus (*enJSRV*) in goats [44] or *Fv4* in murid rodents [45]. Such sequences might also be categorised as 'decaying *syncytins*', which began to lose their functionality over time after their function was picked up by a newly co-opted, functional version [47, 48]. Since bats are one of the oldest and a very divergent

order of mammals, additional work could show whether they contain a higher number of such sequences than other taxa.

Syncytins seem to also play a large role in myoblast fusion and muscle development [42]. Bats are one of the few mammalian species where the female is bigger than the male due to the necessity of flying with additional weight in the form of newborn pups. Due to this it will be interesting to see a study examining bat-*syncytin* knock-out gene study to ascertain whether *syncytins* would have a similar function in these mammals, but more visible in females rather than males in contrast to the mouse model [42].

Our study concentrated on *in-silico* analysis which confirmed the presence of conserved sequences in a number of species in a potentially functional state, while subjected to selective pressure with low rates of dN/dS, which is one of the canonical characteristics of *syncytins* [30]. We designed a laboratory *ex-vivo* assays using identified best candidate sequences to ascertain two other canonical *syncytin* characteristics - fusogenicity, and expression in placenta at a higher level than other tissues [30]. Unfortunately, due to scarcity of bat samples from multiple species, and especially bat placentas, this types of experiment is designed only to test *Yinpterachiroptera R.aegyptiacus* sequences (**Subsection 2.6**).

6.1.2 Chapter 3: Integrated retroviruses in six new bat genomes

While our first study was based on low-quality incomplete genome assemblies, we next investigated six new reference-quality bat genomes sequenced and assembled by the Bat1K Consortium. In this instance we have searched for all classes of ERVs and compared their numbers to those detected in mammalian reference species. Bats have one of the smallest genomes among extant mammals [1], but despite this we have found a

surprising number of retroviral sequences, in most cases greatly surpassing the number seen in mammalian reference genomes of a much higher genomic mass. We detected sequences from all classes of ERVs, including complex retroviruses which endogenize extremely rarely [71, 171]. Our results prove that bats and viruses have co-evolved for a very long time with bats acquiring a large number of integrations over time but not expanding the size of their genome, which suggests the existence of mechanisms which work to balance the gains and losses of genomic elements [130]. This is especially interesting if we compare bats to birds which are also a very old class of animals, but which also maintain a very constrained genome size, which might suggest some connection between that characteristic and the ability of powered flight which both birds and bats possess [130].

In addition to these findings, we detected a completely novel type of retroviral recombination with an alpha-like *env*, which may be evidence of a new retroviral class. Alpharetroviruses have never previously been found in mammalian genomes and until now were recorded only in avian species [16]. This finding alone confirms the special role of bats as viral reservoirs. *Env* genes have a key role in determining the range of viral hosts, while cross-species transmissions are believed to be very rare [16, 76]. This integration might be evidence of a retrovirus with the ability to jump between mammalian and avian species, and possibly evidence that infectious retroviruses can break the barrier between classes of animals, which might be an important factor to take into consideration in potential pandemic prevention strategies.

Our work showed that the existence of high-quality genomes of such a potent viral reservoir as bats might enable the discovery and classification of potential novel retroviral classes and viruses which endogenize rarely, some of them with the potential to express functions as host stand-alone genes, EDIs, or genomic pathogens. The discovery of new

previously unknown EDIs could help us to explain the low susceptibility of bats to some viral diseases that are typically dangerous in other mammals. While our study is based on only six genomes, considering the diversity of bat species, we should expect an even higher number of similar discoveries with the accumulation of new high-quality genome assemblies. Cross-comparison with other bat and mammalian genomes might enable us to discover more about the special relationship between bats and viruses, and possibly better predict potential zoonotic spillover in key species.

6.1.3 Chapter 4: Toll-like receptors in bat species

The long-term co-evolution between bats and viruses has not only impacted their genomic architecture but also altered the bat immune system and its receptors [138, 262, 429, 430, 431]. In our next study we examined TLRs, innate immunity receptors, which are the first line of defence against infection from different types of pathogens, and which are highly conserved in vertebrate species [432, 433]. Research on viral-sensing TLRs in bat species has previously been performed [152, 153, 362] but never on a full TLR profile of several species of *Chiroptera* or as a cross-comparison to the various representatives of other mammalian orders.

Bats are one of the biggest and oldest orders of mammals, consisting of multiple families that have evolved separately for millions of years [1]. Our analysis has elucidated the long-term evolutionary differences in a class of some of the most ancient immune receptors, showing that bat species differ between each other based on the number of TLRs which they possess. Some of these receptors are pseudogenised or missing in certain species or families, or quite the opposite: TLR classes previously mostly associated with rodents are still present in some of the analysed bat species [278, 277]. These differences

might also have shaped bat-pathogen relationships, displayed as either losses or perseverance of particular receptors depending on the presence or absence of the pathogen in the habitat of the specific bat. Our work shows patterns of positive selection in several bat TLRs which might be a sign of the differences in their responses to infections with specific types of pathogens, in addition to zoonotic pathogens and those which result in a high mortality rate in other mammals.

With the previous work performed in mice showing that TLRs can control the expression of ERVs it would be interesting to see if similar effects can also be observed in bats [288]. Our previous studies showed a large number of endogenous sequences in bats, with higher levels only having been observed in the rodent genome. It was also shown before that bats might be the only group of mammals which still have active families of transposons in their genomes. Whether their expansion is controlled by TLRs, or TLRs which have lost their controlling function which allowed these sequences to disperse in the bat genome might be a potential research avenue to explore further.

Bats are possible sources of novel zoonotic diseases which can have a devastating impact on entire countries, or even the whole world as we have seen in the past years. Our analysis makes an important contribution to the field of immunology. Urgent research is necessary into the connections between pathogens and their reservoir species and our work in enhancing our knowledge of the understudied field of bat immunology is vital to this.

6.1.4 Chapter 5: The change of receptor affinity and emergence of the new zoonotic pathogens: ACE2 receptor genetic differences amongst animals

Finally, we looked at the co-evolution between viruses and their hosts from the angle of the virus binding host receptor. The SARS-CoV 2 pandemic showed how difficult it is to identify reservoir hosts of newly emerging zoonotic pathogens. Our comparative study on a large number of different vertebrates from multiple taxa has shown that changes in key residues of the receptors might differ even among closely related families. This was especially emphasized between representatives of bat species which branched with different mammalian orders in five separate groups. Our study also confirmed that the analysis of these key residues is more important than the analysis of the whole gene. This is especially important in the case of well-conserved genes, analysis of which can return results based on the host phylogenetic relationships between taxa, rather than actual positive or purifying selection at the certain residues of interest [386].

Comparing our findings to the early studies on ACE2 receptors, our results have differed from the ones presented by smaller studies [411]. Our work has proved that comparative analyses should be performed on sequences from many diverse species to receive representative results. We are also pointing out that genomic studies can be treated mostly as supporting studies and need to be followed up with additional functional work and investigation of other processes connected with the receptor which might alter viral binding. What is more, identifying a wildlife reservoir would require broad scale ecological sampling and cannot be replaced by *in-silico* or *ex-vivo* studies [62]. Our study pinpointed some species which are known to be susceptible to infection, but also some which have been proven to be at a low risk of infection, like murid rodents. We stress

that there are several other factors which can be determinant in species susceptibility apart from receptor genetic make-up.

6.1.5 General implications and future prospects for the bat-virus interactions and co-evolution

With the exception of the Ebola virus outbreaks, which are from time to time reported in global media, bats have not regularly been attracting the attention of either the science community or the public, despite their status as a potent viral reservoir and living in close proximity to humans. Studies on bat immune systems are lacking and have only started to be more intensive in the last 10-15 years [62, 138, 262]. This trend suddenly shifted after the emergence of SARS-CoV 2 and identifying bats as the most likely origin of the coronavirus pandemic [334], and now the eyes of everyone are on bats and the viruses they host.

There is still a lot we do not know about bats, their immune systems, or even the basics like how flight impacts their broader biological functions. In the research presented in this thesis we attempted to fill in some of the gaps in our knowledge about *Chiroptera*, especially regarding their unique host-pathogen relationship. Bats have evolved unique mechanisms involved in their immune responses, such as the constitutive expression of interferons [155], or stringent controls of inflammatory processes [434]. If we want to learn more about the potential for bats to be viral reservoirs we need to concentrate more on studies involving bat immune systems like for example innate immune pattern recognition receptors e.g. nucleotide oligomerization domain-like receptors (NLRs), or adaptive immune response and T- and B-cells on which studies are still greatly lacking [435].

However, studying bat biology and immunity is not easy, and our research has suffered and has been additionally slowed down by difficulties in obtaining bat tissues or cell lines. In order to confirm *in-silico* findings more functional and general laboratory experiments need to be performed with the use of bat tissues and bat cell-lines. This is currently a big issue with only a few established bat cell lines, as well as very difficult access to tissues, such as placental tissue [138, 262, 435], which has not allowed us to fully explore the nature of the sequences found in our analysis. Many bat species are endangered, and even with the possibility of obtaining all of the needed permissions, capturing wild animals is not a trivial task as bats tend to live in places hardly accessible for humans, are not easy to capture, and for many species their numbers are dwindling [436]. Encountering and capturing wild bats is also risky for researchers due to the danger of rabies or infection from other bat-borne viruses, and maintaining wild animals in laboratory settings to perform *in-vivo* experiments is difficult and costly. The development of bat cell lines from multiple species and tissues should be a priority to limit the contact with wild animals, and be able to look more into virus-host interactions, and infection kinetics. However, establishing such cells is also impacted by its own challenges, e.g. choosing correct tissues with the most suitable kinds of receptors, or developing cell lines which will be viable in conditions similar to bat physiology [435]. Substantial effort should also be put into establishing research bat colonies which can be studied to shed light on the immune response exhibited by these animals, and their susceptibility to infections with no need of invasive sampling of wild animals[179].

Bats in general should not be treated as a single species, and in this thesis we have stressed multiple times that bats are an ancient and extremely diverse order which has led to their developing traits and characteristics sometimes unique to a specific family, or even species. For example, even though one species of *Rhinolophus* bat has been identified as a potential reservoir of a SARSr virus ancestor, subsequent analysis

has shown that other *Rhinolophus* bats are not susceptible to infection with one or both sarbecoviruses, which might stem from the differences in the receptors of different species having co-evolved alongside some of these viruses in isolation from one another [406]. If we want to get more general answers when studying bats we need to analyse multiple different representatives of various families, but we should also keep in mind the divergence of this order, and that being a virus reservoir for one virus might be true only for a specific species of bat, and that there is a chance that we will also see differences in immune responses or immune receptors between families or even on the level of a single species.

Studying host-pathogen interactions should not only be limited to the immune response but should also concentrate on other factors, such as ecology or biological processes of the host. Stress increases viral shedding [201], similarly to possible hormonal changes which can be linked to increased viral replication [435]. All of these might be important markers to predict periods of time or populations at a higher risk of a virus spillover [435].

It is important to note that until recently, two of the most important model species for laboratory work have been house mice and fruit flies. They have a high reproduction rate and are easy to keep in a laboratory environment, making them ideal for research, but both of these species are related distantly to humans so might not be ideal subjects to study some syndromes or diseases. Due to this, bats may well be an alternative model in some circumstances, especially if we take into consideration the fact that bats are related to *Homo sapiens* at a higher level than mice and live longer than rodents which might be a useful characteristic in ageing studies [1, 437, 438].

Our work on bat receptors clearly shows how divergent the *Chiropteran* order is with multiple families and species evolving separately from each other for millions of years.

Our results show clear, stark differences between bat families in their genetic make-up, their number of ERVs, the presence or absence of innate immunity receptors or different levels of positive selection acting on specific residues in their genes. If we want to get a full picture from virus-host studies and model more generalized bat immune responses and mechanisms we need to perform large-scale comparative studies rather than focusing on single species as a proxy for all bats, and recognise that family-level studies will not encapsulate the diversity of responses across species.

6.2 General future prospects and directions for the host-virus co-evolution studies and implications for the science and global infectious diseases threat

Viruses and their reservoir hosts also show equilibrium - co-existing in mutual, commensal or parasitic relationships. Thanks to genetic and physiological adaptations evolved over long-term virus-host relationships, there may be no pathological symptoms visible in a reservoir host upon infection [62, 66, 374, 439]. Although the public tend to view them as 'agents of death and disease', viruses are not biochemically adapted to compromise their hosts' health and survival, but rather simply to replicate, and long-term co-evolution tends to lead to lower pathogenicity [56]. However, when a virus jumps the species barrier it can become deadly for its new host, which lacks genetic adaptations present in the reservoir host [62, 374]. This virus 'spill over' happens through intrusion into a reservoir species' ecological niche, or the exposure of an infectious agent from a reservoir host or a previously isolated environment [200]. Spillover events may be dead end infections where the virus cannot spread efficiently or at all from the new host, but

they can also lead to epidemics or pandemics [62, 367]. In such situations, the fast development of vaccines and new drugs can prove to be problematic, as has been seen with the Ebola virus (EBOV), Severe Acute Respiratory Syndrome coronaviruses (SARS-CoV and SARS-CoV 2), Hendra virus (HeV) and Nipah virus (NiV) outbreaks; all bat-borne viruses that have proven difficult to treat and control [110, 203, 206, 208, 320, 334].

Zoonotic diseases (zoonoses) are infectious diseases of animals which can be transmitted to humans. They are one of the greatest concerns for global health organizations, as more than 70% of all emerging diseases are zoonotic in origin [200, 440]. Climate change and the loss of biodiversity caused by human activities are the major drivers of infectious disease spill-over from animal reservoirs into human and livestock populations [441, 442]. Wild animals are hosts to many pathogens which regularly cause outbreaks of novel diseases [367], but in recent years we have begun to see outbreaks of new zoonotic pathogens more often [441, 443], with bats being the major source of emerging viruses [62, 444]. The rate of discovered zoonoses has increased over time for several reasons. We are better equipped to detect and differentiate diseases via new diagnostic assays, and can study them in greater detail. There are also ecological reasons - the growth of human populations and the loss of wild areas to human use (i.e. agricultural, urban, or industrial), as well as climate change has resulted in habitat changes [440, 445]. This disturbed equilibrium has increased contact between wild animals, humans, and livestock [62, 199]. This is especially visible in developing regions, which interfere with natural habitats more due to rapid development, causing wild animals to move closer to human dwellings or farm animal shelters, and which have less stringent controls on animal markets and animal trade.

Despite this, spillovers of pathogens into new host populations are relatively rare events. A pathogen needs to pass several barriers to be able to establish a state of active infection,

and even more to cause an epidemic [367]. The probability of that happening can be minimised with appropriate policies in place [443, 446, 447]. But in order to understand emerging pathogen dynamics there is a desperate need for further research into potential host ecology, population biology, and behavioural ecology. With the further growth of agriculture in developing regions we need to pay special attention to the landscape of ecology and agricultural sciences, as well as animal epidemiology and the infectious disease dynamics for the region. The loss of natural habitats and of biodiversity pushes humans and wild animals to live closer together, while climate change forces wild animals to migrate and find other environmental niches, sometimes rapidly changing dynamics and the balance of existing ecosystems, introducing new pathogens with the potential to cause outbreaks of new diseases [448]. We need to understand that these drivers will make the possibility of fatal zoonotic spillover more feasible, and we need to establish surveillance which can monitor possible zoonotic reservoirs and hosts, such as bats, birds or rodents.

Potential zoonotic reservoirs need to be closely monitored globally, and appropriate ways of approaching potential zoonotic disease sources should be discussed and observed by the authorities in order to avoid interferences which might disturb delicate balances in the environment and drive pathogen spillover events [201, 443]. For example, recurring outbreaks of Nipah disease could have been controlled by separating animal enclosures from the forests in which bats roost and eat [201, 449].

Domesticated and wild animal interactions are only one way of new infectious disease emergence into population [442]. Direct human contact with wild animals, wild animal products, or bushmeat allows susceptible humans to be exposed to an array of pathogens previously rarely encountered. An understanding of the cultural and economic importance of practices using animal products in many low income countries is a neglected part

of how infectious diseases can spread and cause outbreaks in susceptible communities [443].

Identifying reservoirs and hosts of zoonotic diseases in wild animals is challenging, and even more so in developing regions which do not have well developed surveillance or the backing of well-equipped diagnostic and research facilities [418, 443]. Many countries with limited resources have resorted to fighting outbreaks of infectious diseases only when the threat was already present in the population. Not performing active observations of the possible sources of the zoonoses has led to delays in source identification, making it extremely difficult or even impossible to further identify, as we have seen in the cases of both SARS-CoV 2 and previously Ebola. The approach of dealing only with the emerging outbreak and not putting preventative measures in place first might seem attractive from the perspective of low-income countries, where the emergence of an epidemic is not something easy to predict at any given time, and where funds are limited [450]. This has been shown to be counterproductive and has, over time, cost far more than running active surveillance which can prevent human and economic losses caused by uncontrolled infectious disease epidemics [451]. As we should implement more pro-active, rather than reactive approaches, we should first be targeting global surveillance in epidemic hotspots as another way of tackling that problem [418].

Pathogens also evolve at different rates, and in the case of viruses, they cannot evolve at all without the cells of their host. There is a high proportion of negative-stranded RNA viruses among zoonotic pathogens for e.g. coronaviruses, or the influenza virus. These viruses have a fast mutation rate and lack the proofreading mechanisms required in order to correct mistakes during replication. That might make them more prone to host jumps with the genomic changes helping them to widen host and tissue tropism with time. These mechanisms are not the only ones which can drive species jumps, and

other factors are still poorly understood [367]. Mathematical modelling of the mutation rates of infectious diseases and spill-over risks can also help in epidemic prediction [452]. Mathematical models are relatively cheap and are a fast way to gain knowledge about the evolution of zoonotic pathogens and potential pandemic progress. When used correctly, they can be one of the main components of epidemic control [452] and the collaboration between infectious disease experts, ecologists and public health scientists might help make them more accurate. Observations of the dynamics of a pathogen in the reservoir or intermediate host can help us to predict the human infection risk and act as an early warning system reducing the risk of an outbreak and the scale of fatalities and economic losses.

The development of fast, cheap and reliable methods of sequencing genomes as well as publicly available global databases of pathogen genome sequences, metagenomic analyses, and high-quality genome assemblies of pathogen reservoirs provide opportunities to devise strategies to provide hope to low-income countries and infectious disease hotspots which do not have established networks of diagnostic facilities. It can also help the scientific community to quickly identify the danger quickly and perform preliminary analyses [453, 454]. Early detection and identification of emerging pathogens might help to stop the spread of infection and prevent a potential epidemic outbreak[418, 455, 456], something which failed in the initial stages of the SARS-CoV 2 pandemic, and which we should try to prevent at all costs in the future.

7 Appendix to Chapter 2

Bat	Genome coverage	N50/L50 ratio	Number of contigs in genome assembly
Myotis lucifugus	7x	64330/8574	72785
Myotis brandtii	120x	23289/24039	325414
Myotis davidii	110x	15182/36107	325280
Eptesicus fuscus	84x	21392/24180	167058
Miniopterus natalensis	77x	29777/17067	114632
Rhinolophus ferrumequinum	17x	11659/49796	290685
Rhinolophus sinicus	146.44x	37803/15841	182995
Pteropus alecto	110x	31841/18318	170164
Pteropus vampyrus	188x	21866/22949	225433
Roussetus aegyptiacus	169.2x	1488988/393	3049
Eidolon helvum	18x	12668/42661	288446
Hipposiderios armiger	218.6x	39863/13917	128956
Pteronotus parnelli	17x	9502/58483	443121
Megaderma lyra	18x	7043/68839	518327

Table 7.1: Bat genome assemblies data

Ortholog set	dN/dS (one-ratio)	Neutral evolution test (M0 ω=1 vs M0)	M1(neutral)	M2(positive selection)	M3(discrete)
VesperOrtho1	0.479	Neutrality rejected(61.389886 df=1)	0.494	0.5518	0.5506
VesperOrtho2	2.1077	Neutrality cannot be rejected (0.49425 df=1)			
VesperOrtho3	0.9698	Neutrality cannot be rejected(0.015048 df=1)	0.8604	1.067	1.067
VesperOrtho4	0.7137	Neutrality cannot be rejected(0.845898 df=1)			
VesperOrtho5	0.70006	Neutrality cannot be rejected(2.04 df=1)	0.6253	0.7782	0.7782
VesperOrtho6	0.94303	Neutrality cannot be rejected (0.11324 df=1)			
VesperOrtho7	0.4374	Neutrality cannot be rejected(3,22 df=1)			
VesperOrtho8	1.2085	Neutrality cannot be rejected(0.23843 df=1)	0.5655	2.1763	2.1763
VesperOrtho9	0.4714	Neutrality rejected(4.93258 df=1)	0.4714	0.4714	0.4714
MegabatOrtho1	0.7799	Neutrality cannot be rejected(2.23 df=1)	0.781	0.7958	0.7958
MegabatOrtho2	0.8264	Neutrality cannot be rejected(0.040764 df=1)			
MegabatOrtho3	0.4561	Neutrality rejected(7.23193 df=1)			
MegabatOrtho4	1.14984	Neutrality cannot be rejected(0.278162 df=1)	0.7876	1.363	1.363
MegabatOrtho5	0.3708	Neutrality cannot be rejected(1.326184 df=1)			
MegabatOrtho6	0.5345	Neutrality rejected(10.974878 df=1)	0.5409	0.5409	0.5403
MegabatOrtho7	1.0502	Neutrality cannot be rejected(0.011084 df=1)			
MegabatOrtho8	2.4412	Neutrality cannot be rejected(0.374048 df=1)			
MegabatOrtho9	2.7831	Neutrality cannot be rejected(1.135522 df=1)			
MegabatOrtho10	4.5747	Neutrality cannot be rejected(3.27657 df=1)			
MegabatOrtho11	0.6593	Neutrality cannot be rejected(0.302754 df=1)			
MegabatOrtho12	1.00741	Neutrality cannot be rejected(0.000512 df=1)	1	1.0049	1.0074
MegabatOrtho13	1.2891	Neutrality cannot be rejected(0.53744 df=1)	1	13.0105	13.5327
MegabatOrtho14	0.9991	Neutrality cannot be rejected(0.cannot be r000008 df=1)			
MegabatOrtho15	0.62078	Neutrality ejected(2.04 df=1)	0.528	1.1675	1.1579
MegabatOrtho16	1.7135	Neutrality cannot be rejected(0.706314 df=1)			
MegabatOrtho17	0.5471	Neutrality cannot be rejected (2.84 df=1)			
MegabatOrtho18	1.9117	Neutrality cannot be rejected(0.389932 df=1)			
MegabatOrtho19	0.3652	Neutrality rejected(8.32724 df=1)			
MegabatOrtho20	0.87176	Neutrality cannot be rejected(0.179014 df=1)	0.8334	1.1649	0.9075
MegabatOrtho21	2.6365	Neutrality cannot be rejected(2.289654 df=1)	1	3.3949	4.0669
RhinoHippoOrtho	0.59204	Neutrality rejected(11.009006 df=1)	0.5505	0.7069	0.7069

Table 7.2: dNdS data. * difference of likelihood ratio test, and degrees of freedom (df) with significance level of 0.05 in brackets

Ortholog set	M7(beta)	M8(beta&w>1)	M0 vs M3*	M1 vs M2*	M7 vs M8*
VesperOrtho1	0.4972	0.5506	Cannot reject M3(63.155254 df=4)	Reject M2(3.793126 df=2)	Reject M8(5.808296 df=2)
VesperOrtho2					
VesperOrtho3	0.9	1.0673	Reject M3(5.024954 df=4)	Reject M2(4.572346 df=2)	Reject M8(4.613872 df=2)
VesperOrtho4					
VesperOrtho5	0.6	0.7782	Reject M3(4.891348 df=4)	Reject M2(2.090008 df=2)	Reject M8(2.130014 df=2)
VesperOrtho6					
VesperOrtho7					
VesperOrtho8	1	2.1759	Cannot reject M3(21.793344 df=4)	Cannot reject M2 (18.656794 df=2)	Cannot reject M8(22.030908 df=2)
VesperOrtho9	0.472	0.472	Reject M3(0? df=4)	Reject M2(0.000016 df=2)	Reject M8(0 df=2)
MegabatOrtho1	0.787	0.7958	Reject M3(0.97114 df=4)	Reject M2(0.042318 df=2)	Reject M8(0.0658 df=2)
MegabatOrtho2					
MegabatOrtho3					
MegabatOrtho4	1	1.363	Reject M3-borderline(9.179122 df=4)	Reject M2(7.946978 df=2)	Reject M8(9.457018 df=2)
MegabatOrtho5					
MegabatOrtho6	0.5406	0.5406	Reject M3(0.144582 df=4)	Reject M2(0 df=2)	Reject M8 (0 df=2)
MegabatOrtho7					
MegabatOrtho8					
MegabatOrtho9					
MegabatOrtho10					
MegabatOrtho11					
MegabatOrtho12	1	1.0074	Reject M3(0 df=4)	Reject M2(0.000334 df=2)	Reject M8(0.000512 df=2)
MegabatOrtho13	1	13.0106	Cannot reject M3(10.270602 df=4)	Cannot reject M2(9.44505 df=2)	Cannot reject M8(9.445052 df=2)
MegabatOrtho14					
MegabatOrtho15	0.5	1.162	Cannot reject M3(13.271402 df=4)	Cannot reject M2(7.709804 df=2)	Cannot reject M8(7.773118 df=2)
MegabatOrtho16					
MegabatOrtho17					
MegabatOrtho18					
MegabatOrtho19					
MegabatOrtho20	0.8021	1.1617	Reject M3(0.476396df=4)	Reject M2 (1.082158 df=2)	Reject M8(1.107912 df=2)
MegabatOrtho21	1	3.3949	Reject M3 (2.782606 df=4)	Reject M2(4.249712 df=2)	Reject M8(4.249716 df=2)
RhinoHippoOrtho	0.6	0.7069	Cannot reject M3(22.034628 df=4)	Reject M2(3.641516 df=2)	Reject M8(4.26906 df=2)

Table 7.2: Continued. dNdS data. * difference of likelihood ratio test, and degrees of freedom (df) with significance level of 0.05 in brackets

Acc. Number	Coordinates	Species	Orthologs
AB758125		**SRV Y**	
NM001305593		**Syn Ory-1**	
NC022517		**BaEV**	
LC005748		**FERV**	
NW017731486	1384156-1387821	H.armiger	
NW017731771	1292829-1295308	H.armiger	
NW005370538	1731705-1734211	M.brandtii	
NW006297066	1048438-1050863	M.davidii	
NW017731633	654135-656422	H.armiger	
KE759579	15195-15785	E.helvum	
NW015494506	1892974-1895335	R.aegyptiacus	
AWHC01221602		E.helvum	
NW015493620	32279-34821	R.aegyptiacus	
NW011889075	2322257-2325027	P.vampyrus	
NW006443128	272787-275470	P.alecto	
NW011888965	2578051-2580425	P.vampyrus	
NW007370661	10546430-10547122	E.fuscus	
NW005871050	reverse20471597-20482376	M.lucifugus	VesperOrtho7
NW005370332	2002252-2005031	M.brandtii	VesperOrtho7
NW007370795	1004921-1007385	E.fuscus	
NW007370868	465368-468069	E.fuscus	
NW007370694	7685056-7687637	E.fuscus	
NW005871067	10757195-10760148	M.lucifugus	
NW006298163	700409-702906	M.davidii	
NW005371748	221641-224590	M.brandtii	
NW007370824	1787483-1790364	E.fuscus	
NW007370825	968567-971407	E.fuscus	
NW007370835	5020-7406	E.fuscus	
NW005356265	18240-10779	M.brandtii	
NW007371078	97272-99631	E.fuscus	
NW007370796	71295-73738	E.fuscus	
NW005871052	7794000-7806226	M.lucifugus	
NW005871618	409080-411541	M.lucifugus	
NW007370783	reverse3543842-3546189	E.fuscus	
NW005871763	reverse228491-221419	M.lucifugus	
NW005370500	1832101-1834628	M.brandtii	VesperOrtho8
NW005871203	497671-508411	M.lucifugus	VesperOrtho8
NW006293809		M.davidii	VesperOrtho8
NW006442107	33814746-33817561	P.alecto	
NW011889432	131405-133959	P.vampyrus	
NW011891576	1-4860	P.vampyrus	
NW017732354	111802-114134	H.armiger	
NW017732555	410291-412779	H.armiger	
NW017731418	2419887-2422450	H.armiger	
NW017733207	15822-18403	H.armiger	
NW017731625	353535-356125	H.armiger	
AWHB01206177		cM.lyra	
AJ862649		**envH3**	
AH007767	**672-3426**	**HERV-H**	
A Y849978		**Syn B**	
NM001013751		**Syn-A**	
M35242		**BLV**	
KI185079	reverse1-2344	R.ferrumequinum	
NW017739489	31929-42660	R.sinicus	
KI149001	reverse6390-13552	R.ferrumequinum	
AWGZ01297559		P.parnelli	
KI142656	reverse108604-98921	R.ferrumequinum	
NM001305586		**Syn Ten1**	
NW006438867	34348-36782	P.alecto	
KI035051	reverse6181-15971	M.lyra	
NW007370663		E.fuscus	VesperOrtho5
NW005871126		M.lucifugus	VesperOrtho5
NW006295056	996449-998799	M.davidii	VesperOrtho5
NW005361430		M.brandtii	VesperOrtho5
NW005871152	reverse2646110-2634045	M.lucifugus	
NW015504242	881382-884275	M.natalensis	VesperOrtho1
KE901973	31825-35825	P.parnelli	VesperOrtho1
NW007370652	14517069-14519713	E.fuscus	VesperOrtho1
NW007370652	reverse14519352-14516404	E.fuscus	VesperOrtho1
NW006290675	1064248-1066949	M.davidii	VesperOrtho1
NW005871049	1115719-1112780	M.lucifugus	VesperOrtho1
NW005357764	2863087-2865791	M.brandtii	VesperOrtho1
KE861073	1-6453	P.parnelli	
AWGZ01357182	reverse	P.parnelli	
NW017731971	201073-203396	H.armiger	
NW017732052	526069-528416	H.armiger	
AWGZ01115423	reverse	P.parnelli	
KC010510		**Syn2**	
NW017739572	reverse26202-18717	R.sinicus	
KI034957	1321-11136	M.lyra	
KI124864	reverse1-7837	M.lyra	
XM019829500	**950-3686HEMO**		
NW005370332	1492862-1516197	M.brandtii	VesperOrtho6
NW005871050	reverse20981557-20952777	M.lucifugus	VesperOrtho6
NW006287456	reverse422660-395499	M.davidii	VesperOrtho6
NW011888789	6348765-6367280	P.vampyrus	MegabatOrtho21
NW015493243	1133188-1151911	R.aegyptiacus	MegabatOrtho21
NW006439192	reverse9938456-9918864	P.alecto	MegabatOrtho21
NW006442490	1930705-1933244	P.alecto	

Table 7.3: Sequence positions on a phylogenetic tree

Acc. Number	Coordinates	Species	Orthologs
NW015494725	1017350-1021742	R.aegyptiacus	MegabatOrtho6
NW006436282	2435569-2438936	P.alecto	MegabatOrtho6
NW011888785	reverse13649700-13630959	P.vampyrus	MegabatOrtho6
KI093314	1-5655	M.lyra	
JX412964		**Syn Rum-1**	
JN587095		**Syn Car-1**	
AWGZ01325063	reverse	P.parnelli	
AANH01003690	**37666-38871**	**PercomORF**	
NW017738992	4477580-4488083	R.sinicus	RhinoHippoOrtho
KI133510	reverse14594-25097	R.ferrumequinum	RhinoHippoOrtho
NW017731683	1259826-1262416	H.armiger	RhinoHippoOrtho
NW017732017	112837-115400	H.armiger	
NW017732303	309480-311956	H.armiger	
NW015494520		R.aegyptiacus	MegabatOrth12
KE789057	5734-26051	E.helvum	
NW011888918	5615756-5644635	P.vampyrus	MegabatOrth12
NW006443138	27428382-27430720	P.alecto	MegabatOrth12
NW006432447	7434415-7436972	P.alecto	MegabatOrtho4
NW015493698	reverse3678563-3665377	R.aegyptiacus	MegabatOrtho4
DQ256474		**Syn1**	
AWHB01021880		M.lyra	
NW015494499	667229-671979	R.aegyptiacus	
KI003248	reverse22611-29468	M.lyra	
KI118423	1-5842	M.lyra	
KI075674	reverse5544-16379	M.lyra	
KI003286	reverse73482-79390	M.lyra	
KI055226	reverse1-4954	M.lyra	
AWHB01154537		M.lyra	
NW017739425	reverse909513-917901	R.sinicus	
KI144406	1-6261	R.ferrumequinum	
NW017739743	237446-248814	R.sinicus	
NW017740013	60828-71571	R.sinicus	
KI132377	1-8368	R.ferrumequinum	
KI165136	reverse1-7334	R.ferrumequinum	
NW017739579	reverse394674-402897	R.sinicus	
NW017739744	61254-71778	R.sinicus	
NW017739983	reverse191891-199427	R.sinicus	
NW017739983	reverse20371-212489	R.sinicus	
NW017739401	reverse603782-611521	R.sinicus	
NW017739256	reverse1672065-1679553	R.sinicus	
AWH01130638		R.ferrumequinum	
NW017739098	2020491-2031692	R.sinicus	
NW017739775	reverse92108-99515	R.sinicus	
NW017739775	reverse89108-99515	R.sinicus	
KE876944	reverse1-1993	P.parnelli	
KE893668	reverse2960-8850	P.parnelli	
NW015504588	187509-198675	M.natalensis	
NW015504718	429473-440003	M.natalensis	
NW015504328	2176484-2187857	M.natalensis	
NW015504384	1842123-1853363	M.natalensis	
NW006296618	143317-146327	M.davidii	
NW005354228	591954-595039	M.brandtii	
NW005365764	1794941-1797798	M.brandtii	
NW005871055	14580351-14588011	M.lucifugus	
NW006286818	787503-790099	M.davidii	
NW005354889	390414-392854	M.brandtii	
NW005355427	45791-48156	M.brandtii	
NW006285046	78495-81295	M.davidii	
NW005872103	25066-27707	M.lucifugus	
NW005353538	165644-1168120	M.brandtii	
NW005371136	165046-167432	M.brandtii	
NW006296619	11258-13674	M.davidii	
NW005871061	5947530-5955121	M.lucifugus	
NW005370000	121260-124045	M.brandtii	
NW005355184	45026-47913	M.brandtii	
NW005871052	10565171-10568195	M.lucifugus	
NW006298038	59074-61742	M.davidii	
NW007371030	2360-4977	E.fuscus	
NW005872753	8851-11684	M.lucifugus	
NW005369676	105946-108425	M.brandtii	
NW005871633	353991-356887	M.lucifugus	
NW005872425	12192-14635	M.lucifugus	
NW005871777	42389-45383	M.lucifugus	
NW005871069	10418424-10421266	M.lucifugus	
NW005372150	795167-797667	M.brandtii	
NW005355904	409904-412557	M.brandtii	
NW005370460	1696086-1698841	M.brandtii	
NW007370749	1277989-1280516	E.fuscus	
NW005871089	reverse5046365-5036362	M.lucifugus	
NW006291250	6884-9744	M.davidii	
NW006288994	3448717-3451187	M.davidii	
NW005364953	1832231-1835286	M.brandtii	
NW005371548	23828-26613	M.brandtii	
NW007370763	1179967-1183016	E.fuscus	
NW005367587	940065-942757	M.brandtii	
NW005871236	1032681-1035460	M.lucifugus	
NW005872767	reverse18052-28841	M.lucifugus	
NW005372107	1-5563	M.brandtii	
NW005872364	61185-63790	M.lucifugus	
NW005871122	reverse429009-439842	M.lucifugus	
NW015495010	1088378-1092782	R.aegyptiacus	
NW015494692	1482710-1484873	R.aegyptiacus	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW015494988	2461797-2466444	R.aegyptiacus	
NW015493370	107901-112317	R.aegyptiacus	
NW015493368	55361-57779	R.aegyptiacus	
NW015493620	1988-4130	R.aegyptiacus	
NW017739719	reverse316184-324080	R.sinicus	
NW017739641	reverse145632-153411	R.sinicus	
AWH01107092	reverse	R.ferrumequinum	
NW007370994	185209-187532	E.fuscus	
NW007370723	6865117-6868199	E.fuscus	
NW007371538	3585-6061	E.fuscus	
NW007370914	502242-504967	E.fuscus	
NW007372880	394-2921	E.fuscus	
NW007370921	852847-855278	E.fuscus	
NW007371243	52116-54529	E.fuscus	
NW007370848	489685-492317	E.fuscus	
NW007370665	28153075-28155488	E.fuscus	
NW007370780	3183840-3186466	E.fuscus	
NW005369676	51667-54092	M.brandtii	
NW006298624	759456-761794	M.davidii	
NW005357691	1091751-1094041	M.brandtii	
NW005364087	1-2376	M.brandtii	
NW005871316	983337-986050	M.lucifugus	
NW005370772	45839-48255	M.brandtii	
NW006292345	100635-102925	M.davidii	
NW005357697	124088-126507	M.brandtii	
NW005357049	4000-5532	M.brandtii	
NW005354392	21833-24330	M.brandtii	
NW006292913	10030-12734	M.davidii	
NW005354232	12699-15097	M.brandtii	
NW005871341	1363398-1366369	M.lucifugus	
NW005367486	1580-14218	M.brandtii	
NW005871077	9703967-9706314	M.lucifugus	
NW005882424	1-3618	M.lucifugus	
NW005871704	reverse305161-316170	M.lucifugus	
NW005362155	1496730-1499131	M.brandtii	
NW005353525	246287-248700	M.brandtii	
NW005872168	63799-66194	M.lucifugus	
NW006284323	1158819-1161229	M.davidii	
NW005356088	362763-365137	M.brandtii	
NW006289392	3554002-3556406	M.davidii	
NW005872243	reverse10247-2869	M.lucifugus	
NW005873320	reverse25840-19980	M.lucifugus	
NW007370767	250596-253687	E.fuscus	
NW007370812	2041405-2044889	E.fuscus	
NW005872682	13807-17447	M.lucifugus	
NW006286783	34684-38018	M.davidii	
NW005871646	54708-66530	M.lucifugus	
NW015494843	3511453-3514910	R.aegyptiacus	MegabatOrtho1
A WHC01027759		E.helvum	MegabatOrtho1
NW006432945	2927364-2930824	P.alecto	MegabatOrtho1
NW011888879	461068-464528	P.vampyrus	MegabatOrtho1
NW006293535	272760-275611	M.davidii	
NW006286818	1658120-1659488	M.davidii	
NW006293541	1187457-1191020	M.davidii	
NW005876031	6265-9928	M.lucifugus	
NW005871123	661537-664574	M.lucifugus	
NW005876780	1-5000	M.lucifugus	
NW005871098	1528668-1531264	M.lucifugus	
NW005880363	reverse331-6404	M.lucifugus	
NW005872581	reverse6723-9646	M.lucifugus	
NW005871222	52775-63245	M.lucifugus	
NW005873457	4436-7140	M.lucifugus	
NW005877734	4288-7019	M.lucifugus	
NW005871067	11180477-11183418	M.lucifugus	
NW005871185	107079-109693	M.lucifugus	
NW005872083	88270-91427	M.lucifugus	
NW005873484	17831-20989	M.lucifugus	
NW005871154	1040740-1043855	M.lucifugus	
NW005871257	893724-896714	M.lucifugus	
NW005871141	2531616-2534662	M.lucifugus	
NW005871052	20361390-20364445	M.lucifugus	
NW005871853	52095-55036	M.lucifugus	
NW007370748	1569672-1572259	E.fuscus	
KF547869		**BRTV**	
NW011888821	3602959-3606308	P.vampyrus	
NW006434839	20135239-20138390	P.alecto	
NW011888987	1221787-1225381	P.vampyrus	
NW006436823	27148-30005	P.alecto	
NW006434830	1-4154	P.alecto	
NW011888837	2727858-2730655	P.vampyrus	
NW006438869	1004365-1006742	P.alecto	MegabatOrtho7
NW006432943	10115905-10118936	P.alecto	
NW011914678	133599-136510	P.vampyrus	
NW011889065	1317625-1320140	P.vampyrus	
NW006443129	1053810-1057027	P.alecto	
NW011888802	8325144-8327956	P.vampyrus	
NW011889106	722996-726462	P.vampyrus	
NW006431805	1948202-1973353	P.alecto	MegabatOrtho18
NW011888844	5574434-5577612	P.vampyrus	MegabatOrtho18
NW006434897	7217725-7237749	P.alecto	MegabatOrtho16
NW011893897	1-5556	P.vampyrus	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW011888806	1775112-1777561	P.vampyrus	MegabatOrtho16
NW011889418	99959-203576	P.vampyrus	
NW011889301	1-4000	P.vampyrus	
NW011889301	1-5000	P.vampyrus	
NW011888793	4324025-4327478	P.vampyrus	
NW011890313	6810-9379	P.vampyrus	
NW015493855	10529-15029	R.aegyptiacus	
NW011888863	262600-266827	P.vampyrus	
NW011888814	320010-322309	P.vampyrus	
NW011888937	681663-684787	P.vampyrus	
NW011888787	11722603-11725073	P.vampyrus	
NW011888782	26224676-26228532	P.vampyrus	MegabatOrtho4
NW006441389	121649-124161	P.alecto	
NW006442484	20366779-20369609	P.alecto	
NW006435778	1270936-1273424	P.alecto	
NW011889109	reverse304080-300255	P.vampyrus	
NW011889267	296304-299269	P.vampyrus	
NW011889462	1155270-1159147	P.vampyrus	
NW011888822	7171186-7174070	P.vampyrus	
NW011891060	78736-81974	P.vampyrus	
NW006434181	97329-100003	P.alecto	
NW006439190	515554-519020	P.alecto	
NW011889322	643595-647505	P.vampyrus	
NW011888819	8781892-8785695	P.vampyrus	
NW011888782	42109756-42115045	P.vampyrus	
NW006442497	2863081-2865479	P.alecto	
NW006439568	1437818-1440792	P.alecto	
A_WHC01006168		E.helvum	
NW015493348	110226-114613	R.aegyptiacus	
NW011889421	248791-252427	P.vampyrus	
NW006443119	230586-234210	P.alecto	
NW011889130	2349847-2353736	P.vampyrus	MegabatOrtho10
NW006442597	368974-372614	P.alecto	MegabatOrtho10
NW006429864	1154928-1157776	P.alecto	
NW011889290	2234321-2237963	P.vampyrus	MegabatOrtho14
NW006443569	550898-554567	P.alecto	MegabatOrtho14
NW011889134	708562-711104	P.vampyrus	
NW015494932	3719342-3721776	R.aegyptiacus	
NW006429686	60962-64299	P.alecto	
NW015494594	1349597-1352465	R.aegyptiacus	
NW015494691	247238-251928	R.aegyptiacus	
NW011888910	969697-972134	P.vampyrus	
NW015494964	reverse875372-868948	R.aegyptiacus	
NW011888965	3765969-3794050	P.vampyrus	MegabatOrtho9
NW006440131	reverse674461-650114	P.alecto	MegabatOrtho9
NW015494121	2110339-2130367	R.aegyptiacus	MegabatOrtho15
NW015494667	1415786-1420241	R.aegyptiacus	
NW015493388	2513182-2515841	R.aegyptiacus	
NW011888789	1857242-1873754	P.vampyrus	MegabatOrtho20
NW006439192	reverse14249608-14234728	P.alecto	MegabatOrtho20
NW015495290	469365-486280	R.aegyptiacus	MegabatOrtho20
NW015493194	16292-19010	R.aegyptiacus	
NW015494470	1658915-1663637	R.aegyptiacus	
NW015493264	32504-38183	R.aegyptiacus	
NW015494931	802028-806889	R.aegyptiacus	
KE803860	reverse24156-29833	E.helvum	
AWHC01052730	reverse1-2425	E.helvum	
NW015494904	720862-724092	R.aegyptiacus	
NW015493302	138843-141307	R.aegyptiacus	
NW006443138	30955984-30958580	P.alecto	MegabatOrtho13
NW011888918	1972288-1993050	P.vampyrus	MegabatOrtho13
NW006436282	16783506-16785934	P.alecto	MegabatOrtho5
NW011888850	reverse1368437-1348447	P.vampyrus	MegabatOrtho5
NW015494598	304805-309236	R.aegyptiacus	
NW006442109	34971150-34974010	P.alecto	
NW006494614	35358001-35361008	P.alecto	
NW015493033	1685103-1707096	R.aegyptiacus	MegabatOrtho13
NW015494547	682059-686379	R.aegyptiacus	
NW006435824	31025-34022	P.alecto	
NW006494610	6595021-6597524	P.alecto	
NW006433378	16227869-16230453	P.alecto	
AWHC01091567	reverse10719-17021	E.helvum	
NW015494582	970988-975753	R.aegyptiacus	
NW011888831	3194593-3196925	P.vampyrus	
NW015493621	2540941-2543339	R.aegyptiacus	
NW011889168	1357150-1360244	P.vampyrus	
NW011912080	reverse3135-1	P.vampyrus	
NW015492809	3298421-3302709	R.aegyptiacus	
NW006440638	4447928-4450209	P.alecto	
NW006431917	793527-795898	P.alecto	
NW006434897	705898-708326	P.alecto	MegabatOrtho15
NW011888806	8627185-8629646	P.vampyrus	MegabatOrtho15
NW006494609	3296755-3299087	P.alecto	
NW011892609	1-5215	P.vampyrus	
NW011888909	4929096-4931524	P.vampyrus	
NW006439231	78143-80784	P.alecto	
NW011888838	7712724-7715089	P.vampyrus	
NW015495087	732501-736911	R.aegyptiacus	
NW011888971	1820650-1830053	P.vampyrus	MegabatOrtho7
NW011888942	584419-587786	P.vampyrus	
NW006436810	1442072-1444851	P.alecto	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW015494585	1625726-1630154	R.aegyptiacus	
NW015495162	59630-64008	R.aegyptiacus	
NW015493513	23996-26424	R.aegyptiacus	
NW011889073	444622-454428	P.vampyrus	
NW011888984	1209838-1212269	P.vampyrus	
NW011889002	1588162-1590533	P.vampyrus	
NW006434369	reverse10614095-10587925	P.alecto	
NW006442508	15486576-15490233	P.alecto	
NW011888967	1441671-1444630	P.vampyrus	
NW011888916	1351737-1354186	P.vampyrus	
NW006442497	128443-130859	P.alecto	
NW006431922	925384-928467	P.alecto	
NW011888885	1480874-1483299	P.vampyrus	
NW011888831	9810643-9813071	P.vampyrus	
NW011888831	reverse9826071-9804643	P.vampyrus	MegabatOrtho17
NW006443570	17035310-17051873	P.alecto	MegabatOrtho17
NW011889760	1-10534	P.vampyrus	
NW011888817	3839041-3841508	P.vampyrus	
NW006494613	15965268-15967696	P.alecto	
NW011888807	3093865-3096335	P.vampyrus	
NW017732982	30344-36173	H.armiger	
NW017731771	867756-871425	H.armiger	
NW017732767	1383-14930	H.armiger	
NW017733047	21623-25233	H.armiger	
NW017731368	4872055-4875266	H.armiger	
NW017733483	1-3361	H.armiger	
NW017732303	545860-548818	H.armiger	
NW017731971	165711-168849	H.armiger	
NW017731971	155475-157864	H.armiger	
NW017731989	1335909-1339336	H.armiger	
NW017731660	1336428-1339255	H.armiger	
NW017732243	166922-170430	H.armiger	
NW017731377	1855893-1859728	H.armiger	
NW017731614	1093304-1096962	H.armiger	
NW017731488	1-5181	H.armiger	
NW017731499	947357-950595	H.armiger	
NW017732163	185722-188069	H.armiger	
NW006282670	27207-30487	M.davidii	
NW005368827	1-2781	M.brandtii	
NW005370427	262843-266592	M.brandtii	
NW006296624	732410-736188	M.davidii	VesperOrtho3
NW005369934	2691645-2695423	M.brandtii	VesperOrtho3
NW005871219	reverse1707414-1695635	M.lucifugus	VesperOrtho3
KI092417	reverse1-8944	M.lyra	
NW015504325	2371932-2382903	M.natalensis	
NW015504615	880565-892186	M.natalensis	
NW015504444	797921-808934	M.natalensis	
NW015504414	1672983-1684482	M.natalensis	
NW005358524	188866-194606	M.brandtii	VesperOrtho9
NW005367925	1-3687	M.brandtii	
NW006289419		M.davidii	VesperOrtho9
NW005871745	37715-40059	M.lucifugus	VesperOrtho9
NW005871892	104562-107335	M.lucifugus	
NW005372055	513778-517511	M.brandtii	
NW006286949	49153-52559	M.davidii	
KI004798	reverse14593-25387	M.lyra	
KI055751	reverse2545-13729	M.lyra	
MG254888		**Syn-Mab1**	
AB600229		**HTLV**	
KE790611	1-2355	E.helvum	
NW011888798	7059050-7064269	P.vampyrus	
NW006442484	16762062-16765084	P.alecto	
NW011892495	1-1620	P.vampyrus	
NW011888798	12314453-12317010	P.vampyrus	
A WHC01016235	reverse870-6529	E.helvum	
NW015494770	122717-127310	R.aegyptiacus	
NW015495187	6309837-6314968	R.aegyptiacus	
NW015495211	19029-21770	R.aegyptiacus	
A WHC01042660	reverse10829-17661	E.helvum	
NW015493184	154731-159541	R.aegyptiacus	
NW011889109	232080-235981	P.vampyrus	
NW011889892	13730-16296	P.vampyrus	
NW006439715	109007-111588	P.alecto	
NW011888828	5164717-5167133	P.vampyrus	
NW006440618	334138-336959	P.alecto	
NW006432669	7107085-7109825	P.alecto	
NW006433949	7727157-7731037	P.alecto	
NW011888804	11028164-11032086	P.vampyrus	
NW006439296	1-3420	P.alecto	MegabatOrtho8
NW011888916	1817870-1820664	P.vampyrus	MegabatOrtho8
NW011888986	2153896-2156804	P.vampyrus	
NW011888931	3930069-3932662	P.vampyrus	
NW011889003	1322522-1325313	P.vampyrus	
NW006439568	19523100-19525891	P.alecto	
NW011888816	1521281-1524225	P.vampyrus	
NW006431922	1954647-1956632	P.alecto	MegabatOrtho2
NW011901642	35319-37636	P.vampyrus	MegabatOrtho2
NW011889353	935371-938051	P.vampyrus	
NW015492988	reverse986816-982281	R.aegyptiacus	
AWHC01140484		E.helvum	
NW015494254	3253115-3258132	R.aegyptiacus	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW006433952	51876-55416	P.alecto	
NW006443131	438245-440937	P.alecto	MegabatOrtho11
NW011888869	5661233-5665009	P.vampyrus	
NW011918376	44579-47917	P.vampyrus	
NW011888798	3444483-3446986	P.vampyrus	
NW011889121	3020694-3023899	P.vampyrus	
NW006442512	reverse1246513-1226927	P.alecto	MegabatOrth19
NW006433947	38891-41403	P.alecto	
NW011889332	204927-208691	P.vampyrus	
NW011888965	2990675-2994158	P.vampyrus	
NW011889243	1092558-1096408	P.vampyrus	MegabatOrth19
NW011889051	967611-970747	P.vampyrus	
NW006442484	47392152-47394900	P.alecto	
NW011888783	12894912-12898290	P.vampyrus	
NW011888905	466314-468868	P.vampyrus	MegabatOrtho11
NW011888782	3294305-3297921	P.vampyrus	
NW006433941	5325528-5328160	P.alecto	
NW011888907	6371-9357	P.vampyrus	
NW011888838	3878948-3881277	P.vampyrus	
NW006433940	136031-138342	P.alecto	
NW006435780	31595-34785	P.alecto	
NW011889551	45538-153939	P.vampyrus	
NW006441068	550057-553508	P.alecto	MegabatOrtho3
NW011889269	reverse181316-172960	P.vampyrus	MegabatOrtho3
NW011921536	1-2745	P.vampyrus	
NW011920979	1-4061	P.vampyrus	
NW011920104	1-6434	P.vampyrus	
NW011919991	1-6904	P.vampyrus	
NW011920854	1-4598	P.vampyrus	
NW011920658	1-5801	P.vampyrus	
NW011917761	60114-63970	P.vampyrus	
NW011917761	42369-45478	P.vampyrus	
NW011917761	23620-26729	P.vampyrus	
NW011920581	1-3186	P.vampyrus	
NW011920491	1-2891	P.vampyrus	
NW011917485	1-2576	P.vampyrus	
NW011920716	1-5009	P.vampyrus	
NW011920610	1-5696	P.vampyrus	
NW011920992	1-2544	P.vampyrus	
NW011920560	1-6565	P.vampyrus	
NW011920712	1-3351	P.vampyrus	
NW011921300	1-3204	P.vampyrus	
NW011920517	1-6986	P.vampyrus	
NW011920853	1-4606	P.vampyrus	
NW011921045	1-3397	P.vampyrus	
NW011921659	1-2219	P.vampyrus	
NW011922242	1-1877	P.vampyrus	
NW011920400	1-5578	P.vampyrus	
NW011922220	1-1912	P.vampyrus	
NW011921168	1-3491	P.vampyrus	
NW011918276	16669-20513	P.vampyrus	
NW011921165	1-3495	P.vampyrus	
NW011916912	5010-8744	P.vampyrus	
NW011918961	1-5320	P.vampyrus	
NW011921545	1-2735	P.vampyrus	
NW006441948	708820-711869	P.alecto	
NW011921645	1-2587	P.vampyrus	
NW011920467	1-3369	P.vampyrus	
NW011918401	24959-28803	P.vampyrus	
NW011921294	1-3213	P.vampyrus	
NW011922122	1-2002	P.vampyrus	
NW011920346	1-4246	P.vampyrus	
NW011915754	1-6264	P.vampyrus	
NW011920653	reverse5829-1	P.vampyrus	
NW011918687	6499-10322	P.vampyrus	
NW011922050	1-2078	P.vampyrus	
NW011918872	1-3825	P.vampyrus	
NW011918257	4414-8123	P.vampyrus	
NW011921533	1-2749	P.vampyrus	
NW011921611	1-2011	P.vampyrus	
NW011921742	1-2466	P.vampyrus	
NW011915114	113293-117122	P.vampyrus	
NW011917313	1-3741	P.vampyrus	
NW011921660	1-2568	P.vampyrus	
NW011915316	13497-15714	P.vampyrus	
NW011920248	4995-7630	P.vampyrus	
NW011919297	1-4400	P.vampyrus	
NW011920284	5619-8254	P.vampyrus	
NW011917026	1-2864	P.vampyrus	
NW011920931	1-3286	P.vampyrus	
NW011917482	12130-14765	P.vampyrus	
NW011921693	1-1942	P.vampyrus	
NW011920532	1-2366	P.vampyrus	
NW006290668	330225-333031	M.davidii	
NW005370153	54724-57248	M.brandtii	
NW005355815	65172-67663	M.brandtii	
NW005872590	reverse40094-27422	M.lucifugus	
NW005871067	10770951-10773610	M.lucifugus	
NW005871098	2098531-2101214	M.lucifugus	
NW007370821	669149-673000	E.fuscus	
NW007370681	633412-1636473	E.fuscus	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW007370786	94699-97193	E.fuscus	
NW007370857	840343-843425	E.fuscus	
NW007370885	113694-117624	E.fuscus	
NW007370731	4477936-4481176	E.fuscus	
NW005871055	10991732-10995495	M.lucifugus	
NW005871362	914419-918371	M.lucifugus	
NW006283772	12097-15870	M.davidii	
NW005355554	20699-24135	M.brandtii	
NW005871204	1674444-1677073	M.lucifugus	
NW005871259	279242-1282845	M.lucifugus	
NW005872670	4000-5416	M.lucifugus	
NW005355769	153240-156454	M.brandtii	
NW005370921	78917-82459	M.brandtii	
NW006291791	reverse5936-1	M.davidii	
NW006298272	1-2228	M.davidii	
NW005871067	6793692-6797365	M.lucifugus	
NW005871067	reverse4832439-4821884	M.lucifugus	
NW005871067	4825814-4828425	M.lucifugus	
NW005871067	10462761-10465348	M.lucifugus	
NW005871069	12020901-12024889	M.lucifugus	VesperOrtho2
NW005357847	521227-523526	M.brandtii	VesperOrtho2
NW005358516	138127-140612	M.brandtii	
NW005359255	86910-89809	M.brandtii	
NW005871111	576790-1580775	M.lucifugus	
NW005871119	2835844-2839832	M.lucifugus	
NW005871069	11490131-11501214	M.lucifugus	
NW005872323	reverse20194-8006	M.lucifugus	
NW006288972	323981-326322	M.davidii	
NW005871526	214725-218707	M.lucifugus	VesperOrtho4
NW005370460	reverse136598-125105	M.brandtii	VesperOrtho4
NW005370460	125713-129744	M.brandtii	
NW006300281	4000-5919	M.brandtii	
NW007371018	reverse8703-1	E.fuscus	
NW007372419	1-2140	E.fuscus	
NW007370786	3529187-3532092	E.fuscus	
NW006298367	754446-757255	M.davidii	
NW005359102	591073-594953	M.brandtii	
NW005871534	reverse174780-186425	M.lucifugus	
NW005879130	reverse6979-1	M.lucifugus	
NW005871292	35149-39003	M.lucifugus	
NW005871247	2252876-2255826	M.lucifugus	
NW005355679	1-3315	M.brandtii	
NW005872429	reverse43387-32252	M.lucifugus	
NW005365191	116014-119639	M.brandtii	
NW006289833	115027-117995	M.davidii	
NW015504389	670645-682056	M.natalensis	
NW015504246	3274536-3285348	M.natalensis	
NW015504381	5669775-5680089	M.natalensis	
NW006296051	168499-170582	M.davidii	
NW007370664	23689803-23691228	E.fuscus	
NW007370758	2817667-2821415	E.fuscus	
NW005365553	2978410-2981141	M.brandtii	
NW005371333	876739-879556	M.brandtii	
NW006298311	35716-38579	M.davidii	
NW006286862	5640-9380	M.davidii	
NW005872014	reverse84907-70878	M.lucifugus	
NW006286275	217892-221040	M.davidii	
NW006297224	798583-801506	M.davidii	
NW005871128	reverse277197-289633	M.lucifugus	
NW006291614	22561-25403	M.davidii	
NW005353490	1110866-1113372	M.brandtii	
NW005873255	1-4000	M.lucifugus	
NW005871089	reverse4190320-4177528	M.lucifugus	
NW005356519	reverse233431-223859	M.brandtii	
NW006288962	9136-21510	M.davidii	
NW005872297	reverse47167-33238	M.lucifugus	
NW005357271	390417-393358	M.brandtii	
NW006296450	15681-19321	M.davidii	
NW005356437	1322987-1325898	M.brandtii	
NW005356437	1343245-1346333	M.brandtii	
AWGZ01040320	reverse	P.parnelli	
AWGZ01058436	reverse	P.parnelli	
KI004384	reverse1-6499	M.lyra	
KI020117	1-7223	M.lyra	
NW006299170	4000-8146	M.davidii	
NW005307959	1-806	M.brandtii	
NW015504996	35124-46308	M.natalensis	
NW015504668	577957-588310	M.natalensis	
NW015503982	4511-10492	M.natalensis	
NW015504567	798138-809772	M.natalensis	
NW015504523	1291160-1298049	M.natalensis	
NW015504724	197544-208420	M.natalensis	
NW015505144	12950-19732	M.natalensis	
NM001308377		**Syn-Opo1**	
NW017739594	1-6602	R.sinicus	
NW017740071	1-9029	R.sinicus	
NW017739307	1082871-1094760	R.sinicus	
NW017739687	1-4660	R.sinicus	
NW017739290	1044866-1055708	R.sinicus	
NW005363262	9072-11668	M.brandtii	
NW005370452	213640-216005	M.brandtii	

Table 7.3: Continued. Sequence positions on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW007370766	724037-727166	E.fuscus	
NW007371135	54644-57297	E.fuscus	
NW007371065	196312-199286	E.fuscus	
NW007370833	1248985-1252753	E.fuscus	
NW005371618	31455-35350	M.brandtii	
NW005871339	reverse20591-31337	M.lucifugus	
NW006287902	15941-19802	M.davidii	
NW006293003	1510623-1514191	M.davidii	
NW007370909	104290-107953	E.fuscus	
NW005872168	42859-45908	M.lucifugus	
NW005872235	reverse27861-39693	M.lucifugus	
NW005871818	206724-210562	M.lucifugus	
NW005872014	12680-16504	M.lucifugus	
NW005355347	596893-600740	M.brandtii	
NW005873420	7506-11344	M.lucifugus	
NW005357414	3210050-3214096	M.brandtii	
NW005871115	443349-445853	M.lucifugus	
NW005871835	reverse162851-174898	M.lucifugus	
NW005872087	98071-101984	M.lucifugus	
NW005873795	1455-13017	M.lucifugus	
AF055063		**GaLV**	
NC021704		**KoERV**	
GU980187	**6185-8167**	**PERV**	
AF053745	**5754-7775**	**MdERV**	
AWHC01145997		E.helvum	
NW005368113	5982-10000	M.brandtii	
NW006299713	29416-32984	M.davidii	
NW006287809	892720-893294	M.davidii	
NW005353371	237846-241834	M.brandtii	
NW005372000	15872-119878	M.brandtii	
AY277737		**MuLV**	
NC001940		**FeLV**	
NW005872310	58000-69954	M.lucifugus	
NW007370817	2399686-2403653	E.fuscus	
NW007370686	11099661-11103325	E.fuscus	
NW007370751	568467-572434	E.fuscus	
NW005872791	8490-22451	M.lucifugus	
NW005878789	1-7415	M.lucifugus	
NW005872398	reverse60983-48951	M.lucifugus	
NW005874540	16208-15293	M.lucifugus	
NW005873850	4942-8900	M.lucifugus	
NW015493296	86474-89687	R.aegyptiacus	
NW006432447	5182690-5185853	P.alecto	
NW006440820	1408437-1411369	P.alecto	
NW006438090	329575-332771	P.alecto	
NW015492981	2997-7857	R.aegyptiacus	
NW015494823	898742-903849	R.aegyptiacus	
NW015494735	570801-575949	R.aegyptiacus	
NW015493368	743-6480	R.aegyptiacus	
NW015493345	9469-13036	R.aegyptiacus	
NW015493443	731011-734970	R.aegyptiacus	
NW015494944	978333-983956	R.aegyptiacus	
NW015495043	1055393-1058753	R.aegyptiacus	
NW015494789	338448-342406	R.aegyptiacus	
NW015495031	9086-12855	R.aegyptiacus	
NW015495065	2509174-2511823	R.aegyptiacus	
NW015493647	48908-54515	R.aegyptiacus	
NW015493898	341968-346228	R.aegyptiacus	
NW015495065	reverse2514135-2509465	R.aegyptiacus	
NW015494629	1592829-1598373	R.aegyptiacus	
NW015495031	3404-7392	R.aegyptiacus	
NW015492915	407367-411397	R.aegyptiacus	
NW015495050	783331-787227	R.aegyptiacus	
NW015493524	13678-16805	R.aegyptiacus	
NW015493539	59511-64644	R.aegyptiacus	
NW015494744	858437-861570	R.aegyptiacus	
NW015495189	8505-11748	R.aegyptiacus	
NW015494027	11550-17152	R.aegyptiacus	
NW015494768	1259588-1264202	R.aegyptiacus	
NW015493524	19382-22257	R.aegyptiacus	
NW015494867	266531-271230	R.aegyptiacus	
NW015493308	138372-141948	R.aegyptiacus	
NW015494225	1746-4245	R.aegyptiacus	
NW015494506	219445-222121	R.aegyptiacus	
NW015494235	1980-4859	R.aegyptiacus	
NW015493667	28396-31817	R.aegyptiacus	
NW015493732	135249-139147	R.aegyptiacus	
NW015493766	27562-30973	R.aegyptiacus	
NW015494687	1318811-1326090	R.aegyptiacus	
NW015494546	714239-719835	R.aegyptiacus	
NW011920700	1-3058	P.vampyrus	
NW011920577	1-2337	P.vampyrus	
NW011919931	1-3638	P.vampyrus	
NW011917310	1-5440	P.vampyrus	
NW011921079	1-3625	P.vampyrus	
NW006432943	6166085-6169392	P.alecto	
NW011921211	1-2616	P.vampyrus	
NW011917469	1-2439	P.vampyrus	
NW011920689	1-2894	P.vampyrus	
NW011921801	1-2397	P.vampyrus	
NW011921089	1-3740	P.vampyrus	

Table 7.3: Continued. Sequence position on a phylogenetic tree.

Acc. Number	Coordinates	Species	Orthologs
NW006442512	8672202-8674978	P.alecto	
NW011920829	1-4549	P.vampyrus	
NW011918723	7797-21509	P.vampyrus	
NW011920893	1-4423	P.vampyrus	
NW011921798	1-2400	P.vampyrus	
NW011921507	1-2540	P.vampyrus	
NW011921008	1-3587	P.vampyrus	
NW011917399	1-6857	P.vampyrus	

Table 7.3: Continued. Sequences positions on a phylogenetic tree.

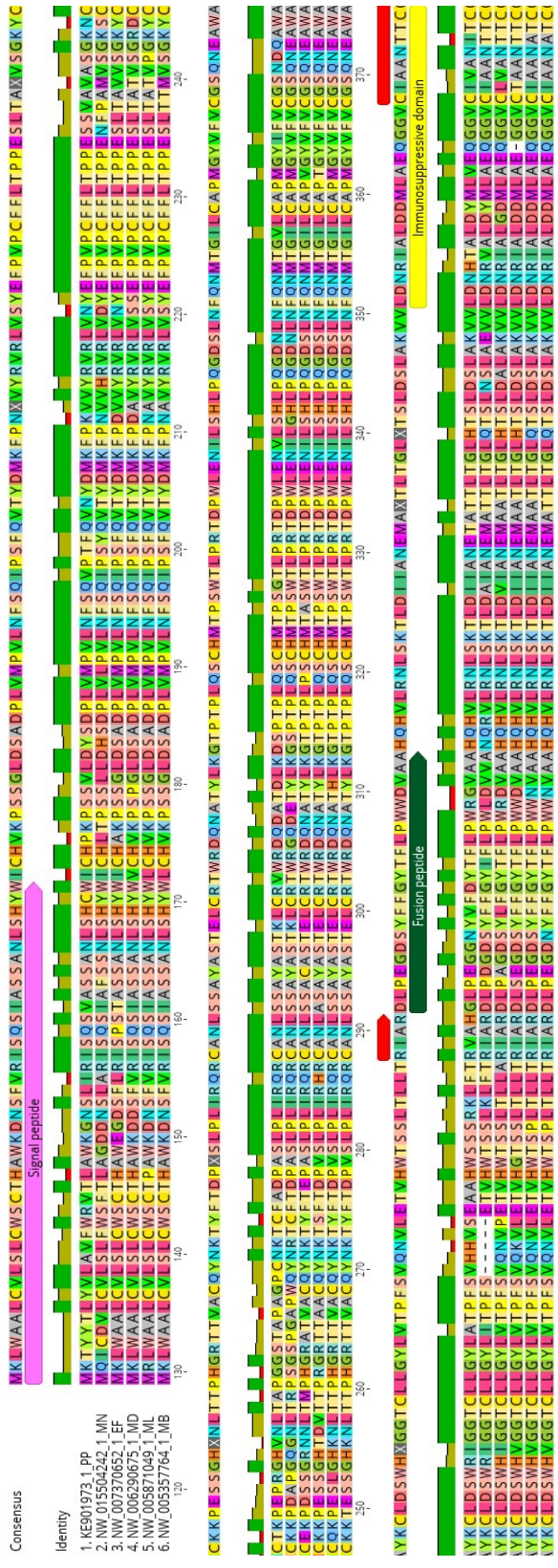


Figure 7.1: VesperOrtho1 alignment with annotations, consensus sequence and identity across the sequence.

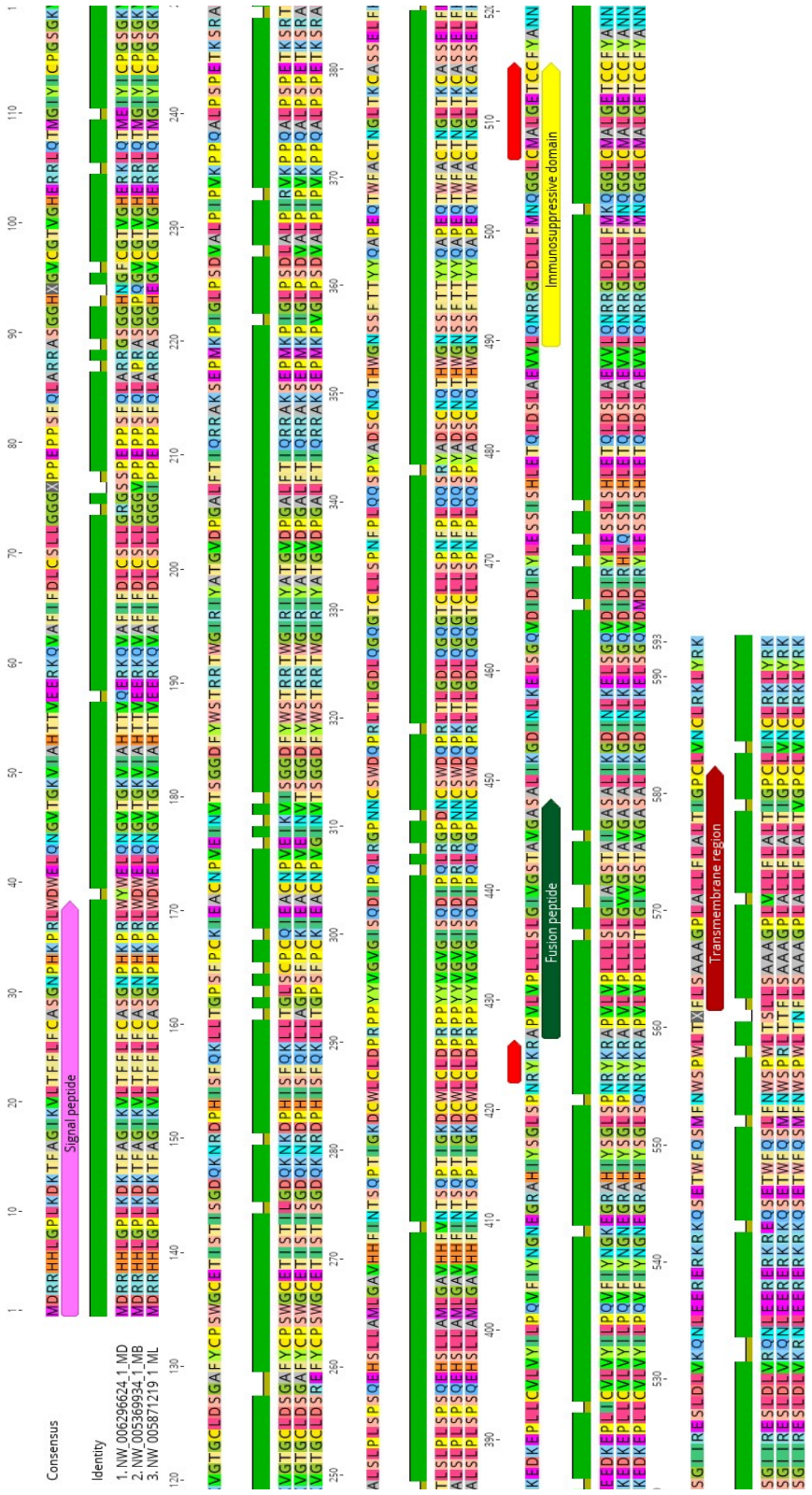


Figure 7.2: VesperOrtho3 alignment with annotations, consensus sequence and identity across the sequence.

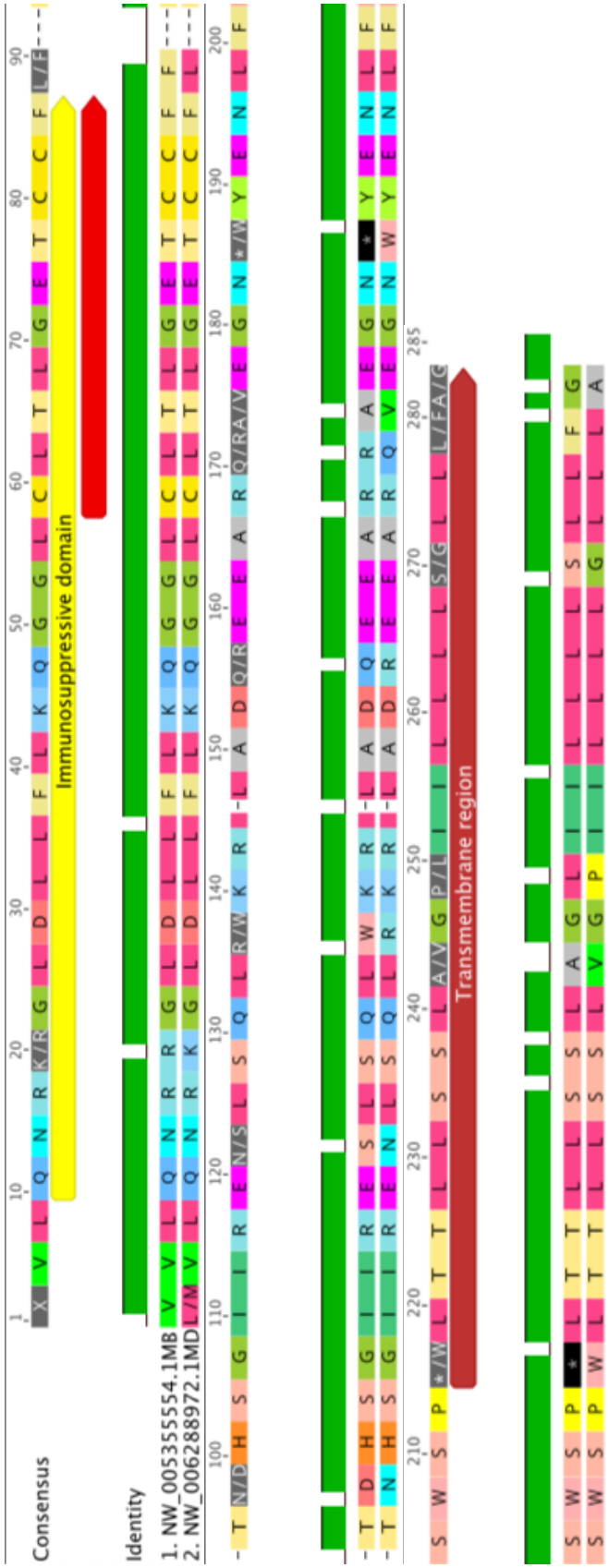


Figure 7.3: VesperOrtho6 alignment with annotations, consensus sequence and identity across the sequence.

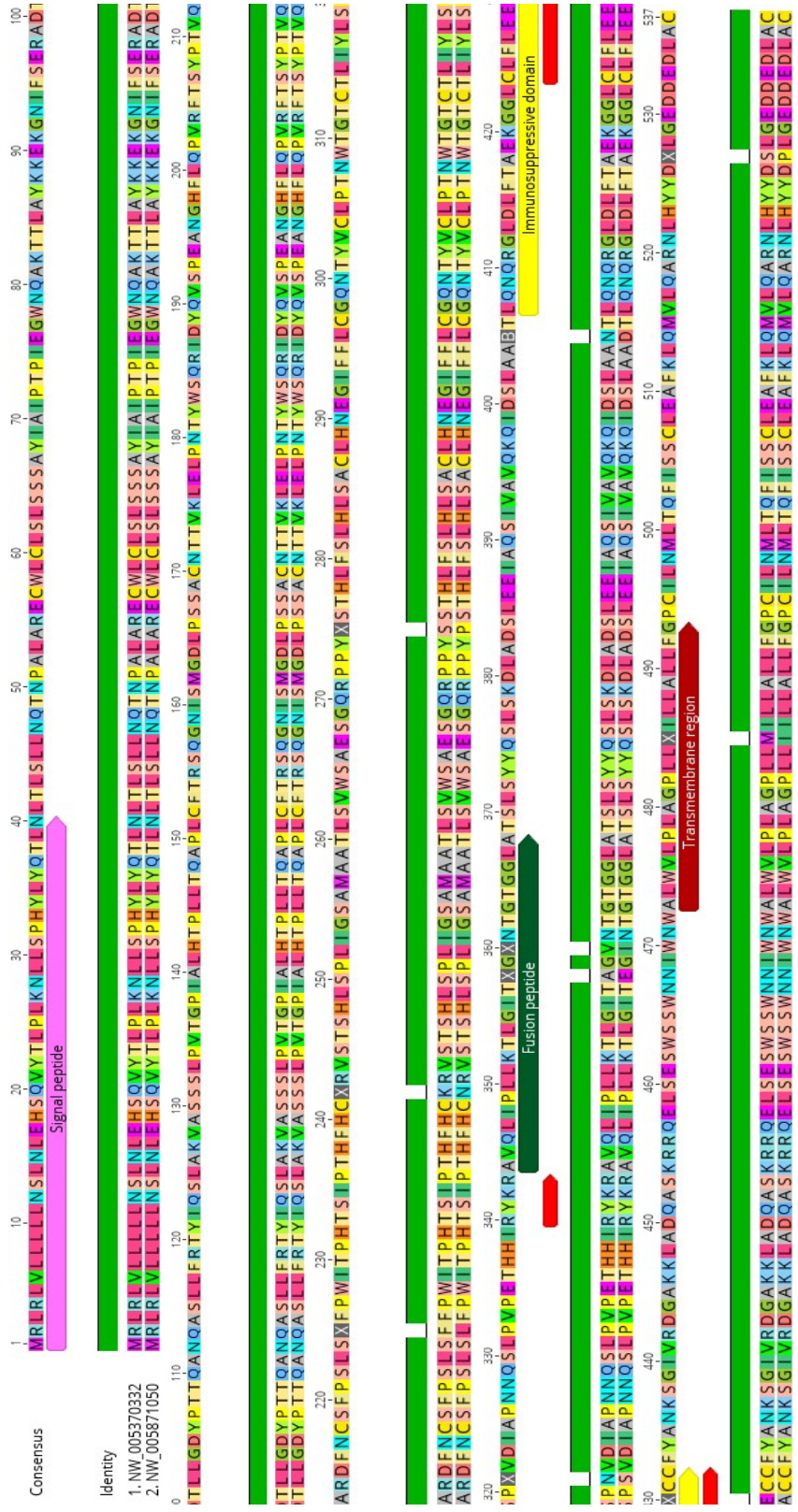


Figure 7.4: VesperOrtho7 alignment with annotations, consensus sequence and identity across the sequence.

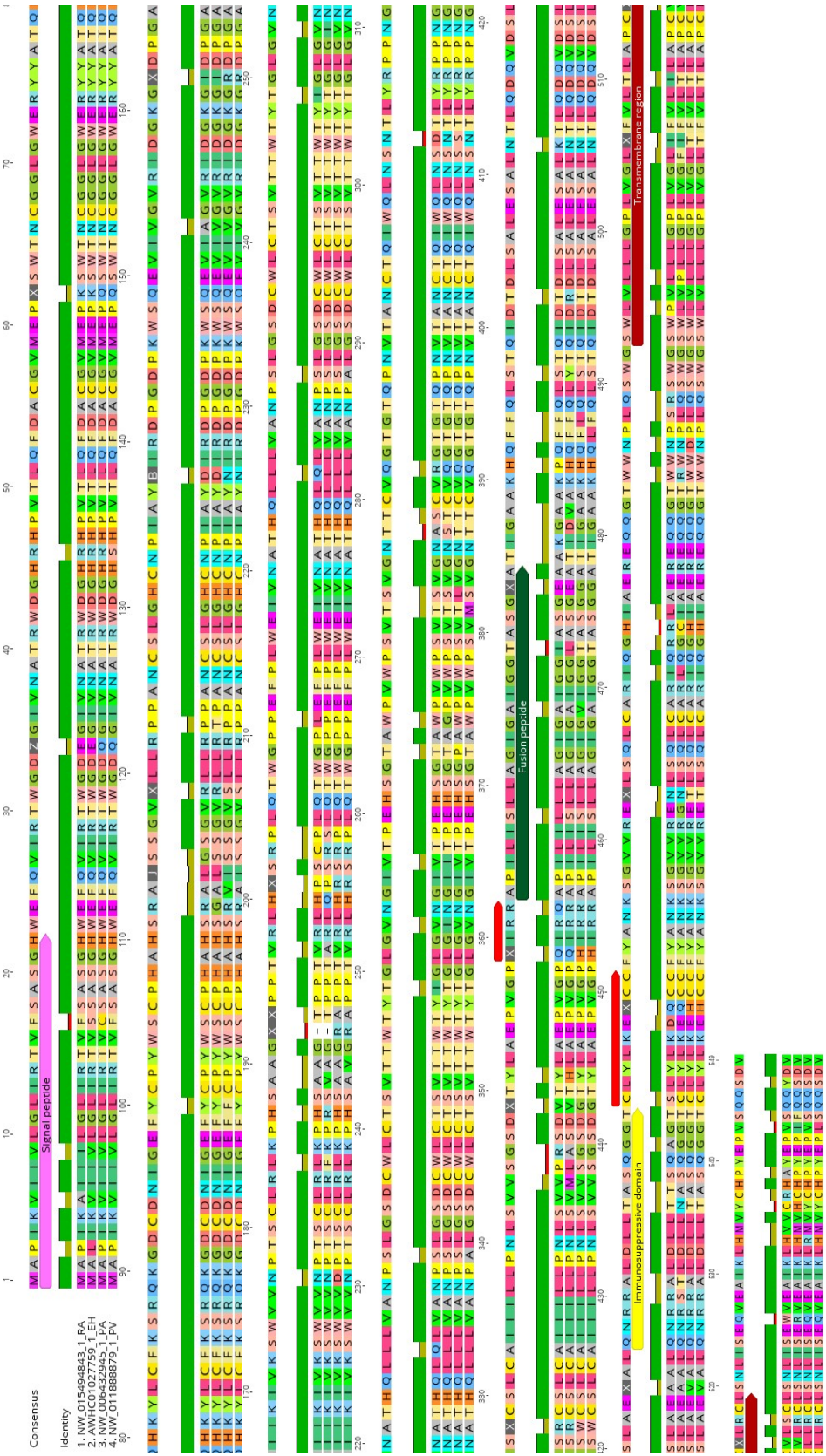


Figure 7.5: MegabatOrtho1 alignment with annotations, consensus sequence and identity across sequence.

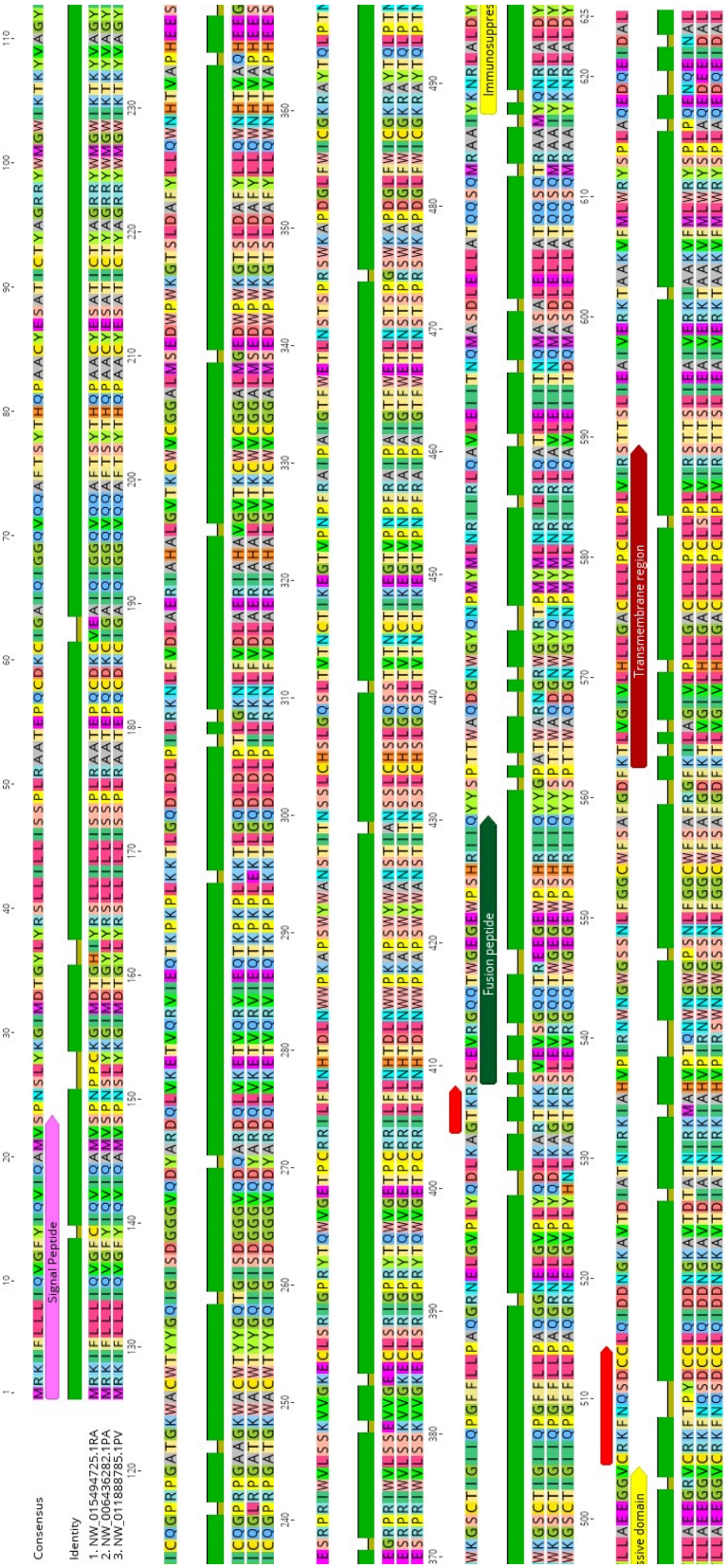


Figure 7.6: MegabatOrtho6 alignment with annotations, consensus sequence and identity across sequence.

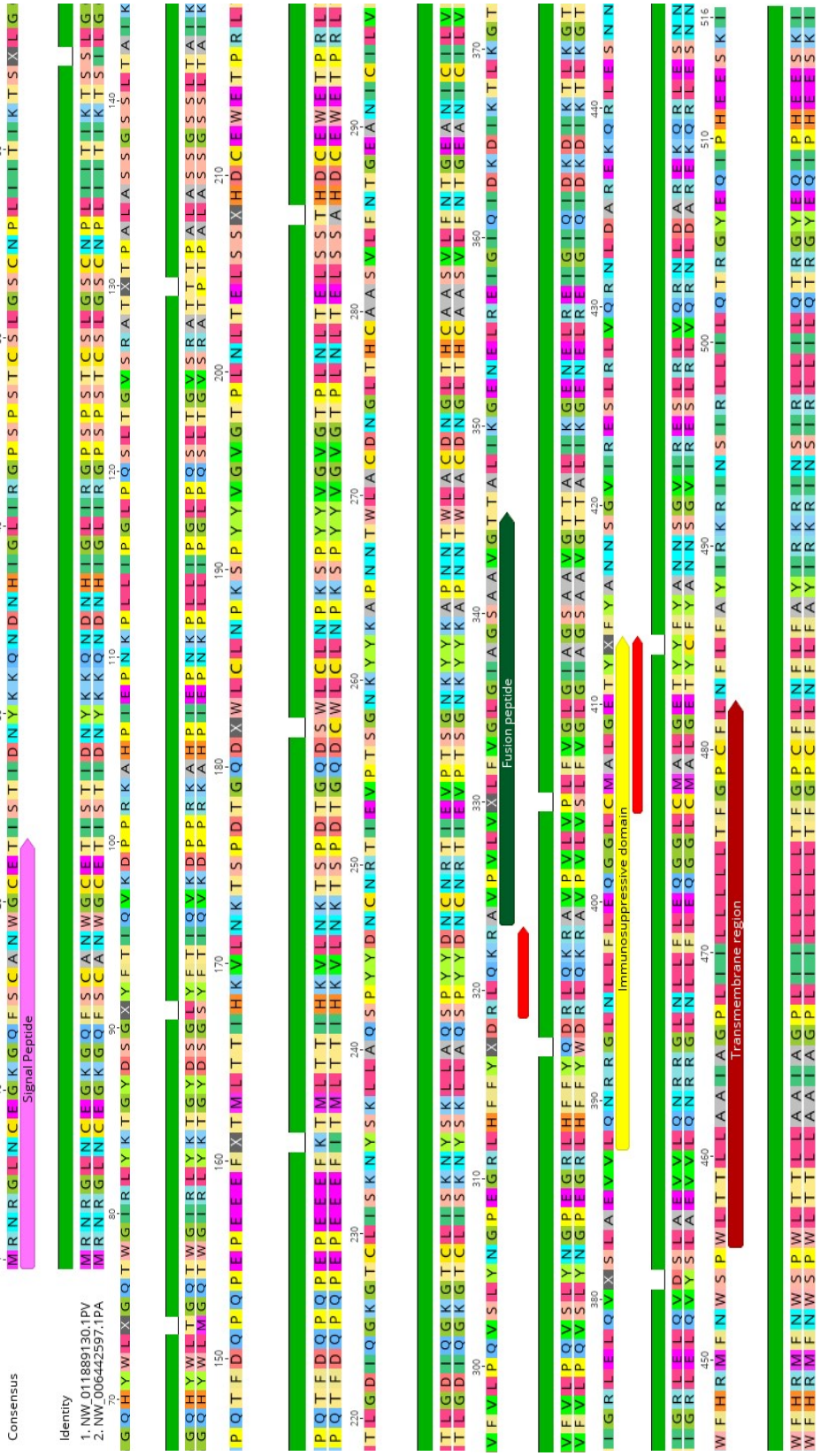


Figure 7.7: MegabatOrtho10 alignment with annotations, consensus sequence and identity across sequence.

8 Appendix to Chapter 3

Family	Abbreviation	Name	POL	ENV	GAG
Alpharetroviruses	ALV	Avian Leukosis Virus	AJG42161.1	AJG42162.1	AJG42160.1
	RSV	Rous Sarcoma Virus	CAA48535.1	CAA48536.1	CAA48534.1
Betaretroviruses	SRV	Simian retrovirus 2	ATN28189.1	ATN28190.1	ATN28187.1
	JSRV	Jaagsiekte sheep retrovirus	AAD45226.1	NP_041188.1	AAA89180.1
	DrERV	Desmodus rotundus endogenous retrovirus	AJR27940.1	AJR27937.1	AJR27933.1
	PoERV	Porcine endogenous retrovirus	AAL38193.1	CAA76583.1	ADG27335.1
Gammaretroviruses	KoRV	Koala Retrovirus	YP_009513211.1	YP_009513212.1	AAF15097.1
Deltaretrovirus	BLV	Bovine leukemia virus	BAA00544.1	ALB75304.1	AAC82585.1
	HTLV	Human T-lymphotropic virus 2	AAD34842.1	AAD34843.1	AAB59884.1
Epsilonretroviruses	WEHV	Walleye epidermal hyperplasia virus 1	AAD30048.1	AAD30049.1	AAD30047.1
Lentiviruses	HIV-1	Human immunodeficiency virus 1	NP_789740.1	AAC82596.1	AAC82593.1
	FIV	Feline immunodeficiency virus	CAA40318.1	AAB59940.1	AAB59936.1
Spumaretroviruses	BFV	Bovine foamy virus	AFR79239.1	AAB68771.1	AAB68769.1
	FFV	Feline foamy virus	CAA11581.1	CAA70076.1	CAA70074.1

Table 8.1: Accession numbers for exogenous and endogenous viral genes used in the ERV BLAST search and analysis

Mollosus mollosus		
gag	**pol**	**env**
Beta_gag—scaffold_m16_p_21—14064531—14063149	Beta_pol—scaffold_m16_p_21—14062353—14059741	None
Beta_gag—scaffold_m16_p_3—50720551—50722587	Beta_pol—scaffold_m16_p_3—50723327—50725903	None
None	Beta_pol—scaffold_m16_p_30—117997—120063	BETA_ENV—scaffold_m16_p_30—121275—122381
None	Gamma_pol—scaffold_m16_p_1—42703620—42707033	GAMMA_ENV—scaffold_m16_p_1—42708426—42709118
Beta_gag—scaffold_m16_p_2—59663508—59661991	Beta_pol—scaffold_m16_p_2—59661245—59658657	None
None	Beta_pol—scaffold_m16_p_12—65386429—65384606	BETA_ENV—scaffold_m16_p_12—65384007—65382412
None	Beta_pol—scaffold_m16_p_4—71823820—71826420	BETA_ENV—scaffold_m16_p_4—71826527—71828203
Beta_gag—scaffold_m16_p_19—41066026—41064512	Beta_pol—scaffold_m16_p_19—41063767—41061074	None
None	Beta_pol—scaffold_m16_p_9—42409896—42411587	BETA_ENV—scaffold_m16_p_9—42411643—42413319
Beta_gag—scaffold_m16_p_6—50199714—50200658	Beta_pol—scaffold_m16_p_6—50201398—50203752	None
Beta_gag—scaffold_m16_p_13—19594833—19596869	Beta_pol—scaffold_m16_p_13—19597609—19600188	None
Beta_gag—scaffold_m16_p_4—52356023—52358077	Beta_pol—scaffold_m16_p_4—52358817—52361396	BETA_ENV—scaffold_m16_p_4—52362328—52362978
Beta_gag—scaffold_m16_p_12—7547736—7546861	Beta_pol—scaffold_m16_p_12—7546613—7544319	None
Beta_gag—scaffold_m16_p_5—51576672—51575641	Beta_pol—scaffold_m16_p_5—51574388—51572745	None
Beta_gag—scaffold_m16_p_31—2409601—2410233	Beta_pol—scaffold_m16_p_31—2412156—2413688	None
Beta_gag—scaffold_m16_p_18—186965—188920	Beta_pol—scaffold_m16_p_18—189667—192360	None
None	Beta_pol—scaffold_m16_p_2—110606024—110608432	BETA_ENV—scaffold_m16_p_2—110609207—110610187
None	Gamma_pol—scaffold_m16_p_1—128018712—128015842	GAMMA_ENV—scaffold_m16_p_1—128015370—128014627
Beta_gag—scaffold_m16_p_30—490030—491643	Beta_pol—scaffold_m16_p_30—492646—495294	None
Beta_gag—scaffold_m16_p_14—34287940—34289046	Beta_pol—scaffold_m16_p_14—34289792—34292440	None
Gamma_gag—scaffold_m16_p_32—506139—507071	None	GAMMA_ENV—scaffold_m16_p_32—509485—510660
Beta_gag—scaffold_m16_p_34—13793—14902	Beta_pol—scaffold_m16_p_34—15655—16968	None
Beta_gag—scaffold_m16_p_10—75551790—75552701	Beta_pol—scaffold_m16_p_10—75553757—75555697	None
Beta_gag—scaffold_m16_p_2—70331044—70331802	Beta_pol—scaffold_m16_p_2—70332905—70334344	None
Beta_gag—scaffold_m16_p_3—132372033—132373841	Beta_pol—scaffold_m16_p_3—132374049—132376355	None
Beta_gag—scaffold_m16_p_15—67387293—67385554	Beta_pol—scaffold_m16_p_15—67384821—67382245	BETA_ENV—scaffold_m16_p_15—67381443—67380661
None	Gamma_pol—scaffold_m16_p_1—186741934—186745527	GAMMA_ENV—scaffold_m16_p_1—186746210—186747226
Beta_gag—scaffold_m16_p_9—98631276—98630401	Beta_pol—scaffold_m16_p_9—98630434—98628194	None
Beta_gag—scaffold_m16_p_5—99319901—99321010	Beta_pol—scaffold_m16_p_5—99322037—99324325	None
None	Gamma_pol—scaffold_m16_p_12—9648696—9645619	GAMMA_ENV—scaffold_m16_p_12—9645466—9644423
Beta_gag—scaffold_m16_p_13—23063532—23064626	Beta_pol—scaffold_m16_p_13—23066030—23067343	None
Beta_gag—scaffold_m16_p_23—4322225—4320777	Beta_pol—scaffold_m16_p_23—4320128—4319832	None
Gamma_gag—scaffold_m16_p_46—82887—81916	None	GAMMA_ENV—scaffold_m16_p_46—77601—76372
Delta_gag—scaffold_m16_p_9—8679506—8680705	Delta_pol—scaffold_m16_p_9—8676507—8679566	Delta_env—scaffold_m16_p_9—8674906—8676474

Table 8.2: List of all integrations found in *Mollosus mollosus* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Myotis myotis		
gag	**pol**	**env**
Beta_gag—scaffold_m19_p_1—21038619—21039884	Beta_pol—scaffold_m19_p_1—21040202—21043153	None
Beta_gag—scaffold_m19_p_2—46127560—46126631	Beta_pol—scaffold_m19_p_2—46125972—46123036	None
Beta_gag—scaffold_m19_p_1—126300324—126301901	Beta_pol—scaffold_m19_p_1—126302917—126304923	None
Beta_gag—scaffold_m19_p_4—75364036—75365298	Beta_pol—scaffold_m19_p_4—75366734—75368584	None
Beta_gag—scaffold_m19_p_4—20084799—20083537	Beta_pol—scaffold_m19_p_4—20083204—20080271	None
Beta_gag—scaffold_m19_p_3—65171049—65172311	Beta_pol—scaffold_m19_p_3—65172629—65175580	None
Beta_gag—scaffold_m19_p_2—215521444—215522709	Beta_pol—scaffold_m19_p_2—215523042—215525978	None
Beta_gag—scaffold_m19_p_15—37402270—37403532	Beta_pol—scaffold_m19_p_15—37403865—37406822	None
Beta_gag—scaffold_m19_p_5—16945964—16947226	Beta_pol—scaffold_m19_p_5—16947544—16950495	None
Beta_gag—scaffold_m19_p_2—94377679—94378959	Beta_pol—scaffold_m19_p_2—94379277—94382228	None
Beta_gag—scaffold_m19_p_6—44486799—44488061	Beta_pol—scaffold_m19_p_6—44488394—44491330	None
Gamma_gag—scaffold_m19_p_4—96326017—96327354	Gamma_pol—scaffold_m19_p_4—96327052—96330561	Gamma_env—scaffold_m19_p_4—96330977—96331522
Beta_gag—scaffold_m19_p_64—321928—323550	Beta_pol—scaffold_m19_p_64—324534—326828	None
Beta_gag—scaffold_m19_p_2—139219781—139218516	Beta_pol—scaffold_m19_p_2—139218198—139215250	None
Beta_gag—scaffold_m19_p_3—146917344—146916100	None	Gamma_env—scaffold_m19_p_3—146920790—146922391
None	Gamma_pol—scaffold_m19_p_4—77152795—77156292	Gamma_env—scaffold_m19_p_4—77156708—77157664
Beta_gag—scaffold_m19_p_7—7171900—7173165	Beta_pol—scaffold_m19_p_7—7173483—7176434	None
Beta_gag—scaffold_m19_p_3—172779019—172777754	Beta_pol—scaffold_m19_p_3—172777421—172774485	None
Beta_gag—scaffold_m19_p_3—18182323—18183588	Beta_pol—scaffold_m19_p_3—18183921—18186857	None
None	Gamma_pol—scaffold_m19_p_25—4632860—4631697	Gamma_env—scaffold_m19_p_25—4631410—4630499
Beta_gag—scaffold_m19_p_4—120507236—120506220	Beta_pol—scaffold_m19_p_4—120505526—120503829	None
Beta_gag—scaffold_m19_p_12—47436393—47435131	Beta_pol—scaffold_m19_p_12—47434798—47431862	None
Beta_gag—scaffold_m19_p_22—7153780—7152599	None	Beta_env—scaffold_m19_p_22—7149441—7147849
Beta_gag—scaffold_m19_p_19—6504397—6503135	Beta_pol—scaffold_m19_p_19—6502802—6499866	None
Beta_gag—scaffold_m19_p_22—2370655—2369393	Beta_pol—scaffold_m19_p_22—2369075—2366124	None
None	Beta_pol—scaffold_m19_p_1—125300145—125301833	Beta_env—scaffold_m19_p_1—125302120—125303829
Beta_gag—scaffold_m19_p_2—111974931—111973669	Beta_pol—scaffold_m19_p_2—111972299—111971100	None
Beta_gag—scaffold_m19_p_4—99528344—99529606	Beta_pol—scaffold_m19_p_4—99529939—99532875	None
Beta_gag—scaffold_m19_p_4—70705043—70706308	None	Gamma_env—scaffold_m19_p_4—70712714—70713334
Beta_gag—scaffold_m19_p_1—119889083—119887821	Beta_pol—scaffold_m19_p_1—119887488—119884552	None
Beta_gag—scaffold_m19_p_9—25198555—25199820	Beta_pol—scaffold_m19_p_9—25200138—25203089	None
Beta_gag—scaffold_m19_p_3—11413099—11414361	Beta_pol—scaffold_m19_p_3—11416078—11417628	None
Beta_gag—scaffold_m19_p_4—43584839—43583577	Beta_pol—scaffold_m19_p_4—43583244—43581283	None
Beta_gag—scaffold_m19_p_1—26968371—26967106	Beta_pol—scaffold_m19_p_1—26966773—26963840	None
None	Gamma_pol—scaffold_m19_p_19—11609646—11613230	Gamma_env—scaffold_m19_p_19—11614249—11615127
Beta_gag—scaffold_m19_p_4—70178300—70177035	Beta_pol—scaffold_m19_p_4—70176702—70173775	None
Beta_gag—scaffold_m19_p_1—44831804—44833066	Beta_pol—scaffold_m19_p_1—44833384—44836335	None
Beta_gag—scaffold_m19_p_12—18563618—18562353	Beta_pol—scaffold_m19_p_12—18562035—18559084	None

Table 8.3: List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Myotis myotis		
gag	**pol**	**env**
Beta_gag—scaffold_m19_p_7—13651612—13652874	Beta_pol—scaffold_m19_p_7—13653207—13656143	None
Beta_gag—scaffold_m19_p_4—98754348—98753083	Beta_pol—scaffold_m19_p_4—98752750—98749814	None
Beta_gag—scaffold_m19_p_8—17809520—17810782	Beta_pol—scaffold_m19_p_8—17811115—17813676	None
Beta_gag—scaffold_m19_p_4—90712649—90713911	Beta_pol—scaffold_m19_p_4—90714244—90717180	None
Beta_gag—scaffold_m19_p_1—55104973—55103711	Beta_pol—scaffold_m19_p_1—55103365—55101314	None
Beta_gag—scaffold_m19_p_4—90304038—90305300	Beta_pol—scaffold_m19_p_4—90306050—90308566	None
Beta_gag—scaffold_m19_p_3—97198466—97199587	Beta_pol—scaffold_m19_p_3—97200862—97202883	None
Beta_gag—scaffold_m19_p_6—11022585—11023847	Beta_pol—scaffold_m19_p_6—11024180—11027113	None
Beta_gag—scaffold_m19_p_4—90724325—90725488	Beta_pol—scaffold_m19_p_4—90726224—90728776	None
Beta_gag—scaffold_m19_p_3—54699704—54700966	Beta_pol—scaffold_m19_p_3—54701299—54704229	None
Beta_gag—scaffold_m19_p_7—28207623—28208885	Beta_pol—scaffold_m19_p_7—28209891—28212176	None
Beta_gag—scaffold_m19_p_10—50134240—50135502	Beta_pol—scaffold_m19_p_10—50136217—50138790	None
Beta_gag—scaffold_m19_p_1—18449237—18450499	Beta_pol—scaffold_m19_p_1—18450832—18453768	None
Beta_gag—scaffold_m19_p_1—80796021—80794759	Beta_pol—scaffold_m19_p_1—80794426—80791490	None
Beta_gag—scaffold_m19_p_13—6751968—6753230	Beta_pol—scaffold_m19_p_13—6754896—6756518	None
Beta_gag—scaffold_m19_p_1—11610029—11611291	Beta_pol—scaffold_m19_p_1—11611609—11614068	None
Beta_gag—scaffold_m19_p_11—47820048—47818783	Beta_pol—scaffold_m19_p_11—47818450—47815514	None
Beta_gag—scaffold_m19_p_2—20009977—20011239	Beta_pol—scaffold_m19_p_2—20011572—20014508	None
Beta_gag—scaffold_m19_p_8—3307953—3306691	Beta_pol—scaffold_m19_p_8—3306358—3303422	None
Beta_gag—scaffold_m19_p_5—30996058—30997320	Beta_pol—scaffold_m19_p_5—30997638—31000589	None
Beta_gag—scaffold_m19_p_4—71964633—71965913	Beta_pol—scaffold_m19_p_4—71966231—71969173	None
Beta_gag—scaffold_m19_p_2—213773256—213774518	Beta_pol—scaffold_m19_p_2—213774845—213777787	None
Beta_gag—scaffold_m19_p_2—48027527—48028789	Beta_pol—scaffold_m19_p_2—48029107—48032058	None
Beta_gag—scaffold_m19_p_5—87715024—87716286	Beta_pol—scaffold_m19_p_5—87716604—87719555	None
Beta_gag—scaffold_m19_p_6—73481502—73480384	Beta_pol—scaffold_m19_p_6—73479664—73477079	None
Beta_gag—scaffold_m19_p_1—6722045—6720801	Beta_pol—scaffold_m19_p_1—6720483—6717538	None
Beta_gag—scaffold_m19_p_9—7905786—7907048	Beta_pol—scaffold_m19_p_9—7907366—7910317	None
Beta_gag—scaffold_m19_p_8—29047624—29048886	Beta_pol—scaffold_m19_p_8—29049626—29052157	None
Beta_gag—scaffold_m19_p_22—5026780—5028042	Beta_pol—scaffold_m19_p_22—5028375—5031311	None
Beta_gag—scaffold_m19_p_2—22272521—22273783	Beta_pol—scaffold_m19_p_2—22274116—22277049	None
Beta_gag—scaffold_m19_p_7—81693642—81694922	Beta_pol—scaffold_m19_p_7—81695255—81698191	None
Beta_gag—scaffold_m19_p_52—717111—715849	Beta_pol—scaffold_m19_p_52—715516—712580	None
Beta_gag—scaffold_m19_p_6—70062451—70063713	Beta_pol—scaffold_m19_p_6—70064046—70067003	None
Beta_gag—scaffold_m19_p_3—8353266—8354528	Beta_pol—scaffold_m19_p_3—8354861—8357797	None
Beta_gag—scaffold_m19_p_2—27715743—27717005	Beta_pol—scaffold_m19_p_2—27717323—27720274	None
Beta_gag—scaffold_m19_p_20—18648999—18650261	Beta_pol—scaffold_m19_p_20—18650579—18653530	None
Beta_gag—scaffold_m19_p_19—29990513—29991775	Beta_pol—scaffold_m19_p_19—29992093—29995044	None
Beta_gag—scaffold_m19_p_7—23042488—23041226	Beta_pol—scaffold_m19_p_7—23040893—23037957	None
Beta_gag—scaffold_m19_p_3—139151169—139153322	Beta_pol—scaffold_m19_p_3—139154062—139156641	Beta_env—scaffold_m19_p_3—139157028—139158638
None	Beta_pol—scaffold_m19_p_4—76554438—76556792	Beta_env—scaffold_m19_p_4—76557071—76558807

Table 8.4: Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Myotis myotis		
gag	**pol**	**env**
Beta_gag—scaffold_m19_p_2—83641110—83642372	Beta_pol—scaffold_m19_p_2—83642705—83645641	None
Gamma_gag—scaffold_m19_p_2—106805168—106803810	None	Gamma_env—scaffold_m19_p_2—106801934—106801140
Beta_gag—scaffold_m19_p_4—67704086—67702824	Beta_pol—scaffold_m19_p_4—67702491—67699555	None
Beta_gag—scaffold_m19_p_31—12786—11524	Beta_pol—scaffold_m19_p_31—10364—8259	None
Beta_gag—scaffold_m19_p_3—123372515—123373813	Beta_pol—scaffold_m19_p_3—123374131—123377061	None
Beta_gag—scaffold_m19_p_1—4581385—4582644	Beta_pol—scaffold_m19_p_1—4582962—4585913	None
Beta_gag—scaffold_m19_p_12—15842801—15843967	Beta_pol—scaffold_m19_p_12—15844746—15847403	None
Beta_gag—scaffold_m19_p_1—58601355—58602599	Beta_pol—scaffold_m19_p_1—58602622—58603770	None
Beta_gag—scaffold_m19_p_10—64050046—64051305	Beta_pol—scaffold_m19_p_10—64052093—64054561	None
Beta_gag—scaffold_m19_p_2—147322460—147321198	Beta_pol—scaffold_m19_p_2—147320880—147317929	None
Beta_gag—scaffold_m19_p_11—66934790—66936052	Beta_pol—scaffold_m19_p_11—66936370—66939318	None
Beta_gag—scaffold_m19_p_1—213026340—213027599	Beta_pol—scaffold_m19_p_1—213028390—213029850	None
Beta_gag—scaffold_m19_p_2—66936003—66934741	Beta_pol—scaffold_m19_p_2—66934408—66932495	None
Beta_gag—scaffold_m19_p_10—72583999—72582737	Beta_pol—scaffold_m19_p_10—72582419—72579471	None
Beta_gag—scaffold_m19_p_3—157733682—157734944	Beta_pol—scaffold_m19_p_3—157735262—157738213	None
Beta_gag—scaffold_m19_p_2—104323637—104322528	Beta_pol—scaffold_m19_p_2—104322043—104319107	None
Gamma_gag—scaffold_m19_p_40—1805465—1806940	Gamma_pol—scaffold_m19_p_40—1807311—1810154	Gamma_env—scaffold_m19_p_40—1810831—1812405
Beta_gag—scaffold_m19_p_10—60130317—60131579	Beta_pol—scaffold_m19_p_10—60131897—60134848	None
Beta_gag—scaffold_m19_p_4—98740019—98738754	Beta_pol—scaffold_m19_p_4—98738421—98735485	None
Beta_gag—scaffold_m19_p_10—64617485—64618519	Beta_pol—scaffold_m19_p_10—64619258—64620937	None
Beta_gag—scaffold_m19_p_2—108251918—108250653	Beta_pol—scaffold_m19_p_2—108249038—108247386	None
Beta_gag—scaffold_m19_p_7—1812260—1813522	Beta_pol—scaffold_m19_p_7—1813852—1816791	None
Beta_gag—scaffold_m19_p_3—84689830—84691092	Beta_pol—scaffold_m19_p_3—84691425—84694361	None
Beta_gag—scaffold_m19_p_2—73322582—73323661	None	Gamma_env—scaffold_m19_p_2—73327309—73329075
Beta_gag—scaffold_m19_p_9—47439620—47438358	Beta_pol—scaffold_m19_p_9—47438040—47435089	None
Beta_gag—scaffold_m19_p_3—21290950—21292212	Beta_pol—scaffold_m19_p_3—21292545—21295481	None
Beta_gag—scaffold_m19_p_5—60853516—60852254	Beta_pol—scaffold_m19_p_5—60851921—60848988	None
Beta_gag—scaffold_m19_p_1—37061036—37062298	Beta_pol—scaffold_m19_p_1—37062616—37065567	None
Beta_gag—scaffold_m19_p_3—174806174—174804912	Beta_pol—scaffold_m19_p_3—174804594—174801643	None
Beta_gag—scaffold_m19_p_3—102264448—102265710	Beta_pol—scaffold_m19_p_3—102266043—102268979	None
Beta_gag—scaffold_m19_p_4—34311644—34312837	Beta_pol—scaffold_m19_p_4—34314785—34316104	None
Beta_gag—scaffold_m19_p_2—2414950—2413688	Beta_pol—scaffold_m19_p_2—2413370—2410419	None
Beta_gag—scaffold_m19_p_1—192228003—192229265	Beta_pol—scaffold_m19_p_1—192229598—192232555	None
Beta_gag—scaffold_m19_p_2—111868393—111869655	Beta_pol—scaffold_m19_p_2—111869988—111872924	None
Beta_gag—scaffold_m19_p_10—43151737—43150475	Beta_pol—scaffold_m19_p_10—43148667—43147216	None
Beta_gag—scaffold_m19_p_1—29153022—29151760	Beta_pol—scaffold_m19_p_1—29151442—29148491	None
Beta_gag—scaffold_m19_p_4—74019072—74020337	Beta_pol—scaffold_m19_p_4—74021487—74023598	None
Beta_gag—scaffold_m19_p_5—4002113—4003375	Beta_pol—scaffold_m19_p_5—4003708—4006644	None
Gamma_gag—scaffold_m19_p_28—3329034—3330083	Gamma_pol—scaffold_m19_p_28—3331300—3333657	None
Beta_gag—scaffold_m19_p_1—143499868—143498606	Beta_pol—scaffold_m19_p_1—143498273—143495316	None

Table 8.5: Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Myotis myotis		
gag	**pol**	**env**
Beta_gag—scaffold_m19_p_6—12249472—12250737	Beta_pol—scaffold_m19_p_6—12251070—12254006	None
Beta_gag—scaffold_m19_p_11—31189137—31190402	Beta_pol—scaffold_m19_p_11—31190735—31193671	None
Beta_gag—scaffold_m19_p_3—190460351—190459089	Beta_pol—scaffold_m19_p_3—190458756—190455820	None
Beta_gag—scaffold_m19_p_2—111789076—111790320	Beta_pol—scaffold_m19_p_2—111790638—111793607	None
Beta_gag—scaffold_m19_p_1—4112680—4111418	Beta_pol—scaffold_m19_p_1—4111091—4108431	None
Beta_gag—scaffold_m19_p_3—130183508—130184770	Beta_pol—scaffold_m19_p_3—130185088—130188039	None
Beta_gag—scaffold_m19_p_4—80261369—80260107	Beta_pol—scaffold_m19_p_4—80259789—80256838	None
Beta_gag—scaffold_m19_p_1—33908029—33909291	Beta_pol—scaffold_m19_p_1—33909624—33912560	None
None	Gamma_pol—scaffold_m19_p_3—101584898—101588407	Gamma_env—scaffold_m19_p_3—101588707—101590284
Beta_gag—scaffold_m19_p_11—55656344—55657606	Beta_pol—scaffold_m19_p_11—55657939—55660875	None
Beta_gag—scaffold_m19_p_2—163796621—163797883	Beta_pol—scaffold_m19_p_2—163798201—163801152	None
Beta_gag—scaffold_m19_p_2—202958162—202959427	Beta_pol—scaffold_m19_p_2—202959760—202962696	None
Beta_gag—scaffold_m19_p_1—34951459—34950197	Beta_pol—scaffold_m19_p_1—34949864—34946928	None
Beta_gag—scaffold_m19_p_1—54383771—54385033	Beta_pol—scaffold_m19_p_1—54385351—54388302	None
Beta_gag—scaffold_m19_p_1—107420337—107419075	Beta_pol—scaffold_m19_p_1—107418728—107415807	None
Beta_gag—scaffold_m19_p_23—2882744—2881482	Beta_pol—scaffold_m19_p_23—2881149—2878213	None
Beta_gag—scaffold_m19_p_9—9132488—9133750	Beta_pol—scaffold_m19_p_9—9134083—9137010	None
Beta_gag—scaffold_m19_p_1—35981565—35980303	Beta_pol—scaffold_m19_p_1—35979585—35977039	None
Beta_gag—scaffold_m19_p_6—38168995—38170260	Beta_pol—scaffold_m19_p_6—38172130—38174028	None
Beta_gag—scaffold_m19_p_3—29021191—29019929	Beta_pol—scaffold_m19_p_3—29019596—29016660	None
Beta_gag—scaffold_m19_p_5—42368130—42369392	Beta_pol—scaffold_m19_p_5—42369710—42372661	None
Beta_gag—scaffold_m19_p_1—19803626—19802361	Beta_pol—scaffold_m19_p_1—19802043—19799092	None
Beta_gag—scaffold_m19_p_4—8536495—8537760	Beta_pol—scaffold_m19_p_4—8539069—8541051	None
Beta_gag—scaffold_m19_p_1—17708652—17709914	Beta_pol—scaffold_m19_p_1—17710247—17713183	None
Beta_gag—scaffold_m19_p_19—8410173—8411435	Beta_pol—scaffold_m19_p_19—8412594—8414696	None
Beta_gag—scaffold_m19_p_18—4594—5832	None	Gamma_env—scaffold_m19_p_18—15600—14806

Table 8.6: Continued. List of all integrations found in *Myotis myotis* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Phyllostomus discolor			
gag	**pol**	**env**	**Viral ma**
None	gamma_pol—scaffold_m19_p_5—65905131—65902696	gamma_env—scaffold_m19_p_5—65901673—65900858	Avian endogenous
beta_gag—scaffold_m19_p_15—1504060—1506243	None	beta_env—scaffold_m19_p_15—1506780—1508210	
None	gamma_pol—scaffold_m19_p_5—65374988—65377579	gamma_env—scaffold_m19_p_5—65377909—65379477	
None	gamma_pol—scaffold_m19_p_13—654202—657774	gamma_env—scaffold_m19_p_13—658807—659682	
gamma_gag—scaffold_m19_p_21—390880—391491	None	gamma_env—scaffold_m19_p_21—395095—395889	
None	beta_pol—scaffold_m19_p_10—43372683—43375286	beta_env—scaffold_m19_p_10—43375291—43375890	
beta_gag—scaffold_m19_p_4—124893685—124892999	beta_pol—scaffold_m19_p_4—124891085—124889769	None	
beta_gag—scaffold_m19_p_1—102961441—102962439	None	beta_env—scaffold_m19_p_1—102962661—102963500	
beta_gag—scaffold_m19_p_3—124654134—124656227	None	beta_env—scaffold_m19_p_3—124656314—124657849	
beta_gag—scaffold_m19_p_3—8663896—8664894	None	beta_env—scaffold_m19_p_3—8665430—8666860	
beta_gag—scaffold_m19_p_2—78385821—78387770	None	beta_env—scaffold_m19_p_2—78387925—78388794	
beta_gag—scaffold_m19_p_11—9958061—9959941	beta_pol—scaffold_m19_p_11—9960157—9962082	beta_env—scaffold_m19_p_11—9963184—9963894	
beta_gag—scaffold_m19_p_6—76710526—76711452	beta_pol—scaffold_m19_p_6—76712509—76714671	None	
beta_gag—scaffold_m19_p_7—4785526—4786668	beta_pol—scaffold_m19_p_7—4787507—4789216	None	
beta_gag—scaffold_m19_p_15—50420283—50422466	beta_pol—scaffold_m19_p_15—50423215—50425833	beta_env—scaffold_m19_p_15—50425820—50426422	
beta_gag—scaffold_m19_p_13—43371465—43369261	None	beta_env—scaffold_m19_p_13—43368725—43367295	
beta_gag—scaffold_m19_p_11—9958061—9959941	beta_pol—scaffold_m19_p_11—9960157—9962082	beta_env—scaffold_m19_p_11—9962180—9963166	
beta_gag—scaffold_m19_p_7—92354458—92352377	None	beta_env—scaffold_m19_p_7—92351749—92350319	
beta_gag—scaffold_m19_p_14—12677594—12675801	None	beta_env—scaffold_m19_p_14—12675265—12673835	
gamma_gag—scaffold_m19_p_5—92487469—92487769	gamma_pol—scaffold_m19_p_5—92483093—92484293	alpha_env—scaffold_m19_p_5—92481255—92481824	

Table 8.7: List of all integrations found in Phyllostomus discolor containing two or more genes with name of the viral family scaffold, and position on the scaffold

Pipistrellus kuhlii		
gag	**pol**	**env**
Beta_gag—scaffold_m20_p_126—297368—298381	None	Beta_env—scaffold_m20_p_126—300467—301129
Beta_gag—scaffold_m20_p_54—427932—427141	Beta_pol—scaffold_m20_p_54—425162—423405	None
Gamma_gag—scaffold_m20_p_47—2309549—2310352	Gamma_pol—scaffold_m20_p_47—2310413—2311867	None
Beta_gag—scaffold_m20_p_2—57523541—57524695	Beta_pol—scaffold_m20_p_2—57525315—57527915	None
None	Gamma_pol—scaffold_m20_p_83—155692—157617	Gamma_env—scaffold_m20_p_83—158529—159557
Gamma_gag—scaffold_m20_p_98—328573—330105	Gamma_pol—scaffold_m20_p_98—330223—331926	None
Beta_gag—scaffold_m20_p_19—8054112—8052895	None	Beta_env—scaffold_m20_p_19—8052722—8051556
Gamma_gag—scaffold_m20_p_119—12928—13665	Gamma_pol—scaffold_m20_p_119—15857—17152	Beta_env—scaffold_m20_p_119—18447—19115
Gamma_gag—scaffold_m20_p_73—61882—62874	Gamma_pol—scaffold_m20_p_73—63474—65747	Gamma_env—scaffold_m20_p_73—66996—67748
Beta_gag—scaffold_m20_p_3—32386000—32386914	Beta_pol—scaffold_m20_p_3—32388492—32389736	None
Gamma_gag—scaffold_m20_p_76—825298—824786	Gamma_pol—scaffold_m20_p_76—823551—821245	None
Beta_gag—scaffold_m20_p_14—47942087—47943211	Beta_pol—scaffold_m20_p_14—47944018—47946486	None
Gamma_gag—scaffold_m20_p_83—355501—354926	Gamma_pol—scaffold_m20_p_83—353095—351749	None
None	Gamma_pol—scaffold_m20_p_14—47261854—47258297	Gamma_env—scaffold_m20_p_14—47258378—47256429
None	Beta_pol—scaffold_m20_p_27—67325—66117	Beta_env—scaffold_m20_p_27—65720—65007
Beta_gag—scaffold_m20_p_58—388311—388919	None	Beta_env—scaffold_m20_p_58—391267—392259
Beta_gag—scaffold_m20_p_54—501391—501987	Beta_pol—scaffold_m20_p_54—503966—505606	None
None	Gamma_pol—scaffold_m20_p_36—3373322—3369762	Gamma_env—scaffold_m20_p_36—3369838—3367898
Beta_gag—scaffold_m20_p_22—882382—881258	Beta_pol—scaffold_m20_p_22—880940—878007	None
Beta_gag—scaffold_m20_p_27—3080233—3081252	Beta_pol—scaffold_m20_p_27—3083377—3084699	None
Beta_gag—scaffold_m20_p_126—297368—298381	None	Beta_env—scaffold_m20_p_126—301152—301940
Beta_gag—scaffold_m20_p_10—8298—8978	None	Beta_env—scaffold_m20_p_10—13508—14485
Beta_gag—scaffold_m20_p_2—189062240—189064219	Beta_pol—scaffold_m20_p_2—189065781—189067664	Beta_env—scaffold_m20_p_2—189071562—189072929
Beta_gag—scaffold_m20_p_1—127411900—127413012	Beta_pol—scaffold_m20_p_1—127413317—127415164	None
None	Beta_pol—scaffold_m20_p_15—2448295—2447162	Beta_env—scaffold_m20_p_15—2445792—2444326
Gamma_gag—scaffold_m20_p_40—204724—205551	Gamma_pol—scaffold_m20_p_40—205812—207551	None
None	Gamma_pol—scaffold_m20_p_122—53464—54921	Gamma_env—scaffold_m20_p_122—59750—60790
Gamma_gag—scaffold_m20_p_8—28302057—28302980	Gamma_pol—scaffold_m20_p_8—28302987—28306553	Gamma_env—scaffold_m20_p_8—28306472—28308421
Beta_gag—scaffold_m20_p_10—8298—8978	None	Gamma_env—scaffold_m20_p_10—12774—13514
Beta_gag—scaffold_m20_p_16—23164984—23164322	None	Gamma_env—scaffold_m20_p_16—23158057—23158638
None	Gamma_pol—scaffold_m20_p_5—4292232—4288666	Gamma_env—scaffold_m20_p_5—4288747—4286798
None	Gamma_pol—scaffold_m20_p_3—61967414—61963857	Gamma_env—scaffold_m20_p_3—61963899—61961989
Epsilon_gag—scaffold_m20_p_183—9343—8612	Epsilon_pol—scaffold_m20_p_183—8339—6975	None
Beta_gag—scaffold_m20_p_154—158669—157545	Beta_pol—scaffold_m20_p_154—156895—154727	None
None	Gamma_pol—scaffold_m20_p_18—8271722—8274901	Gamma_env—scaffold_m20_p_18—8275514—8277181
Beta_gag—scaffold_m20_p_9—35349460—35348708	None	Beta_env—scaffold_m20_p_9—35348418—35347678
Beta_gag—scaffold_m20_p_2—189062240—189064219	Beta_pol—scaffold_m20_p_2—189065781—189067664	Beta_env—scaffold_m20_p_2—189067781—189069427

Table 8.8: List of all integrations found in *Pipistrellus kuhlii* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Rhinolophus ferrumequinum		
gag	**pol**	**env**
Beta_gag—scaffold_m29_p_1—30952981—30951770	Beta_pol—scaffold_m29_p_1—30951020—30948552	None
None	Gamma_pol—scaffold_m29_p_14—5298067—5295458	gamma_env—scaffold_m29_p_14—5295447—5293600
None	Gamma_pol—scaffold_m29_p_24—40780669—40784148	gamma_env—scaffold_m29_p_24—40784168—40786024
Beta_gag—scaffold_m29_p_1—30288137—30287073	Beta_pol—scaffold_m29_p_1—30285367—30283601	None
Beta_gag—scaffold_m29_p_1—52734312—52735499	Beta_pol—scaffold_m29_p_1—52736593—52738461	None
None	Gamma_pol—scaffold_m29_p_2—51341912—51344458	gamma_env—scaffold_m29_p_2—51344478—51346334
Beta_gag—scaffold_m29_p_23—6638760—6640760	Beta_pol—scaffold_m29_p_23—6641500—6644100	None
Beta_gag—scaffold_m29_p_13—28005275—28004127	Beta_pol—scaffold_m29_p_13—28002728—28000908	None
None	Beta_pol—scaffold_m29_p_18—30707192—30708496	beta_env—scaffold_m29_p_18—30709621—30710697
Beta_gag—scaffold_m29_p_11—29691354—29690038	Beta_pol—scaffold_m29_p_11—29689298—29686701	None
Beta_gag—scaffold_m29_p_1—92463849—92465018	Beta_pol—scaffold_m29_p_1—92466635—92468332	None
Beta_gag—scaffold_m29_p_4—14034013—14035329	Beta_pol—scaffold_m29_p_4—14036069—14038654	None
None	Gamma_pol—scaffold_m29_p_4—45763726—45767214	gamma_env—scaffold_m29_p_4—45767225—45769081
Gamma_gag—scaffold_m29_p_34—1863932—1862937	Gamma_pol—scaffold_m29_p_34—1862798—1859310	gamma_env—scaffold_m29_p_34—1859299—1857449
Beta_gag—scaffold_m29_p_1—22538490—22540475	Beta_pol—scaffold_m29_p_1—22541557—22542939	None
Beta_gag—scaffold_m29_p_12—1313361—1315361	Beta_pol—scaffold_m29_p_12—1316288—1317700	gamma_env—scaffold_m29_p_12—1319403—1320377
Beta_gag—scaffold_m29_p_1—27453770—27454954	Beta_pol—scaffold_m29_p_1—27455697—27458003	None
Beta_gag—scaffold_m29_p_1—41533703—41534914	Beta_pol—scaffold_m29_p_1—41535666—41538257	None
None	Gamma_pol—scaffold_m29_p_4—9652287—9655775	gamma_env—scaffold_m29_p_4—9655786—9657642
Beta_gag—scaffold_m29_p_12—248014—249933	Beta_pol—scaffold_m29_p_12—250682—253264	None
Beta_gag—scaffold_m29_p_2—3980402—3979248	None	beta_env—scaffold_m29_p_2—3977350—3976418
Beta_gag—scaffold_m29_p_22—9937578—9938315	Beta_pol—scaffold_m29_p_22—9938457—9939710	None
Beta_gag—scaffold_m29_p_13—46036481—46037692	Beta_pol—scaffold_m29_p_13—46038426—46041029	None
Beta_gag—scaffold_m29_p_8—60491912—60493069	Beta_pol—scaffold_m29_p_8—60494120—60495601	None
None	Gamma_pol—scaffold_m29_p_24—9633397—9636885	gamma_env—scaffold_m29_p_24—9636896—9638752
Gamma_gag—scaffold_m29_p_34—1815438—1816112	None	beta_env—scaffold_m29_p_34—1817288—1817872
Beta_gag—scaffold_m29_p_4—67364189—67365505	Beta_pol—scaffold_m29_p_4—67366245—67368578	None
Beta_gag—scaffold_m29_p_12—15421608—15419674	Beta_pol—scaffold_m29_p_12—15418194—15416335	beta_env—scaffold_m29_p_12—15416249—15414570
Beta_gag—scaffold_m29_p_6—32085603—32086772	None	beta_env—scaffold_m29_p_6—32088111—32089382
Beta_gag—scaffold_m29_p_4—11558620—11557304	Beta_pol—scaffold_m29_p_4—11556564—11554231	None
None	Gamma_pol—scaffold_m29_p_14—53421171—53424659	gamma_env—scaffold_m29_p_14—53424670—53426526
Gamma_gag—scaffold_m29_p_12—9023938—9024717	None	gamma_env—scaffold_m29_p_12—9028587—9029393
Beta_gag—scaffold_m29_p_18—595942—594626	Beta_pol—scaffold_m29_p_18—593886—591289	None
Beta_gag—scaffold_m29_p_1—39599280—39598093	Beta_pol—scaffold_m29_p_1—39596584—39595142	None
Beta_gag—scaffold_m29_p_1—36028915—36030915	Beta_pol—scaffold_m29_p_1—36032396—36034252	None
None	Beta_pol—scaffold_m29_p_46—29275—27233	beta_env—scaffold_m29_p_46—27222—25654
None	Gamma_pol—scaffold_m29_p_15—10846404—10844893	gamma_env—scaffold_m29_p_15—10844882—10843026
None	Gamma_pol—scaffold_m29_p_14—14535359—14538847	gamma_env—scaffold_m29_p_14—14538858—14540714
Beta_gag—scaffold_m29_p_12—83509921—83509199	Beta_pol—scaffold_m29_p_12—83508012—83506363	None
Beta_gag—scaffold_m29_p_25—40504104—40502893	Beta_pol—scaffold_m29_p_25—40502141—40499550	None
None	Gamma_pol—scaffold_m29_p_14—27394654—27397155	gamma_env—scaffold_m29_p_14—27398154—27400010
Beta_gag—scaffold_m29_p_9—82588300—82587200	None	beta_env—scaffold_m29_p_9—82585960—82584362
Beta_gag—scaffold_m29_p_12—1313361—1315361	Beta_pol—scaffold_m29_p_12—1316288—1317700	beta_env—scaffold_m29_p_12—1318769—1319404
None	Gamma_pol—scaffold_m29_p_27—5270643—5274131	gamma_env—scaffold_m29_p_27—5274142—5275998
Beta_gag—scaffold_m29_p_7—34902298—34904253	None	beta_env—scaffold_m29_p_7—34905592—34907091
Beta_gag—scaffold_m29_p_19—17693275—17695059	Beta_pol—scaffold_m29_p_19—17696754—17698403	beta_env—scaffold_m29_p_19—17698489—17700162
Beta_gag—scaffold_m29_p_4—45686571—45684571	Beta_pol—scaffold_m29_p_4—45683089—45681230	beta_env—scaffold_m29_p_4—45681144—45679465
Beta_gag—scaffold_m29_p_19—11646287—11648281	Beta_pol—scaffold_m29_p_19—11649982—11651235	beta_env—scaffold_m29_p_19—11651321—11653000
Beta_gag—scaffold_m29_p_10—41810855—41809785	Beta_pol—scaffold_m29_p_10—41808193—41806724	None
Beta_gag—scaffold_m29_p_31—3751249—3752565	Gamma_pol—scaffold_m29_p_31—3746155—3747882	gamma_env—scaffold_m29_p_31—3755995—3756816
Beta_gag—scaffold_m29_p_6—54551027—54552334	Beta_pol—scaffold_m29_p_6—54553074—54555671	None
Beta_gag—scaffold_m29_p_12—56595204—56593882	Beta_pol—scaffold_m29_p_12—56593142—56590545	None
Beta_gag—scaffold_m29_p_1—40994640—40993714	Beta_pol—scaffold_m29_p_1—40992502—40990337	None

Table 8.9: List of all integrations found in *Rhinolophus ferrumequinum* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Rousettus aegyptiacus		
gag	**pol**	**env**
Beta_gag—scaffold_m13_p_4—47458509—47456452	None	Gamma_env—scaffold_m13_p_4—47456000—47455293
Beta_gag—scaffold_m13_p_11—95230531—95232588	Beta_pol—scaffold_m13_p_11—95233343—95235979	None
Beta_gag—scaffold_m13_p_8—83899451—83901505	Beta_pol—scaffold_m13_p_8—83902054—83904849	Gamma_env—scaffold_m13_p_8—83904951—83906858
Beta_gag—scaffold_m13_p_5—88368397—88370340	Beta_pol—scaffold_m13_p_5—88371204—88373825	None
Beta_gag—scaffold_m13_p_9—82855519—82857156	None	Gamma_env—scaffold_m13_p_9—82857167—82858549
Beta_gag—scaffold_m13_p_9—10974394—10972340	Beta_pol—scaffold_m13_p_9—10971585—10968997	Gamma_env—scaffold_m13_p_9—10968895—10966988
None	Beta_pol—scaffold_m13_p_2—164974108—164975742	Gamma_env—scaffold_m13_p_2—164976115—164977290
Beta_gag—scaffold_m13_p_4—149373520—149375577	Beta_pol—scaffold_m13_p_4—149376332—149378968	None
Beta_gag—scaffold_m13_p_19—1077538—1079595	Beta_pol—scaffold_m13_p_19—1080350—1082971	Gamma_env—scaffold_m13_p_19—1083040—1084956
Beta_gag—scaffold_m13_p_17—22889927—22891972	Beta_pol—scaffold_m13_p_17—22892727—22895363	None
Beta_gag—scaffold_m13_p_1—169767650—169765596	Beta_pol—scaffold_m13_p_1—169764841—169762313	Gamma_env—scaffold_m13_p_1—169761345—169760242
Beta_gag—scaffold_m13_p_11—44419597—44421651	Beta_pol—scaffold_m13_p_11—44422406—44424994	Gamma_env—scaffold_m13_p_11—44425096—44427003
Beta_gag—scaffold_m13_p_7—111441134—111443191	Beta_pol—scaffold_m13_p_7—111443946—111446240	None
Beta_gag—scaffold_m13_p_1—173310304—173312361	Beta_pol—scaffold_m13_p_1—173313116—173314816	None
Beta_gag—scaffold_m13_p_9—81685542—81687620	Beta_pol—scaffold_m13_p_9—81688399—81690939	Gamma_env—scaffold_m13_p_9—81691317—81693203
Beta_gag—scaffold_m13_p_8—37098961—37101018	Beta_pol—scaffold_m13_p_8—37102125—37104395	Gamma_env—scaffold_m13_p_8—37104464—37106380
Beta_gag—scaffold_m13_p_2—135183998—135181941	None	Gamma_env—scaffold_m13_p_2—135181200—135180334
None	Beta_pol—scaffold_m13_p_1—182218940—182220136	Gamma_env—scaffold_m13_p_1—182220881—182222779
Beta_gag—scaffold_m13_p_16—31939543—31941597	Beta_pol—scaffold_m13_p_16—31942352—31944940	Gamma_env—scaffold_m13_p_16—31945042—31946949
None	Beta_pol—scaffold_m13_p_1—180437359—180439947	Gamma_env—scaffold_m13_p_1—180440049—180441959
Beta_gag—scaffold_m13_p_7—111747748—111749841	Beta_pol—scaffold_m13_p_7—111750620—111753166	Gamma_env—scaffold_m13_p_7—111753544—111755430
Beta_gag—scaffold_m13_p_9—26430008—26430817	None	Gamma_env—scaffold_m13_p_9—26432827—26434611
Beta_gag—scaffold_m13_p_13—12461471—12463525	Beta_pol—scaffold_m13_p_13—12464280—12466868	None
Beta_gag—scaffold_m13_p_9—84797595—84799652	Beta_pol—scaffold_m13_p_9—84800538—84802589	None
None	Beta_pol—scaffold_m13_p_17—23271086—23268936	beta_env—scaffold_m13_p_17—23268869—23267196
None	Beta_pol—scaffold_m13_p_4—107372859—107370235	Gamma_env—scaffold_m13_p_4—107369950—107368058
Beta_gag—scaffold_m13_p_7—66056687—66057553	None	Gamma_env—scaffold_m13_p_7—66061189—66063063
Beta_gag—scaffold_m13_p_4—4214583—4216637	Beta_pol—scaffold_m13_p_4—4217392—4219980	Gamma_env—scaffold_m13_p_4—4220082—4221989
Beta_gag—scaffold_m13_p_14—47191230—47189173	Beta_pol—scaffold_m13_p_14—47188418—47185797	Gamma_env—scaffold_m13_p_14—47185728—47183812
Beta_gag—scaffold_m13_p_11—36391349—36389694	None	Gamma_env—scaffold_m13_p_11—36387701—36386175
Beta_gag—scaffold_m13_p_9—87429573—87431630	None	Gamma_env—scaffold_m13_p_9—87433236—87434120
Beta_gag—scaffold_m13_p_3—169180916—169182973	Beta_pol—scaffold_m13_p_3—169183728—169186364	None
Beta_gag—scaffold_m13_p_2—41769836—41767875	Beta_pol—scaffold_m13_p_2—41767120—41764499	Gamma_env—scaffold_m13_p_2—41764430—41762523
Beta_gag—scaffold_m13_p_2—165731312—165733369	Beta_pol—scaffold_m13_p_2—165734124—165736760	None
Beta_gag—scaffold_m13_p_2—5436362—5438329	Beta_pol—scaffold_m13_p_2—5439084—5441705	Gamma_env—scaffold_m13_p_2—5441774—5443690
Beta_gag—scaffold_m13_p_4—114852347—114850293	Beta_pol—scaffold_m13_p_4—114849538—114846950	Gamma_env—scaffold_m13_p_4—114846848—114844941
None	Beta_pol—scaffold_m13_p_1—171413759—171412068	Gamma_env—scaffold_m13_p_1—171411819—171409921
Beta_gag—scaffold_m13_p_6—63542550—63544607	Beta_pol—scaffold_m13_p_6—63545362—63547998	None
Beta_gag—scaffold_m13_p_9—88903906—88902821	None	Gamma_env—scaffold_m13_p_9—88902091—88901267
Beta_gag—scaffold_m13_p_2—80409874—80411529	Beta_pol—scaffold_m13_p_2—80411743—80413038	None

Table 8.10: List of all integrations found in *Roussetus aegyptiacus* containing two or more genes with name of the viral family, scaffold, and position on the scaffold

Rousettus aegyptiacus		
gag	**pol**	**env**
None	Beta_pol—scaffold_m13_p_2—10427358—10429886	Gamma_env—scaffold_m13_p_2—10430688—10431872
Beta_gag—scaffold_m13_p_2—50767584—50769638	Beta_pol—scaffold_m13_p_2—50771523—50772980	Gamma_env—scaffold_m13_p_2—50773082—50774989
Beta_gag—scaffold_m13_p_7—62199343—62197286	Beta_pol—scaffold_m13_p_7—62196737—62193909	Gamma_env—scaffold_m13_p_7—62193840—62191924
Beta_gag—scaffold_m13_p_11—81519693—81521747	Beta_pol—scaffold_m13_p_11—81522502—81525090	None
Beta_gag—scaffold_m13_p_10—84689494—84691548	Beta_pol—scaffold_m13_p_10—84692303—84694891	None
Beta_gag—scaffold_m13_p_3—103046634—103048688	Beta_pol—scaffold_m13_p_3—103049443—103051143	None
Beta_gag—scaffold_m13_p_9—26144978—26142921	None	Gamma_env—scaffold_m13_p_9—26142381—26141497
None	Beta_pol—scaffold_m13_p_9—12909068—12910759	Gamma_env—scaffold_m13_p_9—12911008—12912906
Beta_gag—scaffold_m13_p_11—14692166—14690511	None	Gamma_env—scaffold_m13_p_11—14688684—14687806
None	Gamma_pol—scaffold_m13_p_22—156401—158926	Gamma_env—scaffold_m13_p_22—159507—160847
Beta_gag—scaffold_m13_p_7—59717473—59716217	None	Gamma_env—scaffold_m13_p_7—59715226—59714567
Beta_gag—scaffold_m13_p_9—69711958—69714012	Beta_pol—scaffold_m13_p_9—69714767—69717355	None
None	Beta_pol—scaffold_m13_p_11—49193196—49194803	Gamma_env—scaffold_m13_p_11—49195052—49196950
Beta_gag—scaffold_m13_p_12—42012112—42014169	Beta_pol—scaffold_m13_p_12—42015589—42017544	None
Beta_gag—scaffold_m13_p_2—27961708—27963765	None	Gamma_env—scaffold_m13_p_2—27964067—27965929
Beta_gag—scaffold_m13_p_11—86433545—86435602	None	Gamma_env—scaffold_m13_p_11—86435904—86437766
Beta_gag—scaffold_m13_p_19—147095—149248	None	Gamma_env—scaffold_m13_p_19—152651—154429
Beta_gag—scaffold_m13_p_5—83642024—83644078	Beta_pol—scaffold_m13_p_5—83644833—83647421	None
Beta_gag—scaffold_m13_p_8—101975689—101977746	Beta_pol—scaffold_m13_p_8—101978501—101981137	None
Beta_gag—scaffold_m13_p_7—55452388—55450832	Beta_pol—scaffold_m13_p_7—55449333—55447063	Gamma_env—scaffold_m13_p_7—55446994—55445078
Beta_gag—scaffold_m13_p_2—176351122—176349068	Beta_pol—scaffold_m13_p_2—176348313—176345725	Gamma_env—scaffold_m13_p_2—176345623—176343716
None	Beta_pol—scaffold_m13_p_11—60663760—60665451	Gamma_env—scaffold_m13_p_11—60665700—60667598
Beta_gag—scaffold_m13_p_3—107696793—107697809	Beta_pol—scaffold_m13_p_3—107699593—107701347	None
Beta_gag—scaffold_m13_p_5—56746161—56747132	Beta_pol—scaffold_m13_p_5—56748784—56750388	None
Beta_gag—scaffold_m13_p_2—72987059—72988789	None	Gamma_env—scaffold_m13_p_2—72988921—72990906
Beta_gag—scaffold_m13_p_5—66970422—66972317	Beta_pol—scaffold_m13_p_5—66973072—66975708	None
Beta_gag—scaffold_m13_p_1—88544010—88545536	Beta_pol—scaffold_m13_p_1—88546291—88548912	Gamma_env—scaffold_m13_p_1—88548981—88550897
Beta_gag—scaffold_m13_p_11—6016062—6014008	None	Gamma_env—scaffold_m13_p_11—6013454—6011562
Beta_gag—scaffold_m13_p_3—117930975—117933032	Beta_pol—scaffold_m13_p_3—117933787—117936408	None
Beta_gag—scaffold_m13_p_9—51273649—51272480	Beta_pol—scaffold_m13_p_9—51271172—51269586	None
Beta_gag—scaffold_m13_p_8—48246523—48244682	Beta_pol—scaffold_m13_p_8—48243927—48241339	Gamma_env—scaffold_m13_p_8—48241237—48239330
Beta_gag—scaffold_m13_p_9—65367917—65366829	Beta_pol—scaffold_m13_p_9—65366074—65363438	None
Beta_gag—scaffold_m13_p_11—14673647—14671584	Beta_pol—scaffold_m13_p_11—14669753—14668242	Gamma_env—scaffold_m13_p_11—14667882—14665999
Beta_gag—scaffold_m13_p_9—78688216—78689025	Beta_pol—scaffold_m13_p_9—78691064—78692380	None
None	Beta_pol—scaffold_m13_p_8—109824152—109825840	Gamma_env—scaffold_m13_p_8—109826492—109827814
Beta_gag—scaffold_m13_p_8—109494113—109496167	Beta_pol—scaffold_m13_p_8—109496922—109499510	Gamma_env—scaffold_m13_p_8—109499612—109501333
Beta_gag—scaffold_m13_p_20—3—1532	Beta_pol—scaffold_m13_p_20—3086—4690	None
Beta_gag—scaffold_m13_p_2—177602139—177600070	Beta_pol—scaffold_m13_p_2—177599075—177597738	None
None	Beta_pol—scaffold_m13_p_7—93130465—93132993	Gamma_env—scaffold_m13_p_7—93133795—93134979
Beta_gag—scaffold_m13_p_9—82751532—82753592	Beta_pol—scaffold_m13_p_9—82754347—82756968	Gamma_env—scaffold_m13_p_9—82757037—82758953

Table 8.11: Continued. List of all integrations found in Roussetus aegyptiacus containing two or more genes with name of the viral family, scaffold, and position on the scaffold

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_9 98631276 98630401	565	NC_001550.1	Mason-Pfizer monkey virus
2	scaffold_m16_p_3 118530351 118529326	877	NC_001503.1	Mouse mammary tumor virus
3	scaffold_m16_p_10 52340864 52338828	680	NC_001550.1	Mason-Pfizer monkey virus
4	scaffold_m16_p_15 67387293 67385554	581	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
5	scaffold_m16_p_7 69646590 69647669	658	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
6	scaffold_m16_p_11 70578795 70579532	468	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
7	scaffold_m16_p_10 75551790 75552701	305	NC_001550.1	Mason-Pfizer monkey virus
8	scaffold_m16_p_7 75696541 75695462	660	NC_001550.1	Mason-Pfizer monkey virus
9	scaffold_m16_p_15 79656406 79654373	679	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
10	scaffold_m16_p_15 75566485 75567372	297	NC_001550.1	Mason-Pfizer monkey virus
11	scaffold_m16_p_11 100239175 100238099	360	NC_001550.1	Mason-Pfizer monkey virus
12	scaffold_m16_p_21 12483220 12484269	351	NC_001550.1	Mason-Pfizer monkey virus
13	scaffold_m16_p_21 14064531 14063149	462	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
14	scaffold_m16_p_17 16712729 16710687	682	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
15	scaffold_m16_p_3 50720551 50722587	680	NC_031326.1	Mason-Pfizer monkey virus
16	scaffold_m16_p_5 99319901 99321010	371	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
17	scaffold_m16_p_5 102317391 102316069	442	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
18	scaffold_m16_p_5 98392395 98393144	389	NC_001550.1	Mason-Pfizer monkey virus
19	scaffold_m16_p_9 75105835 75103805	678	NC_001550.1	Mason-Pfizer monkey virus
20	scaffold_m16_p_9 92375126 92375899	259	NC_001550.1	Mason-Pfizer monkey virus
21	scaffold_m16_p_3 132372033 132373841	860	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
22	scaffold_m16_p_5 13102695 13100719	956	NC_001550.1	Mason-Pfizer monkey virus
23	scaffold_m16_p_16 4203764 4201842	642	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
24	scaffold_m16_p_14 26430032 26431489	487	NC_001550.1	Mason-Pfizer monkey virus
25	scaffold_m16_p_14 34287940 34289046	370	NC_001550.1	Mason-Pfizer monkey virus
26	scaffold_m16_p_31 2409601 2410233	212	NC_004994.2	Enzootic nasal tumour virus of goats
27	scaffold_m16_p_23 4322225 4320777	470	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
28	scaffold_m16_p_22 982298 983569	425	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
29	scaffold_m16_p_28 2485312 2485986	226	NC_001494.1	Jaagsiekte sheep retrovirus
30	scaffold_m16_p_25 3311540 3313225	563	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
31	scaffold_m16_p_30 348232 349521	431	NC_001550.1	Mason-Pfizer monkey virus
32	scaffold_m16_p_30 455680 453908	592	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
33	scaffold_m16_p_30 490030 491643	539	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
34	scaffold_m16_p_38 993549 994286	247	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
35	scaffold_m16_p_5 57761201 57762220	341	NC_001550.1	Mason-Pfizer monkey virus
36	scaffold_m16_p_5 65342261 65344297	680	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
37	scaffold_m16_p_5 51576672 51575641	944	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
38	scaffold_m16_p_9 35071985 35071110	565	NC_001550.1	Mason-Pfizer monkey virus
39	scaffold_m16_p_9 35537645 35536566	983	NC_001550.1	Mason-Pfizer monkey virus
40	scaffold_m16_p_19 41066026 41064512	506	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
41	scaffold_m16_p_6 50199714 50200658	316	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
42	scaffold_m16_p_8 112790634 112788598	680	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
43	scaffold_m16_p_6 57782360 57781335	656	NC_001550.1	Mason-Pfizer monkey virus
44	scaffold_m16_p_20 22742247 22742978	245	NC_001550.1	Mason-Pfizer monkey virus
45	scaffold_m16_p_8 17474099 17472063	680	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012

Table 8.12: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
46	scaffold_m16_p_7 61763989 61763174	450	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
47	scaffold_m16_p_15 69407883 69408686	750	NC_001550.1	Mason-Pfizer monkey virus
48	scaffold_m16_p_13 19594833 19596869	680	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
49	scaffold_m16_p_13 23063532 23064626	919	NC_001503.1	Mouse mammary tumor virus
50	scaffold_m16_p_2 47219327 47220430	666	NC_001550.1	Mason-Pfizer monkey virus
51	scaffold_m16_p_2 47291941 47290019	642	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
52	scaffold_m16_p_4 52356023 52358077	686	NC_001550.1	Mason-Pfizer monkey virus
53	scaffold_m16_p_2 83868733 83866808	643	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
54	scaffold_m16_p_2 101781624 101780035	531	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
55	scaffold_m16_p_8 81473756 81471816	616	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
56	scaffold_m16_p_12 7547736 7546861	565	NC_001550.1	Mason-Pfizer monkey virus
57	scaffold_m16_p_27 3512328 3511219	659	NC_001550.1	Mason-Pfizer monkey virus
58	scaffold_m16_p_39 70415 68745	558	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
59	scaffold_m16_p_39 361001 361807	422	NC_001503.1	Mouse mammary tumor virus
60	scaffold_m16_p_34 13793 14902	668	NC_001550.1	Mason-Pfizer monkey virus
61	scaffold_m16_p_1 123907524 123908597	1039	NC_001550.1	Mason-Pfizer monkey virus
62	scaffold_m16_p_1 126357268 126356261	337	NC_001550.1	Mason-Pfizer monkey virus
63	scaffold_m16_p_1 214533803 214534912	668	NC_001550.1	Mason-Pfizer monkey virus
64	scaffold_m16_p_20 76456 77253	413	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
65	scaffold_m16_p_20 161461 162432	940	NC_022518.1	Human endogenous retrovirus K113
66	scaffold_m16_p_20 1416483 1417109	342	NC_001503.1	Mouse mammary tumor virus
67	scaffold_m16_p_18 186965 188920	653	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
68	scaffold_m16_p_2 70331044 70331802	730	NC_001550.1	Mason-Pfizer monkey virus
69	scaffold_m16_p_2 34148286 34147456	471	NC_001503.1	Mouse mammary tumor virus
70	scaffold_m16_p_4 52844582 52845658	360	NC_001550.1	Mason-Pfizer monkey virus
71	scaffold_m16_p_1 159621654 159622466	754	NC_001550.1	Mason-Pfizer monkey virus
72	scaffold_m16_p_1 18287422 18286385	624	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
73	scaffold_m16_p_2 48736931 48737875	316	NC_022518.1	Human endogenous retrovirus K113
74	scaffold_m16_p_2 59663508 59661991	507	NC_001550.1	Mason-Pfizer monkey virus
75	scaffold_m16_p_2 65293211 65291940	425	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
76	scaffold_m16_p_2 78503225 78503932	451	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
77	scaffold_m16_p_2 83875263 83873338	643	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
78	scaffold_m16_p_2 90599563 90598760	509	NC_001550.1	Mason-Pfizer monkey virus
79	scaffold_m16_p_1 122677274 122676021	770	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
80	scaffold_m16_p_1 123955676 123954228	484	NC_001550.1	Mason-Pfizer monkey virus
81	scaffold_m16_p_1 33128289 33127420	531	NC_001550.1	Mason-Pfizer monkey virus

Table 8.12: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_13 23066030 23067343	788	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m16_p_12 3385426 3386751	838	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m16_p_13 32900148 32901827	994	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m16_p_1 102149966 102148416	1333	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m16_p_1 128907159 128909084	1528	NC_022518.1	Human endogenous retrovirus K113
6	scaffold_m16_p_1 238967663 238968787	372	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m16_p_1 49847473 49845818	1030	NC_022518.1	Human endogenous retrovirus K113
8	scaffold_m16_p_1 109883882 109886830	678	NC_022518.1	Human endogenous retrovirus K113
9	scaffold_m16_p_12 258672 260351	1041	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m16_p_10 23674928 23678074	1309	NC_022518.1	Human endogenous retrovirus K113
11	scaffold_m16_p_54 5601 5335	90	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
12	scaffold_m16_p_9 94380859 94379429	465	NC_022518.1	Human endogenous retrovirus K113
13	scaffold_m16_p_29 2894993 2893899	353	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
14	scaffold_m16_p_6 28910250 28908598	538	NC_022518.1	Human endogenous retrovirus K113
15	scaffold_m16_p_2 81595136 81593088	1769	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
16	scaffold_m16_p_2 111106692 111104218	799	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
17	scaffold_m16_p_31 2412156 2413688	500	NC_001494.1	Jaagsiekte sheep retrovirus
18	scaffold_m16_p_1 122368061 122369596	505	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
19	scaffold_m16_p_30 117997 120063	1281	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
20	scaffold_m16_p_5 99322037 99324325	1514	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
21	scaffold_m16_p_4 71823820 71826420	1488	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
22	scaffold_m16_p_9 42409896 42411587	932	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
23	scaffold_m16_p_15 67384821 67382245	858	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
24	scaffold_m16_p_4 52358817 52361396	859	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
25	scaffold_m16_p_8 17471323 17468735	863	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
26	scaffold_m16_p_6 50201398 50203752	786	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
27	scaffold_m16_p_21 14062353 14059741	829	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
28	scaffold_m16_p_13 19597609 19600188	861	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
29	scaffold_m16_p_3 50723327 50725903	860	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
30	scaffold_m16_p_8 112787857 112785278	861	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
31	scaffold_m16_p_3 132374049 132376355	1480	NC_001550.1	Mason-Pfizer monkey virus
32	scaffold_m16_p_9 30654488 30653094	814	NC_001550.1	Mason-Pfizer monkey virus
33	scaffold_m16_p_30 961750 960365	1276	NC_001550.1	Mason-Pfizer monkey virus
34	scaffold_m16_p_52 45876 44251	920	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
35	scaffold_m16_p_5 51574388 51572745	1022	NC_001550.1	Mason-Pfizer monkey virus
36	scaffold_m16_p_12 65386429 65384606	605	NC_001550.1	Mason-Pfizer monkey virus
37	scaffold_m16_p_2 110606024 110608432	1549	NC_001550.1	Mason-Pfizer monkey virus
38	scaffold_m16_p_52 33390 35420	1910	NC_001550.1	Mason-Pfizer monkey virus
39	scaffold_m16_p_7 59746282 59744345	1227	NC_001550.1	Mason-Pfizer monkey virus
40	scaffold_m16_p_2 42451822 42453072	414	NC_001550.1	Mason-Pfizer monkey virus

Table 8.13: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
41	scaffold_m16_p_2 59661245 59658657	861	NC_001550.1	Mason-Pfizer monkey virus
42	scaffold_m16_p_40 395646 396521	289	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
43	scaffold_m16_p_40 396942 397196	86	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
44	scaffold_m16_p_10 75553757 75555697	1387	NC_001550.1	Mason-Pfizer monkey virus
45	scaffold_m16_p_13 49457376 49459148	591	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
46	scaffold_m16_p_13 49438825 49440597	591	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
47	scaffold_m16_p_19 41063767 41061074	2395	NC_001550.1	Mason-Pfizer monkey virus
48	scaffold_m16_p_5 13099920 13097557	789	NC_001550.1	Mason-Pfizer monkey virus
49	scaffold_m16_p_14 27128705 27127089	1061	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
50	scaffold_m16_p_34 15655 16968	863	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
51	scaffold_m16_p_12 7546613 7544319	761	NC_001550.1	Mason-Pfizer monkey virus
52	scaffold_m16_p_2 70332905 70334344	1358	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
53	scaffold_m16_p_16 4199978 4198578	916	NC_001550.1	Mason-Pfizer monkey virus
54	scaffold_m16_p_21 4565248 4564037	405	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
55	scaffold_m16_p_30 492646 495294	1709	NC_001550.1	Mason-Pfizer monkey virus
56	scaffold_m16_p_30 452906 450258	1700	NC_001550.1	Mason-Pfizer monkey virus
57	scaffold_m16_p_20 29192678 29194009	1227	NC_001550.1	Mason-Pfizer monkey virus
58	scaffold_m16_p_7 75195304 75193637	557	NC_001550.1	Mason-Pfizer monkey virus
59	scaffold_m16_p_2 47282163 47280571	1007	NC_001550.1	Mason-Pfizer monkey virus
60	scaffold_m16_p_11 105434860 105432257	1707	NC_001550.1	Mason-Pfizer monkey virus
61	scaffold_m16_p_9 110638033 110640507	1419	NC_001550.1	Mason-Pfizer monkey virus
62	scaffold_m16_p_2 83879091 83876752	1549	NC_001550.1	Mason-Pfizer monkey virus
63	scaffold_m16_p_11 50654307 50651842	2958	NC_001550.1	Mason-Pfizer monkey virus
64	scaffold_m16_p_14 34289792 34292440	1715	NC_001550.1	Mason-Pfizer monkey virus
65	scaffold_m16_p_2 83872592 83869944	1704	NC_001550.1	Mason-Pfizer monkey virus
66	scaffold_m16_p_9 98630434 98628194	2205	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
67	scaffold_m16_p_9 75103059 75100411	1704	NC_001550.1	Mason-Pfizer monkey virus
68	scaffold_m16_p_2 83865076 83863415	1032	NC_001550.1	Mason-Pfizer monkey virus
69	scaffold_m16_p_18 189667 192360	1697	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
70	scaffold_m16_p_6 52264364 52261716	1708	NC_001550.1	Mason-Pfizer monkey virus
71	scaffold_m16_p_23 4320128 4319832	100	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.13: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_15 80205191 80205925	457	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
2	scaffold_m16_p_6 48242098 48241511	197	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
3	scaffold_m16_p_7 73331556 73332134	194	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
4	scaffold_m16_p_4 54671173 54670571	202	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
5	scaffold_m16_p_4 51036650 51037423	259	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m16_p_5 90856922 90857524	202	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
7	scaffold_m16_p_1 84305307 84304663	216	NC_001514.1	Squirrel monkey retrovirus - HLB
8	scaffold_m16_p_4 52362328 52362978	218	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
9	scaffold_m16_p_15 67381443 67380661	262	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
10	scaffold_m16_p_21 14616896 14618551	553	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m16_p_30 121275 122381	370	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
12	scaffold_m16_p_9 42411643 42413319	560	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m16_p_4 71826527 71828203	560	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
14	scaffold_m16_p_2 110609207 110610187	328	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m16_p_3 119932906 119934039	379	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
16	scaffold_m16_p_12 65384007 65382412	533	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m16_p_2 68370255 68371121	290	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
18	scaffold_m16_p_1 101480965 101480195	258	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
19	scaffold_m16_p_17 57016161 57015487	226	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
20	scaffold_m16_p_15 80917449 80916889	188	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
21	scaffold_m16_p_21 2256243 2255347	300	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
22	scaffold_m16_p_13 88149618 88150283	223	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
23	scaffold_m16_p_15 80181529 80180756	259	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
24	scaffold_m16_p_4 55869197 55870066	291	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
25	scaffold_m16_p_7 81013777 81013052	243	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
26	scaffold_m16_p_1 120884247 120885014	257	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
27	scaffold_m16_p_2 74592492 74591884	204	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
28	scaffold_m16_p_2 34529498 34528884	206	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
29	scaffold_m16_p_11 48446769 48446029	248	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
30	scaffold_m16_p_4 89791081 89792001	308	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
31	scaffold_m16_p_4 48339843 48340640	267	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
32	scaffold_m16_p_3 119644654 119645223	191	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
33	scaffold_m16_p_4 50210198 50209479	241	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
34	scaffold_m16_p_5 100049752 100049084	224	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
35	scaffold_m16_p_2 20351158 20350337	275	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
36	scaffold_m16_p_3 86150668 86149964	236	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.14: Reciprocal BLAST hits for the betaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_6 48407391 48406051	687	NC_001885.3	Gibbon ape leukemia virus
2	scaffold_m16_p_55 28719 27637	362	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
3	scaffold_m16_p_24 13315102 13314401	235	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
4	scaffold_m16_p_46 82887 81916	325	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
5	scaffold_m16_p_32 506139 507071	312	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
6	scaffold_m16_p_45 167768 167136	212	NC_003059.1	Porcine endogenous retrovirus E
7	scaffold_m16_p_4 15664371 15663547	276	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
8	scaffold_m16_p_1 122757054 122758424	816	NC_001885.3	Gibbon ape leukemia virus

Table 8.15: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_13 50562164 50559261	969	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m16_p_12 9648696 9645619	1787	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
3	scaffold_m16_p_1 67212458 67208865	3379	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
4	scaffold_m16_p_1 128018712 128015842	1806	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
5	scaffold_m16_p_1 186741934 186745527	2283	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
6	scaffold_m16_p_1 123066929 123063420	2199	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
7	scaffold_m16_p_1 42703620 42707033	2094	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
8	scaffold_m16_p_1 171535989 171538067	2044	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
9	scaffold_m16_p_23 2928839 2932193	3355	NC_001362.1	Friend murine leukemia virus FB29

Table 8.16: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_5 31445463 31444771	363	NC_022517.1	Baboon endogenous virus strain M7
2	scaffold_m16_p_2 117980870 117981433	189	NC_009889.1	RD114 retrovirus
3	scaffold_m16_p_30 2103284 2102553	245	NC_022517.1	Baboon endogenous virus strain M7
4	scaffold_m16_p_1 121367674 121368258	196	NC_022517.1	Baboon endogenous virus strain M7
5	scaffold_m16_p_8 3171541 3172410	291	NC_009889.1	RD114 retrovirus
6	scaffold_m16_p_4 51146537 51147400	289	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
7	scaffold_m16_p_2 80919690 80919094	200	NC_022517.1	Baboon endogenous virus strain M7
8	scaffold_m16_p_2 107301672 107302268	200	NC_009889.1	RD114 retrovirus
9	scaffold_m16_p_14 70089253 70088567	230	NC_022517.1	Baboon endogenous virus strain M7
10	scaffold_m16_p_2 37088697 37087828	291	NC_022517.1	Baboon endogenous virus strain M7
11	scaffold_m16_p_9 99150781 99150005	260	NC_022517.1	Baboon endogenous virus strain M7
12	scaffold_m16_p_1 133939797 133940384	197	NC_022517.1	Baboon endogenous virus strain M7
13	scaffold_m16_p_9 96931113 96930526	197	NC_009889.1	RD114 retrovirus
14	scaffold_m16_p_18 6159003 6159590	197	NC_022517.1	Baboon endogenous virus strain M7
15	scaffold_m16_p_8 2366766 2365897	291	NC_022517.1	Baboon endogenous virus strain M7
16	scaffold_m16_p_6 61174294 61175139	283	NC_022517.1	Baboon endogenous virus strain M7
17	scaffold_m16_p_4 93941582 93940716	290	NC_022517.1	Baboon endogenous virus strain M7
18	scaffold_m16_p_2 87877416 87876829	197	NC_022517.1	Baboon endogenous virus strain M7
19	scaffold_m16_p_1 9651775 9652362	197	NC_022517.1	Baboon endogenous virus strain M7
20	scaffold_m16_p_28 2243616 2243029	197	NC_022517.1	Baboon endogenous virus strain M7
21	scaffold_m16_p_1 109948284 109947415	291	NC_022517.1	Baboon endogenous virus strain M7
22	scaffold_m16_p_1 48683328 48682552	260	NC_022517.1	Baboon endogenous virus strain M7
23	scaffold_m16_p_2 27661598 27660729	291	NC_022517.1	Baboon endogenous virus strain M7
24	scaffold_m16_p_12 86890195 86890971	260	NC_022517.1	Baboon endogenous virus strain M7
25	scaffold_m16_p_5 109396299 109397168	291	NC_022517.1	Baboon endogenous virus strain M7
26	scaffold_m16_p_1 137200844 137201713	291	NC_022517.1	Baboon endogenous virus strain M7
27	scaffold_m16_p_6 59058469 59057600	291	NC_022517.1	Baboon endogenous virus strain M7
28	scaffold_m16_p_10 39991775 39990906	291	NC_022517.1	Baboon endogenous virus strain M7
29	scaffold_m16_p_2 82486744 82486145	201	NC_022517.1	Baboon endogenous virus strain M7
30	scaffold_m16_p_15 75687283 75688152	291	NC_022517.1	Baboon endogenous virus strain M7
31	scaffold_m16_p_6 68787034 68786447	197	NC_022517.1	Baboon endogenous virus strain M7
32	scaffold_m16_p_20 10547027 10547764	247	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
33	scaffold_m16_p_4 73374431 73375267	280	NC_022517.1	Baboon endogenous virus strain M7
34	scaffold_m16_p_8 4267985 4267152	279	NC_022517.1	Baboon endogenous virus strain M7
35	scaffold_m16_p_22 4159885 4159016	291	NC_022517.1	Baboon endogenous virus strain M7
36	scaffold_m16_p_14 70777391 70777978	197	NC_022517.1	Baboon endogenous virus strain M7
37	scaffold_m16_p_6 47577968 47578582	206	NC_009889.1	RD114 retrovirus
38	scaffold_m16_p_1 180203773 180204876	369	NC_009889.1	RD114 retrovirus
39	scaffold_m16_p_2 120185665 120186408	249	NC_009889.1	RD114 retrovirus
40	scaffold_m16_p_2 120815997 120815179	274	NC_009889.1	RD114 retrovirus
41	scaffold_m16_p_2 66428409 66429221	272	NC_009889.1	RD114 retrovirus
42	scaffold_m16_p_2 71626580 71625729	285	NC_009889.1	RD114 retrovirus
43	scaffold_m16_p_2 75241773 75242408	213	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
44	scaffold_m16_p_1 170410082 170409276	270	NC_009889.1	RD114 retrovirus
45	scaffold_m16_p_4 60770779 60771657	294	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
46	scaffold_m16_p_1 24609487 24608852	213	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
47	scaffold_m16_p_7 78716733 78716122	205	NC_022517.1	Baboon endogenous virus strain M7
48	scaffold_m16_p_7 69619853 69621781	971	NC_001501.1	Moloney murine leukemia virus
49	scaffold_m16_p_13 57642713 57641859	507	NC_001885.3	Gibbon ape leukemia virus
50	scaffold_m16_p_2 74750491 74749619	292	NC_001885.3	Gibbon ape leukemia virus

Table 8.17: Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m16_p_7 66345974 66345396	194	NC_001501.1	Moloney murine leukemia virus
52	scaffold_m16_p_30 2675154 2675978	276	NC_039228.1	Koala retrovirus
53	scaffold_m16_p_12 8999937 9000827	298	NC_001885.3	Gibbon ape leukemia virus
54	scaffold_m16_p_7 80110930 80111847	307	NC_001501.1	Moloney murine leukemia virus
55	scaffold_m16_p_2 23803 24465	222	NC_039228.1	Koala retrovirus
56	scaffold_m16_p_3 94984745 94983999	250	NC_039228.1	Koala retrovirus
57	scaffold_m16_p_7 43666941 43666216	243	NC_039228.1	Koala retrovirus
58	scaffold_m16_p_8 55011773 55012567	266	NC_001501.1	Moloney murine leukemia virus
59	scaffold_m16_p_27 775163 774042	375	NC_001362.1	Friend murine leukemia virus FB29
60	scaffold_m16_p_22 16576153 16576920	257	NC_001885.3	Gibbon ape leukemia virus
61	scaffold_m16_p_4 51404002 51402998	336	NC_001501.1	Moloney murine leukemia virus
62	scaffold_m16_p_6 56661947 56662918	325	NC_039228.1	Koala retrovirus
63	scaffold_m16_p_4 53997320 53998210	298	NC_039228.1	Koala retrovirus
64	scaffold_m16_p_7 29823220 29822267	319	NC_001362.1	Friend murine leukemia virus FB29
65	scaffold_m16_p_2 65719381 65720253	292	NC_039228.1	Koala retrovirus
66	scaffold_m16_p_4 54023331 54022105	410	NC_001501.1	Moloney murine leukemia virus
67	scaffold_m16_p_5 81181565 81180528	347	NC_039228.1	Koala retrovirus
68	scaffold_m16_p_3 79784788 79786005	407	NC_001501.1	Moloney murine leukemia virus
69	scaffold_m16_p_16 30215825 30216568	249	NC_039228.1	Koala retrovirus
70	scaffold_m16_p_2 81461306 81462535	411	NC_001362.1	Friend murine leukemia virus FB29
71	scaffold_m16_p_30 391931 393097	390	NC_039228.1	Koala retrovirus
72	scaffold_m16_p_30 553562 552747	273	NC_039228.1	Koala retrovirus
73	scaffold_m16_p_2 58550716 58551921	403	NC_001362.1	Friend murine leukemia virus FB29
74	scaffold_m16_p_30 955214 954036	394	NC_001362.1	Friend murine leukemia virus FB29
75	scaffold_m16_p_5 88963720 88964802	362	NC_039228.1	Koala retrovirus
76	scaffold_m16_p_16 46400692 46399949	249	NC_039228.1	Koala retrovirus
77	scaffold_m16_p_12 9645466 9644423	349	NC_001362.1	Friend murine leukemia virus FB29
78	scaffold_m16_p_14 45607008 45608219	405	NC_001362.1	Friend murine leukemia virus FB29
79	scaffold_m16_p_8 60478195 60477017	394	NC_001362.1	Friend murine leukemia virus FB29
80	scaffold_m16_p_18 26271684 26272652	324	NC_039228.1	Koala retrovirus
81	scaffold_m16_p_16 5376511 5375372	381	NC_039228.1	Koala retrovirus
82	scaffold_m16_p_1 125841272 125842333	355	NC_001362.1	Friend murine leukemia virus FB29
83	scaffold_m16_p_1 128015370 128014627	249	NC_001501.1	Moloney murine leukemia virus
84	scaffold_m16_p_5 100287700 100286804	300	NC_001362.1	Friend murine leukemia virus FB29
85	scaffold_m16_p_1 83731147 83731890	249	NC_039228.1	Koala retrovirus
86	scaffold_m16_p_2 76676231 76677409	394	NC_001501.1	Moloney murine leukemia virus
87	scaffold_m16_p_6 48090063 48088885	394	NC_001362.1	Friend murine leukemia virus FB29
88	scaffold_m16_p_6 65619823 65618636	397	NC_001501.1	Moloney murine leukemia virus
89	scaffold_m16_p_5 108176215 108175091	376	NC_001362.1	Friend murine leukemia virus FB29
90	scaffold_m16_p_8 4309922 4308846	360	NC_001362.1	Friend murine leukemia virus FB29
91	scaffold_m16_p_3 72319691 72318462	411	NC_001501.1	Moloney murine leukemia virus
92	scaffold_m16_p_11 14061258 14062487	411	NC_001501.1	Moloney murine leukemia virus
93	scaffold_m16_p_9 95482 96672	398	NC_001501.1	Moloney murine leukemia virus
94	scaffold_m16_p_5 93790099 93789356	249	NC_039228.1	Koala retrovirus
95	scaffold_m16_p_11 12484770 12483931	281	NC_001362.1	Friend murine leukemia virus FB29
96	scaffold_m16_p_1 220711592 220710513	361	NC_001362.1	Friend murine leukemia virus FB29
97	scaffold_m16_p_1 121081044 121079866	394	NC_039228.1	Koala retrovirus
98	scaffold_m16_p_9 16733409 16732231	394	NC_001362.1	Friend murine leukemia virus FB29
99	scaffold_m16_p_1 186746210 186747226	340	NC_001501.1	Moloney murine leukemia virus
100	scaffold_m16_p_7 7143380 7144558	394	NC_001362.1	Friend murine leukemia virus FB29

Table 8.17: Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m16_p_1 119907336 119908514	394	NC_039228.1	Koala retrovirus
102	scaffold_m16_p_17 55251447 55252625	394	NC_039228.1	Koala retrovirus
103	scaffold_m16_p_18 60790560 60791573	339	NC_039228.1	Koala retrovirus
104	scaffold_m16_p_14 41563573 41564622	351	NC_001362.1	Friend murine leukemia virus FB29
105	scaffold_m16_p_1 67208190 67206988	402	NC_001362.1	Friend murine leukemia virus FB29
106	scaffold_m16_p_32 509485 510660	393	NC_001362.1	Friend murine leukemia virus FB29
107	scaffold_m16_p_3 86331584 86330355	411	NC_001501.1	Moloney murine leukemia virus
108	scaffold_m16_p_16 20684780 20686009	411	NC_001501.1	Moloney murine leukemia virus
109	scaffold_m16_p_1 9429017 9427797	408	NC_001501.1	Moloney murine leukemia virus
110	scaffold_m16_p_2 27766147 27765110	347	NC_039228.1	Koala retrovirus
111	scaffold_m16_p_15 36848043 36846865	394	NC_039228.1	Koala retrovirus
112	scaffold_m16_p_46 77601 76372	411	NC_001362.1	Friend murine leukemia virus FB29
113	scaffold_m16_p_16 61385787 61384945	282	NC_039228.1	Koala retrovirus
114	scaffold_m16_p_1 42708426 42709118	232	NC_039228.1	Koala retrovirus
115	scaffold_m16_p_55 17853 17116	247	NC_039228.1	Koala retrovirus
116	scaffold_m16_p_8 847877 846648	411	NC_001501.1	Moloney murine leukemia virus
117	scaffold_m16_p_2 114550415 114551644	411	NC_001501.1	Moloney murine leukemia virus
118	scaffold_m16_p_7 57358211 57358957	250	NC_001501.1	Moloney murine leukemia virus
119	scaffold_m16_p_1 116352393 116353481	364	NC_001362.1	Friend murine leukemia virus FB29
120	scaffold_m16_p_7 72171259 72170420	281	NC_001362.1	Friend murine leukemia virus FB29
121	scaffold_m16_p_4 56416206 56415118	364	NC_001362.1	Friend murine leukemia virus FB29
122	scaffold_m16_p_30 171308 172375	357	NC_001362.1	Friend murine leukemia virus FB29

Table 8.17: Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Deltaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_9 8679506 8680705	400	NC_003323.1	Simian T-lymphotropic virus 3
2	scaffold_m16_p_9 8674906 8676474	523	NC_000858.1	Simian T-lymphotropic virus 1
3	scaffold_m16_p_9 8676507 8679566	1020	NC_003323.1	Simian T-lymphotropic virus 3

Table 8.18: Reciprocal BLAST hits for the deltaretrovirus Gag hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Deltaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_9 8676507 8679566	1020	NC_003323.1	Simian T-lymphotropic virus 3

Table 8.19: Reciprocal BLAST hits for the deltaretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Deltaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_9 8674906 8676474	523	NC_000858.1	Simian T-lymphotropic virus 1

Table 8.20: Reciprocal BLAST hits for the deltaretroviruses Env hits in Molossus molossus genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name

	Epsilon-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m16_p_44 43307 43736	215	NC_043194.1	Walleye epidermal hyperplasia virus 1

Table 8.21: Reciprocal BLAST hits for the epsilonretrovirus Pol hits in *Molossus molossus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_6 73481502 73480384	699	NC_004994.2	Enzootic nasal tumour virus of goats
2	scaffold_m19_p_6 94146673 94145471	693	NC_004994.2	Enzootic nasal tumour virus of goats
3	scaffold_m19_p_6 2729642 2730904	713	NC_001503.1	Mouse mammary tumor virus
4	scaffold_m19_p_13 6751968 6753230	713	NC_001503.1	Mouse mammary tumor virus
5	scaffold_m19_p_13 14319255 14320517	713	NC_001503.1	Mouse mammary tumor virus
6	scaffold_m19_p_2 27715743 27717005	713	NC_001503.1	Mouse mammary tumor virus
7	scaffold_m19_p_2 50850352 50851614	697	NC_001503.1	Mouse mammary tumor virus
8	scaffold_m19_p_2 80581035 80582297	697	NC_001503.1	Mouse mammary tumor virus
9	scaffold_m19_p_2 152165337 152166593	706	NC_001503.1	Mouse mammary tumor virus
10	scaffold_m19_p_2 163796621 163797883	713	NC_001503.1	Mouse mammary tumor virus
11	scaffold_m19_p_2 183563411 183564673	713	NC_001503.1	Mouse mammary tumor virus
12	scaffold_m19_p_2 183845311 183846576	715	NC_001503.1	Mouse mammary tumor virus
13	scaffold_m19_p_2 187695443 187696705	696	NC_001503.1	Mouse mammary tumor virus
14	scaffold_m19_p_2 202958162 202959427	715	NC_001503.1	Mouse mammary tumor virus
15	scaffold_m19_p_2 213773256 213774518	713	NC_001503.1	Mouse mammary tumor virus
16	scaffold_m19_p_2 215521444 215522709	715	NC_001503.1	Mouse mammary tumor virus
17	scaffold_m19_p_6 38168995 38170260	697	NC_001503.1	Mouse mammary tumor virus
18	scaffold_m19_p_6 96713361 96714383	637	NC_001503.1	Mouse mammary tumor virus
19	scaffold_m19_p_11 77511818 77512873	605	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
20	scaffold_m19_p_4 90304038 90305300	697	NC_001503.1	Mouse mammary tumor virus
21	scaffold_m19_p_4 127687819 127689081	710	NC_001503.1	Mouse mammary tumor virus
22	scaffold_m19_p_4 130200974 130202236	713	NC_001503.1	Mouse mammary tumor virus
23	scaffold_m19_p_9 44419923 44418658	1107	NC_001503.1	Mouse mammary tumor virus
24	scaffold_m19_p_9 91919171 91920145	942	NC_001503.1	Mouse mammary tumor virus
25	scaffold_m19_p_9 47439620 47438358	1103	NC_001503.1	Mouse mammary tumor virus
26	scaffold_m19_p_9 77206811 77205549	1095	NC_001503.1	Mouse mammary tumor virus
27	scaffold_m19_p_9 7905786 7907048	713	NC_001503.1	Mouse mammary tumor virus
28	scaffold_m19_p_6 12249472 12250737	715	NC_001503.1	Mouse mammary tumor virus
29	scaffold_m19_p_6 100080625 100081803	394	NC_001494.1	Jaagsiekte sheep retrovirus
30	scaffold_m19_p_17 767379 768641	713	NC_001503.1	Mouse mammary tumor virus
31	scaffold_m19_p_17 19390446 19391708	731	NC_001503.1	Mouse mammary tumor virus
32	scaffold_m19_p_26 1554170 1553052	374	NC_004994.2	Enzootic nasal tumour virus of goats
33	scaffold_m19_p_9 9132488 9133750	731	NC_001503.1	Mouse mammary tumor virus
34	scaffold_m19_p_6 11022585 11023847	713	NC_001503.1	Mouse mammary tumor virus
35	scaffold_m19_p_9 20030116 20031378	713	NC_001503.1	Mouse mammary tumor virus
36	scaffold_m19_p_9 25198555 25199820	715	NC_001503.1	Mouse mammary tumor virus
37	scaffold_m19_p_6 26383847 26385109	713	NC_001503.1	Mouse mammary tumor virus
38	scaffold_m19_p_6 70062451 70063713	713	NC_001503.1	Mouse mammary tumor virus
39	scaffold_m19_p_6 44486799 44488061	1061	NC_001503.1	Mouse mammary tumor virus
40	scaffold_m19_p_4 98740019 98738754	1116	NC_001503.1	Mouse mammary tumor virus
41	scaffold_m19_p_11 31189137 31190402	715	NC_001503.1	Mouse mammary tumor virus
42	scaffold_m19_p_11 47820048 47818783	1107	NC_001503.1	Mouse mammary tumor virus
43	scaffold_m19_p_11 55656344 55657606	713	NC_001503.1	Mouse mammary tumor virus
44	scaffold_m19_p_11 66934790 66936052	713	NC_001503.1	Mouse mammary tumor virus
45	scaffold_m19_p_13 429837 428131	570	NC_001494.1	Jaagsiekte sheep retrovirus
46	scaffold_m19_p_15 28821483 28820221	1103	NC_001503.1	Mouse mammary tumor virus
47	scaffold_m19_p_14 29890020 29888758	1095	NC_001503.1	Mouse mammary tumor virus
48	scaffold_m19_p_16 52208568 52207306	1103	NC_001503.1	Mouse mammary tumor virus
49	scaffold_m19_p_15 37402270 37403532	731	NC_001503.1	Mouse mammary tumor virus
50	scaffold_m19_p_4 99528344 99529606	710	NC_001503.1	Mouse mammary tumor virus

Table 8.22: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m19_p_2 83641110 83642372	713	NC_001503.1	Mouse mammary tumor virus
102	scaffold_m19_p_2 94377679 94378959	719	NC_001503.1	Mouse mammary tumor virus
103	scaffold_m19_p_2 111868393 111869655	713	NC_001503.1	Mouse mammary tumor virus
104	scaffold_m19_p_5 16563532 16564794	697	NC_001503.1	Mouse mammary tumor virus
105	scaffold_m19_p_5 16945964 16947226	713	NC_001503.1	Mouse mammary tumor virus
106	scaffold_m19_p_1 20426732 20425470	1103	NC_001503.1	Mouse mammary tumor virus
107	scaffold_m19_p_1 21513375 21511618	769	NC_001503.1	Mouse mammary tumor virus
108	scaffold_m19_p_7 28207623 28208885	713	NC_001503.1	Mouse mammary tumor virus
109	scaffold_m19_p_5 30996058 30997320	713	NC_001503.1	Mouse mammary tumor virus
110	scaffold_m19_p_5 42368130 42369392	713	NC_001503.1	Mouse mammary tumor virus
111	scaffold_m19_p_5 81205896 81207158	713	NC_001503.1	Mouse mammary tumor virus
112	scaffold_m19_p_5 87715024 87716286	697	NC_001503.1	Mouse mammary tumor virus
113	scaffold_m19_p_57 509099 510832	1207	NC_001494.1	Jaagsiekte sheep retrovirus
114	scaffold_m19_p_57 536516 535656	536	NC_004994.2	Enzootic nasal tumour virus of goats
115	scaffold_m19_p_1 68392148 68390886	1103	NC_001503.1	Mouse mammary tumor virus
116	scaffold_m19_p_1 200142859 200141597	1103	NC_001503.1	Mouse mammary tumor virus
117	scaffold_m19_p_1 55104973 55103711	1103	NC_001503.1	Mouse mammary tumor virus
118	scaffold_m19_p_1 119889083 119887821	1103	NC_001503.1	Mouse mammary tumor virus
119	scaffold_m19_p_1 124558093 124559331	713	NC_001503.1	Mouse mammary tumor virus
120	scaffold_m19_p_1 199179766 199178504	1095	NC_004994.2	Enzootic nasal tumour virus of goats
121	scaffold_m19_p_1 171145835 171144999	280	NC_001503.1	Mouse mammary tumor virus
122	scaffold_m19_p_1 18361406 18362668	713	NC_001503.1	Mouse mammary tumor virus
123	scaffold_m19_p_1 33908029 33909291	713	NC_001503.1	Mouse mammary tumor virus
124	scaffold_m19_p_1 44831804 44833066	731	NC_001503.1	Mouse mammary tumor virus
125	scaffold_m19_p_1 47534136 47535398	697	NC_001503.1	Mouse mammary tumor virus
126	scaffold_m19_p_1 51668371 51669489	714	NC_001503.1	Mouse mammary tumor virus
127	scaffold_m19_p_1 54383771 54385033	713	NC_001503.1	Mouse mammary tumor virus
128	scaffold_m19_p_1 58601355 58602599	590	NC_001503.1	Mouse mammary tumor virus
129	scaffold_m19_p_1 72052536 72051274	1095	NC_001503.1	Mouse mammary tumor virus
130	scaffold_m19_p_10 72583999 72582737	1095	NC_001503.1	Mouse mammary tumor virus
131	scaffold_m19_p_1 128453906 128455168	713	NC_001503.1	Mouse mammary tumor virus
132	scaffold_m19_p_1 143499868 143498606	1103	NC_001503.1	Mouse mammary tumor virus
133	scaffold_m19_p_1 192228003 192229265	713	NC_001503.1	Mouse mammary tumor virus
134	scaffold_m19_p_1 213407701 213406439	1103	NC_001503.1	Mouse mammary tumor virus
135	scaffold_m19_p_1 4581385 4582644	694	NC_001503.1	Mouse mammary tumor virus
136	scaffold_m19_p_1 34267697 34268959	713	NC_001503.1	Mouse mammary tumor virus
137	scaffold_m19_p_1 37061036 37062298	731	NC_001503.1	Mouse mammary tumor virus
138	scaffold_m19_p_2 48027527 48028789	713	NC_001503.1	Mouse mammary tumor virus
139	scaffold_m19_p_1 64287108 64288367	731	NC_001503.1	Mouse mammary tumor virus
140	scaffold_m19_p_1 67219338 67218106	1077	NC_001503.1	Mouse mammary tumor virus
141	scaffold_m19_p_1 1116526 1114769	769	NC_001503.1	Mouse mammary tumor virus
142	scaffold_m19_p_5 4002113 4003375	713	NC_001503.1	Mouse mammary tumor virus
143	scaffold_m19_p_1 4112680 4111418	1112	NC_001503.1	Mouse mammary tumor virus
144	scaffold_m19_p_7 7171900 7173165	699	NC_001503.1	Mouse mammary tumor virus
145	scaffold_m19_p_30 1415272 1414010	1103	NC_001503.1	Mouse mammary tumor virus
146	scaffold_m19_p_20 18648999 18650261	713	NC_001503.1	Mouse mammary tumor virus
147	scaffold_m19_p_39 1355537 1356781	1162	NC_007015.1	Ovine enzootic nasal tumour virus
148	scaffold_m19_p_5 52254521 52255741	742	NC_001503.1	Mouse mammary tumor virus
149	scaffold_m19_p_64 321928 323550	1004	NC_001494.1	Jaagsiekte sheep retrovirus
150	scaffold_m19_p_36 2100971 2102059	364	NC_001494.1	Jaagsiekte sheep retrovirus

Table 8.22: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
151	scaffold_m19_p_5 100433289 100432156	379	NC_004994.2	Enzootic nasal tumour virus of goats
152	scaffold_m19_p_31 12786 11524	1112	NC_001503.1	Mouse mammary tumor virus
153	scaffold_m19_p_5 16655928 16654666	1103	NC_001503.1	Mouse mammary tumor virus
154	scaffold_m19_p_5 60853516 60852254	1103	NC_001503.1	Mouse mammary tumor virus
155	scaffold_m19_p_7 83236830 83235541	1055	NC_001503.1	Mouse mammary tumor virus
156	scaffold_m19_p_31 676350 677630	703	NC_001503.1	Mouse mammary tumor virus
157	scaffold_m19_p_2 90579400 90580989	1259	NC_004994.2	Enzootic nasal tumour virus of goats
158	scaffold_m19_p_2 104323637 104322528	970	NC_001503.1	Mouse mammary tumor virus
159	scaffold_m19_p_2 106509549 106510667	374	NC_004994.2	Enzootic nasal tumour virus of goats
160	scaffold_m19_p_2 111789076 111790320	1147	NC_001503.1	Mouse mammary tumor virus
161	scaffold_m19_p_2 2414950 2413688	1112	NC_001503.1	Mouse mammary tumor virus
162	scaffold_m19_p_10 4342440 4341178	1103	NC_001503.1	Mouse mammary tumor virus
163	scaffold_m19_p_10 11314468 11313200	1064	NC_001503.1	Mouse mammary tumor virus
164	scaffold_m19_p_2 22272521 22273783	710	NC_001503.1	Mouse mammary tumor virus
165	scaffold_m19_p_10 37501013 37499748	1107	NC_001503.1	Mouse mammary tumor virus
166	scaffold_m19_p_10 43151737 43150475	1103	NC_004994.2	Enzootic nasal tumour virus of goats
167	scaffold_m19_p_2 46127560 46126631	311	NC_001503.1	Mouse mammary tumor virus
168	scaffold_m19_p_2 55218354 55217092	1103	NC_001503.1	Mouse mammary tumor virus
169	scaffold_m19_p_2 66563541 66562270	1055	NC_001503.1	Mouse mammary tumor virus
170	scaffold_m19_p_10 50134240 50135502	713	NC_001503.1	Mouse mammary tumor virus
171	scaffold_m19_p_10 60130317 60131579	713	NC_001503.1	Mouse mammary tumor virus
172	scaffold_m19_p_10 64050046 64051305	1060	NC_001503.1	Mouse mammary tumor virus
173	scaffold_m19_p_10 64617485 64618519	346	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
174	scaffold_m19_p_1 19803626 19802361	1107	NC_001503.1	Mouse mammary tumor virus
175	scaffold_m19_p_1 35981565 35980303	1112	NC_001503.1	Mouse mammary tumor virus
176	scaffold_m19_p_1 8477038 8478300	697	NC_001503.1	Mouse mammary tumor virus
177	scaffold_m19_p_1 45815155 45813890	1116	NC_001503.1	Mouse mammary tumor virus
178	scaffold_m19_p_1 80796021 80794759	1103	NC_001503.1	Mouse mammary tumor virus
179	scaffold_m19_p_1 107420337 107419075	1103	NC_001503.1	Mouse mammary tumor virus
180	scaffold_m19_p_1 127401343 127400225	374	NC_004994.2	Enzootic nasal tumour virus of goats
181	scaffold_m19_p_1 151663815 151664933	374	NC_004994.2	Enzootic nasal tumour virus of goats
182	scaffold_m19_p_1 126300324 126301901	752	NC_001494.1	Jaagsiekte sheep retrovirus
183	scaffold_m19_p_1 6722045 6720801	769	NC_001503.1	Mouse mammary tumor virus
184	scaffold_m19_p_10 70097418 70095661	769	NC_001503.1	Mouse mammary tumor virus
185	scaffold_m19_p_1 8070975 8072240	733	NC_001503.1	Mouse mammary tumor virus
186	scaffold_m19_p_1 11610029 11611291	697	NC_001503.1	Mouse mammary tumor virus
187	scaffold_m19_p_1 18099231 18100493	695	NC_001503.1	Mouse mammary tumor virus
188	scaffold_m19_p_1 177742542 177740782	772	NC_001503.1	Mouse mammary tumor virus
189	scaffold_m19_p_1 17708652 17709914	710	NC_001503.1	Mouse mammary tumor virus
190	scaffold_m19_p_1 18449237 18450499	713	NC_001503.1	Mouse mammary tumor virus
191	scaffold_m19_p_1 21038619 21039884	733	NC_001503.1	Mouse mammary tumor virus
192	scaffold_m19_p_1 122716370 122715126	735	NC_007015.1	Ovine enzootic nasal tumour virus
193	scaffold_m19_p_1 127766871 127767962	365	NC_001494.1	Jaagsiekte sheep retrovirus
194	scaffold_m19_p_1 93795623 93796825	767	NC_001494.1	Jaagsiekte sheep retrovirus
195	scaffold_m19_p_1 151655842 151657023	395	NC_004994.2	Enzootic nasal tumour virus of goats
196	scaffold_m19_p_1 210914923 210913742	395	NC_004994.2	Enzootic nasal tumour virus of goats
197	scaffold_m19_p_1 26968371 26967106	1107	NC_004994.2	Enzootic nasal tumour virus of goats
198	scaffold_m19_p_1 29153022 29151760	1103	NC_001503.1	Mouse mammary tumor virus
199	scaffold_m19_p_1 34951459 34950197	1103	NC_001503.1	Mouse mammary tumor virus
200	scaffold_m19_p_7 28188678 28187416	1103	NC_001503.1	Mouse mammary tumor virus

Table 8.22: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
201	scaffold_m19_p_22 5026780 5028042	713	NC_001503.1	Mouse mammary tumor virus
202	scaffold_m19_p_22 7153780 7152599	395	NC_001494.1	Jaagsiekte sheep retrovirus
203	scaffold_m19_p_7 58646711 58644990	769	NC_001503.1	Mouse mammary tumor virus
204	scaffold_m19_p_52 717111 715849	1094	NC_007015.1	Ovine enzootic nasal tumour virus
205	scaffold_m19_p_7 11091885 11090623	1112	NC_001503.1	Mouse mammary tumor virus
206	scaffold_m19_p_2 10539601 10538516	363	NC_001503.1	Mouse mammary tumor virus
207	scaffold_m19_p_7 13651612 13652874	710	NC_001503.1	Mouse mammary tumor virus
208	scaffold_m19_p_2 73322582 73323661	1015	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
209	scaffold_m19_p_7 81693642 81694922	719	NC_001503.1	Mouse mammary tumor virus
210	scaffold_m19_p_7 1812260 1813522	1061	NC_001503.1	Mouse mammary tumor virus
211	scaffold_m19_p_5 73701759 73703570	923	NC_001494.1	Jaagsiekte sheep retrovirus
212	scaffold_m19_p_19 8410173 8411435	713	NC_001503.1	Mouse mammary tumor virus
213	scaffold_m19_p_19 29990513 29991775	713	NC_001503.1	Mouse mammary tumor virus
214	scaffold_m19_p_3 102264448 102265710	713	NC_001503.1	Mouse mammary tumor virus
215	scaffold_m19_p_3 106314317 106315579	695	NC_001503.1	Mouse mammary tumor virus
216	scaffold_m19_p_3 117471644 117472906	730	NC_001503.1	Mouse mammary tumor virus
217	scaffold_m19_p_3 123372515 123373813	744	NC_001503.1	Mouse mammary tumor virus
218	scaffold_m19_p_3 130183508 130184770	710	NC_001503.1	Mouse mammary tumor virus
219	scaffold_m19_p_3 200405159 200406421	710	NC_001503.1	Mouse mammary tumor virus
220	scaffold_m19_p_8 795529 796794	733	NC_001503.1	Mouse mammary tumor virus
221	scaffold_m19_p_8 2251963 2253102	653	NC_001503.1	Mouse mammary tumor virus
222	scaffold_m19_p_8 17809520 17810782	713	NC_001503.1	Mouse mammary tumor virus
223	scaffold_m19_p_8 25653691 25654809	374	NC_004994.2	Enzootic nasal tumour virus of goats
224	scaffold_m19_p_8 72006366 72005275	365	NC_001494.1	Jaagsiekte sheep retrovirus
225	scaffold_m19_p_8 90530114 90528357	769	NC_001503.1	Mouse mammary tumor virus
226	scaffold_m19_p_8 3307953 3306691	1103	NC_001503.1	Mouse mammary tumor virus
227	scaffold_m19_p_7 23042488 23041226	1103	NC_001503.1	Mouse mammary tumor virus
228	scaffold_m19_p_8 29047624 29048886	713	NC_001503.1	Mouse mammary tumor virus
229	scaffold_m19_p_8 39011332 39010070	1106	NC_001503.1	Mouse mammary tumor virus
230	scaffold_m19_p_8 47905390 47904128	1104	NC_004994.2	Enzootic nasal tumour virus of goats
231	scaffold_m19_p_8 80023239 80022058	1055	NC_001503.1	Mouse mammary tumor virus
232	scaffold_m19_p_8 80256999 80257988	645	NC_001503.1	Mouse mammary tumor virus
233	scaffold_m19_p_7 13941205 13942467	1060	NC_001503.1	Mouse mammary tumor virus
234	scaffold_m19_p_7 15506180 15507931	585	NC_001494.1	Jaagsiekte sheep retrovirus
235	scaffold_m19_p_22 2370655 2369393	1103	NC_001503.1	Mouse mammary tumor virus
236	scaffold_m19_p_3 1028684 1029949	715	NC_001503.1	Mouse mammary tumor virus
237	scaffold_m19_p_12 9429645 9428383	1103	NC_001503.1	Mouse mammary tumor virus
238	scaffold_m19_p_12 12714652 12713387	1116	NC_001503.1	Mouse mammary tumor virus
239	scaffold_m19_p_12 15842801 15843967	1096	NC_007015.1	Ovine enzootic nasal tumour virus
240	scaffold_m19_p_12 60897902 60898921	655	NC_001503.1	Mouse mammary tumor virus
241	scaffold_m19_p_12 18563618 18562353	1118	NC_001503.1	Mouse mammary tumor virus
242	scaffold_m19_p_12 47436393 47435131	1095	NC_001503.1	Mouse mammary tumor virus
243	scaffold_m19_p_12 52793287 52792025	1103	NC_001503.1	Mouse mammary tumor virus
244	scaffold_m19_p_27 1836482 1835406	360	NC_001494.1	Jaagsiekte sheep retrovirus
245	scaffold_m19_p_3 102541235 102539973	1070	NC_001503.1	Mouse mammary tumor virus
246	scaffold_m19_p_3 128174262 128173000	1112	NC_001503.1	Mouse mammary tumor virus
247	scaffold_m19_p_3 75237443 75236280	904	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
248	scaffold_m19_p_18 4594 5832	733	NC_001503.1	Mouse mammary tumor virus
249	scaffold_m19_p_18 30135 28402	579	NC_001494.1	Jaagsiekte sheep retrovirus
250	scaffold_m19_p_3 96763699 96764874	675	NC_001494.1	Jaagsiekte sheep retrovirus

Table 8.22: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
251	scaffold_m19_p_3 139151169 139153322	719	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
252	scaffold_m19_p_3 29021191 29019929	1103	NC_001503.1	Mouse mammary tumor virus
253	scaffold_m19_p_3 11413099 11414361	728	NC_001503.1	Mouse mammary tumor virus
254	scaffold_m19_p_3 19643427 19644689	713	NC_001503.1	Mouse mammary tumor virus
255	scaffold_m19_p_3 21290950 21292212	697	NC_001503.1	Mouse mammary tumor virus
256	scaffold_m19_p_3 67047444 67046182	1112	NC_001503.1	Mouse mammary tumor virus
257	scaffold_m19_p_3 93402441 93403346	303	NC_001494.1	Jaagsiekte sheep retrovirus
258	scaffold_m19_p_3 40847701 40846631	358	NC_001503.1	Mouse mammary tumor virus
259	scaffold_m19_p_3 97198466 97199587	930	NC_004994.2	Enzootic nasal tumour virus of goats
260	scaffold_m19_p_3 41871120 41869858	1095	NC_001503.1	Mouse mammary tumor virus
261	scaffold_m19_p_3 98924582 98923320	1103	NC_001503.1	Mouse mammary tumor virus
262	scaffold_m19_p_3 105179217 105177955	1114	NC_001503.1	Mouse mammary tumor virus
263	scaffold_m19_p_3 54699704 54700966	697	NC_001503.1	Mouse mammary tumor virus
264	scaffold_m19_p_3 60510158 60508893	1116	NC_001503.1	Mouse mammary tumor virus
265	scaffold_m19_p_3 97806741 97808006	1064	NC_001503.1	Mouse mammary tumor virus
266	scaffold_m19_p_3 106328856 106330118	695	NC_001503.1	Mouse mammary tumor virus
267	scaffold_m19_p_3 157733682 157734944	713	NC_001503.1	Mouse mammary tumor virus
268	scaffold_m19_p_3 190460351 190459089	1097	NC_001503.1	Mouse mammary tumor virus
269	scaffold_m19_p_3 206006920 206005658	1103	NC_001503.1	Mouse mammary tumor virus
270	scaffold_m19_p_3 8353266 8354528	710	NC_001503.1	Mouse mammary tumor virus
271	scaffold_m19_p_3 18182323 18183588	715	NC_001503.1	Mouse mammary tumor virus
272	scaffold_m19_p_3 26671910 26673172	713	NC_001503.1	Mouse mammary tumor virus
273	scaffold_m19_p_3 65171049 65172311	731	NC_001503.1	Mouse mammary tumor virus
274	scaffold_m19_p_3 84689830 84691092	713	NC_001503.1	Mouse mammary tumor virus
275	scaffold_m19_p_3 97820280 97821545	1064	NC_001503.1	Mouse mammary tumor virus
276	scaffold_m19_p_3 130892456 130893658	767	NC_001494.1	Jaagsiekte sheep retrovirus
277	scaffold_m19_p_23 2882744 2881482	1103	NC_001503.1	Mouse mammary tumor virus
278	scaffold_m19_p_19 6504397 6503135	1103	NC_001503.1	Mouse mammary tumor virus
279	scaffold_m19_p_3 146917344 146916100	1049	NC_007015.1	Ovine enzootic nasal tumour virus
280	scaffold_m19_p_3 172779019 172777754	1107	NC_001503.1	Mouse mammary tumor virus
281	scaffold_m19_p_3 174806174 174804912	1103	NC_001503.1	Mouse mammary tumor virus

Table 8.22: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_22 1868423 1870876	1376	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m19_p_10 64619258 64620937	1555	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m19_p_2 66570504 66571679	383	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m19_p_6 83676700 83675342	862	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m19_p_2 105193427 105194821	456	NC_022518.1	Human endogenous retrovirus K113
6	scaffold_m19_p_4 79257164 79255764	787	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m19_p_4 79268195 79266795	787	NC_022518.1	Human endogenous retrovirus K113
8	scaffold_m19_p_14 686929 688596	555	NC_022518.1	Human endogenous retrovirus K113
9	scaffold_m19_p_1 2786385 2783794	1633	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m19_p_9 92370612 92368522	697	NC_022518.1	Human endogenous retrovirus K113
11	scaffold_m19_p_73 97775 99004	411	NC_022518.1	Human endogenous retrovirus K113
12	scaffold_m19_p_3 97200862 97202883	1289	NC_022518.1	Human endogenous retrovirus K113
13	scaffold_m19_p_4 77241733 77244300	852	NC_022518.1	Human endogenous retrovirus K113
14	scaffold_m19_p_6 73479664 73477079	1703	NC_022518.1	Human endogenous retrovirus K113
15	scaffold_m19_p_4 75974866 75977454	1697	NC_022518.1	Human endogenous retrovirus K113
16	scaffold_m19_p_3 87393925 87392270	1086	NC_022518.1	Human endogenous retrovirus K113
17	scaffold_m19_p_6 94689111 94687657	873	NC_022518.1	Human endogenous retrovirus K113
18	scaffold_m19_p_3 148120563 148118917	545	NC_001503.1	Mouse mammary tumor virus
19	scaffold_m19_p_2 42855081 42853507	978	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m19_p_1 121206554 121208449	1573	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
21	scaffold_m19_p_4 57393561 57394931	892	NC_001503.1	Mouse mammary tumor virus
22	scaffold_m19_p_8 71123423 71121555	1146	NC_001503.1	Mouse mammary tumor virus
23	scaffold_m19_p_4 120505526 120503829	915	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
24	scaffold_m19_p_2 66286701 66289046	783	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
25	scaffold_m19_p_4 88913431 88911200	1812	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
26	scaffold_m19_p_11 1124009 1125388	808	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
27	scaffold_m19_p_1 167533066 167531078	1249	NC_001503.1	Mouse mammary tumor virus
28	scaffold_m19_p_57 258974 260836	1388	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
29	scaffold_m19_p_9 31197222 31199909	1343	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
30	scaffold_m19_p_2 105957283 105958503	1531	NC_007015.1	Ovine enzootic nasal tumour virus
31	scaffold_m19_p_3 97663182 97665737	1194	NC_007015.1	Ovine enzootic nasal tumour virus
32	scaffold_m19_p_64 324534 326828	1481	NC_007015.1	Ovine enzootic nasal tumour virus
33	scaffold_m19_p_4 75581576 75583204	1057	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
34	scaffold_m19_p_29 662756 665098	1155	NC_004994.2	Enzootic nasal tumour virus of goats
35	scaffold_m19_p_1 125300145 125301833	1119	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
36	scaffold_m19_p_4 76554438 76556792	1563	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
37	scaffold_m19_p_3 139154062 139156641	1705	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
38	scaffold_m19_p_12 15844746 15847403	1962	NC_007015.1	Ovine enzootic nasal tumour virus
39	scaffold_m19_p_2 111798315 111799661	868	NC_001503.1	Mouse mammary tumor virus
40	scaffold_m19_p_1 58602622 58603770	382	NC_001503.1	Mouse mammary tumor virus
41	scaffold_m19_p_1 124581048 124583579	1613	NC_001503.1	Mouse mammary tumor virus
42	scaffold_m19_p_4 67849772 67847682	1913	NC_001503.1	Mouse mammary tumor virus
43	scaffold_m19_p_1 126302917 126304923	1254	NC_001503.1	Mouse mammary tumor virus
44	scaffold_m19_p_7 13653207 13656143	1365	NC_001503.1	Mouse mammary tumor virus
45	scaffold_m19_p_1 33909624 33912560	1365	NC_001503.1	Mouse mammary tumor virus
46	scaffold_m19_p_3 21292545 21295481	1365	NC_001503.1	Mouse mammary tumor virus
47	scaffold_m19_p_2 202959760 202962696	1771	NC_001503.1	Mouse mammary tumor virus
48	scaffold_m19_p_2 111790638 111793607	1136	NC_001503.1	Mouse mammary tumor virus
49	scaffold_m19_p_24 2996949 2999885	870	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
50	scaffold_m19_p_4 71966231 71969173	1127	NC_001503.1	Mouse mammary tumor virus

Table 8.23: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m19_p_2 10603501 10601987	971	NC_001494.1	Jaagsiekte sheep retrovirus
52	scaffold_m19_p_3 65172629 65175580	1149	NC_001503.1	Mouse mammary tumor virus
53	scaffold_m19_p_2 215523042 215525978	1762	NC_001503.1	Mouse mammary tumor virus
54	scaffold_m19_p_10 72582419 72579471	2140	NC_001503.1	Mouse mammary tumor virus
55	scaffold_m19_p_2 124005567 124008503	1149	NC_001503.1	Mouse mammary tumor virus
56	scaffold_m19_p_3 54701299 54704229	1739	NC_001503.1	Mouse mammary tumor virus
57	scaffold_m19_p_2 213774845 213777787	1369	NC_001503.1	Mouse mammary tumor virus
58	scaffold_m19_p_4 91454354 91456438	1149	NC_001503.1	Mouse mammary tumor virus
59	scaffold_m19_p_1 37062616 37065567	1147	NC_001503.1	Mouse mammary tumor virus
60	scaffold_m19_p_2 111869988 111872924	1788	NC_001503.1	Mouse mammary tumor virus
61	scaffold_m19_p_2 27717323 27720274	1145	NC_001494.1	Jaagsiekte sheep retrovirus
62	scaffold_m19_p_8 17811115 17813676	1304	NC_001503.1	Mouse mammary tumor virus
63	scaffold_m19_p_8 54907188 54910121	2243	NC_001503.1	Mouse mammary tumor virus
64	scaffold_m19_p_1 18450832 18453768	1741	NC_001503.1	Mouse mammary tumor virus
65	scaffold_m19_p_1 54385351 54388302	1151	NC_001503.1	Mouse mammary tumor virus
66	scaffold_m19_p_20 18650579 18653530	1151	NC_001503.1	Mouse mammary tumor virus
67	scaffold_m19_p_1 143498273 143495316	1960	NC_001503.1	Mouse mammary tumor virus
68	scaffold_m19_p_3 84691425 84694361	1748	NC_001503.1	Mouse mammary tumor virus
69	scaffold_m19_p_3 18183921 18186857	1361	NC_001503.1	Mouse mammary tumor virus
70	scaffold_m19_p_2 66934408 66932495	1358	NC_001494.1	Jaagsiekte sheep retrovirus
71	scaffold_m19_p_1 192229598 192232555	1365	NC_001503.1	Mouse mammary tumor virus
72	scaffold_m19_p_7 1813852 1816791	1751	NC_001503.1	Mouse mammary tumor virus
73	scaffold_m19_p_1 21040202 21043153	1151	NC_001503.1	Mouse mammary tumor virus
74	scaffold_m19_p_5 16947544 16950495	1151	NC_001503.1	Mouse mammary tumor virus
75	scaffold_m19_p_1 4582962 4585913	1151	NC_001503.1	Mouse mammary tumor virus
76	scaffold_m19_p_3 102266043 102268979	1363	NC_001503.1	Mouse mammary tumor virus
77	scaffold_m19_p_1 6720483 6717538	2135	NC_001503.1	Mouse mammary tumor virus
78	scaffold_m19_p_10 64052093 64054561	1437	NC_001503.1	Mouse mammary tumor virus
79	scaffold_m19_p_1 213028390 213029850	970	NC_001503.1	Mouse mammary tumor virus
80	scaffold_m19_p_1 17710247 17713183	1795	NC_001503.1	Mouse mammary tumor virus
81	scaffold_m19_p_4 90714244 90717180	1361	NC_001503.1	Mouse mammary tumor virus
82	scaffold_m19_p_6 12251070 12254006	1363	NC_001503.1	Mouse mammary tumor virus
83	scaffold_m19_p_15 37403865 37406822	1365	NC_001503.1	Mouse mammary tumor virus
84	scaffold_m19_p_2 83642705 83645641	1365	NC_001503.1	Mouse mammary tumor virus
85	scaffold_m19_p_9 9134083 9137010	1366	NC_001503.1	Mouse mammary tumor virus
86	scaffold_m19_p_2 94379277 94382228	1151	NC_001503.1	Mouse mammary tumor virus
87	scaffold_m19_p_1 56092561 56095512	1149	NC_001503.1	Mouse mammary tumor virus
88	scaffold_m19_p_5 42369710 42372661	1151	NC_001503.1	Mouse mammary tumor virus
89	scaffold_m19_p_1 35979585 35977039	1665	NC_001503.1	Mouse mammary tumor virus
90	scaffold_m19_p_3 123374131 123377061	1115	NC_001503.1	Mouse mammary tumor virus
91	scaffold_m19_p_1 11611609 11614068	1151	NC_001503.1	Mouse mammary tumor virus
92	scaffold_m19_p_2 22274116 22277049	1759	NC_001503.1	Mouse mammary tumor virus
93	scaffold_m19_p_11 31190735 31193671	1766	NC_001503.1	Mouse mammary tumor virus
94	scaffold_m19_p_4 61481667 61484603	1365	NC_001503.1	Mouse mammary tumor virus
95	scaffold_m19_p_6 70064046 70067003	1365	NC_001503.1	Mouse mammary tumor virus
96	scaffold_m19_p_7 81695255 81698191	1748	NC_001503.1	Mouse mammary tumor virus
97	scaffold_m19_p_11 55657939 55660875	1365	NC_001503.1	Mouse mammary tumor virus
98	scaffold_m19_p_4 43583244 43581283	1535	NC_001503.1	Mouse mammary tumor virus
99	scaffold_m19_p_5 4003708 4006644	1771	NC_001503.1	Mouse mammary tumor virus
100	scaffold_m19_p_6 11024180 11027113	1365	NC_001503.1	Mouse mammary tumor virus

Table 8.23: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m19_p_2 20011572 20014508	1771	NC_001503.1	Mouse mammary tumor virus
102	scaffold_m19_p_6 44488394 44491330	1365	NC_001503.1	Mouse mammary tumor virus
103	scaffold_m19_p_22 5028375 5031311	1747	NC_001503.1	Mouse mammary tumor virus
104	scaffold_m19_p_2 163798201 163801152	1151	NC_001503.1	Mouse mammary tumor virus
105	scaffold_m19_p_10 60131897 60134848	1151	NC_001503.1	Mouse mammary tumor virus
106	scaffold_m19_p_2 48029107 48032058	1138	NC_001503.1	Mouse mammary tumor virus
107	scaffold_m19_p_4 99529939 99532875	1361	NC_001503.1	Mouse mammary tumor virus
108	scaffold_m19_p_3 8354861 8357797	1747	NC_001503.1	Mouse mammary tumor virus
109	scaffold_m19_p_1 44833384 44836335	1151	NC_001503.1	Mouse mammary tumor virus
110	scaffold_m19_p_9 25200138 25203089	1151	NC_001503.1	Mouse mammary tumor virus
111	scaffold_m19_p_3 130185088 130188039	1151	NC_001503.1	Mouse mammary tumor virus
112	scaffold_m19_p_3 157735262 157738213	1151	NC_001503.1	Mouse mammary tumor virus
113	scaffold_m19_p_5 87716604 87719555	1151	NC_001503.1	Mouse mammary tumor virus
114	scaffold_m19_p_11 66936370 66939318	1151	NC_001503.1	Mouse mammary tumor virus
115	scaffold_m19_p_7 7173483 7176434	1151	NC_001503.1	Mouse mammary tumor virus
116	scaffold_m19_p_5 30997638 31000589	1151	NC_001503.1	Mouse mammary tumor virus
117	scaffold_m19_p_9 7907366 7910317	1151	NC_001503.1	Mouse mammary tumor virus
118	scaffold_m19_p_19 29992093 29995044	1151	NC_001503.1	Mouse mammary tumor virus
119	scaffold_m19_p_11 10210225 10208108	1387	NC_001503.1	Mouse mammary tumor virus
120	scaffold_m19_p_19 8412594 8414696	2766	NC_001503.1	Mouse mammary tumor virus
121	scaffold_m19_p_2 107327800 107325647	1262	NC_001503.1	Mouse mammary tumor virus
122	scaffold_m19_p_1 112688256 112685314	1494	NC_001503.1	Mouse mammary tumor virus
123	scaffold_m19_p_4 70176702 70173775	3339	NC_001503.1	Mouse mammary tumor virus
124	scaffold_m19_p_4 118774857 118771906	2881	NC_001503.1	Mouse mammary tumor virus
125	scaffold_m19_p_1 123424778 123421827	2875	NC_001503.1	Mouse mammary tumor virus
126	scaffold_m19_p_3 60851253 60848302	2875	NC_001503.1	Mouse mammary tumor virus
127	scaffold_m19_p_2 46125972 46123036	4096	NC_001503.1	Mouse mammary tumor virus
128	scaffold_m19_p_3 174804594 174801643	2875	NC_001503.1	Mouse mammary tumor virus
129	scaffold_m19_p_19 6502802 6499866	4101	NC_001503.1	Mouse mammary tumor virus
130	scaffold_m19_p_2 59866084 59863079	3613	NC_001503.1	Mouse mammary tumor virus
131	scaffold_m19_p_1 29151442 29148491	2876	NC_001503.1	Mouse mammary tumor virus
132	scaffold_m19_p_3 29019596 29016660	4088	NC_001503.1	Mouse mammary tumor virus
133	scaffold_m19_p_8 3306358 3303422	3623	NC_001503.1	Mouse mammary tumor virus
134	scaffold_m19_p_22 2369075 2366124	2875	NC_001503.1	Mouse mammary tumor virus
135	scaffold_m19_p_1 34949864 34946928	3497	NC_001503.1	Mouse mammary tumor virus
136	scaffold_m19_p_1 80794426 80791490	3623	NC_001503.1	Mouse mammary tumor virus
137	scaffold_m19_p_4 20083204 20080271	3612	NC_001503.1	Mouse mammary tumor virus
138	scaffold_m19_p_1 26966773 26963840	3386	NC_001503.1	Mouse mammary tumor virus
139	scaffold_m19_p_4 98752750 98749814	4088	NC_001503.1	Mouse mammary tumor virus
140	scaffold_m19_p_4 98738421 98735485	4088	NC_001503.1	Mouse mammary tumor virus
141	scaffold_m19_p_52 715516 712580	4094	NC_001503.1	Mouse mammary tumor virus
142	scaffold_m19_p_1 119887488 119884552	3624	NC_001503.1	Mouse mammary tumor virus
143	scaffold_m19_p_11 47818450 47815514	3623	NC_001503.1	Mouse mammary tumor virus
144	scaffold_m19_p_9 47438040 47435089	2875	NC_001503.1	Mouse mammary tumor virus
145	scaffold_m19_p_3 190458756 190455820	4090	NC_001503.1	Mouse mammary tumor virus
146	scaffold_m19_p_4 80259789 80256838	2875	NC_001503.1	Mouse mammary tumor virus
147	scaffold_m19_p_7 23040893 23037957	4090	NC_001503.1	Mouse mammary tumor virus
148	scaffold_m19_p_23 2881149 2878213	4090	NC_001503.1	Mouse mammary tumor virus
149	scaffold_m19_p_3 172777421 172774485	4090	NC_001503.1	Mouse mammary tumor virus
150	scaffold_m19_p_4 67702491 67699555	3503	NC_001503.1	Mouse mammary tumor virus

Table 8.23: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
151	scaffold_m19_p_12 47434798 47431862	3623	NC_001503.1	Mouse mammary tumor virus
152	scaffold_m19_p_2 139218198 139215250	2881	NC_001503.1	Mouse mammary tumor virus
153	scaffold_m19_p_2 2413370 2410419	2867	NC_001503.1	Mouse mammary tumor virus
154	scaffold_m19_p_12 18562035 18559084	2875	NC_001503.1	Mouse mammary tumor virus
155	scaffold_m19_p_2 147320880 147317929	2870	NC_001503.1	Mouse mammary tumor virus
156	scaffold_m19_p_1 19802043 19799092	2874	NC_001503.1	Mouse mammary tumor virus
157	scaffold_m19_p_4 75366734 75368584	617	NC_001503.1	Mouse mammary tumor virus
158	scaffold_m19_p_17 39645793 39647190	1216	NC_001503.1	Mouse mammary tumor virus
159	scaffold_m19_p_4 59998269 60000359	1165	NC_001503.1	Mouse mammary tumor virus
160	scaffold_m19_p_16 51384076 51386376	1492	NC_001503.1	Mouse mammary tumor virus
161	scaffold_m19_p_4 90306050 90308566	2584	NC_007015.1	Ovine enzootic nasal tumour virus
162	scaffold_m19_p_10 50136217 50138790	858	NC_001503.1	Mouse mammary tumor virus
163	scaffold_m19_p_7 28209891 28212176	1492	NC_001503.1	Mouse mammary tumor virus
164	scaffold_m19_p_2 79219838 79216902	4070	NC_001503.1	Mouse mammary tumor virus
165	scaffold_m19_p_4 34314785 34316104	1206	NC_001503.1	Mouse mammary tumor virus
166	scaffold_m19_p_1 107418728 107415807	3529	NC_001503.1	Mouse mammary tumor virus
167	scaffold_m19_p_4 8539069 8541051	659	NC_001503.1	Mouse mammary tumor virus
168	scaffold_m19_p_1 55103365 55101314	685	NC_001503.1	Mouse mammary tumor virus
169	scaffold_m19_p_2 104322043 104319107	3318	NC_001503.1	Mouse mammary tumor virus
170	scaffold_m19_p_5 60851921 60848988	2764	NC_001503.1	Mouse mammary tumor virus
171	scaffold_m19_p_4 90726224 90728776	2443	NC_001503.1	Mouse mammary tumor virus
172	scaffold_m19_p_4 3878152 3880092	1846	NC_001503.1	Mouse mammary tumor virus
173	scaffold_m19_p_2 108249038 108247386	1531	NC_001503.1	Mouse mammary tumor virus
174	scaffold_m19_p_6 38172130 38174028	1214	NC_001503.1	Mouse mammary tumor virus
175	scaffold_m19_p_1 4111091 4108431	2638	NC_001503.1	Mouse mammary tumor virus
176	scaffold_m19_p_4 74021487 74023598	2004	NC_001503.1	Mouse mammary tumor virus
177	scaffold_m19_p_3 202788872 202786995	627	NC_001503.1	Mouse mammary tumor virus
178	scaffold_m19_p_10 43148667 43147216	1170	NC_001503.1	Mouse mammary tumor virus
179	scaffold_m19_p_3 194782040 194780166	1783	NC_001503.1	Mouse mammary tumor virus
180	scaffold_m19_p_8 29049626 29052157	3273	NC_001503.1	Mouse mammary tumor virus
181	scaffold_m19_p_13 6754896 6756518	541	NC_001503.1	Mouse mammary tumor virus
182	scaffold_m19_p_31 10364 8259	1323	NC_001503.1	Mouse mammary tumor virus
183	scaffold_m19_p_3 2870380 2868203	2092	NC_001503.1	Mouse mammary tumor virus
184	scaffold_m19_p_18 5798 7084	1196	NC_001503.1	Mouse mammary tumor virus
185	scaffold_m19_p_2 111972299 111971100	401	NC_001503.1	Mouse mammary tumor virus
186	scaffold_m19_p_3 11416078 11417628	1359	NC_001503.1	Mouse mammary tumor virus

Table 8.23: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_44 5364 6254	297	NC_007015.1	Ovine enzootic nasal tumour virus
2	scaffold_m19_p_16 4470149 4469343	264	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
3	scaffold_m19_p_3 101794103 101793282	477	NC_001550.1	Mason-Pfizer monkey virus
4	scaffold_m19_p_19 15064221 15065234	550	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
5	scaffold_m19_p_11 58962269 58963015	250	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m19_p_28 3096508 3095039	491	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
7	scaffold_m19_p_57 41887 40268	541	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
8	scaffold_m19_p_1 125302120 125303829	569	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
9	scaffold_m19_p_3 139157028 139158638	538	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
10	scaffold_m19_p_4 76557071 76558807	579	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m19_p_7 93119111 93118137	573	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
12	scaffold_m19_p_3 97884353 97883751	200	NC_001514.1	Squirrel monkey retrovirus - HLB
13	scaffold_m19_p_3 93675466 93674591	289	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
14	scaffold_m19_p_2 109745573 109744839	244	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m19_p_39 1755965 1754238	566	NC_029852.1	Bovine retrovirus CH15
16	scaffold_m19_p_22 7149441 7147849	532	NC_007015.1	Ovine enzootic nasal tumour virus
17	scaffold_m19_p_22 6862813 6862043	258	NC_007015.1	Ovine enzootic nasal tumour virus
18	scaffold_m19_p_55 363087 362362	243	NC_007015.1	Ovine enzootic nasal tumour virus
19	scaffold_m19_p_89 33366 34082	237	NC_007015.1	Ovine enzootic nasal tumour virus

Table 8.24: Reciprocal BLAST hits for the betaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_6 84287221 84285914	443	NC_001362.1	Friend murine leukemia virus FB29
2	scaffold_m19_p_4 96326017 96327354	438	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
3	scaffold_m19_p_11 70534094 70535470	527	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
4	scaffold_m19_p_2 106805168 106803810	451	NC_001885.3	Gibbon ape leukemia virus
5	scaffold_m19_p_28 3329034 3330083	648	NC_039228.1	Koala retrovirus
6	scaffold_m19_p_5 91491592 91492899	432	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
7	scaffold_m19_p_40 1805465 1806940	493	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
8	scaffold_m19_p_25 4664769 4666013	416	NC_003059.1	Porcine endogenous retrovirus E

Table 8.25: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_33 640930 639596	446	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m19_p_42 1545820 1547061	414	NC_009889.1	RD114 retrovirus
3	scaffold_m19_p_2 115050778 115048799	660	NC_009889.1	RD114 retrovirus
4	scaffold_m19_p_69 191647 192957	856	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
5	scaffold_m19_p_28 3331300 3333657	1565	NC_001885.3	Gibbon ape leukemia virus
6	scaffold_m19_p_26 181887 178783	2256	NC_001362.1	Friend murine leukemia virus FB29
7	scaffold_m19_p_4 73748313 73745167	2657	NC_001702.1	Murine type C retrovirus
8	scaffold_m19_p_2 98689690 98691603	1224	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
9	scaffold_m19_p_4 82968195 82964635	3374	NC_001702.1	Murine type C retrovirus
10	scaffold_m19_p_4 96327052 96330561	917	NC_001885.3	Gibbon ape leukemia virus
11	scaffold_m19_p_4 77152795 77156292	1288	NC_001885.3	Gibbon ape leukemia virus
12	scaffold_m19_p_33 643018 641576	865	NC_039228.1	Koala retrovirus
13	scaffold_m19_p_4 79988884 79985297	3370	NC_001702.1	Murine type C retrovirus
14	scaffold_m19_p_1 97468005 97465105	1882	NC_001702.1	Murine type C retrovirus
15	scaffold_m19_p_9 17560694 17564251	2260	NC_001501.1	Moloney murine leukemia virus
16	scaffold_m19_p_10 31723355 31719795	3376	NC_001501.1	Moloney murine leukemia virus
17	scaffold_m19_p_4 61834557 61837943	3116	NC_001702.1	Murine type C retrovirus
18	scaffold_m19_p_14 35322351 35318791	3378	NC_001501.1	Moloney murine leukemia virus
19	scaffold_m19_p_4 104819742 104823302	2265	NC_001501.1	Moloney murine leukemia virus
20	scaffold_m19_p_8 6874177 6877737	2265	NC_001501.1	Moloney murine leukemia virus
21	scaffold_m19_p_2 128846227 128849067	1793	NC_001702.1	Murine type C retrovirus
22	scaffold_m19_p_1 44736453 44740013	2265	NC_001501.1	Moloney murine leukemia virus
23	scaffold_m19_p_7 35019703 35023263	2265	NC_001501.1	Moloney murine leukemia virus
24	scaffold_m19_p_1 180801360 180804236	1878	NC_009889.1	RD114 retrovirus
25	scaffold_m19_p_12 64821833 64824586	1789	NC_001501.1	Moloney murine leukemia virus
26	scaffold_m19_p_5 4359556 4356686	1883	NC_001702.1	Murine type C retrovirus
27	scaffold_m19_p_8 87120862 87123414	1598	NC_001702.1	Murine type C retrovirus
28	scaffold_m19_p_8 93401716 93404436	1745	NC_001501.1	Moloney murine leukemia virus
29	scaffold_m19_p_19 9070738 9067166	2239	NC_001501.1	Moloney murine leukemia virus
30	scaffold_m19_p_40 1807311 1810154	1798	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
31	scaffold_m19_p_3 101584898 101588407	2254	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
32	scaffold_m19_p_34 4557 3229	435	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
33	scaffold_m19_p_25 4632860 4631697	384	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
34	scaffold_m19_p_4 90641016 90637432	2259	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
35	scaffold_m19_p_19 11609646 11613230	2289	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China

Table 8.26: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_2 73327309 73329075	1350	NC_001885.3	Gibbon ape leukemia virus
2	scaffold_m19_p_24 628379 627096	793	NC_001885.3	Gibbon ape leukemia virus
3	scaffold_m19_p_3 125338299 125339483	396	NC_001885.3	Gibbon ape leukemia virus
4	scaffold_m19_p_4 79985258 79983444	606	NC_001940.1	Feline leukemia virus
5	scaffold_m19_p_1 66329047 66327227	608	NC_001940.1	Feline leukemia virus
6	scaffold_m19_p_4 82964596 82962776	607	NC_001940.1	Feline leukemia virus
7	scaffold_m19_p_4 73744462 73742900	521	NC_001885.3	Gibbon ape leukemia virus
8	scaffold_m19_p_26 177529 176471	351	NC_001501.1	Moloney murine leukemia virus
9	scaffold_m19_p_10 45968127 45969929	602	NC_001885.3	Gibbon ape leukemia virus
10	scaffold_m19_p_9 35896895 35895039	620	NC_001885.3	Gibbon ape leukemia virus
11	scaffold_m19_p_1 190881464 190879629	613	NC_001885.3	Gibbon ape leukemia virus
12	scaffold_m19_p_2 22603301 22602684	207	NC_001362.1	Friend murine leukemia virus FB29
13	scaffold_m19_p_12 35773072 35774988	640	NC_001885.3	Gibbon ape leukemia virus
14	scaffold_m19_p_5 63125204 63123381	609	NC_001885.3	Gibbon ape leukemia virus
15	scaffold_m19_p_11 52056523 52054700	609	NC_001885.3	Gibbon ape leukemia virus
16	scaffold_m19_p_2 81596044 81594221	609	NC_001885.3	Gibbon ape leukemia virus
17	scaffold_m19_p_2 191944202 191942379	609	NC_001885.3	Gibbon ape leukemia virus
18	scaffold_m19_p_2 166953680 166955503	609	NC_001885.3	Gibbon ape leukemia virus
19	scaffold_m19_p_18 3288216 3286393	609	NC_001885.3	Gibbon ape leukemia virus
20	scaffold_m19_p_4 73449838 73451661	609	NC_001885.3	Gibbon ape leukemia virus
21	scaffold_m19_p_9 86311897 86309981	640	NC_001885.3	Gibbon ape leukemia virus
22	scaffold_m19_p_3 174682630 174684546	640	NC_001885.3	Gibbon ape leukemia virus
23	scaffold_m19_p_8 91542873 91543808	313	NC_001362.1	Friend murine leukemia virus FB29
24	scaffold_m19_p_5 11748916 11747000	640	NC_001885.3	Gibbon ape leukemia virus
25	scaffold_m19_p_7 73186260 73188176	640	NC_001885.3	Gibbon ape leukemia virus
26	scaffold_m19_p_7 62220249 62222165	640	NC_001885.3	Gibbon ape leukemia virus
27	scaffold_m19_p_7 89437237 89439153	640	NC_001885.3	Gibbon ape leukemia virus
28	scaffold_m19_p_14 29817701 29815785	640	NC_001885.3	Gibbon ape leukemia virus
29	scaffold_m19_p_14 29822559 29820643	640	NC_001885.3	Gibbon ape leukemia virus
30	scaffold_m19_p_1 48904105 48906021	640	NC_001885.3	Gibbon ape leukemia virus
31	scaffold_m19_p_2 148275136 148277052	640	NC_001885.3	Gibbon ape leukemia virus
32	scaffold_m19_p_3 197175067 197176983	640	NC_001885.3	Gibbon ape leukemia virus
33	scaffold_m19_p_7 40335982 40334066	640	NC_001885.3	Gibbon ape leukemia virus
34	scaffold_m19_p_16 19064814 19066730	640	NC_001885.3	Gibbon ape leukemia virus
35	scaffold_m19_p_6 75858287 75860203	640	NC_001885.3	Gibbon ape leukemia virus
36	scaffold_m19_p_5 69847433 69845958	909	NC_001362.1	Friend murine leukemia virus FB29
37	scaffold_m19_p_1 208062920 208061097	609	NC_001885.3	Gibbon ape leukemia virus
38	scaffold_m19_p_14 26913201 26915117	640	NC_001885.3	Gibbon ape leukemia virus
39	scaffold_m19_p_14 29862135 29860219	640	NC_001885.3	Gibbon ape leukemia virus
40	scaffold_m19_p_11 77778470 77776554	640	NC_001885.3	Gibbon ape leukemia virus
41	scaffold_m19_p_18 23344323 23342407	640	NC_001885.3	Gibbon ape leukemia virus
42	scaffold_m19_p_3 11935760 11933844	640	NC_001885.3	Gibbon ape leukemia virus
43	scaffold_m19_p_6 43094786 43093845	315	NC_001362.1	Friend murine leukemia virus FB29
44	scaffold_m19_p_2 182361193 182363133	648	NC_001885.3	Gibbon ape leukemia virus
45	scaffold_m19_p_6 90632275 90634191	640	NC_001885.3	Gibbon ape leukemia virus
46	scaffold_m19_p_8 92699381 92697465	640	NC_001885.3	Gibbon ape leukemia virus
47	scaffold_m19_p_1 35228623 35226707	640	NC_001885.3	Gibbon ape leukemia virus
48	scaffold_m19_p_10 63895408 63893492	640	NC_001885.3	Gibbon ape leukemia virus
49	scaffold_m19_p_12 69455999 69454083	640	NC_001885.3	Gibbon ape leukemia virus
50	scaffold_m19_p_8 90286035 90284119	640	NC_001885.3	Gibbon ape leukemia virus

Table 8.27: Reciprocal BLAST hits for the gammaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m19_p_1 108041157 108043073	640	NC_001885.3	Gibbon ape leukemia virus
52	scaffold_m19_p_1 108884836 108882920	640	NC_001885.3	Gibbon ape leukemia virus
53	scaffold_m19_p_1 163726833 163724917	640	NC_001885.3	Gibbon ape leukemia virus
54	scaffold_m19_p_14 27266060 27264144	640	NC_001885.3	Gibbon ape leukemia virus
55	scaffold_m19_p_14 29852913 29850997	640	NC_001885.3	Gibbon ape leukemia virus
56	scaffold_m19_p_11 67644755 67642839	640	NC_001885.3	Gibbon ape leukemia virus
57	scaffold_m19_p_3 83847725 83849641	640	NC_001885.3	Gibbon ape leukemia virus
58	scaffold_m19_p_6 43995017 43993101	640	NC_001885.3	Gibbon ape leukemia virus
59	scaffold_m19_p_10 71977228 71975312	640	NC_001885.3	Gibbon ape leukemia virus
60	scaffold_m19_p_11 42884341 42882425	640	NC_001885.3	Gibbon ape leukemia virus
61	scaffold_m19_p_10 7527483 7529399	640	NC_001885.3	Gibbon ape leukemia virus
62	scaffold_m19_p_4 19685926 19684010	640	NC_001885.3	Gibbon ape leukemia virus
63	scaffold_m19_p_2 58554162 58552246	640	NC_001885.3	Gibbon ape leukemia virus
64	scaffold_m19_p_2 148264033 148265949	640	NC_001885.3	Gibbon ape leukemia virus
65	scaffold_m19_p_2 188529164 188531080	640	NC_001885.3	Gibbon ape leukemia virus
66	scaffold_m19_p_2 191070999 191069083	640	NC_001885.3	Gibbon ape leukemia virus
67	scaffold_m19_p_4 74972544 74973509	543	NC_001362.1	Friend murine leukemia virus FB29
68	scaffold_m19_p_7 69481477 69479504	1190	NC_039228.1	Koala retrovirus
69	scaffold_m19_p_10 12971294 12969315	661	NC_039228.1	Koala retrovirus
70	scaffold_m19_p_4 73112329 73111466	284	NC_039228.1	Koala retrovirus
71	scaffold_m19_p_40 1810831 1812405	989	NC_001501.1	Moloney murine leukemia virus
72	scaffold_m19_p_13 4354300 4352759	515	NC_001501.1	Moloney murine leukemia virus
73	scaffold_m19_p_3 101588707 101590284	993	NC_001501.1	Moloney murine leukemia virus
74	scaffold_m19_p_19 9176448 9174823	541	NC_001940.1	Feline leukemia virus
75	scaffold_m19_p_17 44810780 44809872	393	NC_001501.1	Moloney murine leukemia virus
76	scaffold_m19_p_25 4631410 4630499	518	NC_001940.1	Feline leukemia virus
77	scaffold_m19_p_4 60113951 60112032	641	NC_001940.1	Feline leukemia virus
78	scaffold_m19_p_4 60124992 60123073	641	NC_001940.1	Feline leukemia virus
79	scaffold_m19_p_4 74653888 74655285	465	NC_001885.3	Gibbon ape leukemia virus
80	scaffold_m19_p_15 49232390 49231800	198	NC_001940.1	Feline leukemia virus
81	scaffold_m19_p_5 107665098 107663701	702	NC_001362.1	Friend murine leukemia virus FB29
82	scaffold_m19_p_14 3138797 3137670	1096	NC_039228.1	Koala retrovirus
83	scaffold_m19_p_3 418345 419013	224	NC_001940.1	Feline leukemia virus
84	scaffold_m19_p_10 41597588 41598313	242	NC_001885.3	Gibbon ape leukemia virus
85	scaffold_m19_p_3 146920790 146922391	1034	NC_001940.1	Feline leukemia virus
86	scaffold_m19_p_4 90636413 90635538	541	NC_001940.1	Feline leukemia virus
87	scaffold_m19_p_19 11614249 11615127	560	NC_001501.1	Moloney murine leukemia virus
88	scaffold_m19_p_3 92530790 92529189	690	NC_001501.1	Moloney murine leukemia virus
89	scaffold_m19_p_7 76429176 76430717	867	NC_001501.1	Moloney murine leukemia virus
90	scaffold_m19_p_4 58281669 58280428	718	NC_001501.1	Moloney murine leukemia virus
91	scaffold_m19_p_4 93602335 93600950	459	NC_001940.1	Feline leukemia virus
92	scaffold_m19_p_8 77130472 77131845	609	NC_001501.1	Moloney murine leukemia virus
93	scaffold_m19_p_4 74518156 74517428	243	NC_001885.3	Gibbon ape leukemia virus
94	scaffold_m19_p_9 89941160 89941948	263	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
95	scaffold_m19_p_19 8352162 8353208	538	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
96	scaffold_m19_p_4 77156708 77157664	320	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
97	scaffold_m19_p_2 106801934 106801140	469	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
98	scaffold_m19_p_4 96330977 96331522	183	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
99	scaffold_m19_p_29 2613281 2612472	270	NC_022517.1	Baboon endogenous virus strain M7 proviral DNA
100	scaffold_m19_p_18 15600 14806	458	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
101	scaffold_m19_p_4 70712714 70713334	208	NC_022517.1	Baboon endogenous virus strain M7 proviral DNA

Table 8.27: Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Myotis myotis* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Epsilonretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_1 215029247 215029993	250	NC_010955.1	Xenopus laevis endogenous retrovirus Xen1

Table 8.28: Reciprocal BLAST hits for the epsilonretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_5 76292844 76292215	211	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m19_p_5 78386624 78387607	329	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m19_p_7 4785526 4786668	896	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m19_p_9 56279644 56280390	609	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m19_p_4 123442149 123442895	465	NC_022518.1	Human endogenous retrovirus K113
6	scaffold_m19_p_4 124893685 124892999	417	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m19_p_3 8663896 8664894	660	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
8	scaffold_m19_p_6 77074607 77075344	247	NC_022518.1	Human endogenous retrovirus K113
9	scaffold_m19_p_3 87021946 87022923	530	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m19_p_7 92354458 92352377	1356	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m19_p_22 255811 256623	272	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
12	scaffold_m19_p_3 124654134 124656227	699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m19_p_10 44624380 44623364	584	NC_022518.1	Human endogenous retrovirus K113
14	scaffold_m19_p_6 76710526 76711452	476	NC_022518.1	Human endogenous retrovirus K113
15	scaffold_m19_p_16 725792 724812	328	NC_022518.1	Human endogenous retrovirus K113
16	scaffold_m19_p_9 57219339 57218566	259	NC_004994.2	Enzootic nasal tumour virus of goats
17	scaffold_m19_p_15 1504060 1506243	729	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m19_p_11 9958061 9959941	1154	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
19	scaffold_m19_p_13 43371465 43369261	736	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m19_p_15 50420283 50422466	729	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
21	scaffold_m19_p_14 12677594 12675801	1075	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
22	scaffold_m19_p_12 86911720 86912682	799	NC_001550.1	Mason-Pfizer monkey virus
23	scaffold_m19_p_4 121456604 121457302	234	NC_022518.1	Human endogenous retrovirus K113
24	scaffold_m19_p_6 77074267 77074905	214	NC_022518.1	Human endogenous retrovirus K113
25	scaffold_m19_p_1 102961441 102962439	662	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
26	scaffold_m19_p_2 78385821 78387770	1295	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824

Table 8.29: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_2 183942102 183938635	2199	NC_009889.1	RD114 retrovirus
2	scaffold_m19_p_1 55054740 55056083	862	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m19_p_14 33947117 33945978	374	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
4	scaffold_m19_p_3 94010948 94012501	505	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
5	scaffold_m19_p_13 45119681 45118584	358	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m19_p_22 364323 362965	822	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m19_p_12 51407069 51409045	1239	NC_022518.1	Human endogenous retrovirus K113
8	scaffold_m19_p_7 4787507 4789216	893	NC_022518.1	Human endogenous retrovirus K113
9	scaffold_m19_p_6 106072145 106070772	1261	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m19_p_6 106192751 106194124	863	NC_022518.1	Human endogenous retrovirus K113
11	scaffold_m19_p_16 262532 263707	384	NC_022518.1	Human endogenous retrovirus K113
12	scaffold_m19_p_5 72689160 72686899	1214	NC_022518.1	Human endogenous retrovirus K113
13	scaffold_m19_p_6 76712509 76714671	1158	NC_022518.1	Human endogenous retrovirus K113
14	scaffold_m19_p_2 90181695 90180412	825	NC_022518.1	Human endogenous retrovirus K113
15	scaffold_m19_p_15 1301007 1298839	1377	NC_001503.1	Mouse mammary tumor virus
16	scaffold_m19_p_11 9960157 9962082	1214	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m19_p_15 50423215 50425833	1723	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m19_p_10 43372683 43375286	1723	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
19	scaffold_m19_p_9 50775224 50776717	487	NC_001503.1	Mouse mammary tumor virus
20	scaffold_m19_p_11 4848363 4846972	452	NC_001503.1	Mouse mammary tumor virus
21	scaffold_m19_p_9 65808267 65809742	1325	NC_001503.1	Mouse mammary tumor virus
22	scaffold_m19_p_4 124891085 124889769	857	NC_001503.1	Mouse mammary tumor virus

Table 8.30: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_4 120530957 120530154	269	NC_001550.1	Mason-Pfizer monkey virus
2	scaffold_m19_p_1 92589598 92590374	260	NC_001550.1	Mason-Pfizer monkey virus
3	scaffold_m19_p_3 124656314 124657849	513	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
4	scaffold_m19_p_13 43368725 43367295	478	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
5	scaffold_m19_p_2 78387925 78388794	291	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m19_p_15 50425820 50426422	202	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
7	scaffold_m19_p_14 12675265 12673835	478	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
8	scaffold_m19_p_15 1506780 1508210	478	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
9	scaffold_m19_p_11 9963184 9963894	238	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
10	scaffold_m19_p_10 43375291 43375890	201	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m19_p_1 102962661 102963500	281	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
12	scaffold_m19_p_3 8665430 8666860	478	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m19_p_11 9962180 9963166	330	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
14	scaffold_m19_p_7 92351749 92350319	478	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m19_p_8 52310641 52309877	444	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
16	scaffold_m19_p_6 104952941 104951937	336	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
17	scaffold_m19_p_4 125238932 125238276	402	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m19_p_7 70448032 70448805	259	NC_001514.1	Squirrel monkey retrovirus - HLB
19	scaffold_m19_p_9 51214491 51215255	256	NC_001514.1	Squirrel monkey retrovirus - HLB
20	scaffold_m19_p_3 109232046 109232897	285	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
21	scaffold_m19_p_5 69972906 69972088	274	NC_001514.1	Squirrel monkey retrovirus - HLB
22	scaffold_m19_p_5 107111133 107110357	260	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
23	scaffold_m19_p_4 122844588 122845373	263	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.31: Reciprocal BLAST hits for the betaretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_24 102835 103383	184	NC_001502.1	Moloney murine sarcoma virus
2	scaffold_m19_p_8 18925236 18925886	546	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
3	scaffold_m19_p_30 12396 13247	546	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
4	scaffold_m19_p_21 390880 391491	205	NC_003059.1	Porcine endogenous retrovirus E
5	scaffold_m19_p_5 68511034 68511412	189	NC_001362.1	Friend murine leukemia virus FB29
6	scaffold_m19_p_ 592487469 92487769	299	NP_955585.1	Moloney Murine Leukemia Virus

Table 8.32: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_5 65374988 65377579	1704	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m19_p_32 115094 113763	445	NC_007815.2	PreXMRV-1 provirus
3	scaffold_m19_p_1 106553048 106555159	705	NC_022517.1	Baboon endogenous virus strain M7 proviral DNA
4	scaffold_m19_p_21 572338 570215	1378	NC_022517.1	Baboon endogenous virus strain M7 proviral DNA
5	scaffold_m19_p_5 65905131 65902696	2380	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
6	scaffold_m19_p_4 207595156 207598281	1043	NC_009889.1	RD114 retrovirus
7	scaffold_m19_p_2 99127557 99125002	1646	NC_009889.1	RD114 retrovirus
8	scaffold_m19_p_1 51163276 51166092	2539	NC_009889.1	RD114 retrovirus
9	scaffold_m19_p_8 84209603 84212953	1276	NC_009889.1	RD114 retrovirus
10	scaffold_m19_p_12 47306899 47310312	1139	NC_009889.1	RD114 retrovirus
11	scaffold_m19_p_5 144829762 144826292	2200	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
12	scaffold_m19_p_12 46997347 46995017	1783	NC_009889.1	RD114 retrovirus
13	scaffold_m19_p_5 76745406 76748876	2240	NC_009889.1	RD114 retrovirus
14	scaffold_m19_p_11 8480449 8483919	2240	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
15	scaffold_m19_p_1 166305618 166302148	2201	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
16	scaffold_m19_p_4 39794586 39791272	1106	NC_009889.1	RD114 retrovirus
17	scaffold_m19_p_4 11080161 11083631	2234	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
18	scaffold_m19_p_6 151715595 151718441	1712	NC_009889.1	RD114 retrovirus
19	scaffold_m19_p_9 16980553 16977083	2201	NC_009889.1	RD114 retrovirus
20	scaffold_m19_p_5 167847505 167850975	2240	NC_009889.1	RD114 retrovirus
21	scaffold_m19_p_2 77393644 77390411	2663	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
22	scaffold_m19_p_10 44570966 44567523	1149	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
23	scaffold_m19_p_10 44559162 44555725	1147	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
24	scaffold_m19_p_13 654202 657774	2295	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
25	scaffold_m19_p_9 78544001 78540429	2254	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
26	scaffold_m19_p_5 92483093 92484293	1210	NP_955577.1	Feline Leukemia Virus

Table 8.33: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m19_p_13 658807 659682	527	NC_001940.1	Feline leukemia virus
2	scaffold_m19_p_9 78540132 78538531	535	NC_001940.1	Feline leukemia virus
3	scaffold_m19_p_5 65377909 65379477	524	NC_001940.1	Feline leukemia virus
4	scaffold_m19_p_2 77389719 77388127	717	NC_001885.3	Gibbon ape leukemia virus
5	scaffold_m19_p_10 44567220 44565799	627	NC_001885.3	Gibbon ape leukemia virus
6	scaffold_m19_p_5 61564493 61563027	490	NC_001885.3	Gibbon ape leukemia virus
7	scaffold_m19_p_6 163691424 163690723	235	NC_001885.3	Gibbon ape leukemia virus
8	scaffold_m19_p_10 44555369 44553813	520	NC_001940.1	Feline leukemia virus
9	scaffold_m19_p_10 46141923 46143479	520	NC_039228.1	Koala retrovirus
10	scaffold_m19_p_14 10216704 10218230	510	NC_001885.3	Gibbon ape leukemia virus
11	scaffold_m19_p_21 395095 395889	266	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
12	scaffold_m19_p_7 22974744 22975364	208	NC_022517.1	Baboon endogenous virus strain M7
13	scaffold_m19_p_2 144890990 144891754	256	NC_009889.1	RD114 retrovirus
14	scaffold_m19_p_5 65901673 65900858	273	NC_009889.1	RD114 retrovirus
15	scaffold_m19_p_12 48667805 48667095	238	NC_009889.1	RD114 retrovirus
16	scaffold_m19_p_4 122309464 122310114	218	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
17	scaffold_m19_p_10 44953102 44952470	212	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
18	scaffold_m19_p_9 50747202 50748086	296	NC_009889.1	RD114 retrovirus
19	scaffold_m19_p_3 140808335 140807199	380	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
20	scaffold_m19_p_6 66156212 66156808	200	NC_009889.1	RD114 retrovirus
21	scaffold_m19_p_5 144746093 144746755	222	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
22	scaffold_m19_p_7 21543685 21544920	413	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
23	scaffold_m19_p_12 61486248 61487066	274	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
24	scaffold_m19_p_9 82688015 82688665	218	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
25	scaffold_m19_p_1 76785979 76787106	377	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China

Table 8.34: Reciprocal BLAST hits for the gammaretrovirus Env hits in *Phyllostomus discolor* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

Alpharetrovirus-like env			
ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
scaffold_m19_p_5 92481255 92481824	191	NC_005947.1	Avian leukemia virus

Table 8.35: Reciprocal BLAST hits for the alpharetrovirus Env hits in *Phyllostomus discolor* genome. The table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match the common name

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_27 3080233 3081252	341	NC_001494.1	Jaagsiekte sheep retrovirus
2	scaffold_m20_p_27 3091494 3092513	341	NC_001494.1	Jaagsiekte sheep retrovirus
3	scaffold_m20_p_3 32386000 32386914	562	NC_001550.1	Mason-Pfizer monkey virus
4	scaffold_m20_p_5 68698163 68698762	201	NC_001494.1	Jaagsiekte sheep retrovirus
5	scaffold_m20_p_33 2486548 2487357	528	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
6	scaffold_m20_p_54 427932 427141	265	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
7	scaffold_m20_p_54 501391 501987	200	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
8	scaffold_m20_p_9 35349460 35348708	409	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
9	scaffold_m20_p_21 8106163 8105225	535	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
10	scaffold_m20_p_38 4599254 4599883	211	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
11	scaffold_m20_p_31 120192 121187	608	NC_004994.2	Enzootic nasal tumour virus of goats
12	scaffold_m20_p_12 36442103 36443557	486	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m20_p_12 51246842 51248005	389	NC_004994.2	Enzootic nasal tumour virus of goats
14	scaffold_m20_p_2 65642603 65643640	347	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m20_p_2 189062240 189064219	661	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
16	scaffold_m20_p_126 297368 298381	555	NC_001494.1	Jaagsiekte sheep retrovirus
17	scaffold_m20_p_58 388311 388919	380	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m20_p_52 2241470 2240535	313	NC_029852.1	Bovine retrovirus CH15
19	scaffold_m20_p_10 8298 8978	368	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m20_p_51 1636743 1636177	190	NC_022518.1	Human endogenous retrovirus K113
21	scaffold_m20_p_7 1750162 1749425	247	NC_004994.2	Enzootic nasal tumour virus of goats
22	scaffold_m20_p_86 322996 321884	372	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
23	scaffold_m20_p_18 26337486 26336428	543	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
24	scaffold_m20_p_16 29318603 29317986	207	NC_004994.2	Enzootic nasal tumour virus of goats
25	scaffold_m20_p_16 30684473 30685396	475	NC_001494.1	Jaagsiekte sheep retrovirus
26	scaffold_m20_p_16 37034644 37035363	241	NC_004994.2	Enzootic nasal tumour virus of goats
27	scaffold_m20_p_154 158669 157545	409	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
28	scaffold_m20_p_22 882382 881258	409	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
29	scaffold_m20_p_19 8054112 8052895	734	NC_004994.2	Enzootic nasal tumour virus of goats
30	scaffold_m20_p_1 76181091 76182245	872	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
31	scaffold_m20_p_1 127411900 127413012	871	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
32	scaffold_m20_p_14 47942087 47943211	872	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
33	scaffold_m20_p_16 23164984 23164322	222	NC_007015.1	Ovine enzootic nasal tumour virus
34	scaffold_m20_p_3 51439390 51438236	405	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
35	scaffold_m20_p_2 57523541 57524695	868	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.36: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_144 67063 68202	41	NC_007654.1	Atlantic salmon swim bladder sarcoma virus
2	scaffold_m20_p_27 3105894 3107072	379	NC_001494.1	Jaagsiekte sheep retrovirus
3	scaffold_m20_p_52 546655 545258	862	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m20_p_102 327651 326278	808	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m20_p_14 42387805 42389352	1351	NC_022518.1	Human endogenous retrovirus K113
6	scaffold_m20_p_3 65755713 65757698	1250	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m20_p_1 127413317 127415164	1118	NC_022518.1	Human endogenous retrovirus K113
8	scaffold_m20_p_27 67325 66117	396	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
9	scaffold_m20_p_27 3083377 3084699	420	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m20_p_3 97518506 97517187	778	NC_007015.1	Ovine enzootic nasal tumour virus
11	scaffold_m20_p_11 21304127 21302637	947	NC_001550.1	Mason-Pfizer monkey virus
12	scaffold_m20_p_2 189065781 189067664	1214	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m20_p_2 159986243 159984858	462	NC_001550.1	Mason-Pfizer monkey virus
14	scaffold_m20_p_144 67063 68202	374	NC_007654.1	Atlantic salmon swim bladder sarcoma virus
15	scaffold_m20_p_3 32388492 32389736	816	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
16	scaffold_m20_p_14 47944018 47946486	3117	NC_022518.1	Human endogenous retrovirus K113
17	scaffold_m20_p_2 180792013 180789794	737	NC_007015.1	Ovine enzootic nasal tumour virus
18	scaffold_m20_p_2 57525315 57527915	3275	NC_022518.1	Human endogenous retrovirus K113
19	scaffold_m20_p_15 2448295 2447162	379	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
20	scaffold_m20_p_3 41328461 41325861	1637	NC_022518.1	Human endogenous retrovirus K113
21	scaffold_m20_p_147 68333 65733	2104	NC_022518.1	Human endogenous retrovirus K113
22	scaffold_m20_p_22 880940 878007	2719	NC_022518.1	Human endogenous retrovirus K113
23	scaffold_m20_p_29 6919584 6917512	685	NC_007015.1	Ovine enzootic nasal tumour virus
24	scaffold_m20_p_3 82425798 82427423	536	NC_022518.1	Human endogenous retrovirus K113
25	scaffold_m20_p_154 156895 154727	1400	NC_022518.1	Human endogenous retrovirus K113
26	scaffold_m20_p_54 503966 505606	1489	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
27	scaffold_m20_p_54 425162 423405	960	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012

Table 8.37: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_126 300467 301129	222	NC_004994.2	Enzootic nasal tumour virus of goats
2	scaffold_m20_p_126 301152 301940	264	NC_007015.1	Ovine enzootic nasal tumour virus
3	scaffold_m20_p_76 579778 578870	304	NC_001494.1	Jaagsiekte sheep retrovirus
4	scaffold_m20_p_27 65720 65007	239	NC_001494.1	Jaagsiekte sheep retrovirus
5	scaffold_m20_p_38 3571772 3572464	232	NC_007015.1	Ovine enzootic nasal tumour virus
6	scaffold_m20_p_19 8052722 8051556	390	NC_001494.1	Jaagsiekte sheep retrovirus
7	scaffold_m20_p_8 33083374 33082802	192	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
8	scaffold_m20_p_8 65733919 65734854	313	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
9	scaffold_m20_p_2 75759023 75758109	306	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
10	scaffold_m20_p_17 8093278 8092688	198	NC_001550.1	Mason-Pfizer monkey virus
11	scaffold_m20_p_1 102175024 102175941	307	NC_001514.1	Squirrel monkey retrovirus
12	scaffold_m20_p_3 108042971 108043795	276	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m20_p_119 18447 19115	224	NC_001550.1	Mason-Pfizer monkey virus
14	scaffold_m20_p_9 35348418 35347678	248	NC_001514.1	Squirrel monkey retrovirus
15	scaffold_m20_p_58 391267 392259	430	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
16	scaffold_m20_p_2 108024511 108023447	356	NC_001514.1	Squirrel monkey retrovirus
17	scaffold_m20_p_10 13508 14485	527	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m20_p_15 2445792 2444326	490	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
19	scaffold_m20_p_2 189071562 189072929	457	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m20_p_2 189067781 189069427	550	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824

Table 8.38: Reciprocal BLAST hits for the betaretrovirus Env hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_15 37097993 37098913	577	NC_001885.3	Gibbon ape leukemia virus
2	scaffold_m20_p_40 204724 205551	277	NC_001940.1	Feline leukemia virus
3	scaffold_m20_p_73 61882 62874	332	NC_039228.1	Koala retrovirus
4	scaffold_m20_p_83 355501 354926	193	NC_022517.1	Baboon endogenous virus strain M7
5	scaffold_m20_p_98 328573 330105	782	NC_001885.3	Gibbon ape leukemia virus
6	scaffold_m20_p_47 2309549 2310352	269	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
7	scaffold_m20_p_83 62529 61414	373	NC_039228.1	Koala retrovirus
8	scaffold_m20_p_29 1055190 1056023	515	NC_003059.1	Porcine endogenous retrovirus E
9	scaffold_m20_p_29 6531721 6532836	373	NC_001885.3	Gibbon ape leukemia virus
10	scaffold_m20_p_29 2353590 2354858	669	NC_039228.1	Koala retrovirus
11	scaffold_m20_p_64 950962 952473	505	NC_001885.3	Gibbon ape leukemia virus
12	scaffold_m20_p_15 3523592 3523023	357	NC_003059.1	Porcine endogenous retrovirus E
13	scaffold_m20_p_8 28302057 28302980	309	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
14	scaffold_m20_p_17 133283 132492	265	NC_009889.1	RD114 retrovirus
15	scaffold_m20_p_93 291776 292360	196	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
16	scaffold_m20_p_60 295853 296446	199	NC_003059.1	Porcine endogenous retrovirus E
17	scaffold_m20_p_80 548553 549314	354	NC_003059.1	Porcine endogenous retrovirus E
18	scaffold_m20_p_76 825298 824786	172	NC_003059.1	Porcine endogenous retrovirus E
19	scaffold_m20_p_45 1774880 1775551	330	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
20	scaffold_m20_p_30 3384800 3383958	454	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
21	scaffold_m20_p_30 4247748 4248389	215	NC_001885.3	Gibbon ape leukemia virus
22	scaffold_m20_p_119 12928 13665	446	NC_003059.1	Porcine endogenous retrovirus E

Table 8.39: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_83 155692 157617	643	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m20_p_83 142769 144694	643	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
3	scaffold_m20_p_83 58897 57377	1008	NC_022517.1	Baboon endogenous virus strain M7
4	scaffold_m20_p_83 353095 351749	834	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
5	scaffold_m20_p_9 15071250 15072818	1028	NC_007815.2	PreXMRV-1 provirus
6	scaffold_m20_p_98 330223 331926	992	NC_001885.3	Gibbon ape leukemia virus
7	scaffold_m20_p_64 952456 955338	1743	NC_009889.1	RD114 retrovirus
8	scaffold_m20_p_7 79377089 79375677	472	NC_003059.1	Porcine endogenous retrovirus E
9	scaffold_m20_p_73 63474 65747	1399	NC_039228.1	Koala retrovirus
10	scaffold_m20_p_76 823551 821245	2041	NC_003059.1	Porcine endogenous retrovirus E
11	scaffold_m20_p_8 28302987 28306553	851	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
12	scaffold_m20_p_53 864153 865721	1028	NC_007815.2	PreXMRV-1 provirus
13	scaffold_m20_p_6 12649511 12650698	397	NC_001501.1	Moloney murine leukemia virus
14	scaffold_m20_p_63 197266 196130	380	NC_022517.1	Baboon endogenous virus strain M7
15	scaffold_m20_p_41 3606680 3607846	390	NC_039228.1	Koala retrovirus
16	scaffold_m20_p_47 2310413 2311867	940	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
17	scaffold_m20_p_47 2317033 2315522	991	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
18	scaffold_m20_p_5 4292232 4288666	3771	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
19	scaffold_m20_p_33 468158 469480	835	NC_009889.1	RD114 retrovirus
20	scaffold_m20_p_36 3373322 3369762	2032	NC_001885.3	Gibbon ape leukemia virus
21	scaffold_m20_p_38 4185504 4181962	2267	NC_001885.3	Gibbon ape leukemia virus
22	scaffold_m20_p_4 76642334 76640343	665	NC_001702.1	Murine type C retrovirus
23	scaffold_m20_p_40 205812 207551	581	NC_039228.1	Koala retrovirus
24	scaffold_m20_p_3 61967414 61963857	3362	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
25	scaffold_m20_p_32 1844155 1845441	430	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
26	scaffold_m20_p_2 125425733 125429290	2272	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
27	scaffold_m20_p_21 15282088 15283746	1039	NC_022517.1	Baboon endogenous virus strain M7
28	scaffold_m20_p_23 66727 68289	522	NC_003059.1	Porcine endogenous retrovirus E
29	scaffold_m20_p_2 95122479 95125460	995	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
30	scaffold_m20_p_16 33103964 33105568	993	NC_009889.1	RD114 retrovirus
31	scaffold_m20_p_17 1567933 1569375	482	NC_003059.1	Porcine endogenous retrovirus E
32	scaffold_m20_p_18 8271722 8274901	2077	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
33	scaffold_m20_p_19 22025296 22024052	793	NC_003059.1	Porcine endogenous retrovirus E
34	scaffold_m20_p_16 18873094 18874455	455	NC_003059.1	Porcine endogenous retrovirus E
35	scaffold_m20_p_119 15857 17152	433	NC_003059.1	Porcine endogenous retrovirus E
36	scaffold_m20_p_122 53464 54921	487	NC_001501.1	Moloney murine leukemia virus
37	scaffold_m20_p_14 47261854 47258297	3362	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
38	scaffold_m20_p_10 7844062 7845198	380	NC_009889.1	RD114 retrovirus
39	scaffold_m20_p_117 61760 60261	501	NC_007815.2	PreXMRV-1 provirus

Table 8.40: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_10 12774 13514	248	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m20_p_12 36455350 36456261	588	NC_001885.3	Gibbon ape leukemia virus
3	scaffold_m20_p_123 136449 135814	213	NC_001885.3	Gibbon ape leukemia virus
4	scaffold_m20_p_122 59750 60790	646	NC_001885.3	Gibbon ape leukemia virus
5	scaffold_m20_p_1 100825545 100827536	665	NC_001885.3	Gibbon ape leukemia virus
6	scaffold_m20_p_3 60992154 60993770	540	NC_001885.3	Gibbon ape leukemia virus
7	scaffold_m20_p_36 3369838 3367898	1187	NC_039228.1	Koala retrovirus
8	scaffold_m20_p_73 66996 67748	252	NC_001885.3	Gibbon ape leukemia virus
9	scaffold_m20_p_3 123494274 123495365	710	NC_001362.1	Friend murine leukemia virus FB29
10	scaffold_m20_p_1 184159571 184161514	1237	NC_001885.3	Gibbon ape leukemia virus
11	scaffold_m20_p_21 15283718 15285661	1237	NC_001885.3	Gibbon ape leukemia virus
12	scaffold_m20_p_83 158529 159557	592	NC_039228.1	Koala retrovirus
13	scaffold_m20_p_2 98732805 98733506	235	NC_039228.1	Koala retrovirus
14	scaffold_m20_p_2 125429209 125431158	1582	NC_001940.1	Feline leukemia virus
15	scaffold_m20_p_8 28306472 28308421	1581	NC_001940.1	Feline leukemia virus
16	scaffold_m20_p_5 4288747 4286798	680	NC_039228.1	Koala retrovirus
17	scaffold_m20_p_3 61963899 61961989	680	NC_039228.1	Koala retrovirus
18	scaffold_m20_p_14 47258378 47256429	680	NC_039228.1	Koala retrovirus
19	scaffold_m20_p_3 93134397 93132973	476	NC_001885.3	Gibbon ape leukemia virus
20	scaffold_m20_p_18 8275514 8277181	908	NC_001501.1	Moloney murine leukemia virus
21	scaffold_m20_p_2 132560291 132558588	735	NC_001885.3	Gibbon ape leukemia virus
22	scaffold_m20_p_2 99060314 99059358	320	NC_001362.1	Friend murine leukemia virus FB29
23	scaffold_m20_p_54 783975 782854	375	NC_001885.3	Gibbon ape leukemia virus
24	scaffold_m20_p_16 23158057 23158638	195	NC_039228.1	Koala retrovirus

Table 8.41: Reciprocal BLAST hits for the gammaretrovirus Env hits in Pipistrellus kuhlii genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name

	Epsilonretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_138 236617 235868	251	NC_010955.1	Xenopus laevis endogenous retrovirus Xen1
2	scaffold_m20_p_183 9343 8612	245	NC_010955.1	Xenopus laevis endogenous retrovirus Xen1

Table 8.42: Reciprocal BLAST hits for the epsilonretrovirus Gag hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Epsilonretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_183 8339 6975	831	NC_001867.1	Walleye dermal sarcoma virus

Table 8.43: Reciprocal BLAST hits for the epsilonretrovirus Pol hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Epsilonretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m20_p_1 1282145 1281498	217	NC_010955.1	Xenopus laevis endogenous retrovirus Xen1

Table 8.44: Reciprocal BLAST hits for the epsilonretrovirus Env hits in *Pipistrellus kuhlii* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_16 32218104 32217301	509	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m29_p_32 3115237 3116121	540	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m29_p_39 705 4	437	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m29_p_39 19884 19228	220	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m29_p_6 54551027 54552334	708	NC_001494.1	Jaagsiekte sheep retrovirus
6	scaffold_m29_p_31 3751249 3752565	717	NC_001494.1	Jaagsiekte sheep retrovirus
7	scaffold_m29_p_27 8896741 8897928	739	NC_007015.1	Ovine enzootic nasal tumour virus
8	scaffold_m29_p_14 34907298 34908125	277	NC_001494.1	Jaagsiekte sheep retrovirus
9	scaffold_m29_p_14 5574902 5576098	753	NC_001494.1	Jaagsiekte sheep retrovirus
10	scaffold_m29_p_24 23809174 23810082	1275	NC_004994.2	Enzootic nasal tumour virus of goats
11	scaffold_m29_p_2 32632995 32631796	722	NC_001494.1	Jaagsiekte sheep retrovirus
12	scaffold_m29_p_25 40504104 40502893	723	NC_001494.1	Jaagsiekte sheep retrovirus
13	scaffold_m29_p_2 78902735 78901857	294	NC_022518.1	Human endogenous retrovirus K113
14	scaffold_m29_p_2 78913116 78912238	294	NC_022518.1	Human endogenous retrovirus K113
15	scaffold_m29_p_2 106728640 106727441	722	NC_001494.1	Jaagsiekte sheep retrovirus
16	scaffold_m29_p_19 11646287 11648281	438	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m29_p_21 27190047 27191231	736	NC_007015.1	Ovine enzootic nasal tumour virus
18	scaffold_m29_p_15 64046469 64044601	1327	NC_001494.1	Jaagsiekte sheep retrovirus
19	scaffold_m29_p_17 7732243 7731074	1131	NC_001550.1	Mason-Pfizer monkey virus
20	scaffold_m29_p_19 17693275 17695059	1265	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
21	scaffold_m29_p_22 4019466 4020584	374	NC_022518.1	Human endogenous retrovirus K113
22	scaffold_m29_p_23 6638760 6640760	442	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
23	scaffold_m29_p_8 60491912 60493069	665	NC_022518.1	Human endogenous retrovirus K113
24	scaffold_m29_p_8 60491912 60493069	628	NC_022518.1	Human endogenous retrovirus K113
25	scaffold_m29_p_8 60491912 60493069	632	NC_022518.1	Human endogenous retrovirus K113
26	scaffold_m29_p_22 9937578 9938315	247	NC_022518.1	Human endogenous retrovirus K113
27	scaffold_m29_p_9 19036680 19035490	715	NC_007015.1	Ovine enzootic nasal tumour virus
28	scaffold_m29_p_5 77464648 77463449	722	NC_001494.1	Jaagsiekte sheep retrovirus
29	scaffold_m29_p_9 92421507 92420320	709	NC_007015.1	Ovine enzootic nasal tumour virus
30	scaffold_m29_p_12 56595204 56593882	1429	NC_001494.1	Jaagsiekte sheep retrovirus
31	scaffold_m29_p_5 94450475 94449159	1421	NC_001494.1	Jaagsiekte sheep retrovirus
32	scaffold_m29_p_12 15421608 15419674	895	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
33	scaffold_m29_p_12 1313361 1315361	442	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
34	scaffold_m29_p_10 78225754 78224567	709	NC_001494.1	Jaagsiekte sheep retrovirus
35	scaffold_m29_p_2 3980402 3979248	1203	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
36	scaffold_m29_p_20 512531 513535	336	NC_022518.1	Human endogenous retrovirus K113
37	scaffold_m29_p_9 82588300 82587200	889	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
38	scaffold_m29_p_9 82588300 82587200	889	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
39	scaffold_m29_p_10 87509859 87508672	715	NC_001494.1	Jaagsiekte sheep retrovirus
40	scaffold_m29_p_10 41810855 41809785	667	NC_007015.1	Ovine enzootic nasal tumour virus
41	scaffold_m29_p_12 83509921 83509199	389	NC_001494.1	Jaagsiekte sheep retrovirus
42	scaffold_m29_p_3 15509067 15510266	748	NC_007015.1	Ovine enzootic nasal tumour virus
43	scaffold_m29_p_3 21475291 21473291	917	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
44	scaffold_m29_p_3 45673246 45672059	714	NC_001494.1	Jaagsiekte sheep retrovirus
45	scaffold_m29_p_3 103617114 103615945	1210	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
46	scaffold_m29_p_6 90490318 90491505	739	NC_007015.1	Ovine enzootic nasal tumour virus
47	scaffold_m29_p_6 32085603 32086772	627	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
48	scaffold_m29_p_6 90622105 90623307	750	NC_007015.1	Ovine enzootic nasal tumour virus
49	scaffold_m29_p_5 103763784 103764956	724	NC_007015.1	Ovine enzootic nasal tumour virus
50	scaffold_m29_p_12 248014 249933	299	NC_007015.1	Ovine enzootic nasal tumour virus

Table 8.45: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m29_p_12 2062326 2061145	714	NC_007015.1	Ovine enzootic nasal tumour virus
52	scaffold_m29_p_4 28572555 28573496	607	NC_007015.1	Ovine enzootic nasal tumour virus
53	scaffold_m29_p_4 59251768 59252700	756	NC_022518.1	Human endogenous retrovirus K113
54	scaffold_m29_p_4 11558620 11557304	1412	NC_001494.1	Jaagsiekte sheep retrovirus
55	scaffold_m29_p_4 14034013 14035329	714	NC_001494.1	Jaagsiekte sheep retrovirus
56	scaffold_m29_p_4 45686571 45684571	917	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
57	scaffold_m29_p_4 67364189 67365505	712	NC_001494.1	Jaagsiekte sheep retrovirus
58	scaffold_m29_p_7 8281562 8279946	623	NC_022518.1	Human endogenous retrovirus K113
59	scaffold_m29_p_7 1119988 1118792	733	NC_001494.1	Jaagsiekte sheep retrovirus
60	scaffold_m29_p_18 20218305 20219501	747	NC_001494.1	Jaagsiekte sheep retrovirus
61	scaffold_m29_p_4 64114681 64115880	748	NC_007015.1	Ovine enzootic nasal tumour virus
62	scaffold_m29_p_7 34902298 34904253	413	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
63	scaffold_m29_p_7 47291117 47292310	736	NC_001494.1	Jaagsiekte sheep retrovirus
64	scaffold_m29_p_18 595942 594626	1779	NC_001494.1	Jaagsiekte sheep retrovirus
65	scaffold_m29_p_11 29691354 29690038	1407	NC_001494.1	Jaagsiekte sheep retrovirus
66	scaffold_m29_p_11 50611623 50612939	714	NC_001494.1	Jaagsiekte sheep retrovirus
67	scaffold_m29_p_11 77225441 77224551	776	NC_022518.1	Human endogenous retrovirus K113
68	scaffold_m29_p_1 27079164 27078025	641	NC_022518.1	Human endogenous retrovirus K113
69	scaffold_m29_p_1 22538490 22540475	444	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
70	scaffold_m29_p_1 27453770 27454954	738	NC_007015.1	Ovine enzootic nasal tumour virus
71	scaffold_m29_p_1 29455928 29457115	746	NC_007015.1	Ovine enzootic nasal tumour virus
72	scaffold_m29_p_1 30069354 30070241	548	NC_004994.2	Enzootic nasal tumour virus of goats
73	scaffold_m29_p_1 30288137 30287073	629	NC_022518.1	Human endogenous retrovirus K113
74	scaffold_m29_p_1 34740104 34741291	739	NC_007015.1	Ovine enzootic nasal tumour virus
75	scaffold_m29_p_1 36614339 36615526	736	NC_007015.1	Ovine enzootic nasal tumour virus
76	scaffold_m29_p_1 41533703 41534914	751	NC_001494.1	Jaagsiekte sheep retrovirus
77	scaffold_m29_p_1 52734312 52735499	739	NC_007015.1	Ovine enzootic nasal tumour virus
78	scaffold_m29_p_1 29380048 29378861	714	NC_001494.1	Jaagsiekte sheep retrovirus
79	scaffold_m29_p_1 30952981 30951770	723	NC_001494.1	Jaagsiekte sheep retrovirus
80	scaffold_m29_p_1 30344482 30343727	454	NC_001494.1	Jaagsiekte sheep retrovirus
81	scaffold_m29_p_1 74527255 74528442	739	NC_007015.1	Ovine enzootic nasal tumour virus
82	scaffold_m29_p_1 36592574 36593320	898	NC_001494.1	Jaagsiekte sheep retrovirus
83	scaffold_m29_p_1 40994640 40993714	609	NC_007015.1	Ovine enzootic nasal tumour virus
84	scaffold_m29_p_1 37476626 37474626	917	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
85	scaffold_m29_p_1 39525140 39523464	630	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
86	scaffold_m29_p_1 26797480 26798604	1083	NC_007015.1	Ovine enzootic nasal tumour virus
87	scaffold_m29_p_1 53469784 53470971	739	NC_007015.1	Ovine enzootic nasal tumour virus
88	scaffold_m29_p_1 30354489 30353302	715	NC_001494.1	Jaagsiekte sheep retrovirus
89	scaffold_m29_p_1 31668953 31667766	715	NC_001494.1	Jaagsiekte sheep retrovirus
90	scaffold_m29_p_1 34999884 34998697	715	NC_001494.1	Jaagsiekte sheep retrovirus
91	scaffold_m29_p_1 33006743 33005574	1125	NC_001550.1	Mason-Pfizer monkey virus
92	scaffold_m29_p_1 35887096 35885909	715	NC_007015.1	Ovine enzootic nasal tumour virus
93	scaffold_m29_p_1 35997911 35996868	349	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
94	scaffold_m29_p_1 39599280 39598093	715	NC_007015.1	Ovine enzootic nasal tumour virus
95	scaffold_m29_p_1 39207134 39205350	700	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
96	scaffold_m29_p_1 43702964 43704964	442	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
97	scaffold_m29_p_1 45538590 45537421	1206	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
98	scaffold_m29_p_1 73107867 73109315	1146	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
99	scaffold_m29_p_1 38829375 38828497	491	NC_001494.1	Jaagsiekte sheep retrovirus
100	scaffold_m29_p_1 42017319 42016423	500	NC_001494.1	Jaagsiekte sheep retrovirus

Table 8.45: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m29_p_1 45642045 45640858	715	NC_001494.1	Jaagsiekte sheep retrovirus
102	scaffold_m29_p_1 45981779 45982948	622	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
103	scaffold_m29_p_1 51984701 51983514	728	NC_004994.2	Enzootic nasal tumour virus of goats
104	scaffold_m29_p_1 84022092 84020905	714	NC_001494.1	Jaagsiekte sheep retrovirus
105	scaffold_m29_p_1 36028915 36030915	444	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
106	scaffold_m29_p_1 55528254 55530065	438	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
107	scaffold_m29_p_1 45670945 45672729	1265	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
108	scaffold_m29_p_1 50363938 50365092	612	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
109	scaffold_m29_p_1 73432600 73430603	912	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
110	scaffold_m29_p_1 92463849 92465018	618	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
111	scaffold_m29_p_13 43919585 43920772	736	NC_007015.1	Ovine enzootic nasal tumour virus
112	scaffold_m29_p_13 46036481 46037692	755	NC_001494.1	Jaagsiekte sheep retrovirus
113	scaffold_m29_p_13 28005275 28004127	674	NC_022518.1	Human endogenous retrovirus K113
114	scaffold_m29_p_13 43004707 43003520	714	NC_001494.1	Jaagsiekte sheep retrovirus
115	scaffold_m29_p_13 40720658 40721602	882	NC_007015.1	Ovine enzootic nasal tumour virus
116	scaffold_m29_p_13 44759595 44758408	709	NC_001494.1	Jaagsiekte sheep retrovirus
117	scaffold_m29_p_13 45201812 45200625	715	NC_007015.1	Ovine enzootic nasal tumour virus
118	scaffold_m29_p_13 46090832 46089645	714	NC_001494.1	Jaagsiekte sheep retrovirus
119	scaffold_m29_p_13 44211819 44210650	1131	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
120	scaffold_m29_p_13 42604326 42605795	1151	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.45: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_8 60494120 60495601	487	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m29_p_25 36640476 36643031	1522	NC_022518.1	Human endogenous retrovirus K113
3	scaffold_m29_p_13 50302685 50304400	1025	NC_022518.1	Human endogenous retrovirus K113
4	scaffold_m29_p_22 9938457 9939710	415	NC_022518.1	Human endogenous retrovirus K113
5	scaffold_m29_p_1 30285367 30283601	817	NC_022518.1	Human endogenous retrovirus K113
6	scaffold_m29_p_22 373882 371540	1534	NC_022518.1	Human endogenous retrovirus K113
7	scaffold_m29_p_13 28002728 28000908	1124	NC_022518.1	Human endogenous retrovirus K113
8	scaffold_m29_p_11 77205003 77203693	1198	NC_022518.1	Human endogenous retrovirus K113
9	scaffold_m29_p_13 19610452 19609184	420	NC_022518.1	Human endogenous retrovirus K113
10	scaffold_m29_p_18 53127826 53124797	1451	NC_007015.1	Ovine enzootic nasal tumour virus
11	scaffold_m29_p_13 29526630 29528234	881	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
12	scaffold_m29_p_1 31098581 31101337	816	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
13	scaffold_m29_p_1 31114887 31117898	1314	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
14	scaffold_m29_p_1 22541557 22542939	462	NC_001550.1	Mason-Pfizer monkey virus
15	scaffold_m29_p_18 30707192 30708496	866	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
16	scaffold_m29_p_1 92466635 92468332	1495	NC_001550.1	Mason-Pfizer monkey virus
17	scaffold_m29_p_19 11649982 11651235	785	NC_001550.1	Mason-Pfizer monkey virus
18	scaffold_m29_p_1 44389987 44387981	1310	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
19	scaffold_m29_p_1 44255016 44253010	1310	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
20	scaffold_m29_p_12 1316288 1317700	941	NC_001550.1	Mason-Pfizer monkey virus
21	scaffold_m29_p_19 17696754 17698403	1499	NC_001550.1	Mason-Pfizer monkey virus
22	scaffold_m29_p_1 36032396 36034252	2365	NC_001550.1	Mason-Pfizer monkey virus
23	scaffold_m29_p_23 6641500 6644100	1663	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
24	scaffold_m29_p_4 45683089 45681230	1187	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
25	scaffold_m29_p_12 15418194 15416335	1200	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
26	scaffold_m29_p_1 83969091 83967703	915	NC_001550.1	Mason-Pfizer monkey virus
27	scaffold_m29_p_12 250682 253264	861	NC_007015.1	Ovine enzootic nasal tumour virus
28	scaffold_m29_p_1 40992502 40990337	1232	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
29	scaffold_m29_p_4 56651605 56653023	1314	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
30	scaffold_m29_p_1 41535666 41538257	860	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012

Table 8.46: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
31	scaffold_m29_p_25 40502141 40499550	864	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
32	scaffold_m29_p_21 27582125 27579978	1397	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
33	scaffold_m29_p_1 30951020 30948552	1640	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
34	scaffold_m29_p_1 28176588 28178612	1461	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
35	scaffold_m29_p_1 31079121 31077739	461	NC_007015.1	Ovine enzootic nasal tumour virus
36	scaffold_m29_p_1 27455697 27458003	1401	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
37	scaffold_m29_p_13 46038426 46041029	1667	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
38	scaffold_m29_p_12 56593142 56590545	1648	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
39	scaffold_m29_p_4 67366245 67368578	1532	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
40	scaffold_m29_p_4 14036069 14038654	1637	NC_001494.1	Jaagsiekte sheep retrovirus
41	scaffold_m29_p_6 54553074 54555671	1649	NC_001494.1	Jaagsiekte sheep retrovirus
42	scaffold_m29_p_25 2916977 2919562	1647	NC_001494.1	Jaagsiekte sheep retrovirus
43	scaffold_m29_p_11 29689298 29686701	1647	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
44	scaffold_m29_p_4 11556564 11554231	1520	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
45	scaffold_m29_p_18 593886 591289	1648	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
46	scaffold_m29_p_1 30515386 30513683	1087	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
47	scaffold_m29_p_1 35214054 35211907	1379	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
48	scaffold_m29_p_1 40703592 40705967	1563	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
49	scaffold_m29_p_1 27158226 27156049	1422	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
50	scaffold_m29_p_1 29050786 29048921	622	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
51	scaffold_m29_p_1 41360952 41363114	787	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
52	scaffold_m29_p_1 39596584 39595142	481	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
53	scaffold_m29_p_17 46543392 46541932	812	NC_007015.1	Ovine enzootic nasal tumour virus
54	scaffold_m29_p_46 29275 27233	3344	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
55	scaffold_m29_p_1 52736593 52738461	622	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
56	scaffold_m29_p_10 41808193 41806724	818	NC_007015.1	Ovine enzootic nasal tumour virus
57	scaffold_m29_p_1 31088452 31090077	1572	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
58	scaffold_m29_p_1 31104728 31106380	1461	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
59	scaffold_m29_p_21 25721081 25722934	1195	NC_007015.1	Ovine enzootic nasal tumour virus
60	scaffold_m29_p_12 83508012 83506363	1052	NC_007015.1	Ovine enzootic nasal tumour virus

Table 8.46: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_18 30709621 30710697	360	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
2	scaffold_m29_p_7 34905592 34907091	501	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
3	scaffold_m29_p_46 27222 25654	524	NC_004994.2	Enzootic nasal tumour virus of goats
4	scaffold_m29_p_25 36325437 36326159	242	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
5	scaffold_m29_p_9 82585960 82584362	534	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
6	scaffold_m29_p_2 987271 988485	406	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
7	scaffold_m29_p_2 3977350 3976418	312	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
8	scaffold_m29_p_4 45681144 45679465	561	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
9	scaffold_m29_p_6 32088111 32089382	425	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
10	scaffold_m29_p_12 13648246 13646987	421	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
11	scaffold_m29_p_12 15416249 15414570	561	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
12	scaffold_m29_p_12 4791720 4791091	211	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
13	scaffold_m29_p_34 1817288 1817872	196	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
14	scaffold_m29_p_7 3282217 3283107	298	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
15	scaffold_m29_p_8 6525003 6525800	267	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
16	scaffold_m29_p_12 1318769 1319404	213	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m29_p_19 11651321 11653000	561	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
18	scaffold_m29_p_19 17698489 17700162	559	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
19	scaffold_m29_p_21 5358921 5360600	561	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.47: Reciprocal BLAST hits for the betaretrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_34 1815438 1816112	226	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m29_p_34 1863932 1862937	333	NC_001885.3	Gibbon ape leukemia virus
3	scaffold_m29_p_6 98432115 98431339	374	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
4	scaffold_m29_p_12 9023938 9024717	392	NC_001362.1	Friend murine leukemia virus FB29
5	scaffold_m29_p_9 18073 19440	482	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus
6	scaffold_m29_p_13 46309047 46308052	384	NC_001362.1	Friend murine leukemia virus FB29

Table 8.48: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_17 33326080 33328842	2555	NC_009889.1	RD114 retrovirus
2	scaffold_m29_p_22 1326479 1323456	1977	NC_009889.1	RD114 retrovirus
3	scaffold_m29_p_2 51341912 51344458	849	NC_007815.2	PreXMRV-1 provirus
4	scaffold_m29_p_17 19935249 19937669	1431	NC_009889.1	RD114 retrovirus
5	scaffold_m29_p_14 5298067 5295458	868	NC_007815.2	PreXMRV-1 provirus
6	scaffold_m29_p_19 25558194 25554976	1073	NC_009889.1	RD114 retrovirus
7	scaffold_m29_p_3 20117369 20120221	1860	NC_009889.1	RD114 retrovirus
8	scaffold_m29_p_24 40780669 40784148	1160	NC_009889.1	RD114 retrovirus
9	scaffold_m29_p_9 29595841 29592353	1163	NC_009889.1	RD114 retrovirus
10	scaffold_m29_p_21 28820626 28818299	1501	NC_022517.1	Baboon endogenous virus strain M7
11	scaffold_m29_p_11 89067644 89071132	1164	NC_009889.1	RD114 retrovirus
12	scaffold_m29_p_20 38842640 38839452	1064	NC_009889.1	RD114 retrovirus
13	scaffold_m29_p_27 5270643 5274131	1164	NC_009889.1	RD114 retrovirus
14	scaffold_m29_p_20 37396286 37393098	1064	NC_009889.1	RD114 retrovirus
15	scaffold_m29_p_15 10846404 10844893	504	NC_029853.1	Mus musculus mobilized endogenous polytropic provirus clone 15
16	scaffold_m29_p_22 3638817 3635329	1164	NC_009889.1	RD114 retrovirus
17	scaffold_m29_p_19 30334260 30331693	1690	NC_009889.1	RD114 retrovirus
18	scaffold_m29_p_14 27394654 27397155	1476	NC_009889.1	RD114 retrovirus
19	scaffold_m29_p_34 1862798 1859310	1164	NC_009889.1	RD114 retrovirus
20	scaffold_m29_p_14 53433676 53430572	1036	NC_009889.1	RD114 retrovirus
21	scaffold_m29_p_4 9652287 9655775	1164	NC_009889.1	RD114 retrovirus
22	scaffold_m29_p_17 20301024 20304512	1164	NC_009889.1	RD114 retrovirus
23	scaffold_m29_p_14 53421171 53424659	1164	NC_009889.1	RD114 retrovirus
24	scaffold_m29_p_14 14410274 14406786	1163	NC_009889.1	RD114 retrovirus
25	scaffold_m29_p_13 36130856 36127368	1164	NC_009889.1	RD114 retrovirus
26	scaffold_m29_p_4 45763726 45767214	1164	NC_009889.1	RD114 retrovirus
27	scaffold_m29_p_24 9633397 9636885	1164	NC_009889.1	RD114 retrovirus
28	scaffold_m29_p_14 14535359 14538847	1164	NC_009889.1	RD114 retrovirus
29	scaffold_m29_p_13 21882592 21879104	1164	NC_009889.1	RD114 retrovirus
30	scaffold_m29_p_7 95139900 95136412	1164	NC_009889.1	RD114 retrovirus
31	scaffold_m29_p_5 97019647 97023135	1164	NC_009889.1	RD114 retrovirus
32	scaffold_m29_p_5 1270897 1267409	1164	NC_009889.1	RD114 retrovirus
33	scaffold_m29_p_11 73017779 73021267	1162	NC_009889.1	RD114 retrovirus
34	scaffold_m29_p_3 9727177 9723689	1164	NC_009889.1	RD114 retrovirus
35	scaffold_m29_p_1 61399012 61402500	1164	NC_009889.1	RD114 retrovirus
36	scaffold_m29_p_3 7379147 7382620	1159	NC_009889.1	RD114 retrovirus
37	scaffold_m29_p_15 10848386 10846443	649	NC_009889.1	RD114 retrovirus
38	scaffold_m29_p_31 3746155 3747882	1472	NC_003059.1	Porcine endogenous retrovirus E
39	scaffold_m29_p_12 7994056 7991768	1226	NC_001362.1	Friend murine leukemia virus FB29
40	scaffold_m29_p_16 64118594 64116963	1064	NC_003059.1	Porcine endogenous retrovirus E
41	scaffold_m29_p_22 4157642 4155567	1330	NC_003059.1	Porcine endogenous retrovirus E
42	scaffold_m29_p_18 41568665 41567232	940	NC_003059.1	Porcine endogenous retrovirus E
43	scaffold_m29_p_17 5480731 5483538	1833	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China

Table 8.49: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_7 95136401 95134545	1135	NC_001885.3	Gibbon ape leukemia virus
2	scaffold_m29_p_9 92435492 92434344	707	NC_039228.1	Koala retrovirus
3	scaffold_m29_p_7 58665036 58666892	1138	NC_001885.3	Gibbon ape leukemia virus
4	scaffold_m29_p_22 33774519 33772663	1135	NC_001885.3	Gibbon ape leukemia virus
5	scaffold_m29_p_31 3755995 3756816	275	NC_009889.1	RD114 retrovirus
6	scaffold_m29_p_8 60935019 60936566	953	NC_001885.3	Gibbon ape leukemia virus
7	scaffold_m29_p_21 28816161 28815271	298	NC_039228.1	Koala retrovirus
8	scaffold_m29_p_6 98493906 98495762	1144	NC_001885.3	Gibbon ape leukemia virus
9	scaffold_m29_p_21 15141614 15143470	1138	NC_001885.3	Gibbon ape leukemia virus
10	scaffold_m29_p_24 40784168 40786024	1130	NC_001885.3	Gibbon ape leukemia virus
11	scaffold_m29_p_20 6453713 6451857	1135	NC_001885.3	Gibbon ape leukemia virus
12	scaffold_m29_p_14 14538858 14540714	1138	NC_001885.3	Gibbon ape leukemia virus
13	scaffold_m29_p_14 14406775 14404919	1135	NC_001885.3	Gibbon ape leukemia virus
14	scaffold_m29_p_14 27398154 27400010	1136	NC_001885.3	Gibbon ape leukemia virus
15	scaffold_m29_p_14 53424670 53426526	1144	NC_001885.3	Gibbon ape leukemia virus
16	scaffold_m29_p_14 5295447 5293600	1128	NC_001885.3	Gibbon ape leukemia virus
17	scaffold_m29_p_14 15928568 15930424	1138	NC_001885.3	Gibbon ape leukemia virus
18	scaffold_m29_p_14 53430176 53428320	1136	NC_001885.3	Gibbon ape leukemia virus
19	scaffold_m29_p_14 55021901 55023757	1134	NC_001885.3	Gibbon ape leukemia virus
20	scaffold_m29_p_27 5274142 5275998	1138	NC_001885.3	Gibbon ape leukemia virus
21	scaffold_m29_p_27 9031675 9029819	1132	NC_001362.1	Friend murine leukemia virus FB29
22	scaffold_m29_p_15 10844882 10843026	1149	NC_001885.3	Gibbon ape leukemia virus
23	scaffold_m29_p_24 9636896 9638752	1135	NC_001885.3	Gibbon ape leukemia virus
24	scaffold_m29_p_2 3320256 3322112	1138	NC_001885.3	Gibbon ape leukemia virus
25	scaffold_m29_p_2 51344478 51346334	1146	NC_001885.3	Gibbon ape leukemia virus
26	scaffold_m29_p_9 29591320 29590487	503	NC_039228.1	Koala retrovirus
27	scaffold_m29_p_25 35543609 35545465	1138	NC_001362.1	Friend murine leukemia virus FB29
28	scaffold_m29_p_4 45767225 45769081	1138	NC_001885.3	Gibbon ape leukemia virus
29	scaffold_m29_p_19 25554609 25552753	1136	NC_001885.3	Gibbon ape leukemia virus
30	scaffold_m29_p_19 30330762 30328912	1130	NC_001885.3	Gibbon ape leukemia virus
31	scaffold_m29_p_12 52228485 52226629	1135	NC_001362.1	Friend murine leukemia virus FB29
32	scaffold_m29_p_7 10904469 10906325	1153	NC_001885.3	Gibbon ape leukemia virus
33	scaffold_m29_p_34 1859299 1857449	1130	NC_001885.3	Gibbon ape leukemia virus
34	scaffold_m29_p_22 3635318 3633462	1136	NC_001885.3	Gibbon ape leukemia virus
35	scaffold_m29_p_12 9028587 9029393	270	NC_022517.1	Baboon endogenous virus strain M7
36	scaffold_m29_p_4 9655786 9657642	1138	NC_001885.3	Gibbon ape leukemia virus
37	scaffold_m29_p_6 6640598 6638742	1137	NC_001885.3	Gibbon ape leukemia virus
38	scaffold_m29_p_12 1319403 1320377	442	NC_009889.1	RD114 retrovirus

Table 8.50: Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Alpharetrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_23 15398933 15399532	201	NC_001408.1	Avian leukosis virus - RSA
2	scaffold_m29_p_17 30499439 30498843	198	NC_001408.1	Avian leukosis virus - RSA
3	scaffold_m29_p_13 24821733 24822323	358	NC_005947.1	Avian endogenous retrovirus EAV-HP
4	scaffold_m29_p_1 98456109 98456780	223	NC_001408.1	Avian leukosis virus - RSA
5	scaffold_m29_p_5 38325916 38326485	189	NC_008094.1	Y73 sarcoma virus
6	scaffold_m29_p_11 8417924 8417181	367	NC_001408.1	Avian leukosis virus - RSA
7	scaffold_m29_p_3 75623361 75623927	344	NC_008094.1	Y73 sarcoma virus
8	scaffold_m29_p_10 81727143 81726520	394	NC_008094.1	Y73 sarcoma virus
9	scaffold_m29_p_13 32920736 32920164	192	NC_008094.1	Y73 sarcoma virus
10	scaffold_m29_p_23 10356828 10357397	191	NC_008094.1	Y73 sarcoma virus

Table 8.51: Reciprocal BLAST hits for the alpharetrovirus Env hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Spumaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m29_p_33 1104848 1107442	1848	NC_039023.1	Brown greater galago prosimian foamy virus
2	scaffold_m29_p_16 65640774 65639380	1710	NC_039025.1	Eastern chimpanzee simian foamy virus clone pHSRV13
3	scaffold_m29_p_31 40806 38494	1395	NC_043194.1	Walleye epidermal hyperplasia virus type 1
4	scaffold_m29_p_16 65774700 65772583	1140	NC_010820.1	Simian foamy virus 3
5	scaffold_m29_p_15 2140949 2142739	595	NC_039025.1	Eastern chimpanzee simian foamy virus clone pHSRV13
6	scaffold_m29_p_29 368435 365958	1492	NC_010820.1	Simian foamy virus 3
7	scaffold_m29_p_31 116032 117111	361	NC_039031.1	Yellow-breasted capuchin simian foamy virus isolate Z17
8	scaffold_m29_p_31 101136 102215	361	NC_039031.1	Yellow-breasted capuchin simian foamy virus isolate Z17
9	scaffold_m29_p_26 74044 75168	376	NC_039031.1	Yellow-breasted capuchin simian foamy virus isolate Z17
10	scaffold_m29_p_16 13024244 13025329	363	NC_039022.1	Yellow-breasted capuchin simian foamy virus isolate Z17
11	scaffold_m29_p_15 9820554 9817594	1700	NC_043194.1	Walleye epidermal hyperplasia virus type 1

Table 8.52: Reciprocal BLAST hits for the spumaretrovirus Pol hits in *Rhinolophus ferrumequinum* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_14 38451579 38449522	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
2	scaffold_m13_p_14 47191230 47189173	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
3	scaffold_m13_p_16 45036602 45037615	339	NC_001550.1	Mason-Pfizer monkey virus
4	scaffold_m13_p_17 22889927 22891972	683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
5	scaffold_m13_p_10 26502024 26499967	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m13_p_2 27961708 27963765	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
7	scaffold_m13_p_2 31781801 31779744	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
8	scaffold_m13_p_12 38856406 38858496	698	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
9	scaffold_m13_p_12 42012112 42014169	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
10	scaffold_m13_p_16 31939543 31941597	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m13_p_2 50767584 50769638	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
12	scaffold_m13_p_2 62370876 62368981	633	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m13_p_3 117930975 117933032	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
14	scaffold_m13_p_3 169180916 169182973	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m13_p_3 169489138 169487081	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
16	scaffold_m13_p_3 68350618 68351775	1081	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m13_p_2 165731312 165733369	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m13_p_2 71521042 71520140	591	NC_007015.1	Ovine enzootic nasal tumour virus
19	scaffold_m13_p_2 79946997 79944940	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m13_p_2 72987059 72988789	1125	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
21	scaffold_m13_p_3 107696793 107697809	922	NC_001550.1	Mason-Pfizer monkey virus
22	scaffold_m13_p_4 47458509 47456452	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
23	scaffold_m13_p_6 63542550 63544607	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
24	scaffold_m13_p_4 60008528 60007485	684	NC_007015.1	Ovine enzootic nasal tumour virus
25	scaffold_m13_p_4 96769986 96772043	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
26	scaffold_m13_p_4 75086545 75085664	748	NC_022518.1	Human endogenous retrovirus K113
27	scaffold_m13_p_4 75086545 75085664	748	NC_022518.1	Human endogenous retrovirus K113
28	scaffold_m13_p_8 37098961 37101018	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
29	scaffold_m13_p_8 48246523 48244682	615	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
30	scaffold_m13_p_8 83899451 83901505	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
31	scaffold_m13_p_10 84689494 84691548	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
32	scaffold_m13_p_10 87816065 87818122	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
33	scaffold_m13_p_12 93986270 93984216	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
34	scaffold_m13_p_2 135183998 135181941	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
35	scaffold_m13_p_2 176351122 176349068	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
36	scaffold_m13_p_4 114852347 114850293	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
37	scaffold_m13_p_4 149373520 149375577	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
38	scaffold_m13_p_4 104719000 104717276	1114	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
39	scaffold_m13_p_8 101975689 101977746	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
40	scaffold_m13_p_8 109494113 109496167	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
41	scaffold_m13_p_18 15466387 15467448	355	NC_022518.1	Human endogenous retrovirus K113
42	scaffold_m13_p_5 21210478 21209123	728	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
43	scaffold_m13_p_11 14692166 14690511	978	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
44	scaffold_m13_p_7 55452388 55450832	943	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
45	scaffold_m13_p_11 6016062 6014008	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
46	scaffold_m13_p_11 14631488 14633581	699	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
47	scaffold_m13_p_11 14673647 14671584	689	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
48	scaffold_m13_p_11 44419597 44421651	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
49	scaffold_m13_p_2 41769836 41767875	655	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
50	scaffold_m13_p_7 59717473 59716217	420	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.53: Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m13_p_11 14737959 14738738	731	NC_022518.1	Human endogenous retrovirus K113
52	scaffold_m13_p_11 36391349 36389694	978	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
53	scaffold_m13_p_11 81519693 81521747	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
54	scaffold_m13_p_11 86433545 86435602	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
55	scaffold_m13_p_13 3011373 3009643	1049	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
56	scaffold_m13_p_13 12461471 12463525	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
57	scaffold_m13_p_15 60133200 60131146	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
58	scaffold_m13_p_15 22267046 22267954	1078	NC_022518.1	Human endogenous retrovirus K113
59	scaffold_m13_p_15 28476529 28478259	1125	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
60	scaffold_m13_p_5 31131736 31129679	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
61	scaffold_m13_p_13 35931513 35929456	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
62	scaffold_m13_p_1 36723292 36721238	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
63	scaffold_m13_p_7 40929644 40927590	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
64	scaffold_m13_p_7 53358448 53360541	699	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
65	scaffold_m13_p_7 62199343 62197286	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
66	scaffold_m13_p_7 66056687 66057553	987	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
67	scaffold_m13_p_5 66970422 66972317	1563	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
68	scaffold_m13_p_5 83642024 83644078	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
69	scaffold_m13_p_7 102066053 102063996	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
70	scaffold_m13_p_7 111747748 111749841	699	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
71	scaffold_m13_p_11 50304024 50302129	1226	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
72	scaffold_m13_p_5 88368397 88370340	1252	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
73	scaffold_m13_p_5 56746161 56747132	325	NC_001494.1	Jaagsiekte sheep retrovirus
74	scaffold_m13_p_1 88544010 88545536	510	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
75	scaffold_m13_p_11 95230531 95232588	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
76	scaffold_m13_p_1 109268362 109269840	812	NC_001494.1	Jaagsiekte sheep retrovirus
77	scaffold_m13_p_7 111441134 111443191	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
78	scaffold_m13_p_1 169767650 169765596	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
79	scaffold_m13_p_1 155602040 155600883	696	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
80	scaffold_m13_p_1 173310304 173312361	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
81	scaffold_m13_p_1 130465324 130465926	202	NC_001494.1	Jaagsiekte sheep retrovirus
82	scaffold_m13_p_9 49419842 49419009	522	NC_001550.1	Mason-Pfizer monkey virus
83	scaffold_m13_p_9 78688216 78689025	515	NC_007015.1	Ovine enzootic nasal tumour virus
84	scaffold_m13_p_9 10974394 10972340	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
85	scaffold_m13_p_9 26144978 26142921	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
86	scaffold_m13_p_9 39178005 39177232	259	NC_022518.1	Human endogenous retrovirus K113
87	scaffold_m13_p_9 82751532 82753592	688	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
88	scaffold_m13_p_9 65367917 65366829	1040	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
89	scaffold_m13_p_9 84797595 84799652	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
90	scaffold_m13_p_9 26430008 26430817	1072	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
91	scaffold_m13_p_9 41819688 41820944	420	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
92	scaffold_m13_p_9 51273649 51272480	718	NC_007015.1	Ovine enzootic nasal tumour virus
93	scaffold_m13_p_9 69711958 69714012	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
94	scaffold_m13_p_9 87429573 87431630	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
95	scaffold_m13_p_9 57164234 57165061	497	NC_022518.1	Human endogenous retrovirus K113
96	scaffold_m13_p_9 81685542 81687620	694	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
97	scaffold_m13_p_9 82855519 82857156	547	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
98	scaffold_m13_p_9 88903906 88902821	678	NC_001550.1	Mason-Pfizer monkey virus
99	scaffold_m13_p_3 103046634 103048688	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
100	scaffold_m13_p_2 177602139 177600070	701	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.53: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m13_p_2 108766665 108764602	1244	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
102	scaffold_m13_p_2 80409874 80411529	743	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
103	scaffold_m13_p_6 84747793 84748818	655	NC_001550.1	Mason-Pfizer monkey virus
104	scaffold_m13_p_20 3 1532	511	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
105	scaffold_m13_p_4 4214583 4216637	686	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
106	scaffold_m13_p_6 20495926 20493869	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
107	scaffold_m13_p_6 23854561 23852504	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
108	scaffold_m13_p_19 147095 149248	1725	NC_001550.1	Mason-Pfizer monkey virus
109	scaffold_m13_p_19 1077538 1079595	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
110	scaffold_m13_p_19 1173103 1174044	315	NC_022518.1	Human endogenous retrovirus K113
111	scaffold_m13_p_10 4693394 4691337	687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
112	scaffold_m13_p_2 5436362 5438329	657	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824

Table 8.53: Continued. Reciprocal BLAST hits for the betaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_1 90061591 90060023	902	NC_022518.1	Human endogenous retrovirus K113
2	scaffold_m13_p_9 83322012 83323514	959	NC_001550.1	Mason-Pfizer monkey virus
3	scaffold_m13_p_15 25523338 25524594	809	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
4	scaffold_m13_p_7 111750620 111753166	849	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
5	scaffold_m13_p_9 81688399 81690939	846	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
6	scaffold_m13_p_9 84800538 84802589	1967	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
7	scaffold_m13_p_8 109824152 109825840	1040	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
8	scaffold_m13_p_7 55449333 55447063	3079	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
9	scaffold_m13_p_12 42015589 42017544	1246	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
10	scaffold_m13_p_5 88371204 88373825	1653	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
11	scaffold_m13_p_1 173313116 173314816	1081	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
12	scaffold_m13_p_2 50771523 50772980	960	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
13	scaffold_m13_p_11 95233343 95235979	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
14	scaffold_m13_p_5 66973072 66975708	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
15	scaffold_m13_p_2 79944185 79941549	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
16	scaffold_m13_p_10 4690582 4687958	1687	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
17	scaffold_m13_p_9 65366074 65363438	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
18	scaffold_m13_p_10 26499212 26496576	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
19	scaffold_m13_p_6 63545362 63547998	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
20	scaffold_m13_p_14 38448767 38446131	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
21	scaffold_m13_p_1 36719353 36717848	837	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
22	scaffold_m13_p_2 165734124 165736760	1682	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
23	scaffold_m13_p_3 169183728 169186364	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
24	scaffold_m13_p_3 169486326 169483690	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
25	scaffold_m13_p_17 22892727 22895363	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
26	scaffold_m13_p_2 31778989 31776353	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
27	scaffold_m13_p_6 23851749 23849113	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
28	scaffold_m13_p_13 35928701 35926065	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
29	scaffold_m13_p_11 50301374 50298738	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
30	scaffold_m13_p_2 62368226 62365590	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
31	scaffold_m13_p_4 104716521 104713885	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
32	scaffold_m13_p_7 102063241 102060605	1699	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
33	scaffold_m13_p_4 149376332 149378968	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
34	scaffold_m13_p_4 107372859 107370235	1667	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
35	scaffold_m13_p_8 101978501 101981137	1683	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
36	scaffold_m13_p_7 111443946 111446240	766	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
37	scaffold_m13_p_9 82754347 82756968	1708	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
38	scaffold_m13_p_3 103049443 103051143	1083	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
39	scaffold_m13_p_1 169764841 169762313	844	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
40	scaffold_m13_p_8 83902054 83904849	1233	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
41	scaffold_m13_p_8 37102125 37104395	920	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
42	scaffold_m13_p_2 10427358 10429886	1111	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
43	scaffold_m13_p_7 93130465 93132993	1111	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
44	scaffold_m13_p_3 117933787 117936408	1708	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
45	scaffold_m13_p_5 31128924 31126303	1711	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
46	scaffold_m13_p_7 62196737 62193909	2923	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
47	scaffold_m13_p_1 88546291 88548912	1708	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
48	scaffold_m13_p_19 1080350 1082971	1708	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
49	scaffold_m13_p_2 5439084 5441705	1708	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
50	scaffold_m13_p_14 47188418 47185797	1711	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824

Table 8.54: Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m13_p_2 41767120 41764499	1712	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
52	scaffold_m13_p_11 81522502 81525090	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
53	scaffold_m13_p_9 69714767 69717355	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
54	scaffold_m13_p_8 109496922 109499510	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
55	scaffold_m13_p_9 10971585 10968997	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
56	scaffold_m13_p_2 176348313 176345725	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
57	scaffold_m13_p_11 44422406 44424994	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
58	scaffold_m13_p_1 180437359 180439947	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
59	scaffold_m13_p_13 12464280 12466868	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
60	scaffold_m13_p_5 83644833 83647421	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
61	scaffold_m13_p_4 114849538 114846950	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
62	scaffold_m13_p_8 48243927 48241339	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
63	scaffold_m13_p_5 61446269 61443681	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
64	scaffold_m13_p_10 84692303 84694891	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
65	scaffold_m13_p_16 31942352 31944940	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
66	scaffold_m13_p_4 4217392 4219980	864	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
67	scaffold_m13_p_11 14669753 14668242	809	NC_001514.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
68	scaffold_m13_p_1 73729821 73728247	1380	NC_001514.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
69	scaffold_m13_p_3 167048652 167050757	1385	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
70	scaffold_m13_p_2 177599075 177597738	843	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
71	scaffold_m13_p_10 70382713 70380497	2153	NC_027117.1	Desmodus rotundus endogenous retrovirus isolate 824
72	scaffold_m13_p_20 3086 4690	1004	NC_001514.1	Squirrel monkey retrovirus - HLB
73	scaffold_m13_p_1 66130624 66129080	1011	NC_001514.1	Squirrel monkey retrovirus - HLB
74	scaffold_m13_p_4 94031712 94030021	563	NC_001514.1	Squirrel monkey retrovirus - HLB
75	scaffold_m13_p_13 75675907 75677481	938	NC_001514.1	Squirrel monkey retrovirus - HLB
76	scaffold_m13_p_11 84511658 84510354	435	NC_001514.1	Squirrel monkey retrovirus - HLB
77	scaffold_m13_p_3 117973321 117975012	563	NC_001514.1	Squirrel monkey retrovirus - HLB
78	scaffold_m13_p_12 86510271 86511962	563	NC_001514.1	Squirrel monkey retrovirus - HLB
79	scaffold_m13_p_6 43134930 43136537	535	NC_001514.1	Squirrel monkey retrovirus - HLB
80	scaffold_m13_p_2 164974108 164975742	1494	NC_001514.1	Squirrel monkey retrovirus - HLB
81	scaffold_m13_p_5 107188443 107187148	839	NC_001514.1	Squirrel monkey retrovirus - HLB
82	scaffold_m13_p_5 112666934 112668541	535	NC_001514.1	Squirrel monkey retrovirus - HLB
83	scaffold_m13_p_3 66314719 66316410	563	NC_001514.1	Squirrel monkey retrovirus - HLB
84	scaffold_m13_p_12 76959532 76957832	566	NC_001514.1	Squirrel monkey retrovirus - HLB
85	scaffold_m13_p_10 38385374 38387065	563	NC_001514.1	Squirrel monkey retrovirus - HLB
86	scaffold_m13_p_3 5471996 5473687	563	NC_001514.1	Squirrel monkey retrovirus - HLB
87	scaffold_m13_p_15 56882805 56881114	563	NC_001514.1	Squirrel monkey retrovirus - HLB
88	scaffold_m13_p_6 93864798 93866489	563	NC_001514.1	Squirrel monkey retrovirus - HLB
89	scaffold_m13_p_2 155636971 155638602	543	NC_001514.1	Squirrel monkey retrovirus - HLB
90	scaffold_m13_p_3 93292164 93293672	503	NC_001514.1	Squirrel monkey retrovirus - HLB
91	scaffold_m13_p_2 80411743 80413038	817	NC_001514.1	Squirrel monkey retrovirus - HLB
92	scaffold_m13_p_6 92182504 92181188	439	NC_001514.1	Squirrel monkey retrovirus - HLB
93	scaffold_m13_p_13 36613672 36615363	563	NC_001514.1	Squirrel monkey retrovirus - HLB
94	scaffold_m13_p_7 73466474 73465170	435	NC_001514.1	Squirrel monkey retrovirus - HLB
95	scaffold_m13_p_7 74573205 74574896	563	NC_001514.1	Squirrel monkey retrovirus - HLB
96	scaffold_m13_p_11 60663760 60665451	563	NC_001514.1	Squirrel monkey retrovirus - HLB
97	scaffold_m13_p_12 43052977 43054668	563	NC_001514.1	Squirrel monkey retrovirus - HLB
98	scaffold_m13_p_11 49193196 49194803	535	NC_001514.1	Squirrel monkey retrovirus - HLB
99	scaffold_m13_p_9 48800319 48798712	534	NC_001514.1	Squirrel monkey retrovirus - HLB
100	scaffold_m13_p_16 12133525 12131810	571	NC_001514.1	Squirrel monkey retrovirus - HLB

Table 8.54: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
101	scaffold_m13_p_6 92873119 92874810	563	NC_001514.1	Squirrel monkey retrovirus - HLB
102	scaffold_m13_p_1 182218940 182220136	399	NC_001514.1	Squirrel monkey retrovirus - HLB
103	scaffold_m13_p_7 67891772 67893379	534	NC_001514.1	Squirrel monkey retrovirus - HLB
104	scaffold_m13_p_2 168300700 168302208	503	NC_001514.1	Squirrel monkey retrovirus - HLB
105	scaffold_m13_p_2 137535425 137537032	535	NC_001514.1	Squirrel monkey retrovirus - HLB
106	scaffold_m13_p_4 17838568 17840175	535	NC_001514.1	Squirrel monkey retrovirus - HLB
107	scaffold_m13_p_1 171413759 171412068	562	NC_001514.1	Squirrel monkey retrovirus - HLB
108	scaffold_m13_p_15 16456064 16457755	563	NC_001514.1	Squirrel monkey retrovirus - HLB
109	scaffold_m13_p_9 12909068 12910759	563	NC_001514.1	Squirrel monkey retrovirus - HLB
110	scaffold_m13_p_13 7147226 7148917	563	NC_001514.1	Squirrel monkey retrovirus - HLB
111	scaffold_m13_p_3 107699593 107701347	1152	NC_001550.1	Mason-Pfizer monkey virus
112	scaffold_m13_p_15 29105696 29107987	1491	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
113	scaffold_m13_p_3 19439937 19438429	504	NC_001550.1	Mason-Pfizer monkey virus
114	scaffold_m13_p_17 23271086 23268936	2069	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
115	scaffold_m13_p_9 71919792 71922086	1481	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
116	scaffold_m13_p_14 34192330 34194378	1384	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
117	scaffold_m13_p_9 49638087 49640030	1794	NC_001494.1	Jaagsiekte sheep retrovirus
118	scaffold_m13_p_17 45660601 45658412	1409	NC_001494.1	Jaagsiekte sheep retrovirus
119	scaffold_m13_p_14 38121095 38122333	412	NC_004994.2	Enzootic nasal tumour virus of goats
120	scaffold_m13_p_9 78691064 78692380	1209	NC_004994.2	Enzootic nasal tumour virus of goats
121	scaffold_m13_p_11 17951291 17948913	784	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
122	scaffold_m13_p_15 26908252 26905838	1565	NC_001494.1	Jaagsiekte sheep retrovirus
123	scaffold_m13_p_15 26626514 26628193	1090	NC_004994.2	Enzootic nasal tumour virus of goats
124	scaffold_m13_p_19 269731 268298	1321	NC_001494.1	Jaagsiekte sheep retrovirus
125	scaffold_m13_p_5 56748784 56750388	530	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
126	scaffold_m13_p_7 51178623 51180308	1541	NC_001494.1	Jaagsiekte sheep retrovirus
127	scaffold_m13_p_9 51271172 51269586	527	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
128	scaffold_m13_p_9 54564399 54565991	1017	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
129	scaffold_m13_p_1 100488211 100489455	819	NC_001494.1	Jaagsiekte sheep retrovirus
130	scaffold_m13_p_1 30621347 30619827	508	NC_001494.1	Jaagsiekte sheep retrovirus
131	scaffold_m13_p_6 65028981 65027125	1233	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
132	scaffold_m13_p_8 69384229 69386049	835	NC_004994.2	Enzootic nasal tumour virus of goats
133	scaffold_m13_p_13 80082948 80080894	1236	NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012
134	scaffold_m13_p_6 65557610 65558827	800	NC_004994.2	Enzootic nasal tumour virus of goats
135	scaffold_m13_p_16 26368553 26370694	2167	NC_004994.2	Enzootic nasal tumour virus of goats

Table 8.54: Continued. Reciprocal BLAST hits for the betaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Betaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_17 23268869 23267196	559	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1
2	scaffold_m13_p_9 60990230 60988575	553	NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1

Table 8.55: Reciprocal BLAST hits for the betaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like gag			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_6 6218862 6219560	234	NC_001940.1	Feline leukemia virus
2	scaffold_m13_p_6 4920378 4919821	187	NC_009889.1	RD114 retrovirus

Table 8.56: Reciprocal BLAST hits for the gammaretrovirus Gag hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like pol			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_18 418787 417396	745	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
2	scaffold_m13_p_6 85870059 85866775	2059	NC_009889.1	RD114 retrovirus
3	scaffold_m13_p_22 156401 158926	1950	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China

Table 8.57: Reciprocal BLAST hits for the gammaretrovirus Pol hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_13 2059901 2060479	194	NC_022517.1	Baboon endogenous virus strain M7 proviral DNA
2	scaffold_m13_p_12 8220428 8220961	179	NC_005947.1	Avian endogenous retrovirus EAV-HP
3	scaffold_m13_p_17 8254948 8256315	457	NC_006934.1	Reticuloendotheliosis virus strain HA9901 from China
4	scaffold_m13_p_9 88902091 88901267	788	NC_001885.3	Gibbon ape leukemia virus
5	scaffold_m13_p_1 62018604 62019815	405	NC_001885.3	Gibbon ape leukemia virus
6	scaffold_m13_p_19 152651 154429	1125	NC_001885.3	Gibbon ape leukemia virus
7	scaffold_m13_p_2 80808056 80809027	626	NC_001885.3	Gibbon ape leukemia virus
8	scaffold_m13_p_4 47456000 47455293	464	NC_039228.1	Koala retrovirus
9	scaffold_m13_p_8 109499612 109501333	575	NC_001362.1	Friend murine leukemia virus FB29
10	scaffold_m13_p_9 87433236 87434120	540	NC_039228.1	Koala retrovirus
11	scaffold_m13_p_2 50773082 50774989	1347	NC_001362.1	Friend murine leukemia virus FB29
12	scaffold_m13_p_1 180440049 180441959	1350	NC_001362.1	Friend murine leukemia virus FB29
13	scaffold_m13_p_9 10968895 10966988	667	NC_039228.1	Koala retrovirus
14	scaffold_m13_p_8 83904951 83906858	1347	NC_001362.1	Friend murine leukemia virus FB29
15	scaffold_m13_p_11 44425096 44427003	1345	NC_001362.1	Friend murine leukemia virus FB29
16	scaffold_m13_p_16 31945042 31946949	1345	NC_001362.1	Friend murine leukemia virus FB29
17	scaffold_m13_p_4 4220082 4221989	1345	NC_001362.1	Friend murine leukemia virus FB29
18	scaffold_m13_p_11 6013454 6011562	664	NC_039228.1	Koala retrovirus
19	scaffold_m13_p_8 48241237 48239330	667	NC_039228.1	Koala retrovirus
20	scaffold_m13_p_4 114846848 114844941	667	NC_039228.1	Koala retrovirus
21	scaffold_m13_p_9 26142381 26141497	582	NC_001362.1	Friend murine leukemia virus FB29
22	scaffold_m13_p_2 176345623 176343716	667	NC_039228.1	Koala retrovirus
23	scaffold_m13_p_2 72988921 72990906	1281	NC_001501.1	Moloney murine leukemia virus
24	scaffold_m13_p_1 169761345 169760242	369	NC_039228.1	Koala retrovirus
25	scaffold_m13_p_2 41764430 41762523	669	NC_039228.1	Koala retrovirus
26	scaffold_m13_p_2 33755956 33757674	574	NC_001501.1	Moloney murine leukemia virus
27	scaffold_m13_p_11 14688684 14687806	546	NC_039228.1	Koala retrovirus
28	scaffold_m13_p_2 10430688 10431872	396	NC_001885.3	Gibbon ape leukemia virus
29	scaffold_m13_p_7 93133795 93134979	396	NC_001885.3	Gibbon ape leukemia virus
30	scaffold_m13_p_2 135181200 135180334	570	NC_039228.1	Koala retrovirus
31	scaffold_m13_p_8 109826492 109827814	442	NC_001501.1	Moloney murine leukemia virus
32	scaffold_m13_p_7 62193840 62191924	1056	NC_001362.1	Friend murine leukemia virus FB29
33	scaffold_m13_p_11 36387701 36386175	866	NC_001501.1	Moloney murine leukemia virus
34	scaffold_m13_p_14 47185728 47183812	1047	NC_001362.1	Friend murine leukemia virus FB29
35	scaffold_m13_p_19 1083040 1084956	1155	NC_001362.1	Friend murine leukemia virus FB29
36	scaffold_m13_p_9 82757037 82758953	1155	NC_001362.1	Friend murine leukemia virus FB29
37	scaffold_m13_p_1 88548981 88550897	1155	NC_001362.1	Friend murine leukemia virus FB29
38	scaffold_m13_p_7 55446994 55445078	1056	NC_001362.1	Friend murine leukemia virus FB29
39	scaffold_m13_p_7 111764673 111766463	1113	NC_001362.1	Friend murine leukemia virus FB29
40	scaffold_m13_p_2 5441774 5443690	1155	NC_001362.1	Friend murine leukemia virus FB29
41	scaffold_m13_p_8 37104464 37106380	1155	NC_001362.1	Friend murine leukemia virus FB29
42	scaffold_m13_p_7 66061189 66063063	1308	NC_001362.1	Friend murine leukemia virus FB29
43	scaffold_m13_p_11 14667882 14665999	672	NC_001885.3	Gibbon ape leukemia virus
44	scaffold_m13_p_7 59715226 59714567	221	NC_001885.3	Gibbon ape leukemia virus
45	scaffold_m13_p_1 87105453 87107324	1228	NC_001362.1	Friend murine leukemia virus FB29
46	scaffold_m13_p_9 81691317 81693203	1306	NC_001362.1	Friend murine leukemia virus FB29
47	scaffold_m13_p_9 26432827 26434611	1326	NC_001362.1	Friend murine leukemia virus FB29
48	scaffold_m13_p_7 111753544 111755430	1310	NC_001362.1	Friend murine leukemia virus FB29
49	scaffold_m13_p_6 60497737 60496082	672	NC_001362.1	Friend murine leukemia virus FB29
50	scaffold_m13_p_9 82857167 82858549	1060	NC_001362.1	Friend murine leukemia virus FB29

Table 8.58: Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Gammaretrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
51	scaffold_m13_p_2 164976115 164977290	620	NC_001885.3	Gibbon ape leukemia virus
52	scaffold_m13_p_2 27964067 27965929	1313	NC_001362.1	Friend murine leukemia virus FB29
53	scaffold_m13_p_4 107369950 107368058	634	NC_001885.3	Gibbon ape leukemia virus
54	scaffold_m13_p_11 86435904 86437766	1313	NC_001362.1	Friend murine leukemia virus FB29
55	scaffold_m13_p_8 48092982 48091084	635	NC_001885.3	Gibbon ape leukemia virus
56	scaffold_m13_p_1 182220881 182222779	1327	NC_001362.1	Friend murine leukemia virus FB29
57	scaffold_m13_p_11 49195052 49196950	1327	NC_001362.1	Friend murine leukemia virus FB29
58	scaffold_m13_p_11 60665700 60667598	1327	NC_001362.1	Friend murine leukemia virus FB29
59	scaffold_m13_p_9 12911008 12912906	1327	NC_001362.1	Friend murine leukemia virus FB29
60	scaffold_m13_p_1 171411819 171409921	636	NC_001885.3	Gibbon ape leukemia virus
61	scaffold_m13_p_2 88864347 88863481	503	NC_001885.3	Gibbon ape leukemia virus
62	scaffold_m13_p_6 63205024 63205668	394	NC_001885.3	Gibbon ape leukemia virus
63	scaffold_m13_p_9 61135407 61136216	396	NC_001885.3	Gibbon ape leukemia virus
64	scaffold_m13_p_9 54414406 54415740	699	NC_039228.1	Koala retrovirus
65	scaffold_m13_p_9 36731263 36730052	400	NC_001885.3	Gibbon ape leukemia virus
66	scaffold_m13_p_2 90121298 90119832	490	NC_001885.3	Gibbon ape leukemia virus
67	scaffold_m13_p_9 71633878 71635221	449	NC_001885.3	Gibbon ape leukemia virus
68	scaffold_m13_p_9 77017591 77016470	375	NC_001885.3	Gibbon ape leukemia virus
69	scaffold_m13_p_7 60450281 60448473	1001	NC_001940.1	Feline leukemia virus
70	scaffold_m13_p_22 159507 160847	448	NC_001885.3	Gibbon ape leukemia virus
71	scaffold_m13_p_9 72487374 72488699	443	NC_001885.3	Gibbon ape leukemia virus
72	scaffold_m13_p_4 94333982 94333137	283	NC_001940.1	Feline leukemia virus
73	scaffold_m13_p_8 38225103 38226278	393	NC_001501.1	Moloney murine leukemia virus
74	scaffold_m13_p_2 173175161 173175826	356	NC_001885.3	Gibbon ape leukemia virus
75	scaffold_m13_p_2 140345607 140346659	544	NC_001885.3	Gibbon ape leukemia virus
76	scaffold_m13_p_8 26292479 26291877	202	NC_001885.3	Gibbon ape leukemia virus

Table 8.58: Continued. Reciprocal BLAST hits for the gammaretrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

	Alpharetrovirus-like env			
	ERV assembly position	**aa len**	**Viral match ID**	**Viral match name**
1	scaffold_m13_p_5 38799838 38800560	399	NC_015116.1	Avian leukemia virus isolate SCDY1
2	scaffold_m13_p_10 79037674 79038243	191	NC_015116.1	Avian leukemia virus isolate SCDY1
3	scaffold_m13_p_15 20134926 20135489	189	NC_015116.1	Avian leukemia virus isolate SCDY1
4	scaffold_m13_p_3 70755341 70754763	194	NC_015116.1	Avian leukemia virus isolate SCDY1

Table 8.59: Reciprocal BLAST hits for the alpharetrovirus Env hits in *Rousettus aegyptiacus* genome. Table shows position of the ERV hit in the genome, ERV protein aa length, reciprocal BLAST first viral match ID, and the reciprocal BLAST first viral match common name.

9 Appendix to Chapter 4

Bat species	Assembly	Length of the genome (Mb)	N50/L50 ratio	Number of contigs in genome assembly
Myotis lucifugus	Scaffold	2034.58	64330/8574	72785
Myotis brandtii	Scaffold	2107.24	23289/24039	325414
Myotis davidii	Scaffold	2059.8	15182/36107	325280
Eptesicus fuscus	Scaffold	2026.63	21392/24180	167058
Miniopterus natalensis	Scaffold	1803.1	29777/17067	114632
Rhinolophus ferrumequinum	Chromosome	2075.77	31900497/22	289
Pteropus alecto	Scaffold	1985.98	31841/18318	170164
Pteropus vampyrus	Scaffold	2198.28	21866/22949	225433
Roussetus aegyptiacus	Scaffold	1893.6	21999313/25	271
Hipposiderios armiger	Scaffold	2236.58	39863/13917	128956
Desmodus rotundus	Scaffold	2063.81	80250/6951	84109
Phyllostomus discolor	Chromosome	2108.83	6892556/94	953
Molossus molossus	Scaffold	2315.57	22174888/30	409
Pipistrellus kuhlii	Scaffold	1775.69	10591995/43	596
Myotis myotis	Scaffold	2002.8	12511134/41	631
Reference species				
Equus caballus	Chromosome	2506.97	1502753/462	462
Sus scrofa	Chromosome	2501.91	48231277/15	1118
Bos taurus	Chromosome	2715.85	25896116/32	2597
Canis lupus familiaris	Chromosome	2396.86	12024593/54	951
Felis cattus	Chromosome	2521.86	41915695/19	4909
Homo sapiens	Chromosome	3272.09	57879411/18	1580
Mus musculus	Chromosome	2728.22	59462871/15	306
Rattus norvegicus	Chromosome	2647.92	29198295/27	757

Table 9.1: Bat and reference genomes statistical data

	dN/dS	M1(neutral)	M2(positive selection)	M3(discrete)	M7(beta)	M8(beta&ω>1)
TLR1	0.4464	0.4543	0.4942	0.48	0.4427	0.4799
TLR2	0.4582	0.4346	0.5178	0.5042	0.4288	0.5059
TLR3	0.2764	0.3074	0.3074	0.3054	0.289	0.3057
TLR4	0.5423	0.4604	0.6048	0.6009	0.4646	0.5985
TLR5	0.284	0.3885	0.3885	0.3479	0.3186	0.3384
TLR6	0.416	0.4044	0.4198	0.4177	0.3927	0.416
TLR7	0.3342	0.3153	0.3703	0.3645	0.3081	0.3652
TLR8	0.52491	0.4144	0.6177	0.6188	0.4106	0.6095
TLR9	0.1242	0.2338	0.2517	0.1628	0.1549	0.1627
TLR10	0.465	0.4776	0.5241	0.5201	0.4795	0.5207
TLR11	0.8425	0.7403	0.9369	0.9316	0.7565	0.9336
TLR12	0.6862	0.5615	0.8193	0.836	0.5974	0.8194
TLR13	0.5168	0.4509	0.5687	0.5773	0.4677	0.5705

Table 9.2: The dN/dS ratio for each TLR set in all tested PAML *codeml* models

	Neutral evolution test (M0 ω=1 vs M0)	M0 vs M3	M1 vs M2	M7 vs M8
TLR1	Neutrality cannot be rejected (3.0315, df=1)	Cannot reject M3 (198.833164, df=14)	Cannot reject M2(10.927886, df=2)	Cannot reject M8(14.135404, df=2)
TLR2	Neutrality rejected(5.536632, df=1)	Cannot reject M3(352.67159, df=14)	Cannot reject M2(19.671672, df=2)	Cannot reject M8(30.249506, df=2)
TLR3	Neutrality rejected (44.091464, df=1)	Cannot reject M3(291.09372, df=14)	M2 rejectes(0, df=2)	Cannot reject M8(9.1935, df=2)
TLR4	Neutrality rejected (31.616922, df=1)	Cannot reject M3(485.81394, df=14)	Cannot reject M2(53.053582, df=2)	Cannot reject M8(66.54898, df=2)
TLR5	Neutrality rejected (66.326224, df=1)	Cannot reject M3(808.540536, df=14)	M2 rejected(0, df=2)	Cannot reject M8(18.84862, df=2)
TLR6	Neutrality rejected (178.034216, df=4)	M3 rejected (4.332226, df=11)	M2 rejected(5.341332, df=2)	Cannot reject M8(6.947378, df=2)
TLR7	Neutrality rejected (9.810122, df=1)	Cannot reject M3 (462.299198, df=14)	Cannot reject M2(21.642262, df=2)	Cannot reject M8(30.194572, df=2)
TLR8	Neutrality rejected (35.53166, df=1)	Cannot reject M3(1080.396076, df=14)	Cannot reject M2(130.307792, df=2)	Cannot reject M8(167.022066, df=2)
TLR9	Neutrality rejected (597.908782, df=1)	Cannot reject M3 (1125.224692, df=14)	Cannot reject M2(11.30164, df=2)	Cannot reject M8 (26.01441, df=2)
TLR10	Neutrality rejected (4.634224, df=1)	Cannot reject M3(102.75306, df=14)	Cannot reject M2(6.04523, df=2)	Cannot reject M8(9.97747, df=2)
TLR11	Neutrality rejected (180.4276, df=1)	Cannot reject M3(139.209682, df=14)	Cannot reject M2(47.172806, df=2)	Cannot reject M8(47.683834, df=2)
TLR12	Neutrality rejected (78.211854, df=1)	Cannot reject M3(219.765392, df=14)	Cannot reject M2(94.424182, df=2)	Cannot reject M8(109.186606, df=2)
TLR13	Neutrality rejected (18.065392, df=1)	Cannot reject M3(284.817498, df=14)	Cannot reject M2(41.656522, df=2)	Cannot reject M8(50.259458, df=2)

Table 9.3: *Codeml* model comparison for each set of the TLRs

POSITIVE SELECTION						
	aBSREL	**MEME**	**FUBAR**	**FEL**	**SLAC**	**PAML/M8**
TLR1	1 out of 43	15	0	8	0	3
TLR2	1 out of 43	15	14	33	1	17
TLR3	0	7	7	17	0	0
TLR4	5 out of 43	17	6	24	4	32
TLR5	4 out of 43	0	3	16	5	3
TLR6	2 out of 43	11	1	14	0	2
TLR7	3 out of 43	11	1	12	1	16
TLR8	10 out of 43	32	4	32	7	53
TLR9	6 out of 43	17	0	4	3	3
TLR10	0	12	3	9	0	8
TLR11	0	14	8	22	2	54
TLR12	0	13	8	16	0	21
TLR13	4 out of 31	15	2	14	1	27

Table 9.4: Number of sites under positive selection (FEL, SLAC, FUBAR, *codeml* M8, MEME) and number of branches under positive selection (bsREL) for each analysed model and each bat TLR set.

	BATS				
	PAML M8	FUBAR	SLAC	FEL	MEME
TLR 1	315 317 649	-	-	132 **206** 290 481 492 500 617 669	20 23 92 116 132 149 178 **206** 286 292 351 458 492 653 745
TLR2	**75 112 114** 160 **161 162 265 275 303** 307 **322** 336 338 **385** 454 **549 777**	**45 75 98 114 161 162 2 15 265 275** 293 **322 385 409 777**	**114**	36 **45** 88 **98 112 114** 138 **162** 193 **215** 222 **265 275** 279 281 293 300 **303** 304 **322 325** 332 379 **385 409** 431 524 **549** 555 598 640 **777**	81 **98 114 162** 191 **215** 237 **265 389** 451 **454** 514 573 774 **777**
TLR 3	-	**3 7 258 279 357 582 640**	-	**3 7** 80 87 121 **127 258** 262 **279** 286 344 **357** 431 **531** 567 582 **640**	**3 127** 344 469 **531 582** 883
TLR 4	3 56 239 **267 269** 270 293 320 321 323 340 **342 346 365** 368 370 **373** 383 **393** 459 **463** 473 **499** 509 543 **546** 549 **576 578 632 839**	**267 346 373 393 463 499**	**4 135**	**4 135** 219 **267 269** 291 **300** 318 **342 346 365 373 393** 418 461 **463 499 546 576 578** 608 **632** 635 **839**	**267 269 300 365** 372 385 391 **393** 430 457 **463** 495 **499 632** 642 791 834
TLR 5	**312** 422 530	**167 256 312**	**23** 50 **106 128 167**	**23 106 128** 136 **167** 156 272 279 319 **416 655** 657 **669** 675 684 784	50 154 210 **256** 382 **416** 442 572 596 **655 669** 686 687 789 796 848
TLR 6	**233 663**	**233**	-	26 90 171 208 **233** 237 240 324 365 452 482 493 618 **663**	134 157 **233** 291 416 429 459 493 619 654 **663**
TLR 7	**38** 40 **199** 237 281 **302** 362 393 452 **455** 460 488 **639** 654 **671** 732	**302**	**38**	**38** 59 123 **199** 251 386 437 **455** 457 **639 671** 676	**38** 88 111 117 164 **199** 452 **455** 535 610 736
TLR 8	42 **73 86 103** 107 149 191 193 194 215 238 239 277 **279** 280 281 **290** **311** 312 **338** 355 361 **368** **397** 415 **417 418 421** **427** 441 444 **447** 472 480 481 **506** 531 546 595 607 667 **676 684 685** **688** 698 **711** 722 760 761 762 **763** 791	**338 368 676 688**	**73 86 169 173 271 351 381**	22 **103 169 173** 218 **271 279 290 311** **338 351 368 381 412 427 447 506** 529 563 588 593 632 **676** **684 685 688 711** 724 **763** 794 804 810	39 43 146 156 214 **271** 274 303 **338 351** **368 397** 407 **412 417 418** 419 **421** **427** 439 **447** 505 **506** 508 585 663 674 **676 688** 690 **763**
TLR 9	91 **355 723**	-	143 **270**	632 701 **723** 841	105 116 170 **270** 312 326 **355** 364 378 412 **723** 727 737 807 891 915 945
TLR 10	**47 146 345** 512 **599** 670 **716** 792	**146 345 716**	-	**47** 134 **146 345** 557 563 **716**	6 34 117 **146** 312 401 475 488 594 **599** 603 750
TLR 11	80 95 98 111 115 118 146 **151** 165 **177** 182 **211** 232 255 **296** 307 370 373 420 423 458 463 **465 474** 481 483 484 485 487 492 **493** 509 **512** 514 519 524 **534 538 539** 557 **559** 561 563 583 584 585 **622** 701 721 733 **792** 832 895 913	**177 211 296 465 512 538 539 622**	**5** 26	**5** 143 **151 177 211** 252 322 436 **465 474** 480 **493** 510 **512 539 559 561 585 622** 708 **792** 862	331 368 374 398 462 **465 512 534 585** 624 798 747 **792** 892
TLR 12	**53 90** 100 102 **139 143 161 164** 198 201 269 **293 316** 321 **367** **446** 450 **590 616** 666 668 **681**	**43 90 143 161 164 293 446 616**	-	26 38 **43 53 90 139 143** **161 199 293** 332 **367 446 616 681 709**	19 37 **143 199 293 316 367** 429 **446** 537 **590 616 709**
TLR 13	**58** 81 **123 144** 183 241 264 **286** 296 297 307 **366 373 377** 390 **468 484** 515 517 **544 545** 675 **699** 867 949	**544 545**	**286**	**51 58 123 144 366 377** **468 484** 503 **544 545** 685 **699** 947	31 **51 144** 250 320 361 **373** 387 **468 484** 508 520 **544 545** 948

Table 9.5: Sites under positive selection (FEL, SLAC, FUBAR, *codeml* M8, MEME) for each bat TLR set. Sites found in by other studies underlined in the first column, and sites found by more than one model in bold.

NEGATIVE SELECTION			
	FUBAR	**FEL**	**SLAC**
TLR1	78	158	45
TLR2	140	185	68
TLR3	340	263	115
TLR4	112	167	63
TLR5	371	317	177
TLR6	113	167	52
TLR7	273	240	75
TLR8	273	265	107
TLR9	284	312	110
TLR10	48	132	27
TLR11	53	119	29
TLR12	58	128	21
TLR13	75	159	41

Table 9.6: Number of sites under negative selection for each analysed Datamonkey model (FEL, SLAC, FUBAR) in each bat TLR set.

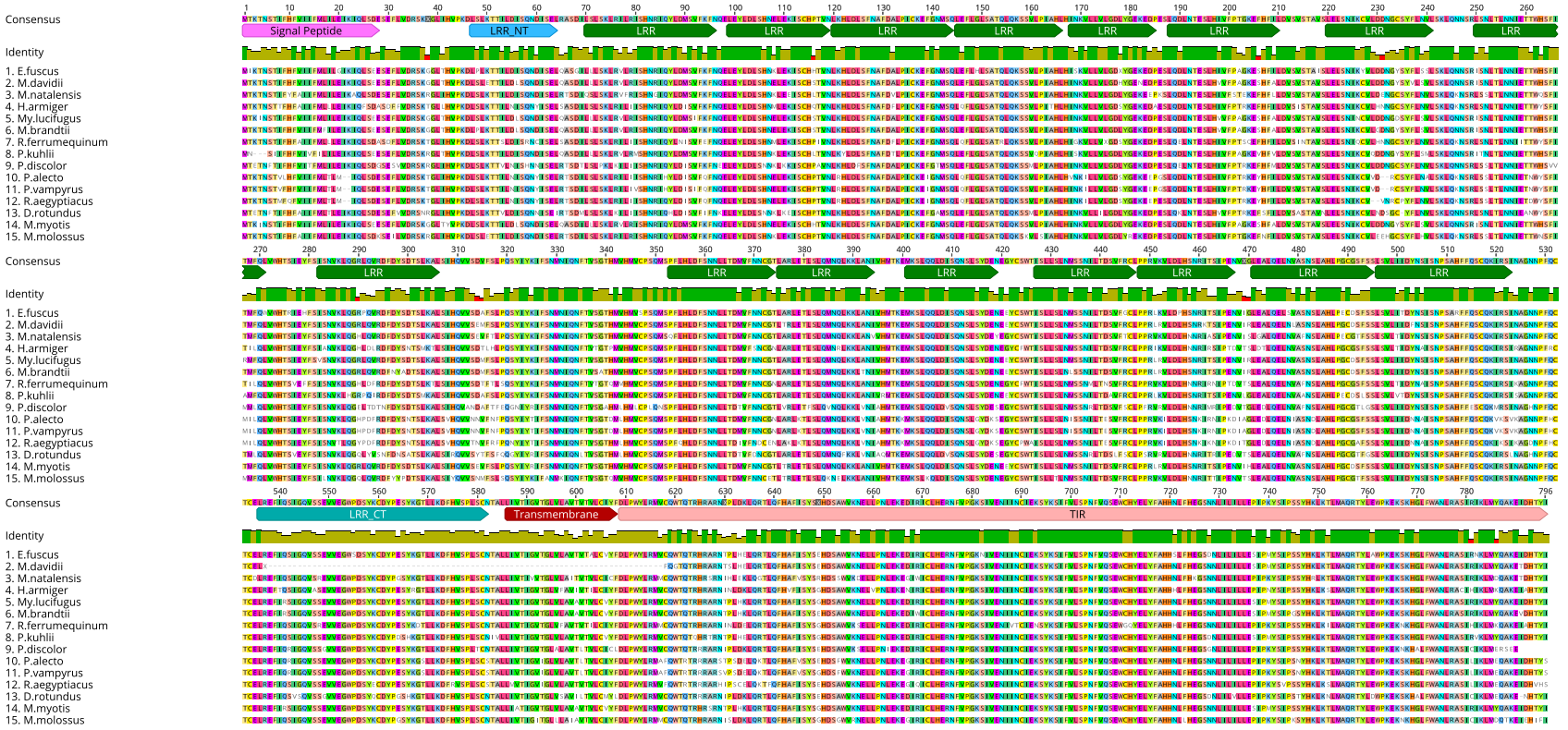


Figure 9.1: TLR1 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).



Figure 9.2: TLR2 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

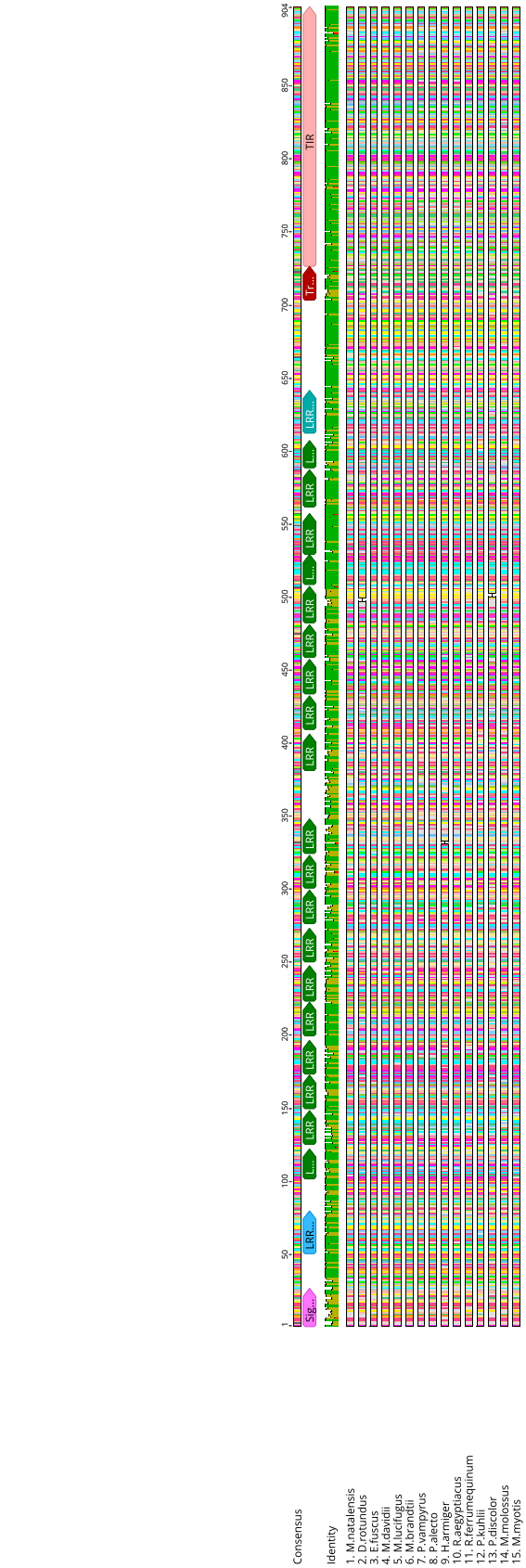


Figure 9.3: TLR3 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

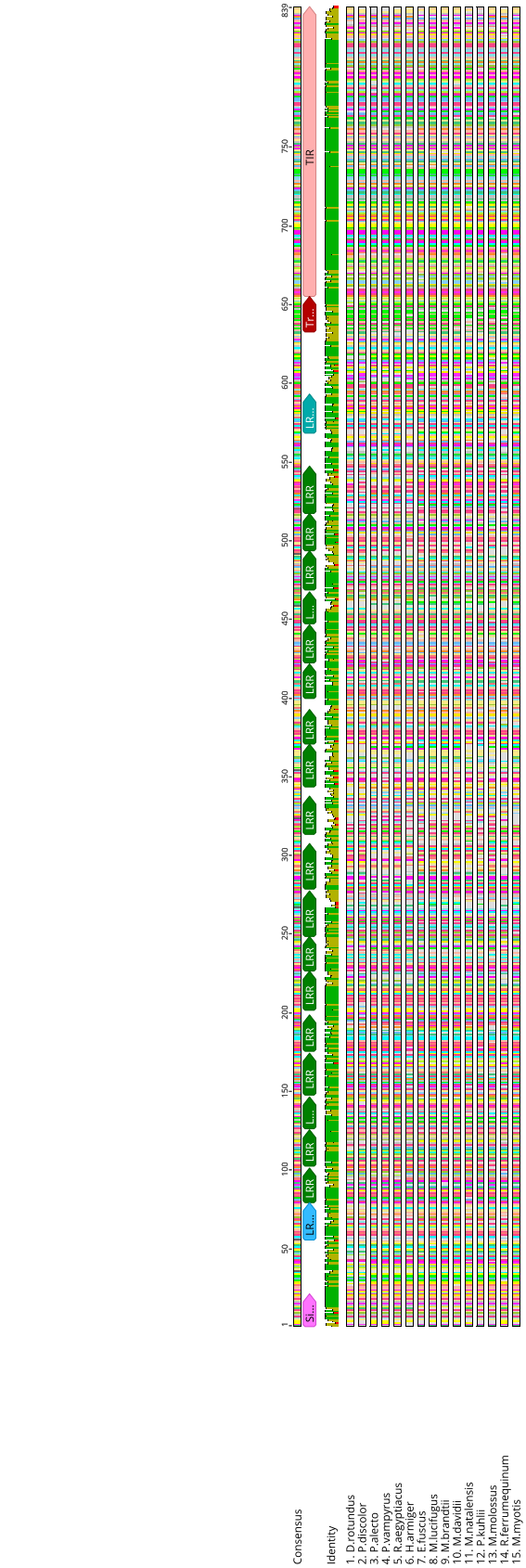


Figure 9.4: TLR4 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

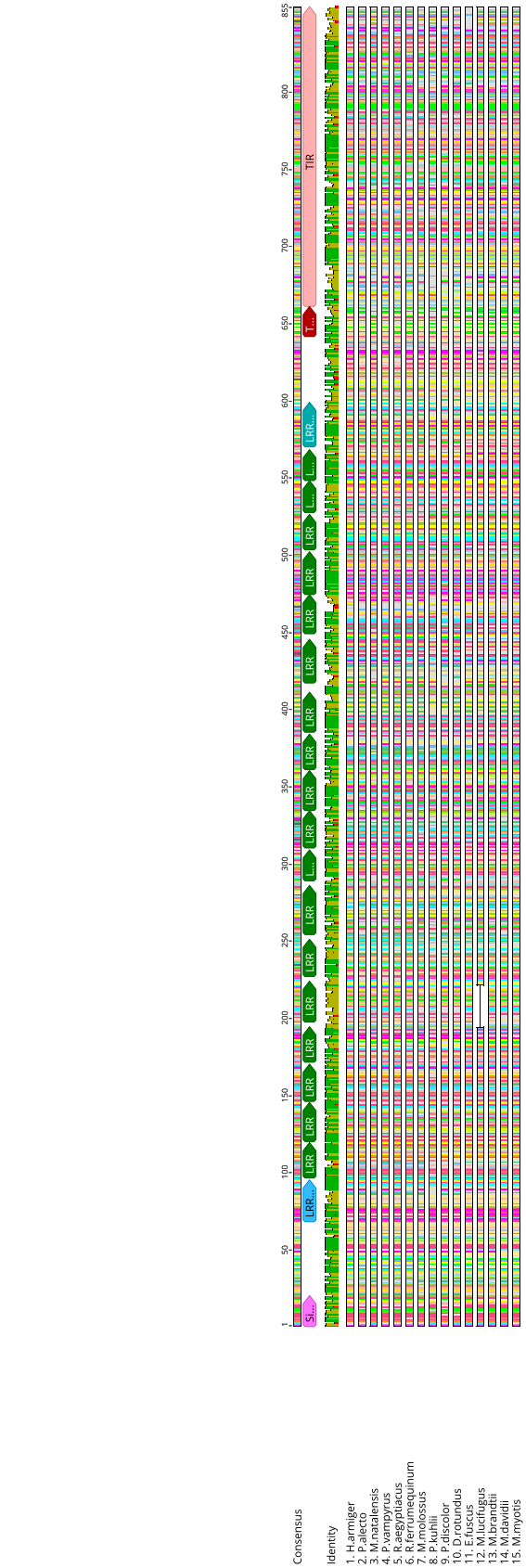
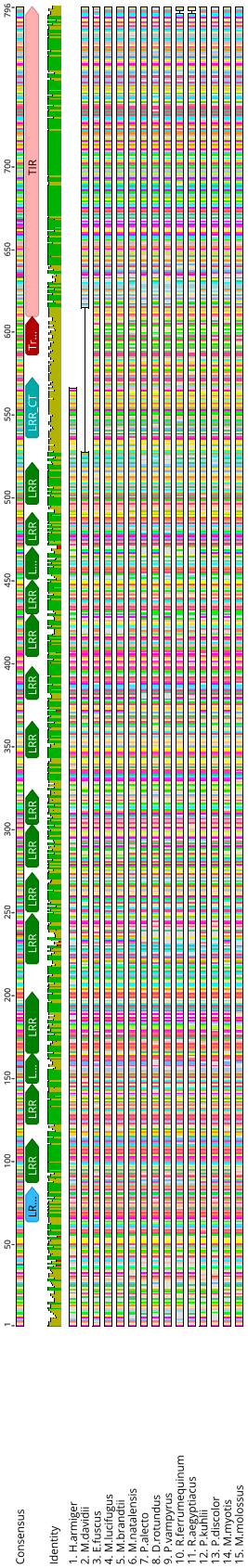


Figure 9.5: TLR5 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).


Consensus
Identity
1. H.armiger
2. M.davidii
3. E.fuscus
4. M.lucifugus
5. M.brandtii
6. M.natalensis
7. P.alecto
8. D.rotundus
9. P.vampyrus
10. R.ferrumequinum
11. R.aegyptiacus
12. P.kuhlii
13. P.discolor
14. M.myotis
15. M.molossus
LR...
LRR
LRR...
LRR
LRR
LRR

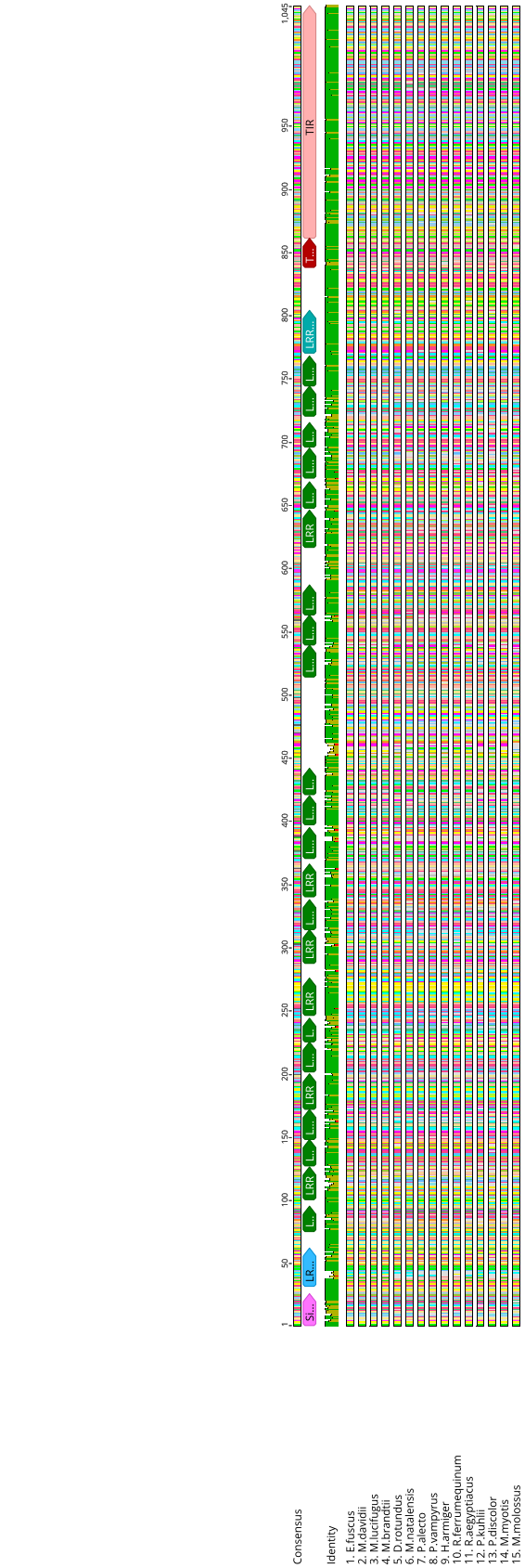
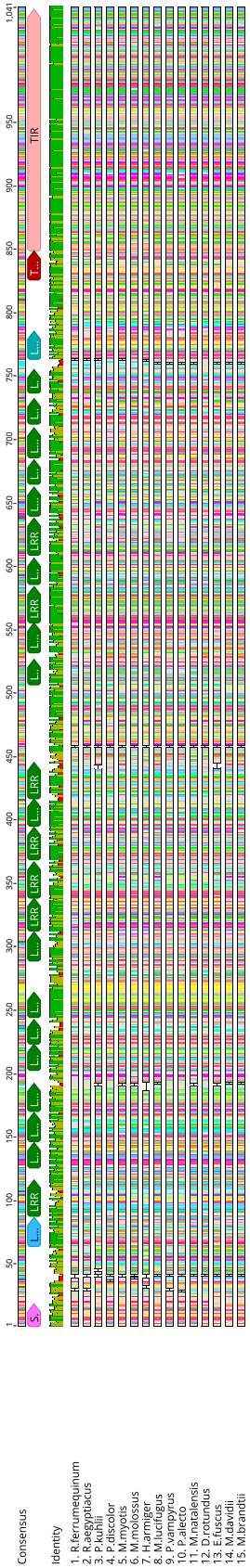


Figure 9.7: TLR7 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).


Consensus
Identity
1. R.ferrumequinum
2. R.aegyptiacus
3. R.sinicus
4. P.discolor
5. M.myotis
6. M.molossus
7. H.armiger
8. M.lucifugus
9. P.vampyrus
10. P.alecto
11. M.natalensis
12. D.rotundus
13. E.fuscus
14. M.davidii
15. M.brandtii
SP
LRR
LRR
LRR

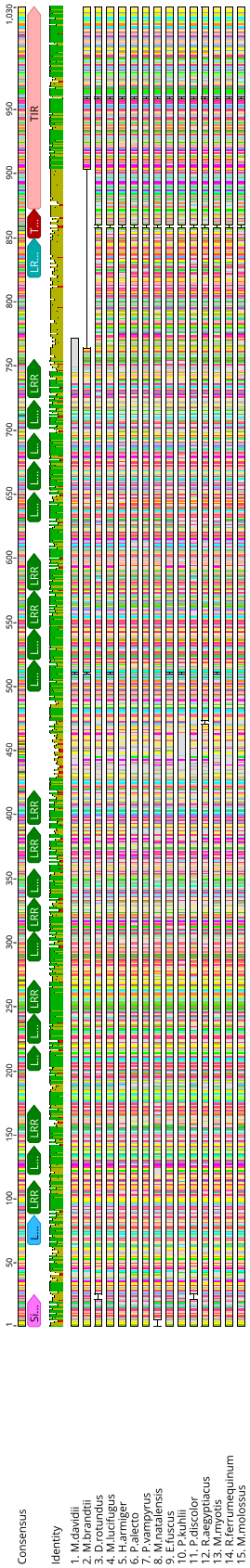


Figure 9.9: TLR9 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive- between 99% and 30%, red - below 30% identity).

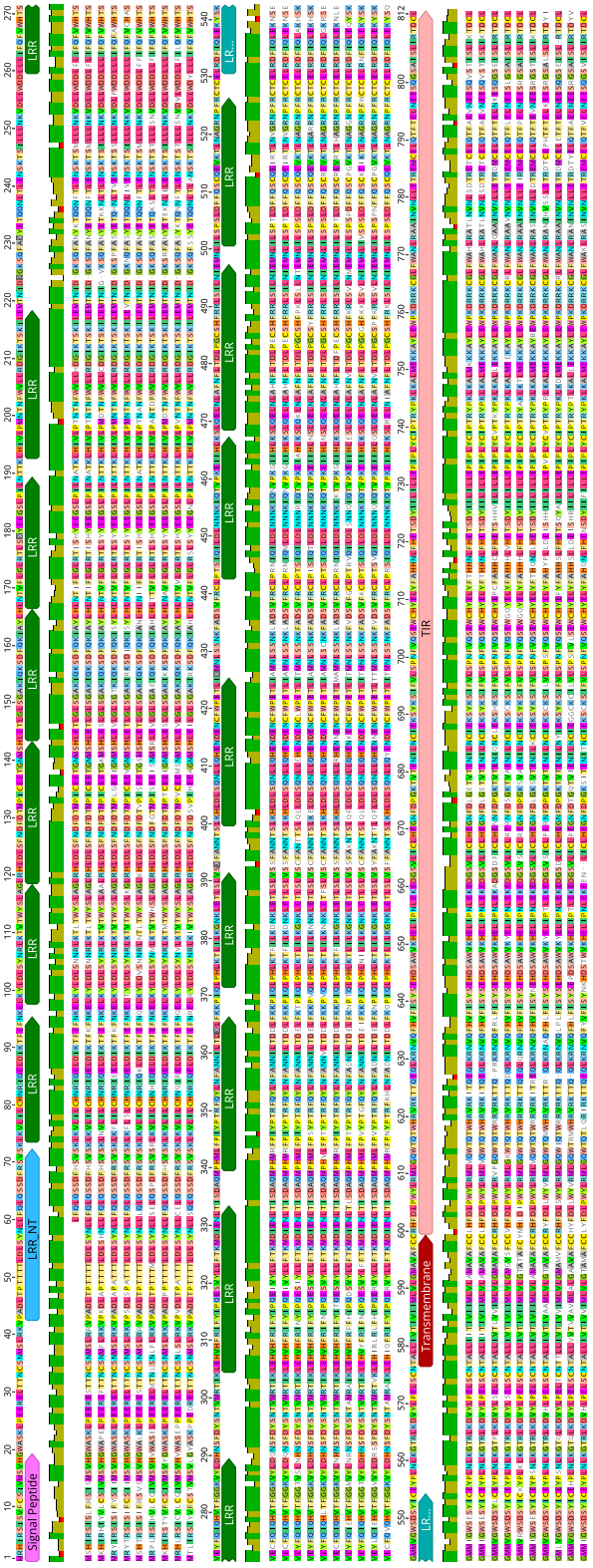


Figure 9.10: TLR10 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

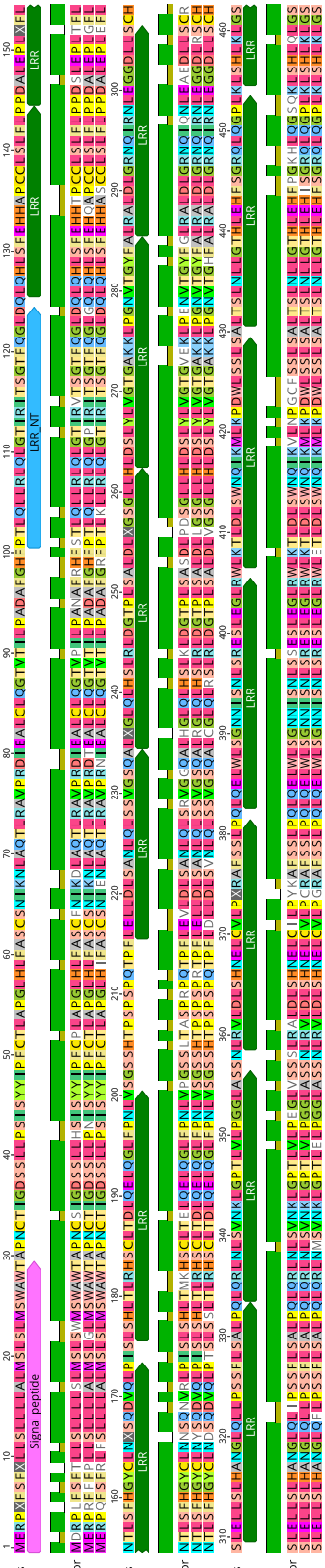


Figure 9.11: TLR11 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

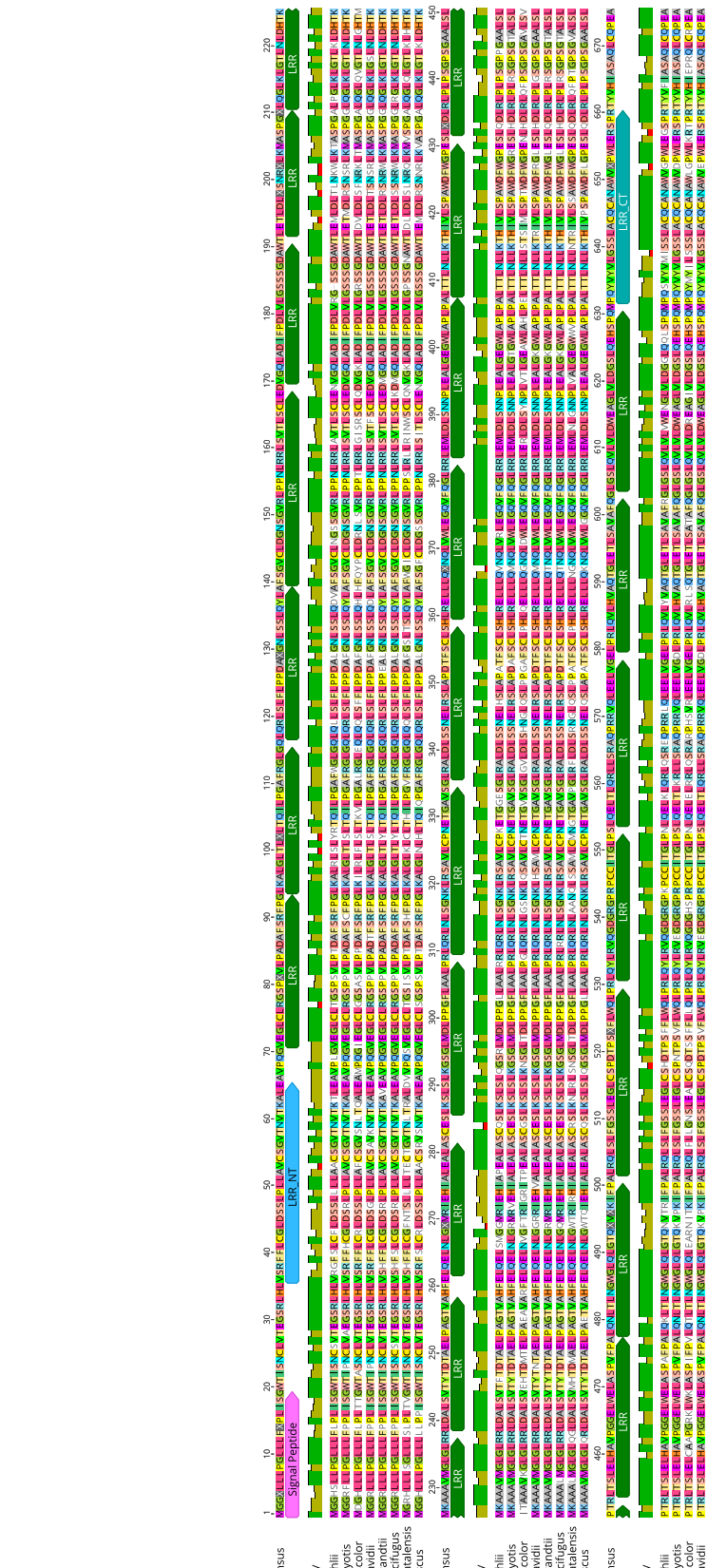


Figure 9.12: TLR12 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

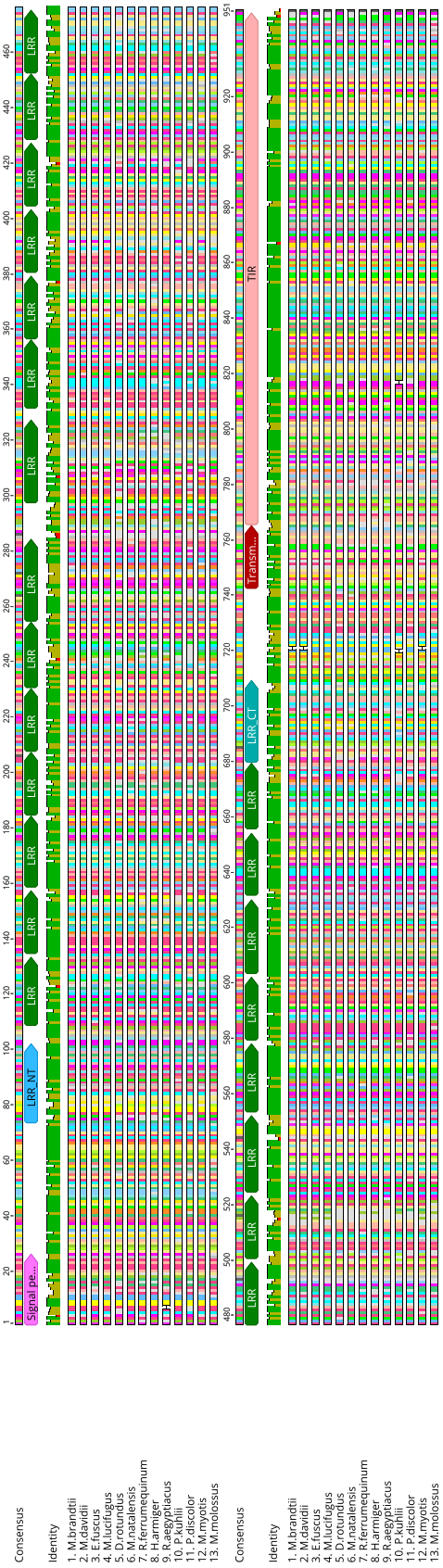


Figure 9.13: TLR13 alignment of all analysed bat sequences with consensus sequence, domain annotations (pink - Signal peptide, dark green - LRR, olive - Transmembrane region, light pink - TIR domain), and mean pairwise identity across the whole sequence (green - 100% identity, olive - between 99% and 30%, red - below 30% identity).

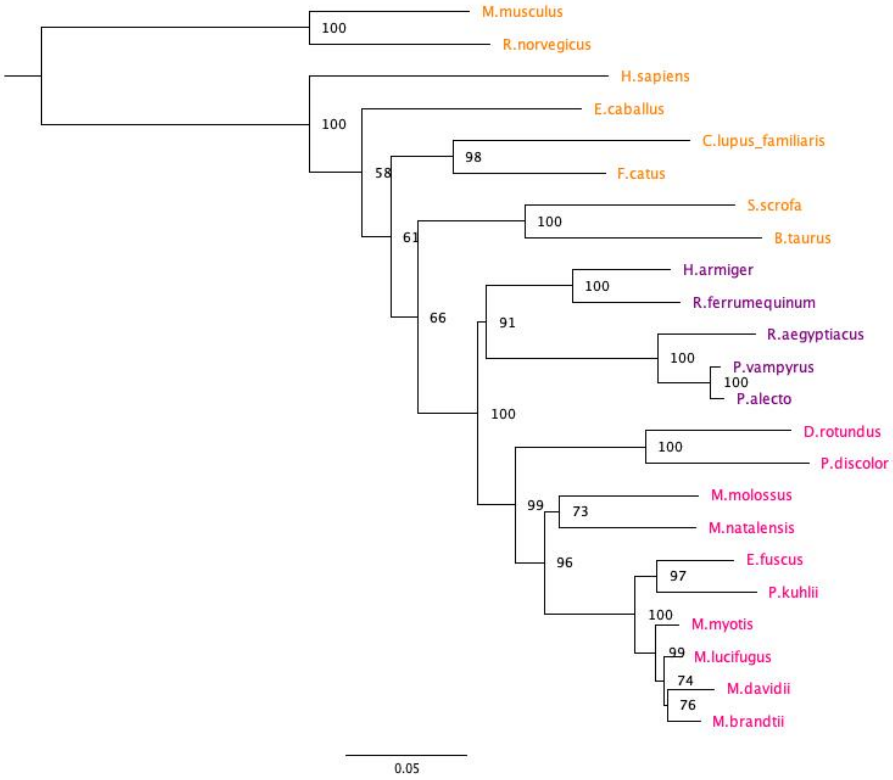


Figure 9.14: TLR1 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

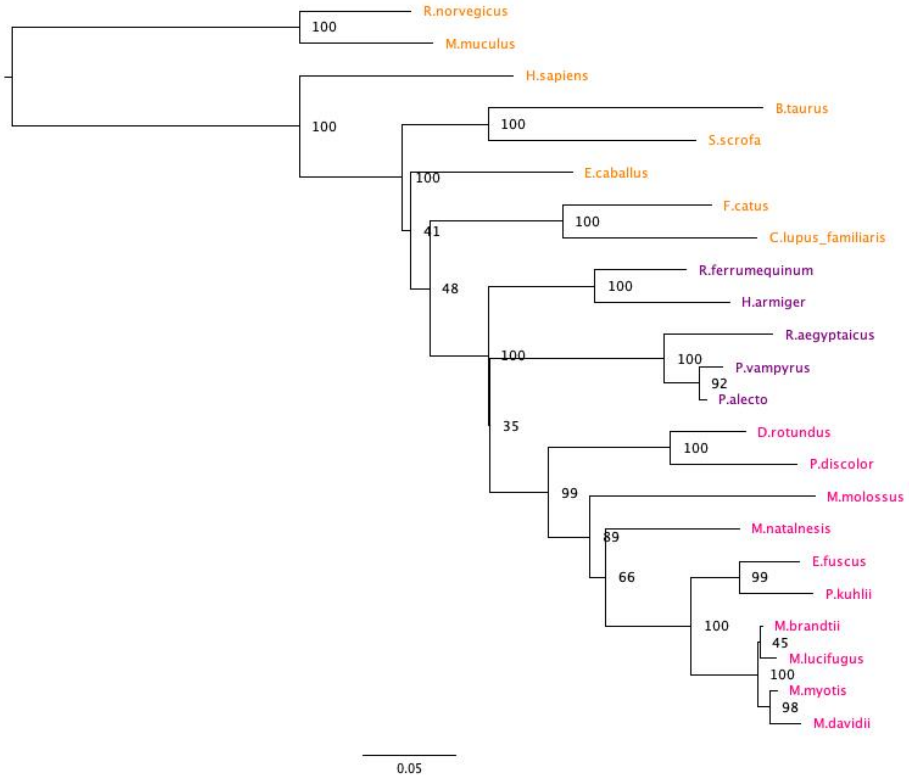


Figure 9.15: TLR2 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

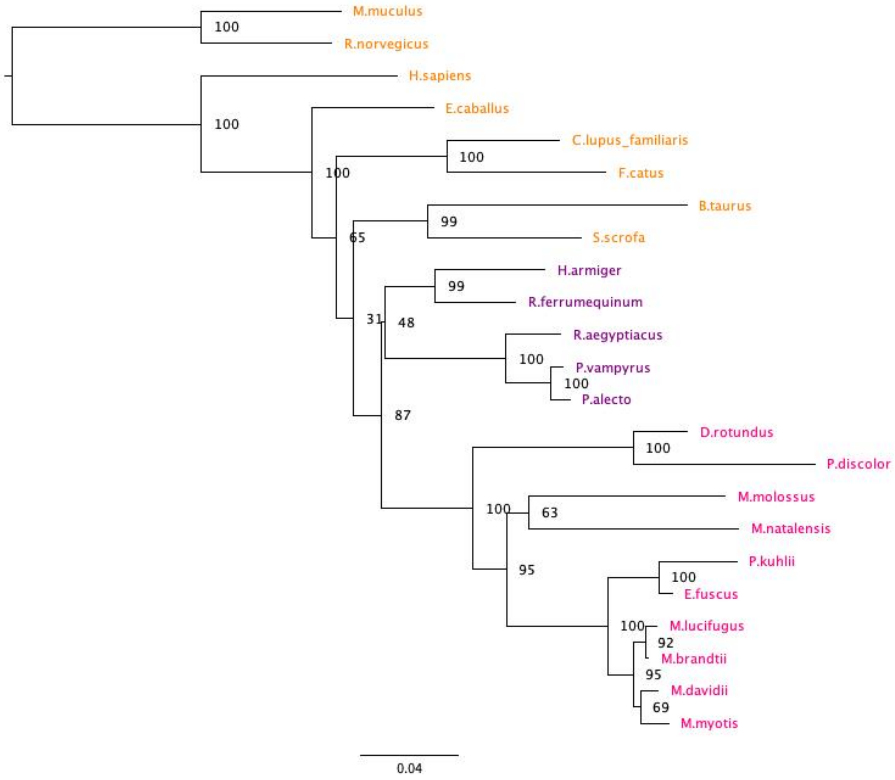


Figure 9.16: TLR3 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

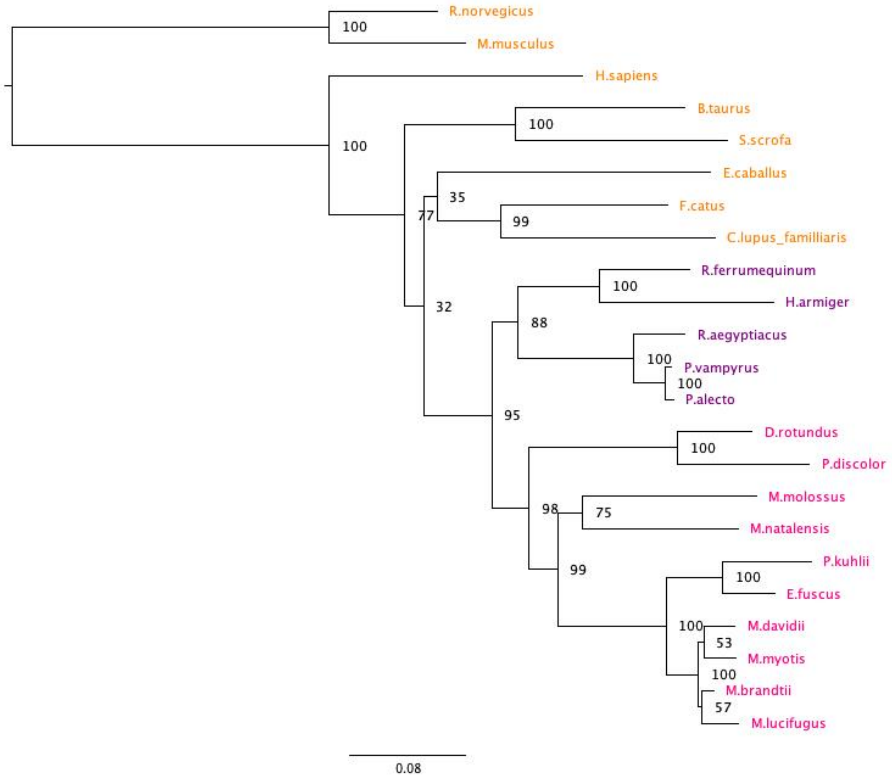


Figure 9.17: TLR4 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

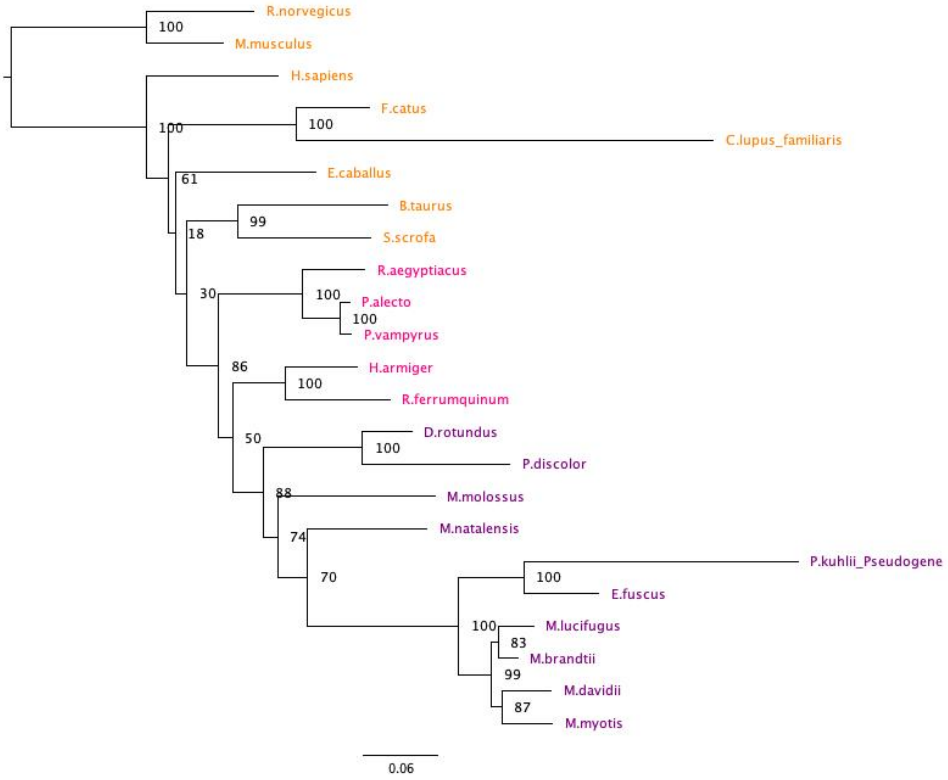


Figure 9.18: TLR5 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

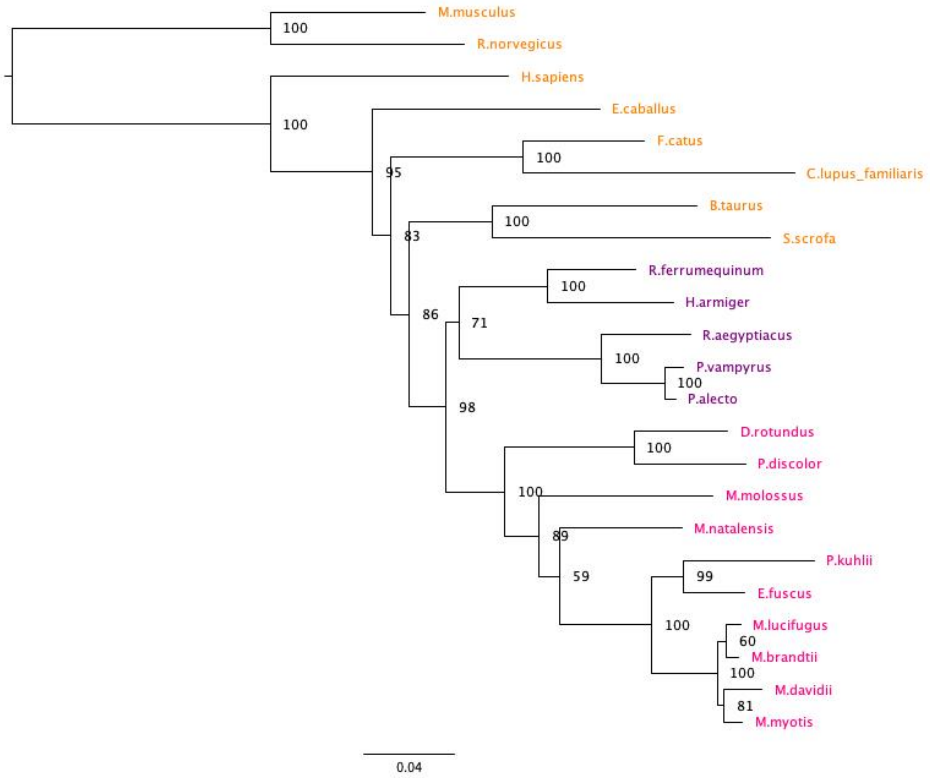


Figure 9.19: TLR6 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

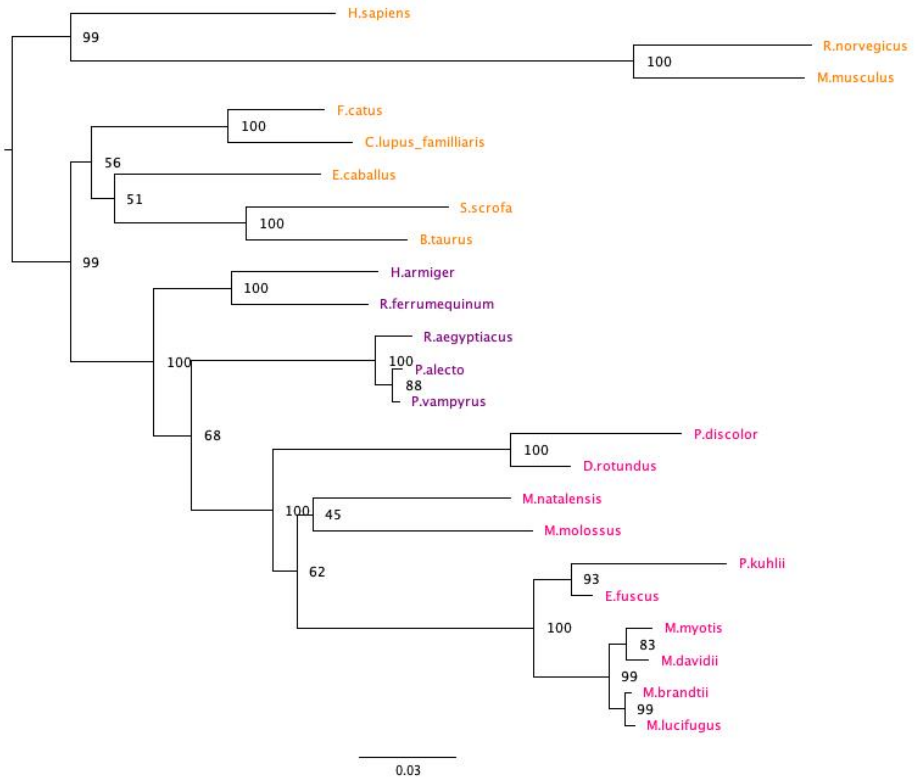


Figure 9.20: TLR7 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

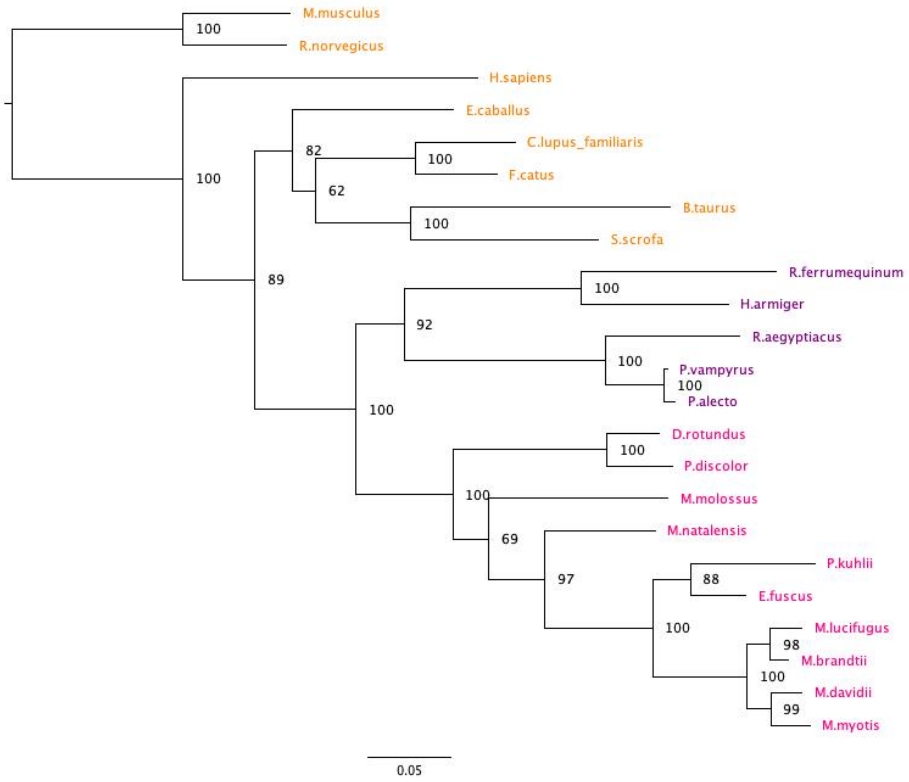


Figure 9.21: TLR8 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

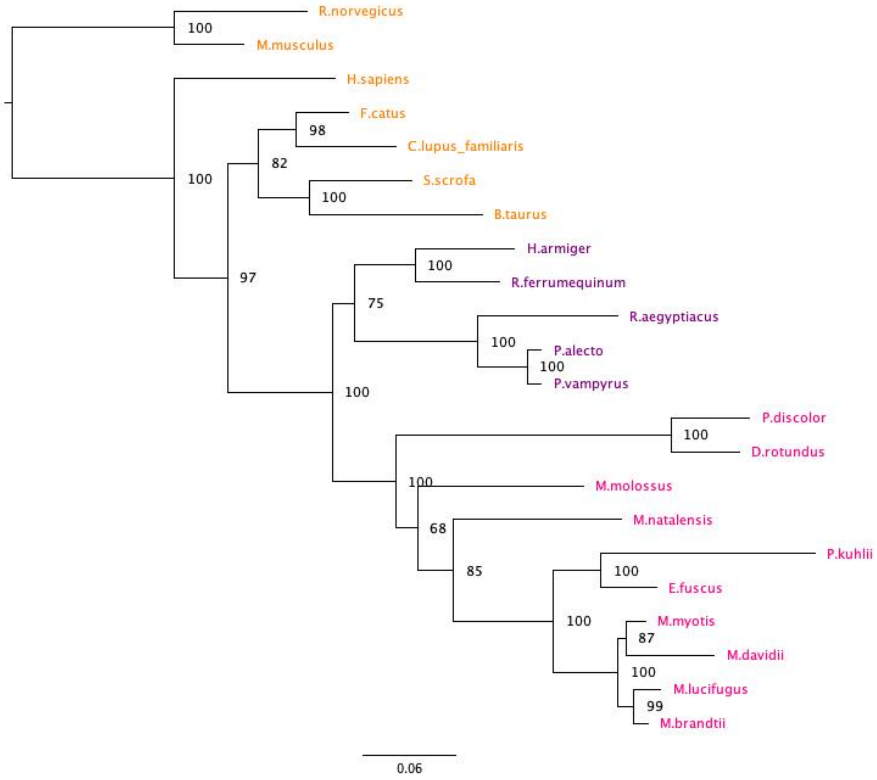


Figure 9.22: TLR9 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

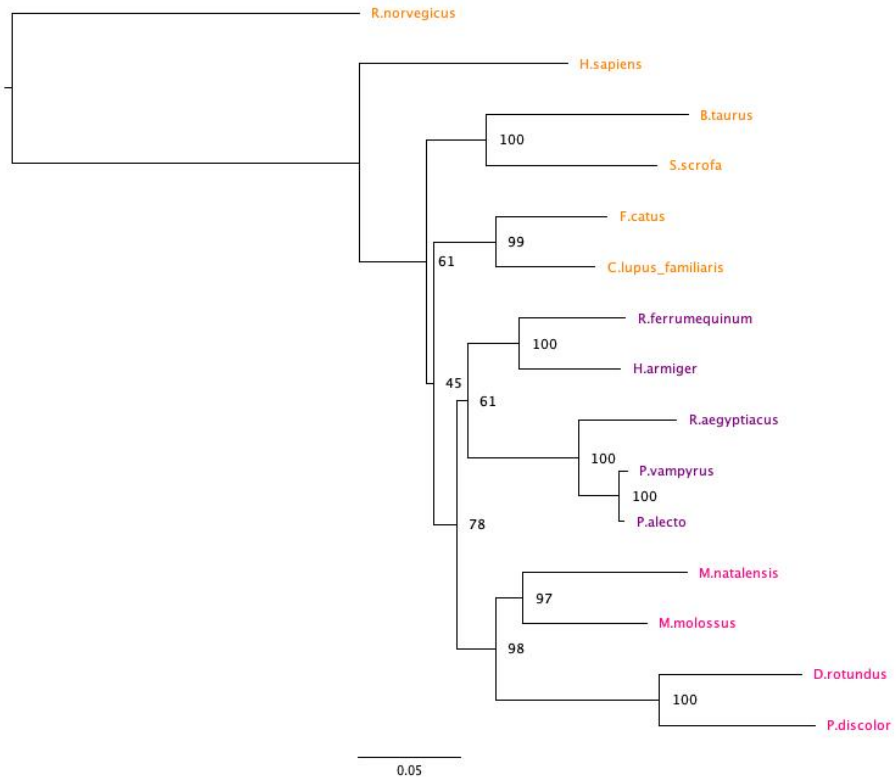


Figure 9.23: TLR10 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

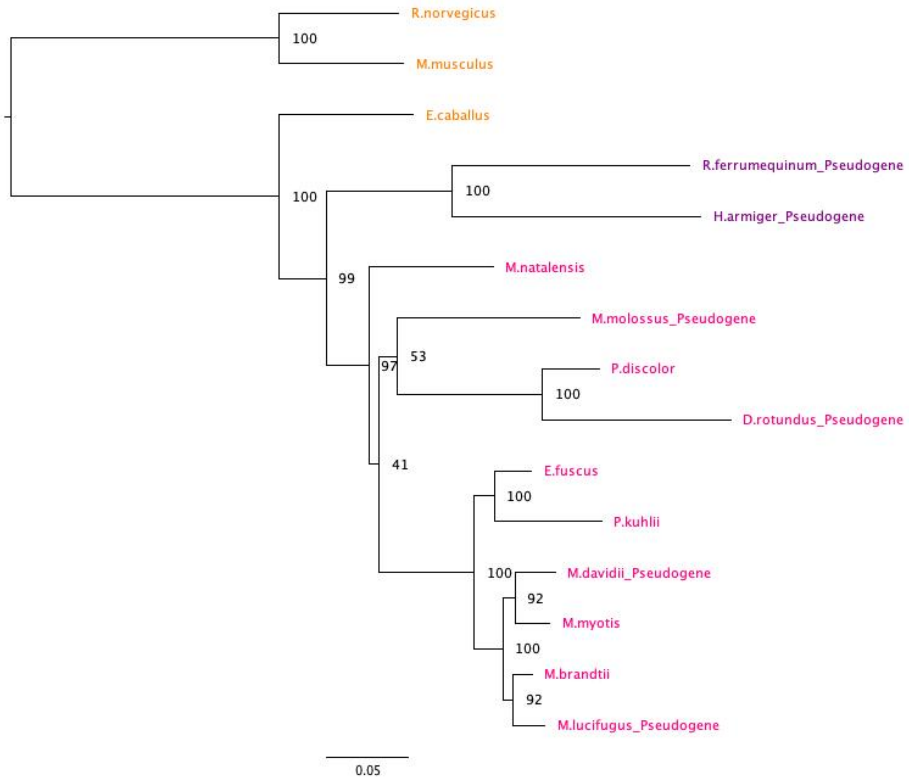


Figure 9.24: TLR11 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

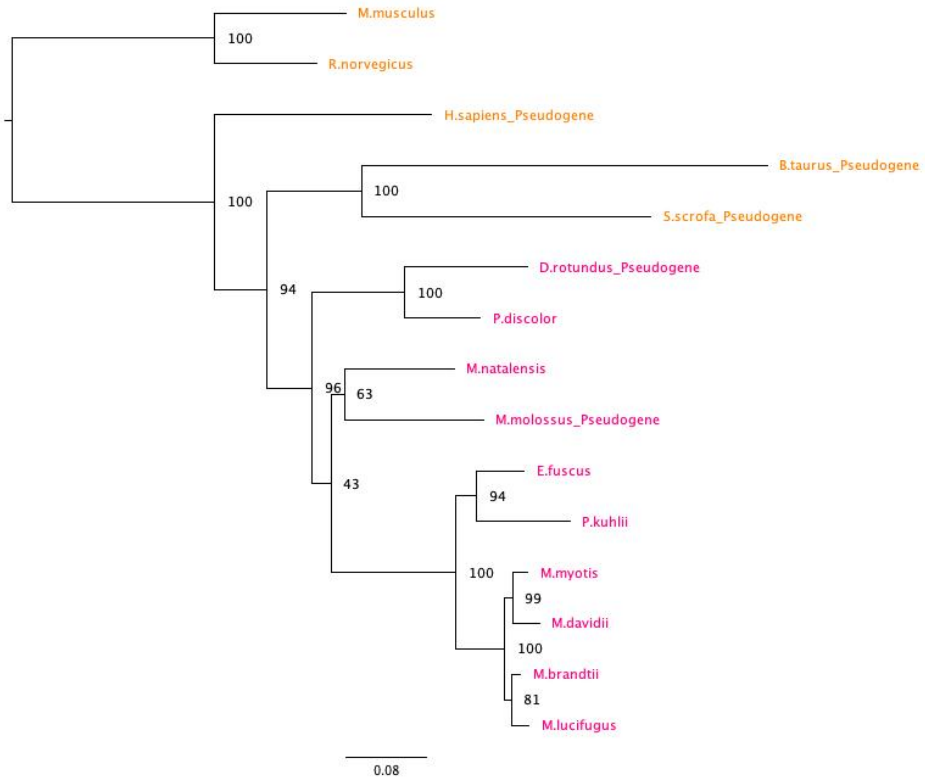


Figure 9.25: TLR12 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

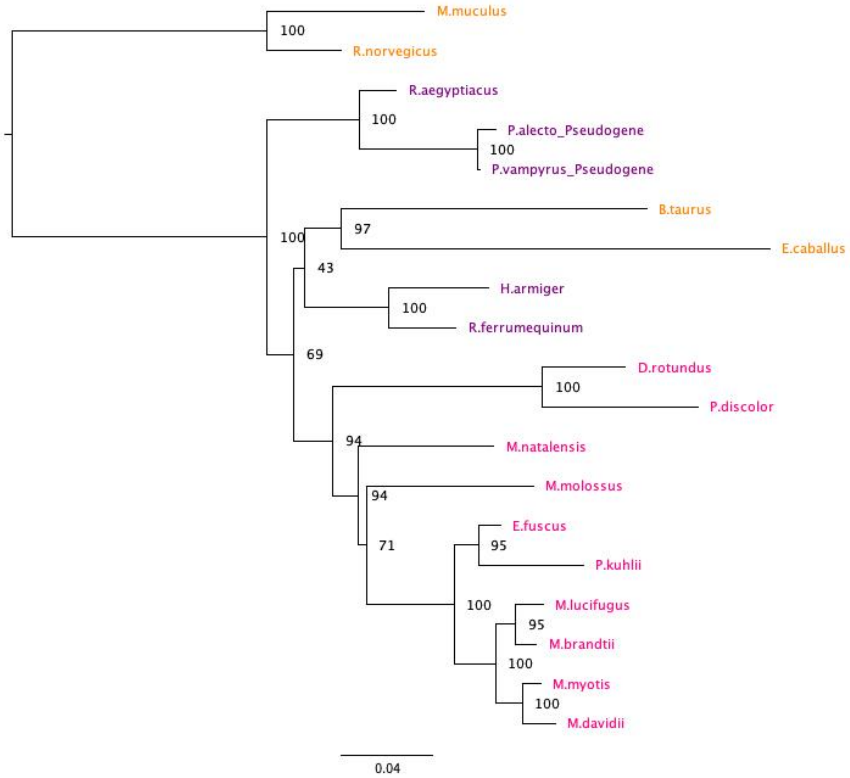


Figure 9.26: TLR13 phylogenetic tree showing bat and reference sequences. Mammalian reference sequences marked with yellow, *Yinpterachiroptera* sequences marked with purple, *Yangochiroptera* sequences marked with pink colour.

10 Appendix to Chapter 5

	Accession number	Latin name	Class	Order	Family	Common Name
1	XM 018563056 .1	Nanorana parkeri	Amphibia	Anura	Dicroglossidae	High Himalaya Frog
2	XM 002938247.4	Xenopus tropicalis	Amphibia	Anura	Pipidae	Tropical Clawed Frog
3	XM 029603226 .1	Rhinatrema bivittatum	Amphibia	Gymnophiona	Rhinatrematidae	Two-Banded Caecillia
4	XM 030202314 .1	Microcaecilia unicolor	Amphibia	Gymnophiona	Siphonophidae	Tiny Cayenne Caecilian
5	XM 011589453 .1	Aquila chrysaetos canadensis	Aves	Accipitriformes	Accipitridae	Golden Eagle
6	XM 010581526 .1	Haliaeetus leucocephalus	Aves	Accipitriformes	Accipitridae	American Bald Eagle
7	XM 011589453 .1	Aquila chrysaetos canadensis	Aves	Accipitriformes	Accipitridae	Golden Eagle
8	XM 009927339 .1	Haliaeetus albicilla	Aves	Accipitriformes	Accipitridae	White-tailed Eagle
9	XM 013183846 .1	Anser cygnoides domesticus	Aves	Anseriformes	Anatidae	Domesticated Swan Goose
10	XM 032202495 .1	Aythya fuligula	Aves	Anseriformes	Anatidae	Tufted Duck
11	XM 013094461 .3	Anas platyrhynchos	Aves	Anseriformes	Anatidae	Mallard
12	XM 009993826 .1	Chaetura pelagica	Aves	Apodiformes	Apodidae	Chimney Swift
13	XM 008494775 .2	Calypte anna	Aves	Apodiformes	Trochilidae	Anna's Hummingbird
14	XM 013950282 .1	Apteryx australis mantelli	Aves	Apterygiformes	Apterygidae	Brown Kiwi
15	XM 026087161 .1	Apteryx rowi	Aves	Apterygiformes	Apterygidae	Okarito Brown Kiwi
16	XM 013950282 .1	Apteryx australis mantelli (Apterys mantelli)	Aves	Apterygiformes	Apterygidae	Northern Island Brown Kiwi
17	XM 010138511 .1	Buceros rhinoceros silvestris	Aves	Bucerotiformes	Bucerotidae	Rhinocerous Hornbill
18	XM 010170936 .1	Antrostomus carolinensis	Aves	Caprimulgiformes	Caprimulgidae	Chuck-wills-widow
19	XM 009705393 .1	Cariama cristata	Aves	Cariamiformes	Cariamidae	Red-legged Seriema
20	XM 026120784 .1	Dromaius novaehollandiae	Aves	Casuariiformes	Dromaiidae	Emu
21	XM 014960219 .1	Calidris pugnax	Aves	Characiformes	Scolopacidae	Ruff
22	XM 009889029 .1	Charadrius vociferus	Aves	Charadriiformes	Chradriidae	Killdeer
23	XM 010207752 .1	Colius striatus	Aves	Coliiformes	Coliidae	Speckled Mousbird
24	XM 021298811 .1	Columba livia	Aves	Columbiformes	Columbidae	Common Pigeon
25	XM 008939271 .1	Merops nubicus	Aves	Coraciiformes	Meropidae	Northern Carmine Bee-eater
26	XM 009565569 .1	Cuculus canorus	Aves	Cuculiformes	Cuculidae	Common Cuckoo
27	XM 010158165 .1	Eurypyga helias	Aves	Eurypygiformes	Eurypygidae	Sunbittern
28	XM 005443036 .2	Falco cherrug	Aves	Falconiformes	Falconidae	Saker Falcon
29	XM 005231927 .2	Falco peregrinus	Aves	Falconiformes	Falconidae	Perigrine
30	XM 021385056 .1	Numida meleagris	Aves	Galliformes	Numididae	Helmeted Guineafowl
31	XM 015886577 .2	Coturnix japonica	Aves	Galliformes	Phasianidae	Japanese Quail
32	XM 416822.5	Gallus gallus	Aves	Galliformes	Phasianidae	Red Junglefowl
33	XM 031596059 .1	Phasianus colchicus	Aves	Galliformes	Phasianidae	Common Pheasant
34	XM 019612009 .2	Meleagris gallopavo	Aves	Galliformes	Phasianidae	Common Turkey
35	XM 009817825 .1	Gavia stellata	Aves	Gaviiformes	Gaviidae	Red-throated Diver
36	XM 009956091 .1	Leptosomus discolor	Aves	Leptosomiformes	Leptosomidae	Cuckoo Roller

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
37	XM 010180401 .1	Mesitornis unicolor	Aves	Mesitornithiformes	Mesitornithidae	Brown Mesite
38	XM 009980113 .1	Tauraco erythrolophus	Aves	Musophagiformes	Musophagidae	Red-crested Turaco
39	XM 009940668 .1	Opisthocomus hoazin	Aves	Opisthocomiformes	Opisthocomidae	Hoatzin
40	XM 010122221 .1	Chlamydotis macqueenii	Aves	Otidiformes	Otididae	Macqueen's Bustard
41	XM 009083902 .1	Acanthisitta chloris	Aves	Passeriformes	Acanthisittidae	Rifleman
42	XM 017728394 .1	Corvus brachyrhynchos	Aves	Passeriformes	Corvidae	American Crow
43	XM 010394433 .3	Corvus cornix cornix	Aves	Passeriformes	Corvidae	Hooded Crow
44	XM 032100703 .1	Corvus moneduloides	Aves	Passeriformes	Corvidae	New Caledonian Crow
45	XM 005491775 .2	Zonotrichia albicollis	Aves	Passeriformes	Emberizidae	White-throated Sparrow
46	XM 002194267 .3	Taeniopygia guttata	Aves	Passeriformes	Estrildidae	Zebra Finch
47	XM 021532351 .2	Lonchura striata domestica	Aves	Passeriformes	Estrildidae	White-rumped Munia
48	XM 009089674 .3	Serinus canaria	Aves	Passeriformes	Fringillidae	Atlantic Canary
49	XM 005037365 .1	Ficedula albicollis	Aves	Passeriformes	Muscicapidae	Collared Flycatcher
50	XM 023918416 .1	Cyanistes caeruleus	Aves	Passeriformes	Paridae	Blue Tit
51	XM 015631329 .1	Parus major	Aves	Passeriformes	Paridae	Great Tit
52	XM 005516655 .1	Pseudopodoces humilis	Aves	Passeriformes	Paridae	Ground Tit
53	XM 017812240 .1	Lepidothrix coronata	Aves	Passeriformes	Pipridae	Blue-crowned Manakin
54	XM 018084005 .2	Manacus vitellinus	Aves	Passeriformes	Pipridae	Golden-collared Manakin
55	XM 027689063 .1	Neopelma chrysocephalum	Aves	Passeriformes	Pipridae	Saffron-crested Neopelma
56	XM 027738173 .1	Pipra filicauda	Aves	Passeriformes	Pipridae	Wire-tailed Manakin
57	XM 027639017 .1	Corapipo altera	Aves	Passeriformes	Pipridae	White-ruffed Manakin
58	XM 014875884 .1	Sturnus vulgaris	Aves	Passeriformes	Sturnidae	Common Starling
59	XM 030955525 .1	Camarhynchus parvulus	Aves	Passeriformes	Thraupidae	Small Tree Finch
60	XM 005426164 .1	Geospiza fortis	Aves	Passeriformes	Thraupidae	Medium Ground Finch
61	XM 027901350 .1	Empidonax traillii	Aves	Passeriformes	Tyrannidae	Traill's Flycatcher
62	XM 009639962 .1	Egretta garzetta	Aves	Pelecaniformes	Ardeidae	Little Egret
63	XM 009480645 .1	Pelecanus crispus	Aves	Pelecaniformes	Pelecanidae	Dalmatian Pelican
64	XM 009476315 .1	Nipponia nippon	Aves	Pelecaniformes	Threskiornithidae	Crested Ibis
65	XM 010291717 .1	Phaethon lepturus	Aves	Phaethontiformes	Phaethontidae	White-tailed Tropicbird
66	XM 009911547 .1	Picoides pubescens	Aves	Piciformes	Picidae	Downy Woodpecker
67	XM 009576601 .1	Fulmarus glacialis	Aves	Procellariiformes	Procellariidae	Fulmar
68	XM 030476779 .1	Strigops habroptila	Aves	Psittaciformes	Psittacidae	Kakapo
69	XM 005151459 .1	Melopsittacus undulatus	Aves	Psittaciformes	Psittacidae	Budgerigar
70	XM 010014179 .1	Nestor notabilis	Aves	Psittaciformes	Psittacidae	Kea
71	XM 010086071 .1	Pterocles gutturalis	Aves	Pteroclidiformes	Pteroclididae	Yellow-throated Sandgrouse

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
72	XM 009276865 .1	Aptenodytes forsteri	Aves	Sphenisciformes	Spheniscidae	Emperor Penguin
73	XM 009325492 .1	Pygoscelis adeliae	Aves	Sphenisciformes	Spheniscidae	Adelie Penguin
74	XM 026849924 .1	Athene cunicularia	Aves	Strigiformes	Strigidae	Burrowing Owl
75	XM 009970907 .1	Tyto alba	Aves	Strigiformes	Tytonidae	Barn Owl
76	XM 009669200 .1	Struthio camelus australis	Aves	Struthioniformes	Struthionidae	South African Ostrich
77	XM 009510775 .1	Phalacrocorax carbo	Aves	Suliformes	Phalacrocoracidae	Cormorant
78	XM 010219282 .1	Tinamus guttatus	Aves	Tinamiformes	Tinamidae	White-throated Tinamou
79	XM 026035320 .1	Nothoprocta perdicaria	Aves	Tinamiformes	Tinamidae	Chilean Tinamou
80	XM 009868754 .1	Apaloderma vittatum	Aves	Trogoniformes	Trogonidae	Bar-tailed Trogon
81	XM 024294863 .1	Oryzias melastigma	Fish (Actinopterygii)	Beloniformes	Adrianichthyidae	A Ricefish
82	XM 030048292 .1	Myripristis murdjan	Fish (Actinopterygii)	Beryciformes	Holocentridae	Big-eye Soldierfish
83	XM 022668208 .1	Astyanax mexicanus	Fish (Actinopterygii)	Characiformes	Characidae	Blind Cavefish
84	XM 017694590 .1	Pygocentrus nattereri	Fish (Actinopterygii)	Characiformes	Serrasalmidae	Batman Crested Loach
85	XM 031558926 .1	Clupea harengus	Fish (Actinopterygii)	Clupeiformes	Clupeidae	Atlantic Herring
86	XM 028981948 .1	Denticeps clupeoides	Fish (Actinopterygii)	Clupeiformes	Denticipitidae	Denticle Herring
87	XM 026275528 .1	Carassius auratus	Fish (Actinopterygii)	Cypriniformes	Cyprinidae	Carp
88	XM 016489839 .1	Sinocyclocheilus anshuiensis	Fish (Actinopterygii)	Cypriniformes	Cyprinidae	A Ray-finned Fish
89	XM 005169359.4	Danio rerio	Fish (Actinopterygii)	Cypriniformes	Cyprinidae	Leopard Danio
90	XM 015371244 .1	Cyprinodon variegatus	Fish (Actinopterygii)	Cyprinodontiformes	Cyprinodontidae	Bahama Sheepshead Minnow
91	XM 021322522 .1	Fundulus heteroclitus	Fish (Actinopterygii)	Cyprinodontiformes	Fundulidae	Killifish
92	XM 015953491 .1	Nothobranchius furzeri	Fish (Actinopterygii)	Cyprinodontiformes	Nothobranchiidae	African Turquoise Killifish
93	XM 028015870 .1	Xiphophorus couchianus	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Monterrey Platyfish
94	XM 008404492 .2	Poecilia reticulata	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Barbados Millions
95	XM 015039827 .1	Poecilia latipinna	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Sailfin Mollie
96	XM 014981539 .1	Poecilia mexicana	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Atlantic Molly
97	XM 007560146 .2	Poecilia formosa	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Amazon Molly
98	XM 005799778 .2	Xiphophorus maculatus	Fish (Actinopterygii)	Cyprinodontiformes	Poeciliidae	Moon
99	XM 017439896 .2	Kryptolebias marmoratus	Fish (Actinopterygii)	Cyprinodontiformes	Rivulidae	Mangrove Rivulus
100	XM 014033474 .1	Austrofundulus limnaeus	Fish (Actinopterygii)	Cyprinodontiformes	Rivulidae	A Killifish
101	XM 010886475.3	Esox lucius	Fish (Actinopterygii)	Esociformes	Escoidae	American Pike
102	XM 030376670 .1	Gadus morhua	Fish (Actinopterygii)	Gadiformes	Gadidae	Atlantic Cod
103	XM 028442074 .1	Gouania willdenowi	Fish (Actinopterygii)	Gobiesociformes	Gobiesocidae	Blunt-snouted Clingfish
104	XM 027011410 .1	Electrophorus electricus	Fish (Actinopterygii)	Gymnotiformes	Gymnotidae	Electric Eel
105	XM 006639122 .2	Lepisosteus oculatus	Fish (Actinopterygii)	Lepisosteiformes	Lepisosteidae	Spotted Gar
106	XM 018729216 .1	Scleropages formosus	Fish (Actinopterygii)	Osteoflossiformes	Osteoglossidae	Arowana

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
107	XM 023823901 .1	Paramormyrops kingsleyae	Fish (Actinopterygii)	Osteoglossiformes	Mormyridae	An Elephantfish
108	XM 028402086 .1	Parambassis ranga	Fish (Actinopterygii)	Perciformes	Ambassidae	Chanda ranga
109	XM 026377646 .1	Anabas testudineus	Fish (Actinopterygii)	Perciformes	Anabantidae	Climbing Perch
110	XM 031846856 .1	Anarrhichthys ocellatus	Fish (Actinopterygii)	Perciformes	Anarhichadidae	Doctorfish
111	XM 030093392 .1	Salarias fasciatus	Fish (Actinopterygii)	Perciformes	Blenniidae	Banded Blenny
112	XM 029427721 .1	Cottoperca gobio	Fish (Actinopterygii)	Perciformes	Bovichtidae	Channel Bull Blenny
113	XM 023401677 .1	Seriola lalandi dorsalis	Fish (Actinopterygii)	Perciformes	Carangidae	A Ray-finned Fish
114	XM 022749333 .1	Seriola dumerili	Fish (Actinopterygii)	Perciformes	Carangidae	Allied Kingfish
115	XM 005943300 .2	Haplochromis burtoni	Fish (Actinopterygii)	Perciformes	Cichlidae	Burton's Haplo
116	XM 031728950 .1	Oreochromis aureus	Fish (Actinopterygii)	Perciformes	Cichlidae	Blue Tilapia
117	XM 005724112 .1	Pundamilia nyererei	Fish (Actinopterygii)	Perciformes	Cichlidae	An African Cichlid
118	XM 030726279 .1	Archocentrus centrarchus	Fish (Actinopterygii)	Perciformes	Cichlidae	Flier Cichlid
119	XM 026164370 .1	Astatotilapia calliptera	Fish (Actinopterygii)	Perciformes	Cichlidae	Eastern Bream
120	XM 003445805.5	Oreochromis niloticus	Fish (Actinopterygii)	Perciformes	Cichlidae	Baringo Tilapia
121	XM 004543425 .2	Maylandia zebra	Fish (Actinopterygii)	Perciformes	Cichlidae	Blue Mbuna
122	XM 029498206 .1	Echeneis naucrates	Fish (Actinopterygii)	Perciformes	Echeneidae	Australian Remora
123	XM 020925939 .1	Boleophthalmus pectinirostris	Fish (Actinopterygii)	Perciformes	Gobiidae	Great Blue Spotted Mudskipper
124	XM 020637971 .2	Labrus bergylta	Fish (Actinopterygii)	Perciformes	Labridae	Ballan Wrasse
125	XM 018683673 .1	Lates calcarifer	Fish (Actinopterygii)	Perciformes	Latidae	Asian Seabass
126	XM 010792153 .1	Notothenia coriiceps	Fish (Actinopterygii)	Perciformes	Nototheniidae	Black Rockcod
127	XM 031306367 .1	Sander lucioperca	Fish (Actinopterygii)	Perciformes	Percidae	European Pike-perch
128	XM 028585562 .1	Perca flavescens	Fish (Actinopterygii)	Perciformes	Percidae	American Perch
129	XM 008292540 .1	Stegastes partitus	Fish (Actinopterygii)	Perciformes	Pomacentridae	Bicolor Damselfish
130	XM 023268388 .1	Amphiprion ocellaris	Fish (Actinopterygii)	Perciformes	Pomacentridae	Anemone Demoiselle
131	XM 022208296 .1	Acanthochromis polyacanthus	Fish (Actinopterygii)	Perciformes	Pomacentridae	Damselfish
132	XM 010731844 .3	Larimichthys crocea	Fish (Actinopterygii)	Perciformes	Sciaenidae	Croceine Croaker
133	XM 030415376 .1	Sparus aurata	Fish (Actinopterygii)	Perciformes	Sparidae	Bream
134	XM 017032425 .2	Cynoglossus semilaevis	Fish (Actinopterygii)	Pleuronectiformes	Cynoglossidae	Tongue Sole
135	XM 020079676 .1	Paralichthys olivaceus	Fish (Actinopterygii)	Pleuronectiformes	Paralichthyidae	Bastard Halibut
136	XM 028799807 .1	Erpetoichthys calabaricus	Fish (Actinopterygii)	Polypteriformes	Polypteridae	Reed
137	XM 014207453 .1	Salmo salar	Fish (Actinopterygii)	Salmoniformes	Salmonidae	Atlantic Salmon
138	XM 021577603 .1	Oncorhynchus mykiss	Fish (Actinopterygii)	Salmoniformes	Salmonidae	Basier
139	XM 024143199 .1	Salvelinus alpinus	Fish (Actinopterygii)	Salmoniformes	Salmonidae	Alpine Char
140	XM 027168723 .1	Tachysurus fulvidraco	Fish (Actinopterygii)	Siluriformes	Bagridae	Banded Catfish
141	XM 017458347 .1	Ictalurus punctatus	Fish (Actinopterygii)	Siluriformes	Ictaluridae	Catfish

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
142	XM 026947809 .1	Pangasianodon hypophthalmus	Fish (Actinopterygii)	Siluriformes	Pangasiidae	Irridescent Shark-Catfish
143	XM 026320164 .1	Mastacembelus armatus	Fish (Actinopterygii)	Synbranchiformes	Mastacembelidae	Marbled Spiny Eel
144	XM 020609990 .1	Monopterus albus	Fish (Actinopterygii)	Synbranchiformes	Synbranchidae	Asian Swamp Eel
145	XM 019887002 .1	Hippocampus comes	Fish (Actinopterygii)	Syngnathiformes	Syngnathidae	Tiger-tailed Seahorse
146	XM 029846414 .1	Takifugu rubripes	Fish (Actinopterygii)	Tetraodontiformes	Tetraodontidae	Japanese Pufferfish
147	XM 007891654 .1	Callorhinchus milii	Fish (Holocephali)	Chimaeriformes	Callorhinchidae	Elephant Fish
148	XM 005997853 .2	Latimeria chalumnae	Fish (Sarcopterygii)	Coelacanthiformes	Latimeriidae	Coelacanth
149	XM 006835610 .1	Chrysochloris asiatica	Mammalia	Afroscoricida	Chrysochloridae	Cape Golden Mole
150	XM 004709945 .1	Echinops telfairi	Mammalia	Afroscoricida	Tenrecidae	Lesser Hedgehog Tenrec
151	NM 001290107 .1	Capra hircus	Mammalia	Artiodactyla	Bovidae	Goat
152	XM 005903111 .1	Bos mutus	Mammalia	Artiodactyla	Bovidae	Wild Yak
153	XM 012106267 .3	Ovis aries	Mammalia	Artiodactyla	Bovidae	Red Sheep
154	XM 010834699 .1	Bison bison bison	Mammalia	Artiodactyla	Bovidae	Bison
155	XM 006041540 .2	Bubalus bubalis	Mammalia	Artiodactyla	Bovidae	Water Buffalo
156	XM 005228428.4	Bos taurus	Mammalia	Artiodactyla	Bovidae	Cattle
157	XM 019956160 .1	Bos indicus	Mammalia	Artiodactyla	Bovidae	Brahama Cattle
158	XM 027533926 .1	Bos indicus x Bos taurus	Mammalia	Artiodactyla	Bovidae	Hybrid Cattle
159	XM 010968001 .1	Camelus bactrianus	Mammalia	Artiodactyla	Camelidae	Bactrian Camel
160	XM 010993415 .2	Camelus dromedarius	Mammalia	Artiodactyla	Camelidae	Arabian Camel
161	XM 006212647 .3	Vicugna pacos	Mammalia	Artiodactyla	Camelidae	Alpaca
162	XM 020913306 .1	Odocoileus virginianus texanus	Mammalia	Artiodactyla	Cervidae	Texan Whitetail
163	NM 001123070 .1	Sus scrofa	Mammalia	Artiodactyla	Suidae	Wild Boar
164	XM 025437140 .1	Canis lupus dingo	Mammalia	Carnivora	Canidae	Dingo
165	NM 001165260 .1	Canis lupus familiaris	Mammalia	Carnivora	Canidae	Domestic Dog
166	XM 025986727 .1	Vulpes vulpes	Mammalia	Carnivora	Canidae	Red Fox
167	XM 025934632 .1	Puma concolor	Mammalia	Carnivora	Felidae	Cougar
168	XM 027054496 .1	Acinonyx jubatus	Mammalia	Carnivora	Felidae	Cheetah
169	XM 023248796 .1	Felis catus	Mammalia	Carnivora	Felidae	Domestic Cat
170	XM 030304979 .1	Lynx canadensis	Mammalia	Carnivora	Felidae	Canada Lynx
171	XM 019417963 .1	Panthera pardus	Mammalia	Carnivora	Felidae	Leopard
172	XM 007090080 .2	Panthera tigris altaica	Mammalia	Carnivora	Felidae	Amur Tiger
173	XM 029930396 .1	Suricata suricatta	Mammalia	Carnivora	Herpestidae	Meerkat
174	XM 032331786 .1	Mustela erminea	Mammalia	Carnivora	Mustelidae	Ermine
175	XM 022518370 .1	Enhydra lutris kenyoni	Mammalia	Carnivora	Mustelidae	Northern Sea Otter
176	NM 001310190 .1	Mustela putorius furo	Mammalia	Carnivora	Mustelidae	Feral Ferret

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
177	XM 004415391 .1	Odobenus rosmarus divergens	Mammalia	Carnivora	Odobenidae	Pacific Walrus
178	XM 025857612 .1	Callorhinus ursinus	Mammalia	Carnivora	Otariidae	Northern Fur Seal
179	XM 028115021 .1	Eumetopias jubatus	Mammalia	Carnivora	Otariidae	Northern Sea Lion
180	XM 027609552 .1	Zalophus californianus	Mammalia	Carnivora	Otariidae	Californian Sealion
181	XM 031030890 .1	Leptonychotes weddellii	Mammalia	Carnivora	Phocidae	Weddell Seal
182	XM 021680805 .1	Neomonachus schauinslandi	Mammalia	Carnivora	Phocidae	Hawaiian Monk Seal
183	XM 032389615 .1	Phoca vitulina	Mammalia	Carnivora	Phocidae	Common Seal
184	XM 008696415 .1	Ursus maritimus	Mammalia	Carnivora	Ursidae	Polar Bear
185	XM 026478080 .1	Ursus arctos horribilis	Mammalia	Carnivora	Ursidae	Brown Bear
186	XM 002930611 .3	Ailuropoda melanoleuca	Mammalia	Carnivora	Ursidae	Bamboo Bear
187	PITB01003962 .1	Paradoxurus hermaphroditus	Mammalia	Carnivora	Viverridae	Asian Palm Civet
188	XM 028164550 .1	Balaenoptera acutorostrata scammoni	Mammalia	Cetacea	Balaenopteridae	North Pacific Minke Whale
189	XM 030848131 .1	Globicephala melas	Mammalia	Cetacea	Delphinidae	Long-finned Pilot Whale
190	XM 027095797 .1	Lagenorhynchus obliquidens	Mammalia	Cetacea	Delphinidae	Pacific White-sided Dolphin
191	XM 019925618 .1	Tursiops truncatus	Mammalia	Cetacea	Delphinidae	Bottle-nosed Dolphin
192	XM 004269657 .1	Orcinus orca	Mammalia	Cetacea	Delphinidae	Killer Whale
193	XM 007466327 .1	Lipotes vexillifer	Mammalia	Cetacea	Lipotidae	Baiji
194	XM 029239971 .1	Monodon monoceros	Mammalia	Cetacea	Monodontidae	Narwal
195	XM 022562652 .2	Delphinapterus leucas	Mammalia	Cetacea	Monodontidae	Beluga
196	XM 024744126 .1	Neophocaena asiaeorientalis asiaeorientalis	Mammalia	Cetacea	Phocoenidae	Yangtze Finless Porpoise
197	XM 024115511 .2	Physeter catodon	Mammalia	Cetacea	Physeteridae	Sperm Whale
198	XM 019667391 .1	Hipposideros armiger	Mammalia	Chiroptera	Hipposideridae	Greater Himalayan Leaf-nosed Bat
199	XM 016202967 .1	Miniopterus natalensis	Mammalia	Chiroptera	Miniopteridae	Natal Long-fngered Bat
200	NW_023425355 .1	Molossus molossus	Mammalia	Chiroptera	Molossidae	Pallas' Free-tailed Bat
201	NC_050198 .1	Phyllostomus discolor	Mammalia	Chiroptera	Phyllostomidae	Pale Spear-nosed Bat
202	XM 024569930 .1	Desmodus rotundus	Mammalia	Chiroptera	Phyllostomidae	Common Vampire Bat
203	XM 006911647 .1	Pteropus alecto	Mammalia	Chiroptera	Pteropodidae	Black Flying Fox
204	NW_023416312 .1	Rousettus aegyptiacus	Mammalia	Chiroptera	Pteropodidae	Egyptian Rousette
205	XM 011362973 .2	Pteropus vampyrus	Mammalia	Chiroptera	Pteropodidae	Large Flying Fox
206	NC_046284 .1	Rhinolophus ferrumequinum	Mammalia	Chiroptera	Rhinolophidae	Greater Horseshoe Bat
207	XM 006775210 .2	Myotis davidii	Mammalia	Chiroptera	Vespertilionidae	Davids Mouse-Eared Bat
208	XM 014544294 .1	Myotis brandtii	Mammalia	Chiroptera	Vespertilionidae	Brandts Mouse-eared Bat
209	NW_023416346 .1	Myotis myotis	Mammalia	Chiroptera	Vespertilionidae	Greater Mouse-eared Bat
210	NW_023425471 .1	Pipistrellus kuhlii	Mammalia	Chiroptera	Vespertilionidae	Kuhls Pipistrelle
211	XM 023753669 .1	Myotis lucifugus	Mammalia	Chiroptera	Vespertilionidae	Little Brown Mouse-eared Bat

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
212	XM 008154928 .2	Eptesicus fuscus	Mammalia	Chiroptera	Vespertilionidae	Big Brown Bat
213	XM 004449067 .3	Dasypus novemcinctus	Mammalia	Cingulata	Dasypodidae	Nine Banded Armadillo
214	XM 007538608 .2	Erinaceus europaeus	Mammalia	Erinaceomorpha	Erinaceidae	West European Hedgehog
215	XM 002719845 .3	Oryctolagus cuniculus	Mammalia	Lagomorpha	Leporidae	European Rabbit
216	XM 004597492 .2	Ochotona princeps	Mammalia	Lagomorpha	Ochotonidae	American Pika
217	XM 006892395 .1	Elephantulus edwardii	Mammalia	Macroscelidae	Macroscelididae	Cape Elephant Shrew
218	XM 014857647 .1	Equus asinus	Mammalia	Perissodactyla	Equidae	African Wild Ass
219	XM 001490191.5	Equus caballus	Mammalia	Perissodactyla	Equidae	Horse
220	XM 008544773 .1	Equus przewalskii	Mammalia	Perissodactyla	Equidae	Prezewalski's Horse
221	XM 004435149 .2	Ceratotherium simum simum	Mammalia	Perissodactyla	Rhinocerotidae	Southern White Rhino
222	JPTV01018035 .1	Manis pentadactyla	Mammalia	Pholidota	Manidae	Chinese Pangolin
223	XM 017650257 .1	Manis javanica	Mammalia	Pholidota	Manidae	Sunda Pangolin
224	SOZM010071730 .1	Manis tricuspis (Phataginus tricuspis)	Mammalia	Pholidota	Minidae	Tree Pangolin
225	XM 012434682 .2	Aotus nancymaae	Mammalia	Primates	Aotidae	Ma's Night Monkey
226	XM 008988993 .1	Callithrix jacchus	Mammalia	Primates	Callitrichidae	Common Marmoset
227	XM 010336623 .1	Saimiri boliviensis boliviensis	Mammalia	Primates	Cebidae	Bolivian Squirrel Monkey
228	XM 032285963 .1	Sapajus apella	Mammalia	Primates	Cebidae	Guianan Brown Capuchin
229	XM 017512376 .1	Cebus capucinus imitator	Mammalia	Primates	Cebidae	Panamanian White-faced Capuchin
230	XM 017888580 .1	Rhinopithecus bieti	Mammalia	Primates	Cercopithecidae	Black Snub-nosed Monkey
231	XM 007991113 .1	Chlorocebus sabaeus	Mammalia	Primates	Cercopithecidae	Green Monkey
232	XM 011995533 .1	Mandrillus leucophaeus	Mammalia	Primates	Cercopithecidae	Drill
233	XM 025372062 .1	Theropithecus gelada	Mammalia	Primates	Cercopithecidae	Gelada
234	XM 011735203 .2	Macaca nemestrina	Mammalia	Primates	Cercopithecidae	Southern Pig-tailed Macaque
235	XM 012035808 .1	Cercocebus atys	Mammalia	Primates	Cercopithecidae	Sooty Mangabey
236	XM 023199053 .2	Piliocolobus tephrosceles	Mammalia	Primates	Cercopithecidae	Ashy Red Colobus
237	XM 021933040 .1	Papio anubis	Mammalia	Primates	Cercopithecidae	Olive Baboon
238	XM 010366065 .2	Rhinopithecus roxellana	Mammalia	Primates	Cercopithecidae	Golden Snub-nosed Monkey
239	XM 005593037 .2	Macaca fascicularis	Mammalia	Primates	Cercopithecidae	Crab-eating Macaque
240	NM 001135696 .1	Macaca mulatta	Mammalia	Primates	Cercopithecidae	Rhesus Macaque
241	XM 020285237 .1	Microcebus murinus	Mammalia	Primates	Cheirogaleidae	Gray Mouse Lemur
242	XM 003791864 .2	Otolemur garnettii	Mammalia	Primates	Galagidae	Garnett's Greater Galago
243	NM 001131132 .2	Pongo abelii	Mammalia	Primates	Hominidae	Sumatran Orangutan
244	NM 001371415 .1	Homo sapiens	Mammalia	Primates	Hominidae	Human
245	XM 016942979 .1	Pan troglodytes	Mammalia	Primates	Hominidae	Chimpanzee
246	XM 019019204 .1	Gorilla gorilla gorilla	Mammalia	Primates	Hominidae	Western Lowland Gorilla

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
247	XM 008974180 .1	Pan paniscus	Mammalia	Primates	Hominidae	Bonobo
248	XM 003261084 .3	Nomascus leucogenys	Mammalia	Primates	Hylobatidae	Northern White-cheeked Crested Gibbon
249	XM 012638731 .1	Propithecus coquereli	Mammalia	Primates	Indriidae	Coquerel's Sifaka
250	XM 008064619 .1	Carlito syrichta	Mammalia	Primates	Tarsiidae	Phillipine Tarsier
251	XM 023555192 .1	Loxodonta africana	Mammalia	Proboscidea	Elephantidae	African Bush Elephant
252	XM 004866100 .3	Heterocephalus glaber	Mammalia	Rodentia	Bathyergidae	Naked Mole-rat
253	XM 010645175 .2	Fukomys damarensis	Mammalia	Rodentia	Bathyergidae	Damara Mole Rat
254	XM 023562040 .1	Cavia porcellus	Mammalia	Rodentia	Caviidae	Domesticated Guinea Pig
255	XM 013506974 .1	Chinchilla lanigera	Mammalia	Rodentia	Chinchillidae	Long-tailed Chinchilla
256	XM 028887776 .1	Peromyscus leucopus	Mammalia	Rodentia	Cricetidae	White-footed Deermouse
257	XM 006973207 .2	Peromyscus maniculatus bairdii	Mammalia	Rodentia	Cricetidae	Deer Mouse
258	XM 005074209 .2	Mesocricetus auratus	Mammalia	Rodentia	Cricetidae	Golden Hamster
259	XM 003503235.4	Cricetulus griseus	Mammalia	Rodentia	Cricetidae	Striped Dwarf Hamster
260	XM 005358761 .3	Microtus ochrogaster	Mammalia	Rodentia	Cricetidae	Prairie Vole
261	XM 004671466 .2	Jaculus jaculus	Mammalia	Rodentia	Dipodidae	Lesser Egyptian Jerboa
262	XM 013032118 .1	Dipodomys ordii	Mammalia	Rodentia	Hetermyidae	Ords Kangaroo Rat
263	NM 027286.4	Mus musculus	Mammalia	Rodentia	Muridae	Eastern House Mouse
264	XM 021188276 .2	Mus pahari	Mammalia	Rodentia	Muridae	Indochinese Shrewlike Mouse
265	XM 031370882 .1	Mastomys coucha	Mammalia	Rodentia	Muridae	Southern African Mastomys
266	XM 028762128 .1	Grammomys surdaster	Mammalia	Rodentia	Muridae	African Thicket Rats
267	XM 008840876 .2	Nannospalax galili	Mammalia	Rodentia	Muridae	Spalax
268	XM 021153479 .2	Mus caroli	Mammalia	Rodentia	Muridae	Ryukyu Mouse
269	XM 023719547 .1	Octodon degus	Mammalia	Rodentia	Octodontidae	Degu
270	XM 015488054 .1	Marmota marmota marmota	Mammalia	Rodentia	Sciruidae	Alpine Marmot
271	XM 027946507 .1	Marmota flaviventris	Mammalia	Rodentia	Sciruidae	Yellow-bellied Marmot
272	XM 005315994 .3	Ictidomys tridecemlineatus	Mammalia	Rodentia	Sciruidae	Thirteen-lined Ground Squirrel
273	XM 026396720 .1	Urocitellus parryii	Mammalia	Rodentia	Sciuridae	Arctic Ground Squirrel
274	XM 015488054 .1	Marmota marmota marmota	Mammalia	Rodentia	Sciuridae	Alpine Marmot
275	XM 006164692 .3	Tupaia chinensis	Mammalia	Scandentia	Tupaiidae	Chinese Tree Shrew
276	XM 004386324 .2	Trichechus manatus latirostris	Mammalia	Sirenia	Trichechdae	Florida Manatee
277	XM 004612209 .1	Sorex araneus	Mammalia	Soricomorpha	Soricidae	Common Shrew
278	XM 012730417 .1	Condylura cristata	Mammalia	Soricomorpha	Talpidae	Star-nosed Mole
279	XM 007952837 .1	Orycteropus afer afer	Mammalia	Tubulidentata	Orycteropodidae	Aardvark
280	XM 031958965 .1	Sarcophilus harrisii	Mammalia (Marsupialia)	Dasyuromorphia	Dasyuridae	Tasmanian Devil
281	XM 007500873 .2	Monodelphis domestica	Mammalia (Marsupialia)	Didelphimorphia	Didelphidae	Grey Short-tailed Oppossum

Table 10.1 continued from previous page

	Accession number	Latin name	Class	Order	Family	Common Name
282	XM 021007494 .1	Phascolarctos cinereus	Mammalia (Marsupialia)	Diprotodontia	Phascolarctide	Koala
283	XM 027835355 .1	Vombatus ursinus	Mammalia (Marsupialia)	Diprotodontia	Vombatidae	Common Wombat
284	XM 001515547.4	Ornithorhynchus anatinus	Mammalia (Marsupialia)	Monotremata	Ornithorhynchidae	Platypus
285	XM 019495142 .1	Alligator mississippiensis	Reptilia	Crocodylia	Alligatoridae	Alligator
286	XM 025210843 .1	Alligator sinensis	Reptilia	Crocodylia	Alligatoridae	Chinese Aligator
287	XM 019529281 .1	Crocodylus porosus	Reptilia	Crocodylia	Crocodylidae	Estuarine Crocodile
288	XM 019525515 .1	Gavialis gangeticus	Reptilia	Crocodylia	Gavialidae	Gavial
289	XM 020786763 .1	Pogona vitticeps	Reptilia	Squamata	Agamidae	Central Bearded Dragon
290	XM 008107248 .2	Anolis carolinensis	Reptilia	Squamata	Dactyloidae	Green Anole
291	XM 026714269 .1	Pseudonaja textilis	Reptilia	Squamata	Elapidae	Eastern Brown Snake
292	XM 026674969 .1	Notechis scutatus	Reptilia	Squamata	Elapidae	Kreft's Tiger Snake
293	XM 015417581 .1	Gekko japonicus	Reptilia	Squamata	Gekkonidae	Schlegel's Japanese Gecko
294	XM 032227043 .1	Thamnophis elegans	Reptilia	Squamata	Natricidae	Western Terrestrial Garter Snake
295	XM 014071461 .1	Thamnophis sirtalis	Reptilia	Squamata	Natricidae	Common Garter Snake
296	XM 007431880 .3	Python bivittatus	Reptilia	Squamata	Phytonidae	Burmese Python
297	XM 029284675 .1	Protobothrops mucrosquamatus	Reptilia	Squamata	Viperidae	Brown Spotted Pitviper
298	XM 007070499 .1	Chelonia mydas	Reptilia	Testudines	Cheloniidae	Green Sea Turtle
299	XM 024108749 .1	Chrysemys picta bellii	Reptilia	Testudines	Emydidae	Western Painted Turtle
300	XM 030552021 .1	Gopherus evgoodei	Reptilia	Testudines	Testudinidae	Goode's Thornscrub Tortoise
301	XM 006122829 .3	Pelodiscus sinensis	Reptilia	Testudines	Trionychidae	Chinese Soft-shelled Turtle

Table 10.1: List of species used in the analysis with accession number, Latin name, class, order, family, and common English name

	Species	Matrix		Species	Matrix
1	**Gorilla gorilla gorilla**	100	45	**Bos taurus**	82.5
2	**Macaca fascicularis**	100	46	**Bubalus bubalis**	82.5
3	**Macaca mulatta**	100	47	**Capra hircus**	82.5
4	**Macaca nemestrina**	100	48	**Oryctolagus cuniculus**	82.5
5	**Pan paniscus**	100	49	**Physeter catodon**	82.5
6	**Pan troglodytes**	100	50	**Camelus dromedarius**	81.25
7	**Cercocebus atys**	98.75	51	**Delphinapterus leucas**	81.25
8	**Chlorocebus sabaeus**	98.75	52	**Lipotes vexillifer**	81.25
9	**Mandrillus leucophaeus**	98.75	53	**Microcebus murinus**	81.25
10	**Nomascus leucogenys**	98.75	54	**Microtus ochrogaster**	81.25
11	**Papio anubis**	98.75	55	**Monodon monoceros**	81.25
12	**Piliocolobus tephrosceles**	98.75	56	**Octodon degus**	81.25
13	**Pongo abelii**	98.75	57	**Odocoileus virginianus texanus**	81.25
14	**Rhinopithecus roxellana**	98.75	58	**Ovis aries**	81.25
15	**Theropithecus gelada**	98.75	59	**Canis lupus dingo**	80.07
16	**Saimiri boliviensis boliviensis**	91.25	60	**Ailuropoda melanoleuca**	80
17	**Aotus nancymaae**	90	61	**Carlito syrichta**	80
18	**Callithrix jacchus**	90	62	**Ceratotherium simum simum**	80
19	**Cebus capucinus imitator**	90	63	**Grammomys surdaster**	80
20	**Sapajus apella**	88.75	64	**Mus pahari**	80
21	**Propithecus coquereli**	87.5	65	**Neophocaena asiaeorientalis**	80
22	**Felis catus**	85	66	**Ochotona princeps**	80
23	**Ictidomys tridecemlineatus**	85	67	**Ursus arctos horribilis**	80
24	**Jaculus jaculus**	85	68	**Ursus maritimusp**	80
25	**Lynx canadensis**	85	69	**Equus asinus**	78.75
26	**Marmota flaviventris**	85	70	**Equus caballus**	78.75
27	**Mesocricetus auratus**	85	71	**Equus przewalskii**	78.75
28	**Panthera pardus**	85	72	**Eumetopias jubatus**	78.75
29	**Panthera tigris altaicap**	85	73	**Mus caroli**	78.75
30	**Puma concolor**	85	74	**Neomonachus schauinslandi**	78.75
31	**Acinonyx jubatus**	83.75	75	**Paradoxurus hermaphroditus**	78.75
32	**Chinchilla lanigera**	83.75	76	**Phoca vitulina**	78.75
33	**Cricetulus griseus**	83.75	77	**Sus scrofa**	78.75
34	**Dipodomys ordii**	83.75	78	**Tursiops truncatus**	78.75
35	**Fukomys damarensis**	83.75	79	**Vicugna pacos**	78.75
36	**Heterocephalus glaber**	83.75	80	**Vulpes vulpes**	78.75
37	**Nannospalax galili**	83.75	81	**Callorhinus ursinus**	77.5
38	**Peromyscus leucopus**	83.75	82	**Cavia porcellus**	77.5
39	**Peromyscus maniculatus bairdii**	83.75	83	**Lagenorhynchus obliquidens**	77.5
40	**Urocitellus parryii**	83.75	84	**Mastomys coucha**	77.5
41	**Balaenoptera acutorostrata scammoni**	82.5	85	**Mus musculus**	77.5
42	**Bison bison bisonp**	82.5	86	**Odobenus rosmarus divergens**	77.5
43	**Bos indicus**	82.5	87	**Orcinus orca**	77.5
44	**Bos mutus**	82.5	88	**Suricata suricatta**	77.5

Table 10.2: List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score.

	Species	Matrix		Species	Matrix
89	**Zalophus californianus**	77.5	133	**Cuculus canorus**	58.75
90	**Canis lupus familiaris**	76.25	134	**Dromaius novaehollandiae**	58.75
91	**Globicephala melas**	76.25	135	**Pelecanus crispus**	58.75
92	**Pteropus alecto**	76.25	136	**Phalacrocorax carbo**	58.75
93	**Pteropus vampyrus**	76.25	137	**Tauraco erythrolophusp**	58.75
94	**R aegyptiacus Bat1k**	76.25	138	**Anser cygnoides domesticus**	57.5
95	**Enhydra lutris kenyoni**	75	139	**Aptenodytes forsteri**	57.5
96	**Manis javanica**	75	140	**Aquila chrysaetos canadensis**	57.5
97	**Manis pentadactyla**	75	141	**Aquila chrysaetos chrysaetos**	57.5
98	**Mustela erminea**	75	142	**Athene cunicularia**	57.5
99	**Mustela putorius furo**	75	143	**Gallus gallus**	57.5
100	**Trichechus manatus latirostris**	75	144	**Gavia stellata**	57.5
101	**Mollosus mollosus**	73.75	145	**Haliaeetus leucocephalus**	57.5
102	**Myotis myotis**	73.75	146	**Leptosomus discolor**	57.5
103	**Myotis brandtii**	73.75	147	**Melopsittacus undulatus**	57.5
104	**Myotis davidii**	73.75	148	**Phaethon lepturus**	57.5
105	**Myotis lucifugus**	73.75	149	**Pterocles gutturalisp**	57.5
106	**Otolemur garnettii**	73.75	150	**Pygoscelis adeliae**	57.5
107	**Dasypus novemcinctus**	72.5	151	**Tinamus guttatus**	57.5
108	**Eptesicus fuscus**	72.5	152	**Pelodiscus sinensis**	56.962
109	**Loxodonta africana**	72.5	153	**Erpetoichthys calabaricus**	56.79
110	**Hipposideros armiger**	71.25	154	**Apteryx rowi**	56.25
111	**Rhinolophus ferrumequinum**	71.25	155	**Calypte anna**	56.25
112	**Tupaia chinensis**	71.25	156	**Cariama cristata**	56.25
113	**Elephantulus edwardii**	70	157	**Corvus moneduloides**	56.25
114	**Miniopterus natalensis**	70	158	**Mesitornis unicolor**	56.25
115	**Orycteropus afer afer**	68.75	159	**Nipponia nippon**	56.25
116	**Phyllostomus discolor**	68.75	160	**Numida meleagris**	56.25
117	**Sarcophilus harrisii**	68.75	161	**Serinus canaria**	56.25
118	**Condylura cristata**	66.25	162	**Struthio camelus australis**	56.25
119	**Erinaceus europaeus**	66.25	163	**Calidris pugnax**	55
120	**Monodelphis domestica**	66.25	164	**Chlamydotis macqueenii**	55
121	**Pipistrellus kuhlii**	66.25	165	**Corvus cornix cornix**	55
122	**Phascolarctos cinereus**	66.25	166	**Empidonax traillii**	55
123	**Vombatus ursinus**	66.25	167	**Gekko japonicus**	55
124	**Chrysochloris asiatica**	65	168	**Haliaeetus albicillap**	55
125	**Desmodus rotundus**	62.5	169	**Meleagris gallopavo**	55
126	**Echinops telfairi**	62.5	170	**Merops nubicus**	55
127	**Sorex araneus**	62.5	171	**Opisthocomus hoazin**	55
128	**Nothoprocta perdicaria**	61.25	172	**Phasianus colchicus**	55
129	**Ornithorhynchus anatinus**	61.25	173	**Strigops habroptila**	55
130	**Anas platyrhynchos**	58.75	174	**Egretta garzetta**	53.807
131	**Aythya fuligula**	58.75	175	**Chaetura pelagica**	53.75
132	**Charadrius vociferus**	58.75	176	**Corapipo altera**	53.75

Table 10.2: Continued. List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score.

Species	Matrix		Species	Matrix		Species	Matrix
Cyanistes caeruleus	**53.75**	221	**Anarrhichthys ocellatus**	**48.148**	265	**Columba livia**	42.857
Falco cherrug	**53.75**	222	**Astatotilapia calliptera**	**48.148**	266	**Nanorana parkeri**	42.5
Fulmarus glacialisp	**53.75**	223	**Carassius auratus**	**48.148**	267	**Clupea harengus**	41.975
Geospiza fortis	**53.75**	224	**Denticeps clupeoides**	**48.148**	268	**Esox lucius**	41.975
Lepidothrix coronata	**53.75**	225	**Haplochromis burtoni**	**48.148**	269	**Gadus morhua**	41.975
Manacus vitellinus	**53.75**	226	**Kryptolebias marmoratus**	**48.148**	270	**Nothobranchius furzeri**	41.975
Neopelma chrysocephalum	**53.75**	227	**Lates calcarifer**	**48.148**	271	**Paralichthys olivaceus**	41.975
Pipra filicauda	**53.75**	228	**Maylandia zebra**	**48.148**	272	**Austrofundulus limnaeus**	40.797
Pseudonaja textilis	**53.75**	229	**Oreochromis aureus**	**48.148**	273	**Boleophthalmus pectinirostris**	40.741
Sturnus vulgaris	**53.75**	230	**Pundamilia nyererei**	**48.148**	274	**Labrus bergylta**	40.741
Taeniopygia guttata	**53.75**	231	**Salarias fasciatus**	**48.148**	275	**Oncorhynchus mykiss**	40.741
Thamnophis elegans	**53.75**	232	**Callorhinchus milii**	**46.914**	276	**Poecilia formosa**	40.741
Protobothrops mucrosquamatus	**53.086**	233	**Danio rerio**	**46.914**	277	**Poecilia latipinna**	40.741
Thamnophis sirtalis	**53.086**	234	**Echeneis naucrates**	**46.914**	278	**Poecilia mexicana**	40.741
Tyto alba	**52.841**	235	**Notothenia coriiceps**	**46.914**	279	**Poecilia reticulata**	40.741
Buceros rhinoceros silvestris	**52.5**	236	**Sander lucioperca**	**46.914**	280	**Xiphophorus couchianus**	40.741
Chelonia mydas	**52.5**	237	**Scleropages formosus**	**46.914**	281	**Xiphophorus maculatus**	40.741
Chrysemys picta bellii	**52.5**	238	**Astyanax mexicanus**	**46.25**	282	**Lepisosteus oculatus**	40
Coturnix japonica	**52.5**	239	**Corvus brachyrhynchos**	**46.25**	283	**Acanthochromis polyacanthus**	39.506
Eurypyga helias	**52.5**	240	**Pygocentrus nattereri**	**46.25**	284	**Salvelinus alpinusp**	38.272
Ficedula albicollis	**52.5**	241	**Cyprinodon variegatus**	**45.679**			
Lonchura striata domestica	**52.5**	242	**Electrophorus electricus**	**45.679**			
Notechis scutatus	**52.5**	243	**Gouania willdenowi**	**45.679**			
Pseudopodoces humilis	**52.5**	244	**Ictalurus punctatus**	**45.679**			
Rhinatrema bivittatum	**52.5**	245	**Latimeria chalumnae**	**45.679**			
Apteryx australis mantelli	**52.174**	246	**Myripristis murdjan**	**45.679**			
Camarhynchus parvulus	**51.852**	247	**Sinocyclocheilus anshuiensis**	**45.679**			
Anolis carolinensis	**51.25**	248	**Sparus aurata**	**45.679**			
Parus major	**51.25**	249	**Xenopus tropicalis**	**45**			
Python bivittatus	**51.25**	250	**Cottoperca gobio**	**44.444**			
Zonotrichia albicollis	**51.25**	251	**Mastacembelus armatus**	**44.444**			
Monopterus albus	**50.617**	252	**Oryzias melastigma**	**44.444**			
Antrostomus carolinensis	**50.455**	253	**Pangasianodon hypophthalmus**	**44.444**			
Falco peregrinus	**50.17**	254	**Seriola dumerili**	**44.444**			
Apaloderma vittatum	**50.057**	255	**Seriola lalandi dorsalis**	**44.444**			
Microcaecilia unicolor	**50**	256	**Stegastes partitus**	**44.444**			
Pogona vitticeps	**50**	257	**Tachysurus fulvidraco**	**44.444**			
Archocentrus centrarchus	**49.383**	258	**Takifugu rubripes**	**44.444**			
Gavialis gangeticus	**49.383**	259	**Amphiprion ocellaris**	**43.21**			
Hippocampus comes	**49.383**	260	**Anabas testudineus**	**43.21**			
Larimichthys crocea	**49.383**	261	**Cynoglossus semilaevis**	**43.21**			
Oreochromis niloticus	**49.383**	262	**Fundulus heteroclitus**	**43.21**			
Alligator sinensis	**48.75**	263	**Parambassis ranga**	**43.21**			
Gopherus evgoodei	**48.75**	264	**Perca flavescens**	**43.21**			

Table 10.2: Continued. List of species used in a SARS-CoV 2 ACE 2 binding site alignment together with % identity matrix score.

M1/M2

Sequences/Test	lnL	2Δ x (lnL1 - lnL0)	Parameters (np)	(np1 - np0)		ω (dN/dS)		Prob. density	Cum. distribution	1-P (<0.05)	Result
ACE2											
0 (M1)	-17368.4604		514			0.34					
1 (M2)	-17360.5313	15.858	515	1		0.33		0.0	1.000	0.000	Significant

M8a/M8

Sequences/Test	lnL	2Δ x (lnL1 - lnL0)	Parameters (np)	(np1 - np0)	p1b	ω (dN/dS)		Prob. density	Cum. distribution	1-P (<0.05)	Result
ACE2											
0 (M8a)	-16589.7888		516								
1 (M8)	-16585.4629	8.652	517	1	0.04023	0.36		0.00	1.00	0.00	Significant

BSAn/BSA

Sequences/Test a	lnL	2Δ x (lnL1 - lnL0)	Parameters (np)	(np1 - np0)	p1b	Background ω (dN/dS)	Foreground ω (dN/dS)	Prob. density	Cum. distribution	1-P (<0.05)	Result
ACE2											
0 (w=1)	-16789.2195		516								
1 (BSA)	-16785.3817	7.676	517	1	0.499	0.14	4.74	∞	0.994	0.006	Significant

Table 10.3: LRT test for the branch-specific model (M1/M2), the site-specific model (M8a/M8), and the branch-site model (BSAn/BSA). p1 is the proportion of sites having $\omega > 1$

Bibliography

[1] E. C. Teeling, S. C. Vernes, L. M. Dávalos, D. A. Ray, M. T. P. Gilbert, and E. a. Myers, "Bat biology, genomes, and the bat1k project: To generate chromosome-level genomes for all living bat species," *Annual Review of Animal Biosciences*, vol. 6, no. 1, pp. 23–46, 2018. PMID: 29166127.

[2] I. Agnarsson, C. M. Zambrana-Torrelio, N. P. Flores-Saldana, and L. J. May-Collado, "A time-calibrated species-level phylogeny of bats (chiroptera, mammalia)," *PLoS Curr*, vol. 3, p. RRN1212, 2011.

[3] Z. Yang, W. S. Wong, and R. Nielsen, "Bayes empirical bayes inference of amino acid sites under positive selection," *Molecular biology and evolution*, vol. 22, no. 4, pp. 1107–1118, 2005.

[4] A. Larsson, "Aliview: a fast and lightweight alignment viewer and editor for large datasets," *Bioinformatics*, vol. 30, no. 22, pp. 3276–3278, 2014.

[5] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond, "Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data," *Bioinformatics*, vol. 28, pp. 1647 – 1649, 2012.

[6] A. Stamatakis, "Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, pp. 1312 – 1313, 2014.

[7] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Meeting modern challenges in visualization and analysis," *Protein Science*, vol. 27, no. 1, pp. 14–25, 2018.

[8] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Structure visualization for researchers, educators, and developers," *Protein Science*, vol. 30, no. 1, pp. 70–82, 2021.

[9] C. Darwin, *The origin of species*. PF Collier & son New York, 1909.

[10] L. Van Valen, "Festschrift.(book reviews: Evolutionary biology. vol. 6)," *Science*, vol. 180, no. 4085, p. 488, 1973.

[11] P. Aiewsakun and A. Katzourakis, "Time-dependent rate phenomenon in viruses," *J Virol*, vol. 90, no. 16, pp. 7184–95, 2016.

[12] S. Kumar and S. Subramanian, "Mutation rates in mammalian genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 803 – 808, 2002.

[13] T. A. Brown, "How genomes evolve," in *Genomes. 2nd edition*, Wiley-Liss, 2002.

[14] A. Katzourakis, "Paleovirology: inferring viral evolution from host genome sequence data," *Philos Trans R Soc Lond B Biol Sci*, vol. 368, no. 1626, p. 20120493, 2013.

[15] J. E. Henzy and W. E. Johnson, "Pushing the endogenous envelope," *Philos Trans R Soc Lond B Biol Sci*, vol. 368, no. 1626, p. 20120506, 2013.

[16] J. E. Henzy, R. J. Gifford, W. E. Johnson, and J. M. Coffin, "A novel recombinant retrovirus in the genomes of modern birds combines features of avian and mammalian retroviruses," *J Virol*, vol. 88, no. 5, pp. 2398–405, 2014.

[17] M. R. Patel, M. Emerman, and H. S. Malik, "Paleovirology—ghosts and gifts of viruses past," *Current opinion in virology*, vol. 1, no. 4, pp. 304–309, 2011.

[18] M. Stremlau, C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, and J. Sodroski, "The cytoplasmic body component trim5alpha restricts hiv-1 infection in old world monkeys," *Nature*, vol. 427, no. 6977, pp. 848–53, 2004.

[19] H. Lahouassa, W. Daddacha, H. Hofmann, D. Ayinde, E. C. Logue, L. Dragin, N. Bloch, C. Maudet, M. Bertrand, T. Gramberg, G. Pancino, S. Priet, B. Canard, N. Laguette, M. Benkirane, C. Transy, N. R. Landau, B. Kim, and F. Margottin-Goguet, "Samhd1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates," *Nat Immunol*, vol. 13, no. 3, pp. 223–8, 2012.

[20] A. A. Compton and M. Emerman, "Convergence and divergence in the evolution of the apobec3g-vif interaction reveal ancient origins of simian immunodeficiency viruses," *PLoS Pathog*, vol. 9, no. 1, p. e1003135, 2013.

[21] R. S. Harris and M. T. Liddament, "Retroviral restriction by apobec proteins," *Nat Rev Immunol*, vol. 4, no. 11, pp. 868–77, 2004.

[22] K. Uriu, Y. Kosugi, J. Ito, and K. Sato, "The battle between retroviruses and apobec3 genes: Its past and present," *Viruses*, vol. 13, no. 1, p. 124, 2021.

[23] J. Ito, R. J. Gifford, and K. Sato, "Retroviruses drive the rapid evolution of mammalian apobec3 genes," *Proceedings of the National Academy of Sciences*, vol. 117, no. 1, pp. 610–618, 2020.

[24] D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S. Winkler, L. S. Jermiin, E. C. Skirmuntt, A. Katzourakis, *et al.*, "Six reference-quality genomes reveal evolution of bat adaptations," *Nature*, vol. 583, no. 7817, pp. 578–584, 2020.

[25] C. Gilbert and C. Feschotte, "Genomic fossils calibrate the long-term evolution of hepadnaviruses," *PLOS Biology*, vol. 8, pp. 1–12, 09 2010.

[26] C. Feschotte and C. Gilbert, "Endogenous viruses: insights into viral evolution and impact on host biology," *Nature Reviews Genetics*, vol. 13, no. 4, pp. 283–296, 2012.

[27] A. Katzourakis and R. J. Gifford, "Endogenous viral elements in animal genomes," *PLoS Genet*, vol. 6, no. 11, p. e1001191, 2010.

[28] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, and W. FitzHugh, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[29] R. H. Waterston, E. S. Lander, and J. E. Sulston, "On the sequencing of the human genome," *Proceedings of the National Academy of Sciences*, vol. 99, no. 6, pp. 3712–3716, 2002.

[30] C. Lavialle, G. Cornelis, A. Dupressoir, C. Esnault, O. Heidmann, C. Vernochet, and T. Heidmann, "Paleovirology of 'syncytins', retroviral env genes exapted for

a role in placentation," *Phil. Trans. R. Soc. B*, vol. 368, no. 1626, p. 20120507, 2013.

[31] A. Pötgens, U. Schmitz, P. Bose, A. Versmold, P. Kaufmann, and H.-G. Frank, "Mechanisms of syncytial fusion: a review," *Placenta*, vol. 23, pp. S107–S113, 2002.

[32] A. M. Carter and A. C. Enders, "Comparative aspects of trophoblast development and placentation," *Reprod Biol Endocrinol*, vol. 2, p. 46, 2004.

[33] E. B. Chuong, "Retroviruses facilitate the rapid evolution of the mammalian placenta," *Bioessays*, vol. 35, no. 10, pp. 853–61, 2013.

[34] J. Harris, "The evolution of placental mammals," *FEBS letters*, vol. 295, no. 1-3, pp. 3–4, 1991.

[35] S. Blaise, N. de Parseval, L. Bénit, and T. Heidmann, "Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution," *Proceedings of the National Academy of Sciences*, vol. 100, no. 22, pp. 13013–13018, 2003.

[36] J.-L. Blond, D. Lavillette, V. Cheynet, O. Bouton, G. Oriol, S. Chapel-Fernandes, B. Mandrand, F. Mallet, and F.-L. Cosset, "An envelope glycoprotein of the human endogenous retrovirus herv-w is expressed in the human placenta and fuses cells expressing the type d mammalian retrovirus receptor," *Journal of virology*, vol. 74, no. 7, pp. 3321–3329, 2000.

[37] K. A. Dunlap, M. Palmarini, M. Varela, R. C. Burghardt, K. Hayashi, J. L. Farmer, and T. E. Spencer, "Endogenous retroviruses regulate periimplantation placen-

tal growth and differentiation," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14390–14395, 2006.

[38] S. Mi, X. Lee, X.-P. Li, G. M. Veldman, H. Finnerty, L. Racie, E. Lavallie, X.-Y. Tang, P. Edouard, and S. Howes, "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis," *Nature*, vol. 403, no. 6771, pp. 785–789, 2000.

[39] I. Knerr, E. Beinder, and W. Rascher, "Syncytin, a novel human endogenous retroviral gene in human placenta: evidence for its dysregulation in preeclampsia and hellp syndrome," *American journal of obstetrics and gynecology*, vol. 186, no. 2, pp. 210–213, 2002.

[40] L. Bénit, P. Dessen, and T. Heidmann, "Identification, phylogeny, and evolution of retroviral elements based on their envelope genes," *Journal of virology*, vol. 75, no. 23, pp. 11709–11719, 2001.

[41] M. Mangeney, M. Renard, G. Schlecht-Louf, I. Bouallaga, O. Heidmann, C. Letzelter, A. Richaud, B. Ducos, and T. Heidmann, "Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins," *Proceedings of the National Academy of Sciences*, vol. 104, no. 51, pp. 20534–20539, 2007.

[42] F. Redelsperger, N. Raddi, A. Bacquin, C. Vernochet, V. Mariot, V. Gache, N. Blanchard-Gutton, S. Charrin, L. Tiret, and J. Dumonceaux, "Genetic evidence that captured retroviral envelope syncytins contribute to myoblast fusion and muscle sexual dimorphism in mice," *PLoS Genet*, vol. 12, no. 9, p. e1006289, 2016.

[43] S. Blaise, N. De Parseval, and T. Heidmann, "Functional characterization of two newly identified human endogenous retrovirus coding envelope genes," *Retrovirology*, vol. 2, no. 1, p. 1, 2005.

[44] F. Arnaud, M. Caporale, M. Varela, R. Biek, B. Chessa, A. Alberti, M. Golder, M. Mura, Y.-p. Zhang, and L. Yu, "A paradigm for virus–host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses," *PLoS pathog*, vol. 3, no. 11, p. e170, 2007.

[45] H. Ikeda and H. Sugimura, "Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties," *J Virol*, vol. 63, no. 12, pp. 5405–12, 1989.

[46] C. Vernochet, O. Heidmann, A. Dupressoir, G. Cornelis, P. Dessen, F. Catzeflis, and T. Heidmann, "A syncytin-like endogenous retrovirus envelope gene of the guinea pig specifically expressed in the placenta junctional zone and conserved in caviomorpha," *Placenta*, vol. 32, no. 11, pp. 885–892, 2011.

[47] C. Esnault, G. Cornelis, O. Heidmann, and T. Heidmann, "Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the envv syncytin, captured for a function in placentation," *PLoS Genet*, vol. 9, no. 3, p. e1003400, 2013.

[48] N. De Parseval and T. Heidmann, "Human endogenous retroviruses: from infectious elements to human genes," *Cytogenetic and genome research*, vol. 110, no. 1-4, pp. 318–332, 2005.

[49] G. Cornelis, C. Vernochet, Q. Carradec, S. Souquere, B. Mulot, F. Catzeflis, M. A. Nilsson, B. R. Menzies, M. B. Renfree, and G. Pierron, "Retroviral envelope gene

captures and syncytin exaptation for placentation in marsupials," *Proceedings of the National Academy of Sciences*, vol. 112, no. 5, pp. E487–E496, 2015.

[50] N. De Parseval and T. Heidmann, "Physiological knockout of the envelope gene of the single-copy erv-3 human endogenous retrovirus in a fraction of the caucasian population," *Journal of virology*, vol. 72, no. 4, pp. 3442–3445, 1998.

[51] H. Ikeda, F. Laigret, M. A. Martin, and R. Repaske, "Characterization of a molecularly cloned retroviral sequence associated with fv-4 resistance," *Journal of virology*, vol. 55, no. 3, pp. 768–777, 1985.

[52] P. Jolicoeur, "The fv-1 gene of the mouse and its control of murine leukemia virus replication," *Curr Top Microbiol Immunol*, vol. 86, pp. 67–122, 1979.

[53] L. C. Samuelson, K. Wiebauer, D. L. Gumucio, and M. H. Meisler, "Expression of the human amylase genes: recent origin of a salivary amylase promoter from an actin pseudogene," *Nucleic acids research*, vol. 16, no. 17, pp. 8261–8276, 1988.

[54] L. C. Samuelson, K. Wiebauer, C. Snow, and M. H. Meisler, "Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution.," *Molecular and cellular biology*, vol. 10, no. 6, pp. 2513–2520, 1990.

[55] J. Ashley, B. Cordy, D. Lucia, L. G. Fradkin, V. Budnik, and T. Thomson, "Retrovirus-like gag protein arc1 binds rna and traffics across synaptic boutons," *Cell*, vol. 172, no. 1-2, pp. 262–274, 2018.

[56] D. J. Taylor, R. W. Leach, and J. Bruenn, "Filoviruses are ancient and integrated into mammalian genomes," *BMC Evolutionary Biology*, vol. 10, no. 1, p. 193, 2010.

[57] D. J. Taylor, K. Dittmar, M. J. Ballinger, and J. A. Bruenn, "Evolutionary maintenance of filovirus-like genes in bat genomes," *BMC evolutionary biology*, vol. 11, no. 1, p. 336, 2011.

[58] K. E. Jones, O. R. P. Bininda-Emonds, and J. L. Gittleman, "Bats, clocks, and rocks: Diversification patterns in chiroptera," *Evolution*, vol. 59, no. 10, pp. 2243–2255, 2005.

[59] E. C. Teeling, M. S. Springer, O. Madsen, P. Bates, S. J. O'brien, and W. J. Murphy, "A molecular phylogeny for bats illuminates biogeography and the fossil record," *Science*, vol. 307, no. 5709, pp. 580–584, 2005.

[60] R. W. Meredith, J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. Simão, T. Stadler, *et al.*, "Impacts of the cretaceous terrestrial revolution and kpg extinction on mammal diversification," *science*, vol. 334, no. 6055, pp. 521–524, 2011.

[61] T. H. Kunz, E. B. de Torrez, D. Bauer, T. Lobova, and T. H. Fleming, "Ecosystem services provided by bats," *Europe*, vol. 31, p. 32, 2011.

[62] C. H. Calisher, J. E. Childs, H. E. Field, K. V. Holmes, and T. Schountz, "Bats: important reservoir hosts of emerging viruses," *Clinical microbiology reviews*, vol. 19, no. 3, pp. 531–545, 2006.

[63] K. J. Olival, P. R. Hosseini, C. Zambrana-Torrelio, N. Ross, T. L. Bogich, and P. Daszak, "Host and viral traits predict zoonotic spillover from mammals," *Nature*, vol. 546, no. 7660, pp. 646–650, 2017.

[64] L.-F. Wang and D. E. Anderson, "Viruses in bats and potential spillover to animals and humans," *Current opinion in virology*, vol. 34, pp. 79–89, 2019.

[65] A. D. Luis, D. T. Hayman, T. J. O'Shea, P. M. Cryan, A. T. Gilbert, J. R. Pulliam, J. N. Mills, M. E. Timonin, C. K. Willis, A. A. Cunningham, *et al.*, "A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?," *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1756, p. 20122753, 2013.

[66] Y. Wang, M. Swiecki, S. A. McCartney, and M. Colonna, "dsrna sensors and plasmacytoid dendritic cells in host defense and autoimmunity," *Immunological reviews*, vol. 243, no. 1, pp. 74–90, 2011.

[67] R. Kurth and N. Bannert, *Retroviruses: molecular biology, genomics and pathogenesis*. Horizon Scientific Press, 2010.

[68] F. Broecker and K. Moelling, "Evolution of immune systems from viruses and transposable elements," *Frontiers in microbiology*, vol. 10, p. 51, 2019.

[69] J. Chen, M. Foroozesh, and Z. Qin, "Transactivation of human endogenous retroviruses by tumor viruses and their functions in virus-associated malignancies," *Oncogenesis*, vol. 8, no. 1, pp. 1–9, 2019.

[70] M. E. Engel and S. W. Hiebert, "The enemy within: dormant retroviruses awaken," *Nature medicine*, vol. 16, no. 5, pp. 517–518, 2010.

[71] R. A. Weiss, "On the concept and elucidation of endogenous retroviruses," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1626, p. 20120494, 2013.

[72] A. Katzourakis, M. Tristem, O. G. Pybus, and R. J. Gifford, "Discovery and analysis of the first endogenous lentivirus," *Proceedings of the National Academy of Sciences*, vol. 104, no. 15, pp. 6261–6265, 2007.

[73] G.-Z. Han and M. Worobey, "A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses.," *Molecular biology and evolution*, vol. 32 1, pp. 211–5, 2015.

[74] A. Hayward, C. K. Cornwallis, and P. Jern, "Pan-vertebrate comparative genomics unmasks retrovirus macroevolution," *Proceedings of the National Academy of Sciences*, vol. 112, no. 2, pp. 464–469, 2015.

[75] W. E. Johnson, "Endogenous retroviruses in the genomics era," *Annual review of virology*, vol. 2, pp. 135–159, 2015.

[76] J. A. Hayward, M. Tachedjian, J. Cui, H. Field, E. C. Holmes, L.-F. Wang, and G. Tachedjian, "Identification of diverse full-length endogenous betaretroviruses in megabats and microbats," *Retrovirology*, vol. 10, no. 1, p. 35, 2013.

[77] G. Zhang, C. Cowled, Z. Shi, Z. Huang, K. A. Bishop-Lilly, X. Fang, J. W. Wynne, Z. Xiong, M. L. Baker, and W. Zhao, "Comparative analysis of bat genomes provides insight into the evolution of flight and immunity," *Science*, vol. 339, no. 6118, pp. 456–460, 2013.

[78] M. Escalera-Zamudio, M. L. Z. Mendoza, F. Heeger, E. Loza-Rubio, E. Rojas-Anaya, M. L. Méndez-Ojeda, B. Taboada, C. J. Mazzoni, C. F. Arias, and A. D. Greenwood, "A novel endogenous betaretrovirus in the common vampire bat (desmodus rotundus) suggests multiple independent infection and cross-species transmission events," *Journal of virology*, vol. 89, no. 9, pp. 5180–5184, 2015.

[79] R. E. Tarlinton, J. Meers, and P. R. Young, "Retroviral invasion of the koala genome," *Nature*, vol. 442, no. 7098, pp. 79–81, 2006.

[80] J. Cui, G. Tachedjian, M. Tachedjian, E. C. Holmes, S. Zhang, and L.-F. Wang, "Identification of diverse groups of endogenous gammaretroviruses in mega-and microbats," *Journal of General Virology*, vol. 93, no. 9, pp. 2037–2045, 2012.

[81] X. Zhuo and C. Feschotte, "Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages," *PLoS Pathog*, vol. 11, no. 11, p. e1005279, 2015.

[82] E. C. Skirmuntt and A. Katzourakis, "The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology," *Virus research*, vol. 270, p. 197645, 2019.

[83] J. Cui, M. Tachedjian, L. Wang, G. Tachedjian, L.-F. Wang, and S. Zhang, "Discovery of retroviral homologs in bats: implications for the origin of mammalian gammaretroviruses," *Journal of virology*, vol. 86, no. 8, pp. 4288–4293, 2012.

[84] X. Zhuo, M. Rho, and C. Feschotte, "Genome-wide characterization of endogenous retroviruses in the bat myotis lucifugus reveals recent and diverse infections," *Journal of virology*, vol. 87, no. 15, pp. 8493–8501, 2013.

[85] J. Cui, G. Tachedjian, and L.-F. Wang, "Bats and rodents shape mammalian retroviral phylogeny," *Scientific reports*, vol. 5, 2015.

[86] R. Gifford, P. Kabát, J. Martin, C. Lynch, and M. Tristem, "Evolution and distribution of class ii-related endogenous retroviruses," *Journal of Virology*, vol. 79, pp. 6478 – 6486, 2005.

[87] H. Farkašová, T. Hron, J. Pačes, P. Hulva, P. Benda, R. J. Gifford, and D. Elleder, "Discovery of an endogenous deltaretrovirus in the genome of long-fingered bats

(chiroptera: Miniopteridae)," *Proceedings of the National Academy of Sciences*, p. 201621224, 2017.

[88] T. Hron, H. Farkašová, R. J. Gifford, P. Benda, P. Hulva, T. Görföl, J. Pačes, and D. Elleder, "Remnants of an ancient deltaretrovirus in the genomes of horseshoe bats (rhinolophidae)," *Viruses*, vol. 10, no. 4, p. 185, 2018.

[89] J. F. Drexler, V. M. Corman, M. A. Müller, G. D. Maganga, P. Vallo, T. Binger, F. Gloza-Rausch, V. M. Cottontail, A. Rasche, S. Yordanov, *et al.*, "Bats host major mammalian paramyxoviruses," *Nature communications*, vol. 3, no. 1, pp. 1–13, 2012.

[90] D. Hayman, R. Bowen, P. Cryan, G. McCracken, T. O'shea, A. Peel, A. Gilbert, C. Webb, and J. Wood, "Ecology of zoonotic infectious diseases in bats: current knowledge and future directions," *Zoonoses and Public Health*, vol. 60, no. 1, pp. 2–21, 2013.

[91] M. Escalera-Zamudio, E. Rojas-Anaya, S.-O. Kolokotronis, B. Taboada, E. Loza-Rubio, M. L. Méndez-Ojeda, C. F. Arias, N. Osterrieder, and A. D. Greenwood, "Bats, primates, and the evolutionary origins and diversification of mammalian gammaherpesviruses," *MBio*, vol. 7, no. 6, 2016.

[92] V. A. Belyi, A. J. Levine, and A. M. Skalka, "Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes," *PLoS Pathog*, vol. 6, no. 7, p. e1001030, 2010.

[93] M. Horie, T. Honda, Y. Suzuki, Y. Kobayashi, T. Daito, T. Oshida, K. Ikuta, P. Jern, T. Gojobori, J. M. Coffin, *et al.*, "Endogenous non-retroviral rna virus elements in mammalian genomes," *Nature*, vol. 463, no. 7277, pp. 84–87, 2010.

[94] M. L. Z. Mendoza, Z. Xiong, M. Escalera-Zamudio, A. K. Runge, J. Thézé, D. Streicker, H. K. Frank, E. Loza-Rubio, S. Liu, O. A. Ryder, *et al.*, "Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat," *Nature ecology & evolution*, vol. 2, no. 4, pp. 659–668, 2018.

[95] M. Horie, Y. Kobayashi, T. Honda, K. Fujino, T. Akasaka, C. Kohl, G. Wibbelt, K. Mühldorfer, A. Kurth, M. A. Müller, *et al.*, "An rna-dependent rna polymerase gene in bat genomes derived from an ancient negative-strand rna virus," *Scientific reports*, vol. 6, p. 25873, 2016.

[96] H. Ludwig and L. Bode, "Borna disease virus: new aspects on infection, disease, diagnosis and epidemiology.," *Revue scientifique et technique (International Office of Epizootics)*, vol. 19, no. 1, pp. 259–288, 2000.

[97] P. M. Kinnunen, A. Palva, A. Vaheri, and O. Vapalahti, "Epidemiology and host spectrum of borna disease virus infections," *Journal of General Virology*, vol. 94, no. 2, pp. 247–262, 2013.

[98] J. Cui and L.-F. Wang, "Genomic mining reveals deep evolutionary relationships between bornaviruses and bats," *Viruses*, vol. 7, no. 11, pp. 5792–5800, 2015.

[99] K. Fujino, M. Horie, T. Honda, D. K. Merriman, and K. Tomonaga, "Inhibition of borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome," *Proceedings of the National Academy of Sciences*, vol. 111, no. 36, pp. 13175–13180, 2014.

[100] T. Geib, C. Sauder, S. Venturelli, C. Hässler, P. Staeheli, and M. Schwemmle, "Selective virus resistance conferred by expression of borna disease virus nucleocapsid components," *Journal of Virology*, vol. 77, no. 7, pp. 4283–4290, 2003.

[101] S. Herzog, K. Frese, and R. Rott, "Studies on the genetic control of resistance of black hooded rats to borna disease," *Journal of general virology*, vol. 72, no. 3, pp. 535–540, 1991.

[102] N. F. Parrish, K. Fujino, Y. Shiromoto, Y. W. Iwasaki, H. Ha, J. Xing, A. Makino, S. Kuramochi-Miyagawa, T. Nakano, H. Siomi, *et al.*, "pirnas derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals," *Rna*, vol. 21, no. 10, pp. 1691–1703, 2015.

[103] J. Zong, X. Yao, J. Yin, D. Zhang, and H. Ma, "Evolution of the rna-dependent rna polymerase (rdrp) genes: duplications and possible losses before and after the divergence of major eukaryotic groups," *Gene*, vol. 447, no. 1, pp. 29–39, 2009.

[104] P. Maillard, C. Ciaudo, A. Marchais, Y. Li, F. Jay, S. Ding, and O. Voinnet, "Antiviral rna interference in mammalian cells," *Science*, vol. 342, no. 6155, pp. 235–238, 2013.

[105] T. Honda and K. Tomonaga, "Endogenous non-retroviral rna virus elements evidence a novel type of antiviral immunity," *Mobile genetic elements*, vol. 6, no. 3, pp. 1548–1554, 2016.

[106] T. F. Booth, D. R. Beniac, M. J. Rabb, and L. L. Lamboo, *FILOVIRUS STRUCTURE AND MORPHOGENESIS*, ch. CHAPTER 17, pp. 427–451.

[107] A. Negredo, G. Palacios, S. Vázquez-Morón, F. González, H. Dopazo, F. Molero, J. Juste, J. Quetglas, N. Savji, M. de la Cruz Martínez, *et al.*, "Discovery of an ebolavirus-like filovirus in europe," *PLoS Pathog*, vol. 7, no. 10, p. e1002304, 2011.

[108] J. H. Kuhn.

[109] X.-L. Yang, C. W. Tan, D. E. Anderson, R.-D. Jiang, B. Li, W. Zhang, Y. Zhu, X. F. Lim, P. Zhou, X.-L. Liu, *et al.*, "Characterization of a filovirus (měnglà virus) from rousettus bats in china," *Nature microbiology*, vol. 4, no. 3, pp. 390–395, 2019.

[110] E. M. Leroy, B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Délicat, J. T. Paweska, J.-P. Gonzalez, and R. Swanepoel, "Fruit bats as reservoirs of ebola virus," *Nature*, vol. 438, no. 7068, pp. 575–576, 2005.

[111] X. Pourrut, M. Souris, J. S. Towner, P. E. Rollin, S. T. Nichol, J.-P. Gonzalez, and E. Leroy, "Large serological survey showing cocirculation of ebola and marburg viruses in gabonese bat populations, and a high seroprevalence of both viruses in rousettus aegyptiacus," *BMC infectious diseases*, vol. 9, no. 1, p. 159, 2009.

[112] A. C. Banyard, J. S. Evans, T. R. Luo, and A. R. Fooks, "Lyssaviruses and bats: emergence and zoonotic threat," *Viruses*, vol. 6, no. 8, pp. 2974–2990, 2014.

[113] T. Kondoh, R. Manzoor, N. Nao, J. Maruyama, W. Furuyama, H. Miyamoto, A. Shigeno, M. Kuroda, K. Matsuno, D. Fujikura, *et al.*, "Putative endogenous filovirus vp35-like protein potentially functions as an ifn antagonist but not a polymerase cofactor," *PloS one*, vol. 12, no. 10, p. e0186450, 2017.

[114] S. F. Cotmore and P. Tattersall, "Parvoviruses: small does not mean simple," *Annual review of virology*, vol. 1, pp. 517–537, 2014.

[115] G. Arriagada and R. J. Gifford, "Parvovirus-derived endogenous viral elements in two south american rodent genomes," *Journal of virology*, vol. 88, no. 20, pp. 12158–12162, 2014.

[116] A. Kapoor, P. Simmonds, and W. I. Lipkin, "Discovery and characterization of mammalian endogenous parvoviruses," *Journal of virology*, vol. 84, no. 24, pp. 12628–12635, 2010.

[117] H. Liu, Y. Fu, J. Xie, J. Cheng, S. A. Ghabrial, G. Li, Y. Peng, X. Yi, and D. Jiang, "Widespread endogenization of densoviruses and parvoviruses in animal and human genomes," *Journal of virology*, vol. 85, no. 19, pp. 9863–9876, 2011.

[118] J. J. Pénzes, S. Marsile-Medun, M. Agbandje-McKenna, and R. J. Gifford, "Endogenous amdoparvovirus-related elements reveal insights into the biology and evolution of vertebrate parvoviruses," *Virus evolution*, vol. 4, no. 2, p. vey026, 2018.

[119] S. K. Lau, S. S. Ahmed, H.-W. Tsoi, H. C. Yeung, K. S. Li, R. Y. Fan, P. S. Zhao, C. C. Lau, C. S. Lam, K. K. Choi, *et al.*, "Bats host diverse parvoviruses as possible origin of mammalian dependoparvoviruses and source for bat–swine interspecies transmission," *Journal of General Virology*, vol. 98, no. 12, pp. 3046–3059, 2017.

[120] C. A. Bill and J. Summers, "Genomic dna double-strand breaks are targets for hepadnaviral dna integration," *Proceedings of the National Academy of Sciences*, vol. 101, no. 30, pp. 11135–11140, 2004.

[121] L. A. Shackelton, C. R. Parrish, U. Truyen, and E. C. Holmes, "High rate of viral evolution associated with the emergence of carnivore parvovirus," *Proceedings of the National Academy of Sciences*, vol. 102, no. 2, pp. 379–384, 2005.

[122] A. Y. Chen and J. Qiu, "Parvovirus infection-induced cell death and cell cycle arrest," *Future virology*, vol. 5, no. 6, pp. 731–743, 2010.

[123] C. Feschotte and E. J. Pritham, "Dna transposons and the evolution of eukaryotic genomes," *Annu. Rev. Genet.*, vol. 41, pp. 331–368, 2007.

[124] E. Rubin, G. Lithwick, and A. A. Levy, "Structure and evolution of the hat transposon superfamily," *Genetics*, vol. 158, no. 3, pp. 949–957, 2001.

[125] I. Grabundzija, S. A. Messing, J. Thomas, R. L. Cosby, I. Bilic, C. Miskey, A. Gogol-Döring, V. Kapitonov, T. Diem, and A. Dalda, "A helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes," *Nature communications*, vol. 7, 2016.

[126] D. A. Ray, H. J. Pagan, M. L. Thompson, and R. D. Stevens, "Bats with hats: evidence for recent dna transposon activity in genus myotis," *Molecular biology and evolution*, vol. 24, no. 3, pp. 632–639, 2007.

[127] R. N. Platt, M. W. Vandewege, C. Kern, C. J. Schmidt, F. G. Hoffmann, and D. A. Ray, "Large numbers of novel mirnas originate from dna transposons and are coincident with a large species radiation in bats," *Molecular biology and evolution*, vol. 31, no. 6, pp. 1536–1545, 2014.

[128] D. A. Ray, C. Feschotte, H. J. Pagan, J. D. Smith, E. J. Pritham, P. Arensburger, P. W. Atkinson, and N. L. Craig, "Multiple waves of recent dna transposon activity in the bat, myotis lucifugus," *Genome research*, 2008.

[129] X. Wu and S. Burgess, "Integration target site selection for retroviruses and transposable elements," *Cellular and Molecular Life Sciences CMLS*, vol. 61, no. 19, pp. 2588–2596, 2004.

[130] A. Kapusta, A. Suh, and C. Feschotte, "Dynamics of genome size evolution in birds and mammals," *Proceedings of the National Academy of Sciences*, vol. 114, pp. E1460 – E1469, 2017.

[131] J. Thomas, C. D. Phillips, R. J. Baker, and E. J. Pritham, "Rolling-circle transposons catalyze genomic innovation in a mammalian lineage," *Genome biology and evolution*, vol. 6, no. 10, pp. 2595–2610, 2014.

[132] A. Hua-Van, A. Le Rouzic, C. Maisonhaute, and P. Capy, "Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences," *Cytogenetic and genome research*, vol. 110, no. 1-4, pp. 426–440, 2005.

[133] L. Yang and J. L. Bennetzen, "Structure-based discovery and description of plant and animal helitrons," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 12832–12837, 2009.

[134] K. Mühldorfer, S. Speck, A. Kurth, R. Lesnik, C. Freuling, T. Müller, S. Kramer-Schadt, and G. Wibbelt, "Diseases and causes of death in european bats: Dynamics in disease susceptibility and infection rates," *PLOS ONE*, vol. 6, pp. 1–12, 12 2011.

[135] E. Picard-Meyer, A. Servat, M. Wasniewski, M. Gaillard, C. Borel, and F. Cliquet, "Bat rabies surveillance in france: first report of unusual mortality among serotine bats," *BMC Veterinary Research*, vol. 13, 2017.

[136] J. M. Yob, H. Field, A. M. Rashdi, C. Morrissy, B. M. van der Heide, P. Rota, A. bin Adzhar, J. White, P. Daniels, A. Jamaluddin, and T. Ksiazek, "Nipah virus infection in bats (order chiroptera) in peninsular malaysia.," *Emerging Infectious Diseases*, vol. 7, pp. 439 – 441, 2001.

[137] M. Baker, T. Schountz, and L.-F. Wang, "Antiviral immune responses of bats: A review," *Zoonoses and Public Health*, vol. 60, pp. 104 – 116, 2013.

[138] A. Banerjee, M. Baker, K. A. Kulcsar, V. Misra, R. Plowright, and K. Mossman, "Novel insights into immune systems of bats," *Frontiers in Immunology*, vol. 11, 2020.

[139] C. Brook and A. Dobson, "Bats as 'special' reservoirs for emerging zoonotic pathogens," *Trends in Microbiology*, vol. 23, pp. 172 – 180, 2015.

[140] T. Schountz, M. Baker, J. Butler, and V. Munster, "Immunological control of viral infections in bats and the emergence of viruses highly pathogenic to humans," *Frontiers in Immunology*, vol. 8, 2017.

[141] T. Omatsu, E.-J. Bak, Y. Ishii, S. Kyuwa, Y. Tohya, H. Akashi, and Y. Yoshikawa, "Induction and sequencing of rousette bat interferon alpha and beta genes," *Veterinary immunology and immunopathology*, vol. 124, p. 169—176, July 2008.

[142] D. Becker, C. Nachtmann, H. Argibay, G. Botto, M. Escalera-Zamudio, J. Carrera, C. Tello, E. Winiarski, A. Greenwood, M. L. Méndez-Ojeda, E. Loza-Rubio, A. Lavergne, B. de Thoisy, G. Czirják, R. Plowright, S. Altizer, and D. Streicker, "Leukocyte profiles reflect geographic range limits in a widespread neotropical bat," *Integrative and Comparative Biology*, vol. 59, pp. 1176 – 1189, 2019.

[143] S. K. Sarkar and A. K. Chakravarty, "Analysis of immunocompetent cells in the bat, pteropus giganteus: isolation and scanning electron microscopic characterization.," *Developmental and comparative immunology*, vol. 15 4, pp. 423–30, 1991.

[144] C. T. Ng, B. Sullivan, J. Teijaro, A. M. Lee, M. Welch, S. Rice, K. Sheehan, R. Schreiber, and M. Oldstone, "Blockade of interferon beta, but not interferon

alpha, signaling controls persistent viral infection.," *Cell host & microbe*, vol. 17 5, pp. 653–61, 2015.

[145] G. Fabozzi, C. S. Nabel, M. Dolan, and N. Sullivan, "Ebolavirus proteins suppress the effects of small interfering rna by direct interaction with the mammalian rna interference pathway," *Journal of Virology*, vol. 85, pp. 2512 – 2523, 2011.

[146] D. Leung, K. C. Prins, D. Borek, M. Farahbakhsh, J. Tufariello, P. Ramanan, J. Nix, L. A. Helgeson, Z. Otwinowski, R. Honzatko, C. Basler, and G. Amarasinghe, "Structural basis for dsrna recognition and interferon antagonism by ebola vp35," *Nature Structural & Molecular Biology*, vol. 17, pp. 165–172, 2010.

[147] W. Xu, M. Edwards, D. Borek, A. R. Feagins, A. Mittal, J. B. Alinger, K. N. Berry, B. Yen, J. Hamilton, T. J. Brett, R. Pappu, D. Leung, C. Basler, and G. Amarasinghe, "Ebola virus vp24 targets a unique nls binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated stat1.," *Cell host & microbe*, vol. 16 2, pp. 187–200, 2014.

[148] S. Subudhi, N. Rapin, and V. Misra, "Immune system modulation and viral persistence in bats: Understanding viral spillover," *Viruses*, vol. 11, 2019.

[149] S. A. Bratsch, N. Wertz, K. Chaloner, T. Kunz, and J. Butler, "The little brown bat, m. lucifugus, displays a highly diverse v h, d h and j h repertoire but little evidence of somatic hypermutation.," *Developmental and comparative immunology*, vol. 35 4, pp. 421–30, 2011.

[150] K. Boehme and T. Compton, "Innate sensing of viruses by toll-like receptors," *Journal of Virology*, vol. 78, pp. 7867 – 7873, 2004.

[151] K. Iha, T. Omatsu, S. Watanabe, N. Ueda, S. Taniguchi, H. Fujii, Y. Ishii, S. Kyuwa, H. Akashi, and Y. Yoshikawa, "Molecular cloning and expression analysis of bat toll-like receptors 3, 7 and 9.," *The Journal of veterinary medical science*, vol. 72 2, pp. 217–20, 2010.

[152] C. Cowled, M. Baker, M. Tachedjian, P. Zhou, D. Bulach, and L. Wang, "Molecular characterisation of toll-like receptors in the black flying fox pteropus alecto," *Developmental and Comparative Immunology*, vol. 35, pp. 7 – 18, 2011.

[153] M. Escalera-Zamudio, M. L. Zepeda-Mendoza, E. Loza-Rubio, E. Rojas-Anaya, M. L. Méndez-Ojeda, C. Arias, and A. Greenwood, "The evolution of bat nucleic acid-sensing toll-like receptors.," *Molecular ecology*, vol. 24 23, pp. 5899–909, 2015.

[154] M. Ahn, J. Cui, A. T. Irving, and L. Wang, "Unique loss of the pyhin gene family in bats amongst mammals: Implications for inflammasome sensing," *Scientific Reports*, vol. 6, 2016.

[155] P. Zhou, M. Tachedjian, J. Wynne, V. A. Boyd, J. Cui, I. Smith, C. Cowled, J. Ng, L. Mok, W. Michalski, I. Mendenhall, G. Tachedjian, L. Wang, and M. Baker, "Contraction of the type i ifn locus and unusual constitutive expression of ifn-? in bats," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 2696 – 2701, 2016.

[156] J. Xie, Y. Li, X. Shen, G. Goh, Y. Zhu, J. Cui, L. Wang, Z. Shi, and P. Zhou, "Dampened sting-dependent interferon activation in bats," *Cell Host & Microbe*, vol. 23, pp. 297 – 301.e4, 2018.

[157] I. Kuzmin, T. M. Schwarz, P. A. Ilinykh, I. Jordan, T. Ksiazek, R. Sachidanandam,

C. Basler, and A. Bukreyev, "Innate immune responses of bat and human cells to filoviruses: Commonalities and distinctions," *Journal of Virology*, vol. 91, 2017.

[158] V. J. Philbin, M. Iqbal, Y. Boyd, M. J. Goodchild, R. K. Beal, N. Bumstead, J. Young, and A. L. Smith, "Identification and characterization of a functional, alternatively spliced toll-like receptor 7 (tlr7) and genomic disruption of tlr8 in chickens," *Immunology*, vol. 114, no. 4, pp. 507–521, 2005.

[159] F. N. Lauw, D. R. Caffrey, and D. Golenbock, "Of mice and man: Tlr11 (finally) finds profilin.," *Trends in immunology*, vol. 26 10, pp. 509–11, 2005.

[160] Bat1K, "Bat1k consortium." `https://bat1k.ucd.ie/Bat1K`, 2019.

[161] M. Eisenstein, "Bat research takes wing," *Lab Animal*, vol. 47, pp. 97 – 100, 2018.

[162] E. Chuong, "The placenta goes viral: Retroviruses control gene expression in pregnancy," *PLoS Biology*, vol. 16, 2018.

[163] Y. Ophinni, U. Palatini, Y. Hayashi, and N. F. Parrish, "pirna-guided crispr-like immunity in eukaryotes.," *Trends in immunology*, 2019.

[164] Z. J. Whitfield, P. T. Dolan, M. Kunitomi, M. Tassetto, and R. Andino, "The diversity, structure, and function of heritable adaptive immunity sequences in the aedes aegypti genome," *Current Biology*, vol. 27, pp. 3511–3519.e7, 2017.

[165] M. Tassetto, M. Kunitomi, Z. J. Whitfield, P. T. Dolan, I. Sánchez-Vargas, M. Garcia-Knight, I. Ribiero, T. Chen, K. Olson, and R. Andino, "Control of rna viruses in mosquito cells through the acquisition of vdna and endogenous viral elements," *eLife*, vol. 8, 2019.

[166] J. N. Mandl, C. Schneider, D. S. Schneider, and M. Baker, "Going to bat(s) for studies of disease tolerance," *Frontiers in Immunology*, vol. 9, 2018.

[167] E. J. Pritham and C. Feschotte, "Massive amplification of rolling-circle transposons in the lineage of the bat myotis lucifugus," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 1895 – 1900, 2007.

[168] J. A. Hayward, M. Tachedjian, C. Kohl, A. Johnson, M. Dearnley, B. Jesaveluk, C. Langer, P. D. Solymosi, G. Hille, A. Nitsche, *et al.*, "Infectious korv-related retroviruses circulating in australian bats," *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, pp. 9529–9536, 2020.

[169] B. M. Hause, E. A. Nelson, and J. Christopher-Hennings, "North american big brown bats (eptesicus fuscus) harbor an exogenous deltaretrovirus," *Msphere*, vol. 5, no. 5, 2020.

[170] E. B. Chuong, N. C. Elde, and C. Feschotte, "Regulatory evolution of innate immunity through co-option of endogenous retroviruses," *Science*, vol. 351, no. 6277, pp. 1083–1087, 2016.

[171] A. Aswad and A. Katzourakis, "Paleovirology and virally derived immunity," *Trends in Ecology & Evolution*, vol. 27, no. 11, pp. 627–636, 2012.

[172] W. Cárdenas, Y. Loo, M. Gale, A. Hartman, C. R. Kimberlin, L. Martínez-Sobrido, E. O. Saphire, and C. Basler, "Ebola virus vp35 protein binds double-stranded rna and inhibits alpha/beta interferon production induced by rig-i signaling," *Journal of Virology*, vol. 80, pp. 5168 – 5178, 2006.

[173] K. Ozato, D. mi Shin, T.-H. Chang, and H. Morse, "Trim family proteins and their emerging roles in innate immunity," *Nature Reviews Immunology*, vol. 8, pp. 849–860, 2008.

[174] J. A. Hayward, M. Tachedjian, J. Cui, A. Z. Cheng, A. Johnson, M. Baker, R. Harris, L. Wang, and G. Tachedjian, "Differential evolution of antiretroviral restriction factors in pteropid bats as revealed by apobec3 gene complexity," *Molecular Biology and Evolution*, vol. 35, pp. 1626 – 1637, 2018.

[175] G. Kassiotis and J. Stoye, "Immune responses to endogenous retroelements: taking the bad with the good," *Nature Reviews Immunology*, vol. 16, pp. 207–219, 2016.

[176] G. Young, M. J. Ploquin, U. Eksmond, M. Wadwa, J. Stoye, and G. Kassiotis, "Negative selection by an endogenous retrovirus promotes a higher-avidity cd4+ t cell response to retroviral infection," *PLoS Pathogens*, vol. 8, 2012.

[177] C. Brook, M. Boots, K. Chandran, A. Dobson, C. Drosten, A. Graham, B. Grenfell, M. Müller, M. Ng, L.-F. Wang, and A. van Leeuwen, "Accelerated viral dynamics in bat cell lines, with implications for zoonotic emergence," *eLife*, vol. 9, 2020.

[178] D. Hayman, "Bat tolerance to viral infections," *Nature Microbiology*, vol. 4, pp. 728 – 729, 2019.

[179] Z. Huang, C. V. Whelan, N. M. Foley, D. Jebb, F. Touzalin, E. Petit, S. Puechmaille, and E. Teeling, "Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats," *Nature Ecology and Evolution*, vol. 3, pp. 1110–1120, 2019.

[180] A. T. Chinwalla, L. L. Cook, K. D. Delehaunty, G. A. Fewell, L. A. Fulton, R. S. Fulton, T. A. Graves, L. W. Hillier, E. R. Mardis, and J. D. McPherson, "Initial

sequencing and comparative analysis of the mouse genome," *Nature*, vol. 420, no. 6915, pp. 520–562, 2002.

[181] W. Lucas, "Viral capsids and envelopes: Structure and function," *eLS*, 2010.

[182] A. Alcami and U. H. Koszinowski, "Viral mechanisms of immune evasion," *Trends Microbiol*, vol. 8, no. 9, pp. 410–8, 2000.

[183] B. B. Finlay and G. McFadden, "Anti-immunology: evasion of the host immune system by bacterial and viral pathogens," *Cell*, vol. 124, no. 4, pp. 767–82, 2006.

[184] H. L. Ploegh, "Viral strategies of immune evasion," *Science*, vol. 280, no. 5361, pp. 248–253, 1998.

[185] M. M. Januszeski, P. M. Cannon, D. Chen, Y. Rozenberg, and W. F. Anderson, "Functional analysis of the cytoplasmic tail of moloney murine leukemia virus envelope protein," *J Virol*, vol. 71, no. 5, pp. 3613–9, 1997.

[186] S. Blaise, M. Mangeney, and T. Heidmann, "The envelope of mason–pfizer monkey virus has immunosuppressive properties," *Journal of General Virology*, vol. 82, no. 7, pp. 1597–1600, 2001.

[187] M. Mangeney and T. Heidmann, "Tumor cells expressing a retroviral envelope escape immune rejection in vivo," *Proc Natl Acad Sci U S A*, vol. 95, no. 25, pp. 14920–5, 1998.

[188] M. Mangeney, N. de Parseval, G. Thomas, and T. Heidmann, "The full-length envelope of an herv-h human endogenous retrovirus has immunosuppressive properties," *Journal of General Virology*, vol. 82, no. 10, pp. 2515–2518, 2001.

[189] G. J. Cianciolo, T. D. Copeland, S. Oroszlan, and R. Snyderman, "Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins," *Science*, vol. 230, no. 4724, pp. 453–5, 1985.

[190] V. A. Morozov, A. V. Morozov, M. Semaan, and J. Denner, "Single mutations in the transmembrane envelope protein abrogate the immunosuppressive property of hiv-1," *Retrovirology*, vol. 9, p. 67, 2012.

[191] G. Schlecht-Louf, M. Mangeney, H. El-Garch, V. Lacombe, H. Poulet, and T. Heidmann, "A targeted mutation within the feline leukemia virus (felv) envelope protein immunosuppressive domain to improve a canarypox virus-vectored felv vaccine," *Journal of virology*, vol. 88, no. 2, pp. 992–1001, 2014.

[192] G. Schlecht-Louf, M. Renard, M. Mangeney, C. Letzelter, A. Richaud, B. Ducos, I. Bouallaga, and T. Heidmann, "Retroviral infection in vivo requires an immune escape virulence factor encrypted in the envelope protein of oncoretroviruses," *Proc Natl Acad Sci U S A*, vol. 107, no. 8, pp. 3782–7, 2010.

[193] P. Sonigo, C. Barker, E. Hunter, and S. Wain-Hobson, "Nucleotide sequence of mason-pfizer monkey virus: an immunosuppressive d-type retrovirus," *Cell*, vol. 45, no. 3, pp. 375–385, 1986.

[194] G. Cornelis, C. Vernochet, S. Malicorne, S. Souquere, A. C. Tzika, S. M. Goodman, F. Catzeflis, T. J. Robinson, M. C. Milinkovitch, and G. Pierron, "Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive afrotherian tenrecs," *Proceedings of the National Academy of Sciences*, vol. 111, no. 41, pp. E4332–E4341, 2014.

[195] G. Cornelis, M. Funk, C. Vernochet, F. Leal, O. A. Tarazona, G. Meurice, O. Heidmann, A. Dupressoir, A. Miralles, M. P. Ramírez-Pinilla, and T. Heidmann, "An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental mabuya lizard," *Proceedings of the National Academy of Sciences*, vol. 114, pp. E10991 – E11000, 2017.

[196] J. E. Henzy, R. J. Gifford, C. P. Kenaley, and W. E. Johnson, "An intact retroviral gene conserved in spiny-rayed fishes for over 100 my," *Molecular Biology and Evolution*, p. msw262, 2016.

[197] D. Haig, "Retroviruses and the placenta," *Current Biology*, vol. 22, pp. R609–R613, 2012.

[198] H. S. Malik, "Retroviruses push the envelope for mammalian placentation," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 2184 – 2185, 2012.

[199] T. Omatsu, S. Watanabe, H. Akashi, and Y. Yoshikawa, "Biological characters of bats in relation to natural reservoir of emerging viruses," *Comparative immunology, microbiology and infectious diseases*, vol. 30, no. 5, pp. 357–374, 2007.

[200] I. Smith and L.-F. Wang, "Bats and their virome: an important source of emerging viruses capable of infecting humans," *Current opinion in virology*, vol. 3, no. 1, pp. 84–91, 2013.

[201] R. K. Plowright, P. Eby, P. J. Hudson, I. L. Smith, D. Westcott, W. L. Bryden, D. Middleton, P. A. Reid, R. A. McFarlane, G. Martin, G. M. Tabor, L. F. Skerratt, D. L. Anderson, G. Crameri, D. Quammen, D. Jordan, P. Freeman, L. F. Wang, J. H. Epstein, G. A. Marsh, N. Y. Kung, and H. McCallum, "Ecological dynamics

of emerging bat virus spillover," *Proc Biol Sci*, vol. 282, no. 1798, p. 20142124, 2015.

[202] G. M. Baer and G. L. Bales, "Experimental rabies infection in the mexican freetail bat," *J Infect Dis*, vol. 117, no. 1, pp. 82–90, 1967.

[203] K. B. Chua, C. L. Koh, P. S. Hooi, K. F. Wee, J. H. Khong, B. H. Chua, Y. P. Chan, M. E. Lim, and S. K. Lam, "Isolation of nipah virus from malaysian island flying-foxes," *Microbes Infect*, vol. 4, no. 2, pp. 145–51, 2002.

[204] D. Chu, L. Poon, Y. Guan, and J. Peiris, "Novel astroviruses in insectivorous bats," *Journal of Virology*, vol. 82, pp. 9107 – 9114, 2008.

[205] H. Field, P. Young, J. M. Yob, J. Mills, L. Hall, and J. Mackenzie, "The natural history of hendra and nipah viruses," *Microbes Infect*, vol. 3, no. 4, pp. 307–14, 2001.

[206] J. N. Hanna, I. K. Carney, G. A. Smith, A. E. Tannenberg, J. E. Deverill, J. A. Botha, I. L. Serafin, B. J. Harrower, P. F. Fitzpatrick, and J. W. Searle, "Australian bat lyssavirus infection: a second human case, with a long incubation period," *Med J Aust*, vol. 172, no. 12, pp. 597–9, 2000.

[207] R. Johnson, "Emerging viral infections of the nervous system," *Journal of NeuroVirology*, vol. 9, pp. 140–147, 2011.

[208] J. S. Mackenzie, H. E. Field, and K. J. Guyatt, "Managing emerging diseases borne by fruit bats (flying foxes), with particular reference to henipaviruses and australian bat lyssavirus," *J Appl Microbiol*, vol. 94 Suppl, pp. 59S–69S, 2003.

[209] D. J. Middleton, C. Morrissy, B. M. van der Heide, G. M. Russell, M. Braun, H. Westbury, K. Halpin, and P. Daniels, "Experimental nipah virus infection in pteropid bats (pteropus poliocephalus).," *Journal of comparative pathology*, vol. 136 4, pp. 266–72, 2007.

[210] S. E. Sulkin and R. Allen, "Virus infections in bats," *Monogr Virol*, vol. 8, no. 0, pp. 1–103, 1974.

[211] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–410, 1990.

[212] A. R. Quinlan, "Bedtools: the swiss-army tool for genome feature analysis," *Current protocols in bioinformatics*, pp. 11.12. 1–11.12. 34, 2014.

[213] A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. Lanczycki, F. Lu, G. Marchler, J. Song, N. Thanki, Z. Wang, R. Yamashita, D. Zhang, C. Zheng, and S. Bryant, "Cdd: Ncbi's conserved domain database," *Nucleic Acids Research*, vol. 43, pp. D222 – D226, 2015.

[214] F. Abascal, R. Zardoya, and D. Posada, "Prottest: selection of best-fit models of protein evolution," *Bioinformatics*, vol. 21 9, pp. 2104–5, 2005.

[215] J. Huelsenbeck and F. Ronquist, "Mrbayes: Bayesian inference of phylogenetic trees," *Bioinformatics*, vol. 17 8, pp. 754–5, 2001.

[216] B. Shapiro, A. Rambaut, and A. Drummond, "Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences.," *Molecular biology and evolution*, vol. 23 1, pp. 7–9, 2006.

[217] Z. Yang, "Paml 4: phylogenetic analysis by maximum likelihood," *Molecular biology and evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.

[218] A. Sinha and W. Johnson, "Retroviruses of the rdr superinfection interference group: ancient origins and broad host distribution of a promiscuous env gene.," *Current opinion in virology*, vol. 25, pp. 105–112, 2017.

[219] M. Anisimova, J. Bielawski, and Z. Yang, "Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution.," *Molecular biology and evolution*, vol. 18 8, pp. 1585–92, 2001.

[220] M. Anisimova, M. Gil, J. Dufayard, C. Dessimoz, and O. Gascuel, "Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes," *Systematic Biology*, vol. 60, pp. 685 – 699, 2011.

[221] O. Heidmann, A. Beguin, J. Paternina, R. Berthier, M. Deloger, O. Bawa, and T. Heidmann, "Hemo, an ancestral endogenous retroviral envelope protein shed in the blood of pregnant women and expressed in pluripotent stem cells and tumors," *Proceedings of the National Academy of Sciences*, vol. 114, pp. E6642 – E6651, 2017.

[222] F. Redelsperger, G. Cornelis, C. Vernochet, B. C. Tennant, F. Catzeflis, B. Mulot, O. Heidmann, T. Heidmann, and A. Dupressoir, "Capture of syncytin-mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the rodentia squirrel-related clade," *Journal of virology*, vol. 88, no. 14, pp. 7915–7928, 2014.

[223] Y. Y. B. Rivera, C. Brütting, C. Schmidt, I. Volkmer, and M. S. Staege, "Endogenous retrovirus 3 ? history, physiology, and pathology," *Frontiers in Microbiology*, vol. 8, 2018.

[224] R. Leiser and B. Koob, "Development and characteristics of placentation in a carnivore, the domestic cat.," *The Journal of experimental zoology*, vol. 266 6, pp. 642–56, 1993.

[225] J. Denner, "Expression and function of endogenous retroviruses in the placenta," *APMIS*, vol. 124, pp. 31 – 43, 2016.

[226] A. Dupressoir, G. Marceau, C. Vernochet, L. Benit, C. Kanellopoulos, V. Sapin, and T. Heidmann, "Syncytin-a and syncytin-b, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in muridae.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102 3, pp. 725–30, 2005.

[227] O. Heidmann, C. Vernochet, A. Dupressoir, and T. Heidmann, "Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new" syncytin" in a third order of mammals," *Retrovirology*, vol. 6, no. 1, p. 107, 2009.

[228] A. Gopalakrishna and K. B. Karim, "Fetal membranes and placentation in chiroptera," *J Reprod Fertil*, vol. 56, no. 1, pp. 417–29, 1979.

[229] F. Lisón, A. Haz, C. González-Revelles, and J. F. Calvo, "Sexual size dimorphism in greater mouse-eared bat myotis myotis (chiroptera: Vespertilionidae) from a mediterranean region," *Acta Zoologica*, vol. 95, no. 2, pp. 137–143, 2014.

[230] P. Myers, "Sexual dimorphism in size of vespertilionid bats," *The American Naturalist*, vol. 112, no. 986, pp. 701–711, 1978.

[231] M. T. O'Mara, K. Bauer, D. Blank, J. W. Baldwin, and D. K. N. Dechmann, "Common noctule bats are sexually dimorphic in migratory behaviour and body size but not wing shape," *PLoS ONE*, vol. 11, 2016.

[232] R. D. Stevens, M. E. Johnson, and E. S. McCulloch, "Absolute and relative secondary-sexual dimorphism in wing morphology: a multivariate test of the 'big mother'hypothesis," *Acta Chiropterologica*, vol. 15, no. 1, pp. 163–170, 2013.

[233] G. Cornelis, O. Heidmann, S. Bernard-Stoecklin, K. Reynaud, G. Véron, B. Mulot, A. Dupressoir, and T. Heidmann, "Ancestral capture of syncytin-car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in carnivora," *Proceedings of the National Academy of Sciences*, vol. 109, no. 7, pp. E432–E441, 2012.

[234] G. Cornelis, O. Heidmann, S. A. Degrelle, C. Vernochet, C. Lavialle, C. Letzelter, S. Bernard-Stoecklin, A. Hassanin, B. Mulot, and M. Guillomot, "Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants," *Proceedings of the National Academy of Sciences*, vol. 110, no. 9, pp. E828–E837, 2013.

[235] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of dna molecules up to several hundred kilobases," *Nature methods*, vol. 6, no. 5, pp. 343–345, 2009.

[236] R. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. Antonarakis, J. Attwood,

R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. Brent, D. Brown, S. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. Chinwalla, D. Church, M. Clamp, C. Clee, F. Collins, L. Cook, R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. Dermitzakis, C. N. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. Dunn, S. Eddy, L. Elnitski, R. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. Fewell, P. Flicek, K. Foley, W. Frankel, L. Fulton, R. Fulton, T. Furey, D. Gage, R. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. Graves, E. Green, S. Gregory, R. Guigó, M. Guyer, R. Hardison, D. Haussler, Y. Hayashizaki, D. W. LaHillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. Jaffe, L. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. Kent, A. Kirby, D. Kolbe, I. F. Korf, R. Kucherlapati, E. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. Maglott, E. Mardis, L. Matthews, E. Mauceli, J. Mayer, M. Mccarthy, W. McCombie, S. McLaren, K. McLay, J. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. Miner, E. Mongin, K. Montgomery, M. Morgan, R. Mott, J. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. O'connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. Pepin, J. Peterson, P. A. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. Roskin, E. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. Singer, G. Slater, A. Smit, D. Smith, B. Spencer, A. Stabenau,

N. Stange-thomann, C. Sugnet, M. Suyama, G. Tesler, J. R. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. Vinson, A. von Niederhausern, C. Wade, M. Wall, R. J. Weber, R. Weiss, M. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. Worley, D. W. Wyman, S. Yang, S. P. Yang, E. Zdobnov, M. Zody, and E. Lander, "Initial sequencing and comparative analysis of the mouse genome," *Nature*, vol. 420, pp. 520–562, 2002.

[237] M. Tristem, "Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database," *Journal of virology*, vol. 74, no. 8, pp. 3715–3730, 2000.

[238] M. Tristem, P. Kabat, L. Lieberman, S. Linde, A. Karpas, and F. Hill, "Characterization of a novel murine leukemia virus-related subgroup within mammals," *Journal of virology*, vol. 70, no. 11, pp. 8241–8246, 1996.

[239] R. P. Subramanian, J. H. Wildschutte, C. Russo, and J. M. Coffin, "Identification, characterization, and comparative genomic distribution of the herv-k (hml-2) group of human endogenous retroviruses," *Retrovirology*, vol. 8, no. 1, p. 90, 2011.

[240] S. K. Wootton, C. L. Halbert, and A. D. Miller, "Sheep retrovirus structural protein induces lung tumours," *Nature*, vol. 434, no. 7035, pp. 904–907, 2005.

[241] J. Zhao, K. Rycaj, S. Geng, M. Li, J. B. Plummer, B. Yin, H. Liu, X. Xu, Y. Zhang, and Y. Yan, "Expression of human endogenous retrovirus type k envelope protein is a novel candidate prognostic marker for human breast cancer," *Genes & cancer*, vol. 2, no. 9, pp. 914–922, 2011.

[242] G. Baillie, L. N. van de Lagemaat, C. Baust, and D. Mager, "Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals," *Journal of Virology*, vol. 78, pp. 5784 – 5798, 2004.

[243] K. Quade, S. Saule, D. Stéhelin, G. Kitchener, and M. Hayman, "Virus gene expression in rat cells transformed by avian myelocytomatosis virus strain mc29 and avian erythroblastosis virus.," *The Journal of general virology*, vol. 64 (Pt 1), pp. 83–94, 1983.

[244] J. Svoboda, P. Chyle, D. Simkovic, and I. Hilgert, "Demonstration of the absence of infectious rous virus in rat tumour xc, whose structurally intact cells produce rous sarcoma when transferred to chicks," *Folia biologica*, vol. 9, pp. 77–81, 1963.

[245] M. Bolisetty, J. Blomberg, F. Benachenhou, G. Sperber, and K. Beemon, "Unexpected diversity and expression of avian endogenous retroviruses," *mBio*, vol. 3, 2012.

[246] Z. Keckesova, L. M. J. Ylinen, G. Towers, R. Gifford, and A. Katzourakis, "Identification of a relik orthologue in the european hare (lepus europaeus) reveals a minimum age of 12 million years for the lagomorph lentiviruses.," *Virology*, vol. 384 1, pp. 7–11, 2009.

[247] J. Cui and E. Holmes, "Endogenous lentiviruses in the ferret genome," *Journal of Virology*, vol. 86, pp. 3383 – 3385, 2012.

[248] G.-Z. Han and M. Worobey, "Endogenous lentiviral elements in the weasel family (mustelidae).," *Molecular biology and evolution*, vol. 29 10, pp. 2905–8, 2012.

[249] R. Gifford, A. Katzourakis, M. Tristem, O. Pybus, M. Winters, and R. Shafer, "A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 20362 – 20367, 2008.

[250] C. Gilbert, D. G. Maxfield, S. M. Goodman, and C. Feschotte, "Parallel germline infiltration of a lentivirus in two malagasy lemurs," *PLoS Genetics*, vol. 5, 2009.

[251] A. Katzourakis, R. Gifford, M. Tristem, M. Gilbert, and O. Pybus, "Macroevolution of complex retroviruses," *Science*, vol. 325, pp. 1512 – 1512, 2009.

[252] G.-Z. Han and M. Worobey, "An endogenous foamy-like viral element in the coelacanth genome," *PLoS Pathogens*, vol. 8, 2012.

[253] "Chapter 23 - retroviruses"," in *Fenner and White's Medical Virology (Fifth Edition)* (C. J. Burrell, C. R. Howard, and F. A. Murphy, eds.), pp. "317 – 344",, London: Academic Press, fifth edition ed.

[254] J. Strauss and E. G. Strauss, "Chapter 6 – viruses whose life cycle uses reverse transcriptase," 2008.

[255] G. Wilkinson and D. M. Adams, "Recurrent evolution of extreme longevity in bats," *Biology Letters*, vol. 15, 2019.

[256] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *The lancet*, vol. 395, no. 10224, pp. 565–574, 2020.

[257] X. Xu, H. Zhao, Z. Gong, and G.-Z. Han, "Endogenous retroviruses of non-avian/mammalian vertebrates illuminate diversity and deep history of retroviruses," *PLoS Pathogens*, vol. 14, 2018.

[258] Z. Wu, X. Ren, L. Yang, Y. Hu, J. Yang, G. He, J. Zhang, J. Dong, L. Sun, J. Du, L. Liu, Y. Xue, J. Wang, F. Yang, S. Zhang, and Q. Jin, "Virome analysis for identification of novel mammalian viruses in bat species from chinese provinces," *Journal of Virology*, vol. 86, pp. 10999 – 11012, 2012.

[259] A. Katzourakis, G. Magiorkinis, A. G. Lim, S. Gupta, R. Belshaw, and R. Gifford, "Larger mammalian body size leads to lower retroviral activity," *PLoS Pathogens*, vol. 10, 2014.

[260] V. Gorbunova, A. Seluanov, and B. K. Kennedy, "The world goes bats: Living longer and tolerating viruses," *Cell Metabolism*, vol. 32, pp. 31 – 43, 2020.

[261] N. Mollentze and D. Streicker, "Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 9423 – 9430, 2020.

[262] E. C. Skirmuntt, M. Escalera-Zamudio, E. C. Teeling, A. Smith, and A. Katzourakis, "The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses," *Annual Review of Virology*, vol. 7, pp. 103–119, 2020.

[263] A. T. Irving, M. Ahn, G. Goh, D. E. Anderson, and L.-F. Wang, "Lessons from the host defences of bats, a unique viral reservoir," *Nature*, vol. 589, no. 7842, pp. 363–370, 2021.

[264] M. Ahn, D. E. Anderson, Q. Zhang, C. W. Tan, B. L. Lim, K. Luko, M. Wen, W. N. Chia, S. Mani, L. C. Wang, *et al.*, "Dampened nlrp3-mediated inflammation in bats and implications for a special viral reservoir host," *Nature Microbiology*, vol. 4, no. 5, pp. 789–799, 2019.

[265] C. A. Janeway, "Approaching the asymptote? evolution and revolution in immunology," in *Cold Spring Harbor symposia on quantitative biology*, vol. 54, pp. 1–13, Cold Spring Harbor Laboratory Press, 1989.

[266] S. S. Diebold, "Activation of dendritic cells by toll-like receptors and c-type lectins," *Dendritic Cells*, pp. 3–30, 2009.

[267] T. Kawai and S. Akira, "The role of pattern-recognition receptors in innate immunity: update on toll-like receptors," *Nature immunology*, vol. 11, no. 5, p. 373, 2010.

[268] S. Akira, K. Takeda, and T. Kaisho, "Toll-like receptors: critical proteins linking innate and acquired immunity," *Nature immunology*, vol. 2, no. 8, pp. 675–680, 2001.

[269] R. Suresh and D. M. Mosser, "Pattern recognition receptors in innate immunity, host defense, and immunopathology," *Advances in physiology education*, vol. 37, no. 4, pp. 284–291, 2013.

[270] H. Tanji, U. Ohto, T. Shibata, K. Miyake, and T. Shimizu, "Structural reorganization of the toll-like receptor 8 dimer induced by agonistic ligands," *Science*, vol. 339, no. 6126, pp. 1426–1429, 2013.

[271] N. Matsushima, T. Tanaka, P. Enkhbayar, T. Mikami, M. Taga, K. Yamada, and Y. Kuroki, "Comparative sequence analysis of leucine-rich repeats (lrrs) within vertebrate toll-like receptors," *BMC genomics*, vol. 8, no. 1, pp. 1–20, 2007.

[272] I. Botos, D. M. Segal, and D. R. Davies, "The structural biology of toll-like receptors," *Structure*, vol. 19, no. 4, pp. 447–459, 2011.

[273] S. Uematsu and S. Akira, "Toll-like receptors and type i interferons*," *Journal of Biological Chemistry*, vol. 282, pp. 15319 – 15323, 2007.

[274] T. Hurst and G. Magiorkinis, "Activation of the innate immune response by endogenous retroviruses.," *The Journal of general virology*, vol. 96 Pt 6, pp. 1207–18, 2015.

[275] S. Lester and K. Li, "Toll-like receptors in antiviral innate immunity," *Journal of Molecular Biology*, vol. 426, 2014.

[276] H. Velová, M. W. Gutowska-Ding, D. W. Burt, and M. Vinkler, "Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection," *Molecular biology and evolution*, vol. 35, no. 9, pp. 2170–2184, 2018.

[277] Y. Liu, J. Li, D. Liu, H. Song, C. Chen, M. Lv, X. Pei, and Z. Hu, "Clinical features and outcomes of 2019 novel coronavirus-infected patients with cardiac injury," *medRxiv*, 2020.

[278] J. Roach, G. Glusman, L. Rowen, A. Kaur, M. Purcell, K. Smith, L. Hood, and A. Aderem, "The evolution of vertebrate toll-like receptors.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102 27, pp. 9577–82, 2005.

[279] T. Eigenbrod and A. H. Dalpke, "Bacterial rna: an underestimated stimulus for innate immune responses," *The Journal of Immunology*, vol. 195, no. 2, pp. 411–418, 2015.

[280] E. A. Kurt-Jones, L. Popova, L. Kwinn, L. M. Haynes, L. P. Jones, R. A. Tripp, E. E. Walsh, M. W. Freeman, D. T. Golenbock, L. J. Anderson, *et al.*, "Pattern recognition receptors tlr4 and cd14 mediate response to respiratory syncytial virus," *Nature immunology*, vol. 1, no. 5, pp. 398–401, 2000.

[281] K. Bieback, E. Lien, I. M. Klagge, E. Avota, J. Schneider-Schaulies, W. P. Duprex, H. Wagner, C. Kirschning, V. ter meulen, and S. Schneider-Schaulies, "Hemagglutinin protein of wild-type measles virus activates toll-like receptor 2 signaling," *Journal of Virology*, vol. 76, pp. 8729 – 8736, 2002.

[282] J. Wang, J. Chai, and H. Wang, "Structure of the mouse toll-like receptor 13 ectodomain in complex with a conserved sequence from bacterial 23s ribosomal rna," *The FEBS journal*, vol. 283, no. 9, pp. 1631–1635, 2016.

[283] X.-D. Li and Z. J. Chen, "Sequence specific detection of bacterial 23s ribosomal rna by tlr13," *elife*, vol. 1, p. e00102, 2012.

[284] K. Tabeta, K. Hoebe, E. M. Janssen, X. Du, P. Georgel, K. Crozat, S. Mudd, N. Mann, S. Sovath, J. Goode, L. Shamel, A. A. Herskovits, D. A. Portnoy, M. Cooke, L. M. Tarantino, T. Wiltshire, B. Steinberg, S. Grinstein, and B. Beutler, "The unc93b1 mutation 3d disrupts exogenous antigen presentation and signaling via toll-like receptors 3, 7 and 9," *Nature Immunology*, vol. 7, pp. 156–164, 2006.

[285] Z. , Z. Cai, A. Sanchez, T. Zhang, S. Wen, J. Wang, J. Yang, S. Fu, and D. Zhang, "A novel toll-like receptor that recognizes vesicular stomatitis virus," *Journal of Biological Chemistry*, vol. 286, no. 6, pp. 4517–4524, 2011.

[286] M. Dewannieux, F. Harper, A. Richaud, C. Letzelter, D. Ribet, G. Pierron, and T. Heidmann, "Identification of an infectious progenitor for the multiple-copy herv-k human endogenous retroelements.," *Genome research*, vol. 16 12, pp. 1548–56, 2006.

[287] G. Magiorkinis, R. Gifford, A. Katzourakis, J. D. Ranter, and R. Belshaw, "Env-less endogenous retroviruses are genomic superspreaders," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 7385 – 7390, 2012.

[288] P. Yu, W. Lübben, H. Slomka, J. Gebler, M. Konert, C. Cai, L. Neubrandt, O. P. da Costa, S. Paul, S. Dehnert, K. Döhne, M. Thanisch, S. Storsberg, L. Wiegand, A. Kaufmann, M. Nain, L. Quintanilla-Martinez, S. Bettio, B. Schnierle, L. Kolesnikova, S. Becker, M. Schnare, and S. Bauer, "Nucleic acid-sensing toll-like receptors are essential for the control of endogenous retrovirus viremia and erv-induced tumors.," *Immunity*, vol. 37 5, pp. 867–79, 2012.

[289] E. Browne, "Toll-like receptor 7 controls the anti-retroviral germinal center response," *PLoS Pathogens*, vol. 7, 2011.

[290] E. Browne, "Toll-like receptor 7 inhibits early acute retroviral infection through rapid lymphocyte responses," *Journal of Virology*, vol. 87, pp. 7357 – 7366, 2013.

[291] S.-I. Saitoh, F. Abe, A. Kanno, N. Tanimura, Y. M. Saitoh, R. Fukui, T. Shibata, K. Sato, T. Ichinohe, M. Hayashi, *et al.*, "Tlr7 mediated viral recognition results in

focal type i interferon secretion by dendritic cells," *Nature communications*, vol. 8, no. 1, pp. 1–12, 2017.

[292] Y. Koh, J. C. Scatizzi, J. D. Gahan, B. Lawson, R. Baccala, K. Pollard, B. Beutler, A. Theofilopoulos, and D. Kono, "Role of nucleic acid–sensing tlrs in diverse autoantibody specificities and anti-nuclear antibody–producing b cells," *The Journal of Immunology*, vol. 190, pp. 4982 – 4990, 2013.

[293] D. Werling, O. C. Jann, V. Offord, E. J. Glass, and T. J. Coffey, "Variation matters: Tlr structure and species-specific pathogen recognition," *Trends in immunology*, vol. 30, no. 3, pp. 124–130, 2009.

[294] J. Schultz, F. Milpetz, P. Bork, and C. P. Ponting, "Smart, a simple modular architecture research tool: identification of signaling domains," *Proceedings of the National Academy of Sciences*, vol. 95, no. 11, pp. 5857–5864, 1998.

[295] I. Letunic, S. Khedkar, and P. Bork, "Smart: recent updates, new developments and status in 2020," *Nucleic Acids Research*, vol. 49, no. D1, pp. D458–D460, 2021.

[296] H. Ellegren, "Comparative genomics and the study of evolution by natural selection," *Molecular ecology*, vol. 17, no. 21, pp. 4586–4596, 2008.

[297] S. Weaver, S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond, "Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes," *Molecular biology and evolution*, vol. 35, no. 3, pp. 773–777, 2018.

[298] S. L. Kosakovsky Pond, A. F. Poon, R. Velazquez, S. Weaver, N. L. Hepler, B. Murrell, S. D. Shank, B. R. Magalis, D. Bouvier, A. Nekrutenko, *et al.*, "Hyphy

2.5—a customizable platform for evolutionary hypothesis testing using phylogenies," *Molecular biology and evolution*, vol. 37, no. 1, pp. 295–299, 2020.

[299] S. L. Kosakovsky Pond and S. D. Frost, "Not so different after all: a comparison of methods for detecting amino acid sites under selection," *Molecular biology and evolution*, vol. 22, no. 5, pp. 1208–1222, 2005.

[300] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S. L. Kosakovsky Pond, and K. Scheffler, "Fubar: a fast, unconstrained bayesian approximation for inferring selection," *Molecular biology and evolution*, vol. 30, no. 5, pp. 1196–1205, 2013.

[301] M. D. Smith, J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, and S. L. Kosakovsky Pond, "Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection," *Molecular biology and evolution*, vol. 32, no. 5, pp. 1342–1353, 2015.

[302] B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. K. Pond, "Detecting individual sites subject to episodic diversifying selection," *PLoS genet*, vol. 8, no. 7, p. e1002764, 2012.

[303] G. Wlasiuk and M. W. Nachman, "Adaptation and constraint at toll-like receptors in primates," *Molecular biology and evolution*, vol. 27, no. 9, pp. 2172–2186, 2010.

[304] H. Areal, J. Abrantes, and P. J. Esteves, "Signatures of positive selection in toll-like receptor (tlr) genes in mammals," *BMC evolutionary biology*, vol. 11, no. 1, pp. 1–14, 2011.

[305] T. Shen, S. Xu, X. Wang, W. Yu, K. Zhou, and G. Yang, "Adaptive evolution and functional constraint at tlr4 during the secondary aquatic adaptation and

diversification of cetaceans," *BMC evolutionary biology*, vol. 12, no. 1, pp. 1–12, 2012.

[306] A. Fornusková, M. Vinkler, M. Pagés, M. Galan, E. Jousselin, F. Cerqueira, S. Morand, N. Charbonnel, J. Bryja, and J. Cosson, "Contrasted evolutionary histories of two toll-like receptors (tlr4 and tlr7) in wild rodents (murinae)," *BMC Evolutionary Biology*, vol. 13, pp. 194 – 194, 2013.

[307] C. E. Grueber, G. P. Wallis, and I. G. Jamieson, "Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes," *PloS one*, vol. 9, no. 3, p. e89632, 2014.

[308] M. Jin, S. Kim, J. Heo, M. E. Lee, H. M. Kim, S. Paik, H. Lee, and J.-O. Lee, "Crystal structure of the tlr1-tlr2 heterodimer induced by binding of a tri-acylated lipopeptide," *Cell*, vol. 130, pp. 1071–1082, 2007.

[309] J. Kang, X.-P. Nan, M. Jin, S. Youn, Y. Ryu, S. Mah, S. Han, H. Lee, S. Paik, and J.-O. Lee, "Recognition of lipopeptide patterns by toll-like receptor 2-toll-like receptor 6 heterodimer.," *Immunity*, vol. 31 6, pp. 873–84, 2009.

[310] J. Garate and C. Oostenbrink, "Lipid a from lipopolysaccharide recognition: Structure, dynamics and cooperativity by molecular dynamics simulations," *Proteins: Structure*, vol. 81, 2013.

[311] T. Paramo, T. Piggot, C. Bryant, and P. Bond, "The structural basis for endotoxin-induced allosteric regulation of the toll-like receptor 4 (tlr4) innate immune receptor*," *The Journal of Biological Chemistry*, vol. 288, pp. 36215 – 36225, 2013.

[312] B. S. Park, D. H. Song, H. M. Kim, B.-S. Choi, H. Lee, and J.-O. Lee, "The structural basis of lipopolysaccharide recognition by the tlr4–md-2 complex," *Nature*, vol. 458, pp. 1191–1195, 2009.

[313] T. Scior, J. Lozano-Aponte, V. Figueroa-Vazquez, J. A. Yunes-Rojas, U. Zahringer, and C. Alexander, "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," *Computational and Structural Biotechnology Journal*, vol. 7, 2013.

[314] Z. Zhang, U. Ohto, T. Shibata, E. Krayukhina, M. Taoka, Y. Yamauchi, H. Tanji, T. Isobe, S. Uchiyama, K. Miyake, and T. Shimizu, "Structural analysis reveals that toll-like receptor 7 is a dual receptor for guanosine and single-stranded rna.," *Immunity*, vol. 45 4, pp. 737–748, 2016.

[315] L. B. Barreiro, M. Ben-Ali, H. Quach, G. Laval, E. Patin, J. K. Pickrell, C. Bouchier, M. Tichit, O. Neyrolles, B. Gicquel, *et al.*, "Evolutionary dynamics of human toll-like receptors and their different contributions to host defense," *PLoS Genet*, vol. 5, no. 7, p. e1000562, 2009.

[316] W. J. Swanson, Z. Yang, M. F. Wolfner, and C. F. Aquadro, "Positive darwinian selection drives the evolution of several female reproductive proteins in mammals," *Proceedings of the National Academy of Sciences*, vol. 98, no. 5, pp. 2509–2514, 2001.

[317] M. Vinkler, H. Bainová, and J. Bryja, "Protein evolution of toll-like receptors 4, 5 and 7 within galloanserae birds," *Genetics Selection Evolution*, vol. 46, no. 1, pp. 1–12, 2014.

[318] M. Vinkler, H. Bainová, A. Bryjová, O. Tomášek, T. Albrecht, and J. Bryja, "Characterisation of toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge," *Genetica*, vol. 143, no. 1, pp. 101–112, 2015.

[319] Z. Świderská, A. Šmídová, L. Buchtová, A. Bryjová, A. Fabiánová, P. Munclinger, and M. Vinkler, "Avian toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[320] S. Wong, S. Lau, P. Woo, and K.-Y. Yuen, "Bats as a continuing source of emerging infections in humans," *Reviews in medical virology*, vol. 17, no. 2, pp. 67–91, 2007.

[321] K. D. Reed, J. K. Meece, J. S. Henkel, and S. K. Shukla, "Birds, migration and emerging zoonoses: West nile virus, lyme disease, influenza a and enteropathogens," *Clinical medicine & research*, vol. 1, no. 1, pp. 5–12, 2003.

[322] A. Ozinsky, D. M. Underhill, J. D. Fontenot, A. M. Hajjar, K. D. Smith, C. B. Wilson, L. Schroeder, and A. Aderem, "The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors," *Proceedings of the National Academy of Sciences*, vol. 97, no. 25, pp. 13766–13771, 2000.

[323] L. H. Travassos, S. E. Girardin, D. J. Philpott, D. Blanot, M.-A. Nahori, C. Werts, and I. G. Boneca, "Toll-like receptor 2-dependent bacterial sensing does not occur via peptidoglycan recognition," *EMBO reports*, vol. 5, no. 10, pp. 1000–1006, 2004.

[324] J. I. Godfroy III, M. Roostan, Y. S. Moroz, I. V. Korendovych, and H. Yin, "Isolated toll-like receptor transmembrane domains are capable of oligomerization," *PloS one*, vol. 7, no. 11, p. e48875, 2012.

[325] L. de Oliviera Nascimento, P. Massari, and L. M. Wetzler, "The role of tlr2 in infection and immunity," *Frontiers in immunology*, vol. 3, p. 79, 2012.

[326] Y. Guan, D. R. E. Ranoa, S. Jiang, S. K. Mutha, X. Li, J. Baudry, and R. I. Tapping, "Human tlrs 10 and 1 share common mechanisms of innate immune sensing but not signaling," *The Journal of Immunology*, vol. 184, no. 9, pp. 5094–5103, 2010.

[327] U. Hasan, C. Chaffois, C. Gaillard, V. Saulnier, E. Merck, S. Tancredi, C. Guiet, F. Brière, J. Vlach, S. Lebecque, *et al.*, "Human tlr10 is a functional receptor, expressed by b cells and plasmacytoid dendritic cells, which activates gene transcription through myd88," *The Journal of Immunology*, vol. 174, no. 5, pp. 2942–2950, 2005.

[328] M. Oosting, S.-C. Cheng, J. M. Bolscher, R. Vestering-Stenger, T. S. Plantinga, I. C. Verschueren, P. Arts, A. Garritsen, H. van Eenennaam, P. Sturm, *et al.*, "Human tlr10 is an anti-inflammatory pattern-recognition receptor," *Proceedings of the National Academy of Sciences*, vol. 111, no. 42, pp. E4478–E4484, 2014.

[329] G. Goh, M. Ahn, F. Zhu, L. B. Lee, D. Luo, A. T. Irving, and L.-F. Wang, "Complementary regulation of caspase-1 and il-1β reveals additional mechanisms of dampened inflammation in bats," *Proceedings of the National Academy of Sciences*, vol. 117, no. 46, pp. 28939–28949, 2020.

[330] T. H. Mogensen and S. R. Paludan, "Reading the viral signature by toll-like receptors and other pattern recognition receptors," *Journal of molecular medicine*, vol. 83, no. 3, pp. 180–192, 2005.

[331] L. Alexopoulou, A. C. Holt, R. Medzhitov, and R. A. Flavell, "Recognition of double-stranded rna and activation of nf-κb by toll-like receptor 3," *Nature*, vol. 413, no. 6857, pp. 732–738, 2001.

[332] S. S. Diebold, T. Kaisho, H. Hemmi, S. Akira, and C. R. e Sousa, "Innate antiviral responses by means of tlr7-mediated recognition of single-stranded rna," *Science*, vol. 303, no. 5663, pp. 1529–1531, 2004.

[333] F. Heil, H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, and S. Bauer, "Species-specific recognition of single-stranded rna via toll-like receptor 7 and 8," *Science*, vol. 303, no. 5663, pp. 1526–1529, 2004.

[334] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H. rui Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Dèng, L.-L. Liu, B. Yan, F. Zhan, Y.-Y. Wang, G. fu Xiao, and Z.-L. Shi, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, pp. 270 – 273, 2020.

[335] T. Dzopalic, A. Dragicevic, S. Vasilijic, D. Vucevic, I. Majstorovic, B. Bozic, B. Balint, and M. Colic, "Loxoribine, a selective toll-like receptor 7 agonist, induces maturation of human monocyte-derived dendritic cells and stimulates their th-1-and th-17-polarizing capability," *International immunopharmacology*, vol. 10, no. 11, pp. 1428–1433, 2010.

[336] H. Hemmi, O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, M. Matsumoto, K. Hoshino, H. Wagner, K. Takeda, *et al.*, "A toll-like receptor recognizes bacterial dna," *Nature*, vol. 408, no. 6813, pp. 740–745, 2000.

[337] H. Wagner, "The immunobiology of the tlr9 subfamily," *Trends in immunology*, vol. 25, no. 7, pp. 381–386, 2004.

[338] H. M. Kim, B. S. Park, J.-I. Kim, S. E. Kim, J. Lee, S. C. Oh, P. Enkhbayar, N. Matsushima, H. Lee, O. J. Yoo, *et al.*, "Crystal structure of the tlr4-md-2 complex with bound endotoxin antagonist eritoran," *Cell*, vol. 130, no. 5, pp. 906–917, 2007.

[339] J.-K. Ryu, S. J. Kim, S.-H. Rah, J. I. Kang, H. E. Jung, D. Lee, H. K. Lee, J.-O. Lee, B. S. Park, T.-Y. Yoon, *et al.*, "Reconstruction of lps transfer cascade reveals structural determinants within lbp, cd14, and tlr4-md2 for efficient lps recognition and transfer," *Immunity*, vol. 46, no. 1, pp. 38–50, 2017.

[340] K. Shinya, M. Ito, A. Makino, M. Tanaka, K. Miyake, A. J. Eisfeld, and Y. Kawaoka, "The tlr4-trif pathway protects against h5n1 influenza virus infection," *Journal of virology*, vol. 86, no. 1, pp. 19–24, 2012.

[341] J. Zhao, C. Wohlford-Lenane, J. Zhao, E. Fleming, T. E. Lane, P. B. McCray, and S. Perlman, "Intranasal treatment with poly (i· c) protects aged mice from lethal respiratory virus infections," *Journal of virology*, vol. 86, no. 21, pp. 11416–11424, 2012.

[342] Y. Imai, K. Kuba, G. G. Neely, R. Yaghubian-Malhami, T. Perkmann, G. van Loo, M. Ermolaeva, R. Veldhuizen, Y. C. Leung, H. Wang, *et al.*, "Identification of oxidative stress and toll-like receptor 4 signaling as a key pathway of acute lung injury," *Cell*, vol. 133, no. 2, pp. 235–249, 2008.

[343] Q. M. Nhu, K. Shirey, J. Teijaro, D. Farber, S. Netzel-Arnett, T. Antalis, A. Fasano, and S. Vogel, "Novel signaling interactions between proteinase-

activated receptor 2 and toll-like receptors in vitro and in vivo," *Mucosal immunology*, vol. 3, no. 1, pp. 29–39, 2010.

[344] S. Tong, X. Zhu, Y. Li, M. Shi, J. Zhang, M. Bourgeois, H. Yang, X. Chen, S. Recuenco, J. Gomez, *et al.*, "New world bats harbor diverse influenza a viruses," *PLoS pathog*, vol. 9, no. 10, p. e1003657, 2013.

[345] K. Ciminski, F. Pfaff, M. Beer, and M. Schwemmle, "Bats reveal the true power of influenza a virus adaptability," *PLoS pathogens*, vol. 16, no. 4, p. e1008384, 2020.

[346] K. Ciminski, W. Ran, M. Gorka, J. Lee, A. Malmlov, J. Schinköthe, M. Eckley, R. A. Murrieta, T. A. Aboellail, C. L. Campbell, *et al.*, "Bat influenza viruses transmit among bats but are poorly adapted to non-bat species," *Nature microbiology*, vol. 4, no. 12, pp. 2298–2309, 2019.

[347] A. C. A. Campos, L. G. B. Góes, A. Moreira-Soto, C. de Carvalho, G. Ambar, A.-L. Sander, C. Fischer, A. R. da Rosa, D. C. de Oliveira, A. P. G. Kataoka, *et al.*, "Bat influenza a (hl18nl11) virus in fruit bats, brazil," *Emerging infectious diseases*, vol. 25, no. 2, p. 333, 2019.

[348] H. Bainová, T. Králová, A. Bryjová, T. Albrecht, J. Bryja, and M. Vinkler, "First evidence of independent pseudogenization of toll-like receptor 5 in passerine birds," *Developmental & Comparative Immunology*, vol. 45, no. 1, pp. 151–155, 2014.

[349] F. Hayashi, K. D. Smith, A. Ozinsky, T. R. Hawn, C. Y. Eugene, D. R. Goodlett, J. K. Eng, S. Akira, D. M. Underhill, and A. Aderem, "The innate immune response to bacterial flagellin is mediated by toll-like receptor 5," *Nature*, vol. 410, no. 6832, pp. 1099–1103, 2001.

[350] J. Yang and H. Yan, "Tlr5: beyond the recognition of flagellin," *Cellular & molecular immunology*, vol. 14, no. 12, pp. 1017–1019, 2017.

[351] W. A. Andrade, M. do Carmo Souza, E. Ramos-Martinez, K. Nagpal, M. S. Dutra, M. B. Melo, D. C. Bartholomeu, S. Ghosh, D. T. Golenbock, and R. T. Gazzinelli, "Combined action of nucleic acid-sensing toll-like receptors and tlr11/tlr12 heterodimers imparts resistance to toxoplasma gondii in mice," *Cell host & microbe*, vol. 13, no. 1, pp. 42–53, 2013.

[352] P. Broz and D. M. Monack, "Newly described pattern recognition receptors team up against intracellular pathogens," *Nature Reviews Immunology*, vol. 13, no. 8, pp. 551–565, 2013.

[353] M. Raetz, A. Kibardin, C. R. Sturge, R. Pifer, H. Li, E. Burstein, K. Ozato, S. Larin, and F. Yarovinsky, "Cooperation of tlr12 and tlr11 in the irf8-dependent il-12 response to toxoplasma gondii profilin," *The Journal of Immunology*, vol. 191, no. 9, pp. 4818–4827, 2013.

[354] F. Yarovinsky, D. Zhang, J. F. Andersen, G. L. Bannenberg, C. N. Serhan, M. S. Hayden, S. Hieny, F. S. Sutterwala, R. A. Flavell, S. Ghosh, *et al.*, "Tlr11 activation of dendritic cells by a protozoan profilin-like protein," *Science*, vol. 308, no. 5728, pp. 1626–1629, 2005.

[355] A. A. Koblansky, D. Jankovic, H. Oh, S. Hieny, W. Sungnak, R. Mathur, M. S. Hayden, S. Akira, A. Sher, and S. Ghosh, "Recognition of profilin by toll-like receptor 12 is critical for host resistance to toxoplasma gondii," *Immunity*, vol. 38, no. 1, pp. 119–130, 2013.

[356]

[357] N. D. Temperley, S. Berlin, I. R. Paton, D. K. Griffin, and D. W. Burt, "Evolution of the chicken toll-like receptor gene family: a story of gene gain and gene loss," *BMC genomics*, vol. 9, no. 1, pp. 1–12, 2008.

[358] E. Ishengoma and M. Agaba, "Evolution of toll-like receptors in the context of terrestrial ungulates and cetaceans diversification," *BMC evolutionary biology*, vol. 17, no. 1, pp. 1–13, 2017.

[359] M. Oldenburg, A. Krüger, R. Ferstl, A. Kaufmann, G. Nees, A. Sigmund, B. Bathke, H. Lauterbach, M. Suter, S. Dreher, *et al.*, "Tlr13 recognizes bacterial 23s rrna devoid of erythromycin resistance–forming modification," *Science*, vol. 337, no. 6098, pp. 1111–1115, 2012.

[360] M. M. Brinkmann, E. Spooner, K. Hoebe, B. Beutler, H. L. Ploegh, and Y.-M. Kim, "The interaction between the er membrane protein unc93b and tlr3, 7, and 9 is crucial for tlr signaling," *Journal of Cell Biology*, vol. 177, no. 2, pp. 265–275, 2007.

[361] Y.-M. Kim, M. M. Brinkmann, M.-E. Paquet, and H. L. Ploegh, "Unc93b1 delivers nucleotide-sensing toll-like receptors to endolysosomes," *Nature*, vol. 452, no. 7184, pp. 234–238, 2008.

[362] J. Schad and C. C. Voigt, "Adaptive evolution of virus-sensing toll-like receptor 8 in bats," *Immunogenetics*, vol. 68, no. 10, pp. 783–795, 2016.

[363] H. Jiang, J. Li, L. Li, X. Zhang, L. Yuan, and J. Chen, "Selective evolution of toll-like receptors 3, 7, 8, and 9 in bats," *Immunogenetics*, vol. 69, no. 4, pp. 271–285, 2017.

[364] M. E. Woolhouse and S. Gowtage-Sequeria, "Host range and emerging and reemerging pathogens," *Emerging infectious diseases*, vol. 11, no. 12, p. 1842, 2005.

[365] W. Karesh, A. Dobson, J. Lloyd-Smith, J. Lubroth, M. Dixon, M. Bennett, S. Aldrich, T. Harrington, P. Formenty, E. H. Loh, C. Machalaba, M. J. Thomas, and D. Heymann, "Ecology of zoonoses: natural and unnatural histories," *Lancet (London, England)*, vol. 380, pp. 1936 – 1945, 2012.

[366] J. Hughes, M. Wilson, B. Pike, K. Saylors, J. Fair, M. Lebreton, U. Tamoufé, C. Djoko, A. Rimoin, and N. Wolfe, "The origin and prevention of pandemics," *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, vol. 50, pp. 1636 – 1640, 2010.

[367] R. Plowright, C. Parrish, H. McCallum, P. Hudson, A. Ko, A. Graham, and J. Lloyd-Smith, "Pathways to zoonotic spillover," *Nature Reviews. Microbiology*, vol. 15, pp. 502 – 510, 2017.

[368] R. Webby, E. Hoffmann, and R. Webster, "Molecular constraints to interspecies transmission of viral pathogens," *Nature Medicine*, vol. 10, pp. S77 – S81, 2004.

[369] J. Paulson and R. P. de Vries, "H5n1 receptor specificity as a factor in pandemic risk.," *Virus research*, vol. 178 1, pp. 99–113, 2013.

[370] R. Hulswit, C. D. de Haan, and B. Bosch, "Coronavirus spike protein and tropism changes," *Advances in Virus Research*, vol. 96, pp. 29 – 57, 2016.

[371] M. Woolhouse, F. Scott, Z. Hudson, R. Howey, and M. Chase-Topping, "Human viruses: discovery and emergence," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, pp. 2864 – 2871, 2012.

[372] F. Li, "Structure, function, and evolution of coronavirus spike proteins.," *Annual review of virology*, vol. 3 1, pp. 237–261, 2016.

[373] H. Feldmann, V. Wahl-Jensen, S. M. Jones, and U. Ströher, "Ebola virus ecology: a continuing mystery," *Trends in microbiology*, vol. 12, no. 10, pp. 433–437, 2004.

[374] A. Banerjee, A. C. Doxey, K. Mossman, and A. T. Irving, "Unravelling the zoonotic origin and transmission of sars-cov-2," *Trends in ecology & evolution*, 2020.

[375] M. Mahdy, W. Younis, and Z. Ewaida, "An overview of sars-cov-2 and animal infection," *Frontiers in Veterinary Science*, vol. 7, 2020.

[376] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li, "Structural basis of receptor recognition by sars-cov-2," *Nature*, vol. 581, no. 7807, pp. 221–224, 2020.

[377] K. Andersen, A. Rambaut, W. Lipkin, E. Holmes, and R. Garry, "The proximal origin of sars-cov-2," *Nature Medicine*, pp. 1 – 3, 2020.

[378] X. Tang, C. Wu, X. Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian, J. Cui, and J. Lu, "On the origin and continuing evolution of sars-cov-2," *National Science Review*, 2020.

[379] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, *et al.*, "Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor," *Nature*, vol. 581, no. 7807, pp. 215–220, 2020.

[380] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, *et al.*, "Angiotensin-

converting enzyme 2 is a functional receptor for the sars coronavirus," *Nature*, vol. 426, no. 6965, pp. 450–454, 2003.

[381] V. B. Patel, J.-C. Zhong, M. B. Grant, and G. Y. Oudit, "Role of the ace2/angiotensin 1–7 axis of the renin–angiotensin system in heart failure," *Circulation research*, vol. 118, no. 8, pp. 1313–1326, 2016.

[382] Y. Feng, X. Yue, H. Xia, S. M. Bindom, P. J. Hickman, C. M. Filipeanu, G. Wu, and E. Lazartigues, "Angiotensin-converting enzyme 2 overexpression in the subfornical organ prevents the angiotensin ii–mediated pressor and drinking responses and is associated with angiotensin ii type 1 receptor downregulation," *Circulation research*, vol. 102, no. 6, pp. 729–736, 2008.

[383] S. Tipnis, N. Hooper, R. Hyde, E. Karran, G. Christie, and A. Turner, "A human homolog of angiotensin-converting enzyme," *The Journal of Biological Chemistry*, vol. 275, pp. 33238 – 33243, 2000.

[384] G. Lu, Q. Wang, and G. F. Gao, "Bat-to-human: spike features determining 'host jump'of coronaviruses sars-cov, mers-cov, and beyond," *Trends in microbiology*, vol. 23, no. 8, pp. 468–478, 2015.

[385] K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, *et al.*, "The mutational constraint spectrum quantified from variation in 141,456 humans," *Nature*, vol. 581, no. 7809, pp. 434–443, 2020.

[386] K. Suryamohan, D. Diwanji, E. Stawiski, R. Gupta, S. Miersch, J. Liu, C. Chen, Y.-P. Jiang, F. A. Fellouse, J. Sathirapongsasuti, P. Albers, T. Deepak, R. Saberianfar, A. Ratan, G. Washburn, M. Mis, D. Santhosh, S. Somasekar, G. H. Hi-

ranjith, D. Vargas, S. Mohan, S. Phalke, B. Kuriakose, A. Antony, M. Ustav, S. Schuster, S. Sidhu, J. R. Junutula, N. Jura, and S. Seshagiri, "Human ace2 receptor polymorphisms and altered susceptibility to sars-cov-2," *Communications Biology*, vol. 4, 2021.

[387] F. Li, "Receptor recognition and cross-species infections of sars coronavirus," *Antiviral research*, vol. 100, no. 1, pp. 246–254, 2013.

[388] F. Li, W. Li, M. Farzan, and S. C. Harrison, "Structure of sars coronavirus spike receptor-binding domain complexed with receptor," *Science*, vol. 309, no. 5742, pp. 1864–1868, 2005.

[389] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z.-Y. Zhang, G. Lu, C. Qiao, Y. Hu, K. Yuen, Q. Wang, H. Zhou, J. Yan, and J. Qi, "Structural and functional basis of sars-cov-2 entry by using human ace2," *Cell*, vol. 181, pp. 894 – 904.e9, 2020.

[390] I. Mercurio, V. Tragni, F. Busto, A. De Grassi, and C. L. Pierri, "Protein structure analysis of the interactions between sars-cov-2 spike protein and the human ace2 receptor: from conformational changes to novel neutralizing antibodies," *Cellular and Molecular Life Sciences*, vol. 78, no. 4, pp. 1501–1522, 2021.

[391] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, "Structural basis for the recognition of sars-cov-2 by full-length human ace2," *Science (New York, N.y.)*, vol. 367, pp. 1444 – 1448, 2020.

[392] Y. Wang, M. Liu, and J. Gao, "Enhanced receptor binding of sars-cov-2 through networks of hydrogen-bonding and hydrophobic interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 13967 – 13974, 2020.

[393] Y. Wei, P. Aris, H. Farookhi, and X. Xia, "Predicting mammalian species at risk of being infected by sars-cov-2 from an ace2 perspective," *Scientific Reports*, vol. 11, 2021.

[394] T. Zhang, Q. Wu, and Z. Zhang, "Probable pangolin origin of sars-cov-2 associated with the covid-19 outbreak," *Current biology*, vol. 30, no. 7, pp. 1346–1351, 2020.

[395] N. Oreshkova, R. Molenaar, S. Vreman, F. Harders, B. O. O. Munnink, R. W. H. van der Honing, N. Gerhards, P. Tolsma, R. Bouwstra, R. Sikkema, M. Tacken, M. M. D. de Rooij, E. Weesendorp, M. Engelsma, C. Bruschke, L. Smit, M. Koopmans, W. V. D. van der Poel, and A. Stegeman, "Sars-cov-2 infection in farmed minks, the netherlands, april and may 2020," *Eurosurveillance*, vol. 25, 2020.

[396] P. Halfmann, M. Hatta, S. Chiba, T. Maemura, S. Fan, M. Takeda, N. Kinoshita, S. Hattori, Y. Sakai-Tagawa, K. Iwatsuki-Horimoto, M. Imai, and Y. Kawaoka, "Transmission of sars-cov-2 in domestic cats.," *The New England journal of medicine*, 2020.

[397] T. Sit, C. J. Brackman, S. Ip, K. W. S. Tam, P. Y. T. Law, E. M. W. To, V. Y. T. Yu, L. Sims, D. Tsang, D. Chu, R. Perera, L. Poon, and M. Peiris, "Infection of dogs with sars-cov-2," *Nature*, pp. 1–3, 2020.

[398] J. Chan, A. J. Zhang, S. Yuan, V. Poon, C. Chan, A. Lee, W. Chan, Z. Fan, H. Tsoi, L. Wen, R. Liang, J. Cao, Y. xia Chen, K. Tang, C. Luo, J. Cai, K. Kok, H. Chu, K. Chan, S. Sridhar, Z. Chen, H. Chen, K. To, and K. Yuen, "Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (covid-19) in golden syrian hamster model: implications for disease pathogenesis and transmissibility," *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 2020.

[399] A. Chandrashekar, J. Liu, A. Martinot, K. McMahan, N. Mercado, L. Peter, L. H. Tostanoski, J. Yu, Z. Maliga, M. D. Nekorchuk, K. Busman-Sahay, M. Terry, L. M. Wrijil, S. Ducat, D. R. Martinez, C. Atyeo, S. Fischinger, J. S. Burke, M. D. Slein, L. Pessaint, A. V. Ry, J. Greenhouse, T. Taylor, K. Blade, A. Cook, B. Finneyfrock, R. Brown, E. Teow, J. Velasco, R. Zahn, F. Wegmann, P. Abbink, E. A. Bondzie, G. Dagotto, M. S. Gebre, X. He, C. Jacob-Dolan, N. Kordana, Z. Li, M. Lifton, S. Mahrokhian, L. F. Maxfield, R. Nityanandam, J. Nkolola, A. Schmidt, A. Miller, R. Baric, G. Alter, P. Sorger, J. Estes, H. Andersen, M. Lewis, and D. Barouch, "Sars-cov-2 infection protects against rechallenge in rhesus macaques," *Science (New York, N.y.)*, vol. 369, pp. 812 – 817, 2020.

[400] B. Rockx, T. Kuiken, S. Herfst, T. Bestebroer, M. Lamers, B. O. O. Munnink, D. de Meulder, G. van Amerongen, J. V. D. van den Brand, N. M. A. Okba, D. Schipper, P. van Run, L. Leijten, R. Sikkema, E. Verschoor, B. Verstrepen, W. Bogers, J. Langermans, C. Drosten, M. F. van Vlissingen, R. Fouchier, R. D. de Swart, M. Koopmans, and B. Haagmans, "Comparative pathogenesis of covid-19, mers, and sars in a nonhuman primate model," *Science (New York, N.y.)*, vol. 368, pp. 1012 – 1015, 2020.

[401] K. Schlottau, M. Rissmann, A. Graaf, J. Schön, J. Sehl, C. Wylezich, D. Höper, T. Mettenleiter, A. Balkema-Buschmann, T. Harder, C. Grund, D. Hoffmann, A. Breithaupt, and M. Beer, "Sars-cov-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study," *The Lancet. Microbe*, vol. 1, pp. e218 – e225, 2020.

[402] M. Richard, A. Kok, D. de Meulder, T. Bestebroer, M. Lamers, N. M. A. Okba, M. F. van Vlissingen, B. Rockx, B. Haagmans, M. Koopmans, R. Fouchier, and

S. Herfst, "Sars-cov-2 is transmitted via contact and via the air between ferrets," *Nature Communications*, vol. 11, 2020.

[403] K. Xiao, J. Zhai, Y. Feng, N. Zhou, X. Zhang, J. Zou, N. Li, Y. Guo, X. bing Li, X. Shen, Z. Zhang, F. Shu, W. Huang, Y. Li, R. ai Chen, Y. Wu, S. ming Peng, M. Huang, W. Xie, Q.-H. Cai, F. Hou, Y. Liu, W. Chen, L. Xiao, and Y. Shen, "Isolation and characterization of 2019-ncov-like coronavirus from malayan pangolins," *bioRxiv*, 2020.

[404] J. Sun, W.-T. He, L. Wang, A. Lai, X. Ji, X. Zhai, G. Li, M. A. Suchard, J. Tian, J. Zhou, *et al.*, "Covid-19: epidemiology, evolution, and cross-disciplinary perspectives," *Trends in molecular medicine*, vol. 26, no. 5, pp. 483–495, 2020.

[405] J. Damas, G. M. Hughes, K. C. Keough, C. A. Painter, N. S. Persky, M. Corbo, M. Hiller, K.-P. Koepfli, A. R. Pfenning, H. Zhao, *et al.*, "Broad host range of sars-cov-2 predicted by comparative and structural analysis of ace2 in vertebrates," *Proceedings of the National Academy of Sciences*, vol. 117, no. 36, pp. 22311–22322, 2020.

[406] H. Yan, H. Jiao, Q. Liu, Z. Zhang, Q. Xiong, B. Wang, X. dong Wang, M. Guo, L.-F. Wang, K. Lan, Y. Chen, and H. Zhao, "Ace2 receptor usage reveals variation in susceptibility to sars-cov and sars-cov-2 infection among bat species.," *Nature ecology & evolution*, 2021.

[407] W. Li, C. Zhang, J. Sui, J. Kuhn, M. Moore, S. Luo, S.-K. Wong, I.-C. Huang, K. Xu, N. Vasilieva, A. Murakami, Y. He, W. Marasco, Y. Guan, H. Choe, and M. Farzan, "Receptor and viral determinants of sars-coronavirus adaptation to human ace2," *The EMBO Journal*, vol. 24, pp. 1634 – 1643, 2005.

[408] A. Melin, M. C. Janiak, F. Marrone, P. Arora, and J. Higham, "Comparative ace2 variation and primate covid-19 risk," *bioRxiv*, 2020.

[409] O. FamBibianaS, P. Vargas-Pinilla, C. E. G. Amorim, V. A. Sortica, and M. Bortolini, "Ace2 diversity in placental mammals reveals the evolutionary strategy of sars-cov-2," *Genetics and Molecular Biology*, vol. 43, 2020.

[410] J. Shi, Z. Wen, G. Zhong, H. Yang, C. Wang, B. Huang, R. Liu, X. He, L. Shuai, Z. Sun, Y. Zhao, P. Liu, L. Liang, P. Cui, J. Wang, X. Zhang, Y. Guan, W. Tan, G. Wu, H. Chen, and Z. Bu, "Susceptibility of ferrets, cats, dogs, and other domesticated animals to sars–coronavirus 2," *Science (New York, N.y.)*, vol. 368, pp. 1016 – 1020, 2020.

[411] K. Sun, L. Gu, L. Ma, and Y. Duan, "Atlas of ace2 gene expression reveals novel insights into transmission of sars-cov-2," *Heliyon*, vol. 7, pp. e05850 – e05850, 2020.

[412] J. V. D. van den Brand, B. Haagmans, L. Leijten, D. van Riel, B. Martina, A. Osterhaus, and T. Kuiken, "Pathology of experimental sars coronavirus infection in cats and ferrets," *Veterinary Pathology*, vol. 45, pp. 551 – 562, 2008.

[413] H. Hui, Y. qian Zhang, X. Yang, X. Wang, B. He, L. Li, H. Li, J. Tian, and Y. dai Chen, "Clinical and radiographic features of cardiac injury in patients with 2019 novel coronavirus pneumonia," *medRxiv*, 2020.

[414] T. Enkirch and V. von Messling, "Ferret models of viral pathogenesis," *Virology*, vol. 479, pp. 259 – 270, 2015.

[415] J. Bartlett, "Virology: Sars virus infection of cats and ferrets," *Infectious Diseases in Clinical Practice*, vol. 12, 2004.

[416] T. Gramberg, H. Hofmann, P. Möller, P. Lalor, A. Marzi, M. Geier, M. Krumbiegel, T. Winkler, F. Kirchhoff, D. Adams, S. Becker, J. Münch, and S. Pöhlmann, "Lsectin interacts with filovirus glycoproteins and the spike protein of sars coronavirus," *Virology*, vol. 340, pp. 224 – 236, 2005.

[417] K. Pepin, S. Lass, J. Pulliam, A. Read, and J. Lloyd-Smith, "Identifying genetic markers of adaptation for surveillance of viral host jumps," *Nature Reviews. Microbiology*, vol. 8, pp. 802 – 813, 2010.

[418] S. Morse, J. Mazet, M. Woolhouse, C. Parrish, D. Carroll, W. Karesh, C. Zambrana-Torrelio, W. Lipkin, and P. Daszak, "Prediction and prevention of the next pandemic zoonosis," *Lancet (London, England)*, vol. 380, pp. 1956 – 1965, 2012.

[419] C. R. Parrish, E. C. Holmes, D. M. Morens, E.-C. Park, D. S. Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif, and P. Daszak, "Cross-species virus transmission and the emergence of new epidemic diseases," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 3, pp. 457–470, 2008.

[420] Z. Wu, L. Yang, X. Ren, G. He, J. Zhang, J. Yang, Z. Qian, J. Dong, L. Sun, Y. Zhu, J. Du, F. Yang, S. Zhang, and Q. Jin, "Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases," *The ISME Journal*, vol. 10, pp. 609 – 620, 2016.

[421] R. Moratelli and C. H. Calisher, "Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses?," *Memórias do Instituto Oswaldo Cruz*, vol. 110, no. 1, pp. 1–22, 2015.

[422] T. Schountz, "Immunology of bats and their viruses: challenges and opportunities," *Viruses*, vol. 6, no. 12, pp. 4880–4901, 2014.

[423] A. Lopez-Baucells, R. Rocha, and A. Fernandez-Llamazares, "When bats go viral: negative framings in virological research imperil bat conservation," *Mammal Review*, vol. 48, pp. 62–66, 2018.

[424] M. B. Kalka, A. R. Smith, and E. K. Kalko, "Bats limit arthropods and herbivory in a tropical forest," *Science*, vol. 320, no. 5872, pp. 71–71, 2008.

[425] J. G. Boyles, P. M. Cryan, G. F. McCracken, and T. H. Kunz, "Economic importance of bats in agriculture," *Science*, vol. 332, no. 6025, pp. 41–42, 2011.

[426] F. Mayen, "Haematophagous bats in brazil, their role in rabies transmission, impact on public health, livestock industry and alternatives to an indiscriminate reduction of bat population," *Journal of Veterinary Medicine, Series B*, vol. 50, no. 10, pp. 469–472, 2003.

[427] G. Wibbelt, M. S. Moore, T. Schountz, and C. C. Voigt, "Emerging diseases in chiroptera: why bats?," *Biol Lett*, vol. 6, no. 4, pp. 438–40, 2010.

[428] P. Pérot, P.-A. Bolze, and F. Mallet, "From viruses to genes: syncytins," in *Viruses: essential agents of life*, pp. 325–361, Springer, 2012.

[429] A. Papenfuss, M. Baker, Z.-P. Feng, M. Tachedjian, G. Crameri, C. Cowled, J. Ng, V. Janardhana, H. Field, and L. Wang, "The immune gene repertoire of an important viral reservoir, the australian black flying fox," *BMC Genomics*, vol. 13, pp. 261 – 261, 2012.

[430] T. Shaw, A. Srivastava, W.-C. Chou, L. Liu, A. Hawkinson, T. Glenn, R. Adams, and T. Schountz, "Transcriptome sequencing and annotation for the jamaican fruit bat (artibeus jamaicensis)," *PLoS ONE*, vol. 7, 2012.

[431] A. Lee, K. A. Kulcsar, O. Elliott, H. Khiabanian, E. R. Nagle, M. Jones, B. Amman, M. Sanchez-Lockhart, J. Towner, G. Palacios, and R. Rabadán, "De novo transcriptome reconstruction and annotation of the egyptian rousette bat," *BMC Genomics*, vol. 16, 2015.

[432] S. Koyama, K. Ishii, C. Coban, and S. Akira, "Innate immune response to viral infection.," *Cytokine*, vol. 43 3, pp. 336–41, 2008.

[433] C. Janeway and R. Medzhitov, "Innate immune recognition.," *Annual review of immunology*, vol. 20, pp. 197–216, 2002.

[434] A. Banerjee, N. Rapin, T. Bollinger, and V. Misra, "Lack of inflammatory gene expression in bats: a unique role for a transcription repressor," *Scientific reports*, vol. 7, no. 1, pp. 1–15, 2017.

[435] A. Banerjee, V. Misra, T. Schountz, and M. L. Baker, "Tools to study pathogen-host interactions in bats," *Virus research*, vol. 248, pp. 5–12, 2018.

[436] K. E. Jones, A. Purvis, and J. L. Gittleman, "Biological correlates of extinction risk in bats," *The American Naturalist*, vol. 161, no. 4, pp. 601–614, 2003.

[437] D. Holmes and D. Kristan, "Comparative and alternative approaches and novel animal models for aging research," *AGE*, vol. 30, pp. 63–73, 2008.

[438] J. Munshi-South and G. Wilkinson, "Bats and birds: Exceptional longevity despite high metabolic rates," *Ageing Research Reviews*, vol. 9, pp. 12–19, 2010.

[439] D. Haydon, S. Cleaveland, L. Taylor, and M. Laurenson, "Identifying reservoirs of infection: A conceptual and practical challenge," *Emerging Infectious Diseases*, vol. 8, pp. 1468 – 1473, 2002.

[440] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, "Global trends in emerging infectious diseases," *Nature*, vol. 451, no. 7181, pp. 990–3, 2008.

[441] R. Gibb, L. Franklinos, D. Redding, and K. Jones, "Ecosystem perspectives are needed to manage zoonotic risks in a changing climate," *The BMJ*, vol. 371, 2020.

[442] A. Engering, L. Hogerwerf, and J. Slingenbergh, "Pathogen–host–environment interplay and disease emergence," *Emerging Microbes & Infections*, vol. 2, 2013.

[443] B. Bird and J. Mazet, "Detection of emerging zoonotic pathogens: An integrated one health approach.," *Annual review of animal biosciences*, vol. 6, pp. 121–139, 2018.

[444] E. Corrales-Aguilar, M. Schwemmle, and D. Hewitt, "Bats and viruses: Current research and future trends," *Emerging Infectious Diseases*, vol. 27, no. 3, p. 993, 2021.

[445] F. Keesing, L. K. Belden, P. Daszak, A. Dobson, C. D. Harvell, R. D. Holt, P. Hudson, A. Jolles, K. E. Jones, C. E. Mitchell, *et al.*, "Impacts of biodiversity on the emergence and transmission of infectious diseases," *Nature*, vol. 468, no. 7324, pp. 647–652, 2010.

[446] R. Coker, J. Rushton, S. Mounier-Jack, E. Karimuribo, P. Lutumba, D. Kambarage, D. Pfeiffer, K. Stärk, and M. Rweyemamu, "Towards a conceptual frame-

work to support one-health research for policy on emerging zoonoses," *The Lancet. Infectious Diseases*, vol. 11, pp. 326 – 331, 2011.

[447] W. Karesh and R. Cook, "The human-animal link," *Foreign Affairs*, vol. 84, pp. 38–50, 2005.

[448] G. Martin, C. Yañez-Arenas, C. Chen, R. Plowright, R. Webb, and L. Skerratt, "Climate change could increase the geographic extent of hendra virus spillover risk," *Ecohealth*, vol. 15, pp. 509 – 525, 2018.

[449] L. Brierley, M. Vonhof, K. Olival, P. Daszak, and K. Jones, "Quantifying global drivers of zoonotic bat viruses: A process-based perspective," *The American Naturalist*, vol. 187, pp. E53 – E64, 2016.

[450] N. Wolfe, P. Daszak, A. M. Kilpatrick, and D. Burke, "Bushmeat hunting, deforestation, and prediction of zoonotic disease," *Emerging Infectious Diseases*, vol. 11, pp. 1822 – 1827, 2005.

[451] P. Torgerson and C. Macpherson, "The socioeconomic burden of parasitic zoonoses: global trends.," *Veterinary parasitology*, vol. 182 1, pp. 79–95, 2011.

[452] K. Royce and F. Fu, "Mathematically modeling spillovers of an emerging infectious zoonosis with an intermediate host," *PLoS ONE*, vol. 15, 2020.

[453] C. Gilchrist, S. Turner, M. Riley, W. Petri, and E. Hewlett, "Whole-genome sequencing in outbreak analysis," *Clinical Microbiology Reviews*, vol. 28, pp. 541 – 563, 2015.

[454] A. Mongan, J. Tuda, and L. Runtuwene, "Portable sequencer in the fight against infectious disease," *Journal of Human Genetics*, vol. 65, pp. 35 – 40, 2019.

[455] W. Lipkin, "Microbe hunting," *Microbiology and Molecular Biology Reviews*, vol. 74, pp. 363 – 377, 2010.

[456] D. Relman, "Microbial genomics and infectious diseases.," *The New England journal of medicine*, vol. 365 4, pp. 347–57, 2011.