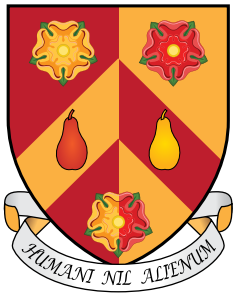


UNIVERSITY OF
OXFORD

Molecular sequence analysis of Crimean-Congo haemorrhagic fever virus and the discovery and characterisation of novel sarbecoviruses



Jack M Crook

Wolfson College

A thesis submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Table of Figures 3
Table of Tables 5
Acknowledgements 6
Abstract 8
1 Introduction 10
1.1 What are viruses? 10
1.2 DNA and RNA viruses 11
1.3 Diversity within an RNA virus genomic population 12
1.4 Recombination and Reassortment 13
1.5 Emerging infectious disease and zoonotic spillover 14
1.6 Reverse zoonosis 17
1.7 A history of pandemics 18
1.8 Bats as a reservoir of viral pathogens 19
1.9 Metagenomics for viral detection and characterisation 22
1.10 Molecular phylogenetics 25
1.11 Phylodynamics, phylogeography and tracking RNA viruses 36
1.12 Project aims 37
2 Materials and Methodology 39
2.1 Crimean-Congo haemorrhagic fever virus sample pre-processing 39
2.2 UK bat dropping collection and pre-processing 40
2.3 Sequence Independent Single Primer Amplification 44
2.4 Targeted amplification 45
2.5 Next generation sequencing 46
2.6 Receptor binding assay methodology 47
2.7 General bioinformatic tools 50
2.8 Computational analysis of amplicon scheme raw data 52
2.9 Computational analysis for the CCHFv dataset 53
2.10 Retrieving NCBI sequences for dataset generation 55
2.11 Phylogenetic analyses 55
2.12 Time calibrated phylogenetic and phylogeographic analyses 56
2.13 Model selection 57
2.14 Discrete trait phylogeographic diffusion 58
2.15 Testing for selection 59
2.16 Testing for recombination 60
2.17 Testing for reassortment 61
2.18 UK Bat bioinformatic analyses 61

3 Molecular sequence analysis of Crimean-Congo haemorrhagic fever virus 64
3.1 Introduction 64
3.2 Results 98
3.3 Discussion 134
3.4 Conclusion 145
4 The discovery and characterisation of novel sarbecoviruses 147
4.1 Introduction 147
4.2 Results 170
4.3 Discussion 185
4.4 Conclusion 191
5 Longitudinal sampling of a lesser horseshoe roost in the UK.....192
5.1 Results 192
5.2 Discussion 220
5.3 Conclusions 226
6 Summary 227
6.1 Molecular sequence analysis of Crimean-Congo haemorrhagic fever virus 227
6.2 Initial detection and longitudinal surveillance of lesser horseshoe bats in the UK 229
Bibliography 232
Appendix 285

Table of Figures

Figure 1-1. Zoonotic spillover.....15
Figure 1-2. Reverse zoonosis.....18
Figure 1-3. mNGS pathways.....23
Figure 1-4. A hypothetical phylogenetic tree.....26
Figure 1-5. The two phylogenetic methods.....28
Figure 1-6. Types of heterogeneity within genomic sequences.....31
Figure 1-7. Methods for relaxing the clock.....33
Figure 1-8. A normally distributed prior for the evolutionary rate.....34
Figure 3-1.The enzootic maintenance cycle.....67
Figure 3-2. The virion and genome.....68
Figure 3-3. The nucleoprotein.....70
Figure 3-4.The GP38 protein.....72
Figure 3-5. The Gc protein.....74
Figure 3-6. The small segment phylogeny.....77
Figure 3-7. Evidence for CCHFv in Africa.....79
Figure 3-8. The Africa 2 and Africa 3 clades.....80
Figure 3-9. Evidence for CCHFv in Senegal and Mauritania.....84
Figure 3-10. Evidence for CCHFv in Asia.....86
Figure 3-11. The Asia 1 clade.....87
Figure 3-12. The Asia 2 clades.....89
Figure 3-13. Evidence for CCHFv in India.....90
Figure 3-14. Evidence for CCHFv in China.....92
Figure 3-15. The European clade.....93
Figure 3-16. Evidence for CCHFv in Europe.....95
Figure 3-17. S, M and L phylogenies.....96
Figure 3-18. The metadata and clustering of novel sequences.....99
Figure 3-19. Phylogenies highlighting novel European sequences.....100
Figure 3-20. Phylogenies for the Africa 2 containing clades.....101
Figure 3-21. Phylogenies for the Asia 2a containing clades.....102
Figure 3-22. A Phylogeny for the Asia 1 clade.....103
Figure 3-23. Small segment recombination candidates.....104
Figure 3-24. Medium segment recombination candidates.....105
Figure 3-25. Phylogenetic incongruence between segments.....106
Figure 3-26. Medium segment reassortment candidates.....107
Figure 3-27. Small segment reassortment candidates.....108

Figure 3-28. Large segment reassortment candidates. 109
Figure 3-29. Candidate reassorted taxa identified by RDP4.....110
Figure 3-30. The nucleoprotein sites under negative selection. 111
Figure 3-31. The nucleoprotein sites under episodic positive selection.....112
Figure 3-32. Sites in the Gc and GP38 protein under positive selection 113
Figure 3-33. Sites in the Gc protein under episodic positive selection. 114
Figure 3-34. Rooting hypotheses and temporal signal investigation. 115
Figure 3-35. Clades with temporal signal.....117
Figure 3-36. Small segment clade estimates.....118
Figure 3-37. Small segment time scaled phylogeny. 120
Figure 3-38. Rate estimates under fixed local clocks. 121
Figure 3-39. Differential coalescence and a time scaled phylogeny.....123
Figure 3-40. Discrete trait reconstruction for the Africa 2 clade.....124
Figure 3-41. Discrete trait reconstruction for the Africa 3 clade.....125
Figure 3-42. Discrete trait reconstruction for the Asia 1 clade. 127
Figure 3-43. Discrete trait reconstruction for the Asia 2a clade. 128
Figure 3-44. Discrete trait reconstruction for the Asia 2b clade. 130
Figure 3-45. Discrete trait reconstruction for the European clade. 131
Figure 3-46. Discrete trait reconstruction for the full small segment. 133
Figure 3-47. Major global bird flyways. 140
Figure 4-1. Spikes with the S1, S2 and RBD highlighted.....154
Figure 4-2. The coronavirus envelope protein. 159
Figure 4-3. The global distribution of sarbecoviruses. 165
Figure 4-4. Read classification at family and genus level. 170
Figure 4-5. Read classification at species level. 171
Figure 4-6. Reads mapped to representative *Sarbecovirus* genomes.....172
Figure 4-7. Sample 46 mapped to the genome assembly. 174
Figure 4-8. The genome organisation of RhGB01. 175
Figure 4-9. Phylogenies for *Rhinolophus* associated sarbecoviruses. 178
Figure 4-10. RhGB01 structural proteins. 179
Figure 4-11. Cleavage sites present in the Spike. 180
Figure 4-12. The RhGB01 and SARS-CoV-2 Spikes.....181
Figure 4-13. hACE2 contact residues in sarbecoviruses from Europe. 182
Figure 4-14. Mammalian ACE2 binding assay.....184
Figure 5-1. Sequencing statistics for mNGS.....195
Figure 5-2. CI63 and IM328 read depths. 196
Figure 5-3. CI19 and PLY24 read depths. 197

Figure 5-4. Blastn and raw reads mapped to the assembly.....198
Figure 5-5. Substitution rates for the different open reading frames.....199
Figure 5-6. Nucleotide and amino acid differences.....200
Figure 5-7. Spike glycoprotein comparison between RhGB01 and CI63.....202
Figure 5-8. RBD differences in RhGB01 and CI63. 204
Figure 5-9. RhGB01 and CI63 nonstructural proteins.....205
Figure 5-10. A phylogeny for *Rhinolophus* associated sarbecoviruses. 206
Figure 5-11. Samples for whole genome tiling approach.....207
Figure 5-12. Minor variant populations from the tiling approach.....208
Figure 5-13. The six RBD genotypes from the Cinderford roost.....211
Figure 5-14. The six genotypes aligned to the RhGB01 RBD. 213
Figure 5-15. Differences between the sarbecoviruses in the UK.....214
Figure 5-16. RBD differences between RhGB01 and CI63. 216
Figure 5-17. RBD differences between RhGB01 and RhGB07. 217
Figure 5-18. UK sarbecoviruses compared to SARS-CoV-2 and its hACE2 binding. 219

Table of Tables

Table 1-1. The Baltimore classification system.....11
Table 1-2. Assays for pathogen detection. 24
Table 1-3. mNGS for pathogen detection. 25
Table 2-1. Turbo DNase reaction constituents. 39
Table 2-2. EzDNase treatment reaction.....43
Table 2-3. qRT-PCR reaction and conditions. 43
Table 2-4. The SISPA protocol. 44
Table 2-5. The amplification reaction for the amplicon scheme PCR. 46
Table 2-6. Metadata associated with the reference sequences.....54
Table 3-1. Metadata for non-turkish novel sequences. 98
Table 3-2. TMRCA for the small segment clades. 119
Table 3-3. Evolutionary rates for the small segment clades. 119
Table 3-4. Supported discrete trait transitions for the small segment. 132
Table 4-1. The functions of the 16 non-structural proteins. 151
Table 4-2. Sequencing metrics for the 2020 faecal samples. 173
Table 5-1. qRT-PCR results across the years for the Cinderford roost. 192
Table 5-2. Faecal samples collected from across Somerset and Devon.....194
Table 5-3. ORFs with 100% recovery. 210

Acknowledgements

Completing this thesis would not have been possible without the support, guidance, and encouragement of numerous individuals and institutions, to whom I extend my greatest thanks.

First and foremost, I would like to express my sincere appreciation to my supervisors, Dr Steven Pullan and Professor Miles Carroll for their invaluable guidance, support, and feedback throughout. Their expertise and encouragement have been instrumental in guiding this work. I would like to thank Daniel Carter for his CL3 training and help in processing faecal samples. I also thank Professor Oliver Pybus and his laboratory for all their help and willingness over the last three and a half years. Additionally, I would like to thank Dr Tom Fletcher for collecting many of the haemorrhagic fever samples which have been used in this thesis. Additionally, I would like to thank the High Containment Microbiology group at Porton Down for retrieval and propagation of archived haemorrhagic fever samples where needed. I would also like to thank Dr Dalan Bailey from the Pirbright Institute for providing a useful training week in his group. Furthermore, I would like to thank Daniel Whitby, whose expertise were paramount in conducting dropping collections. His commitment and willingness to facilitate this aspect of the project were instrumental in its completion and success. Without his contributions this project would not have been feasible.

I extend my thanks to Porton Down for providing the necessary resources, facilities, and opportunities for conducting this research. The conducive academic environment and access to libraries, laboratories, and databases have been indispensable in the completion of this thesis.

I am grateful to my group and friends in the Genomics and Pathogen Characterisation division for their moral support, stimulating discussions, and solidarity throughout this journey. Their encouragement and understanding during challenging times, especially during the pandemic, have been a source of motivation. I am also very grateful to the NIHR for funding the project and my cohort at the HPRU EZI, whose yearly gatherings and group training sessions have been a highlight.

Special thanks are due to my family for their unconditional love, encouragement, and unwavering belief in my abilities. Their unwavering support and sacrifices have been the cornerstone of my academic pursuits. All their support over the last decade of studying at various places, without whom this would not have been possible.

In conclusion, I am profoundly grateful to everyone who has played a role, no matter how small, in the completion of this thesis. Your support and encouragement have been indispensable, and I am deeply appreciative of your contributions.

Abstract

Zoonotic diseases and viral infections impose substantial health and economic challenges globally. Timely prediction, detection, and response to these threats are paramount for zoonoses prevention. Recent strides in next-generation sequencing and computational capabilities have elevated pathogen genomics to its application as a leading method for predication, and both real-time and retrospective tracking of zoonotic infections.

Globally, there are a limited number of facilities which can routinely process and characterise hazard group 4 agents which cause high consequence viral infections. Many HG4 agents are underrepresented by genomic sequences in the public domain which prevents accurate prediction of risk regions and dissemination. Crimean-Congo haemorrhagic fever virus is a globally distributed virus with limited genomic representation compared to other forms of evidence, serological or molecular. This project characterised novel genomes from archival and contemporary clinical samples stored at Porton Down. We characterise the heterogeneous viral population circulating in Turkey in 2015 and corroborate current predictions for the global dissemination and divergence of CCHFv with its origin in Southern Africa.

The COVID-19 pandemic provided a unique opportunity for the characterisation of sarbecoviruses from horseshoe bats in the UK. Prior to 2020, surveillance in horseshoe bats had been focused on populations in China and Hong Kong leaving a largely undefined western limit for sarbecoviruses within horseshoe bat populations. Here we detect and characterise the first *Sarbecovirus* from lesser horseshoe bats in Europe and the first in horseshoe bats in the UK. Additionally,

continued longitudinal surveillance highlighted an accelerated evolutionary rate within the receptor binding motif, potentially highlighting it as a zoonotic risk.

1 Introduction

1.1 What are viruses?

Towards the end of the 19th century, Adolf Mayer, a chemist-turned botanist, was tasked with investigating the etiological agent causing Tobacco mosaic disease. Initial investigations identified infection persistence under environmental conditions known to reduce bacterial infectivity such as humidity and temperature. Subsequent work by Ivanovsky and Beijerinck resulted in the discovery of an infectious agent they termed *contagium vivum fluidum,* "the virus" (Beijerinck, 1898; Bos, 1999; Lechevalier, 1972; Lecoq, 2001). The characterisation of foot and mouth disease virus (FMDV) as the causative agent of FMD by Friedrich Loeffler in 1898 marked the first identification of a virus causing infection in vertebrates (Brown, 2003; Loeffler & Frosch, 1897; Wittmann, 1999). This was closely followed by the characterisation of the first virus capable of human infection, Yellow Fever Virus (YFV) in 1901 by US physician Walter Reed (Reed, 1901).

Now more than a century later, our understanding of viruses has expanded and the massive diversity of the planetary virome has been revealed. With huge variation in viral characteristics including but not limited to, genome structure, replication strategy and virion composition.

The Baltimore classification system was proposed in 1971 to classify viruses into seven categories, focused on genome structure and replication strategy (Baltimore, 1971; Koonin et al., 2021). Briefly, viruses can be grouped by DNA or RNA, and negative or positive polarity, which impacts the hijacking mechanism utilised for replication (Table 1-1).

	Genome structure	Features and examples
BC1	Double stranded DNA	Replicate exactly as host proteins e.g., Herpes
BC2	Single stranded DNA	Requires generation of a dsDNA intermediate for replication e.g., parvoviruses
BC3	Double stranded RNA	Require transcription prior to replication for mRNA generation e.g., Reovirus
BC4	(+) RNA	Translated directly as if host mRNA e.g., Hepatitis C
BC5	(-) RNA	Require transcription prior to translation e.g., Influenza
BC6	RT-RNA	Require a DNA intermediate for package of positive sense RNA which can then be integrated into the host genome e.g., HIV
BC7	RT-dsDNA	Require an RT step for generation of an RNA intermediate

Table 1-1. The Baltimore classification system.

1.2 DNA and RNA viruses

All viruses exist in heterogeneous, fluctuating environments and constantly evolve to remain fit to inhabit them. Mutation rate refers to the replication error rate, whereas substitution rate describes the rate at which mutations are fixed within the genomic population. RNA viruses typically exhibit accelerated mutation rates compared to DNA viruses (10^{-4}-10^{-6} to 10^{-6}-10^{-8} substitutions per nucleotide site per cell infection (s/n/c)) as a result of the polymerase used in replication and the reduce genome size (Peck & Lauring, 2018).

Different polymerase enzymes are employed for RNA and DNA replication with differing fidelity, the incorporation of the correct nucleotide within the nascent RNA molecule. DNA polymerases typically have a proof-reading exonuclease function which removes noncognate nucleotides and reduces the error rate during replication (Choi, 2012). RNA polymerases typically lack this function, and the nascent RNA strand is therefore more likely to contain mis-matched bases, creating variation within the nucleic acid population. Interestingly, Coronaviruses

are the only RNA virus to express proteins exerting an exonuclease function (S. C. Baker & Lai, 1990; Deval & Gurard-Levin, 2022; Robson *et al*., 2020).

A higher mutation rate has enforced the reduction of genome size within the RNA viruses to prevent accumulation of fatal mutations (Belshaw *et al*., 2007), whereas DNA viruses have larger genomes as transcripts encoding many more functional proteins can be replicated without fatal error. This comparatively rapid evolutionary rate, estimated between $1x10^{-3}$ – $1x10^{-6}$ substitutions per site per year, is much faster than mutation rates estimated for bacteria, $1x10^{-6}$ – $1x10^{-9}$ (Domingo *et al*., 2021; Nachman & Crowell, 2000).

1.3 Diversity within an RNA virus genomic population

The erroneous nature of the RNA polymerase enzyme generates a pool of genomic variants upon which evolutionary selection can act, the minor variant population (K. Singh et al., 2023). Another term is commonly used interchangeably to describe this minority variant population, the quasispecies (Eigen, 1993). The quasispecies term is specifically referring to the minor variation as a cloud of diverse variants linked genetically by mutation, which cooperate on a functional level to contribute to the population's characteristics (Lauring & Andino, 2010). In either case, within an infected cell, the erroneous replication process generates a heterogeneous viral population with mutations across the genome present at low levels within the overall genomic population (Sanjuán & Domingo-Calap, 2021; Vignuzzi *et al*., 2006). A large pool of genomic variants to select from results in a viral population which can rapidly evolve in the event of fluctuating selection pressures e.g., host-switching (de Vienne *et al*.,

2013; Sharp & Hahn, 2010) and immune activation (Horton *et al*., 2002). This presents challenges for the development of both prophylactic and therapeutic preventative measures as they exert selection pressure upon the diverse underlying viral population, favouring the proliferation of resistant genotypes (Gerrish & García-Lerma, 2003; K. Singh *et al*., 2023).

1.4 Recombination and Reassortment

In addition to variation because of erroneous replication, co-infection of one host cell by multiple viral genotypes can provide opportunity for recombination and/or reassortment. This often results in accelerated evolutionary change as the resulting genome is a mosaic of the two parental lineages (Pérez-Losada *et al*., 2015).

There are different types of recombination which generate different combinations of parental nucleic acid. Homologous recombination refers to recombination at the same site in parental strands providing an in-frame recombination event. Non-homologous recombination occurs at different sites in the two parental lineages, often giving rise to out of frame, mis-sense nascent RNA sequence (Galli & Bukh, 2014).

Through reassortment segmented viruses can exchange entire segments between co-infecting virions, resulting in new genomes with significantly different segment organisation (Cross & Fields, 1976; Vijaykrishna *et al*., 2015). Reassortment is best studied with regards to the generation of influenza virus variants (Gerber *et al*., 2014; Richman *et al*., 1977), but is also reported commonly for *Bunyaviruses*, such as Crimean-Congo haemorrhagic fever virus

(CCHFv) (Briese *et al*., 2013; Hewson *et al*., 2004). In evolutionary terms, reassorted viruses are comprised of genomic segments with disparate evolutionary timelines. Thus, the majority of reassortment detection methods are constructed upon the hypothesis that phylogenetic incongruence supported by high bootstrap support between phylogenies inferred from different segments of one virus, are often resulting from reassortment (Svinti *et al*., 2013).

1.5 Emerging infectious disease and zoonotic spillover

An infectious disease can be defined as illness resulting from a pathogen or its toxic produce caused by transmission from an infected person, animal or contaminated object to a susceptible host (van Seventer & Hochberg, 2017). Emerging infectious diseases (EID) are defined as those newly appearing in a population or rapidly becoming more prevalent in a previously unaffected area (McArthur, 2019).

Many EIDs result from zoonotic spillover. Zoonotic spillover is the transmission of a pathogen from a reservoir species to a human host, but does not include transmission by arthropod vectors e.g., ticks, mosquitos (Plowright *et al*., 2017). Fundamental to zoonotic spillover are the reservoirs within which agents circulate prior to spilling into the human population. A comprehensive literature review identified 61% of etiological agents causing illness in humans as zoonotic, compared to just 3% which utilise the human population as a reservoir (Taylor *et al*., 2001). Thus, targeting the interactions between the reservoir and human population and understanding how spillover events occur could significantly reduce the public health burden of infectious disease. The term “One Health, One

World" was first termed after the initial SARS-CoV outbreak, where the Manhattan Principles were generated with the aim of elucidating the mutual relationship between human and animal health to reduce the threat presented by spillover events (Mackenzie & Jeggo, 2019).

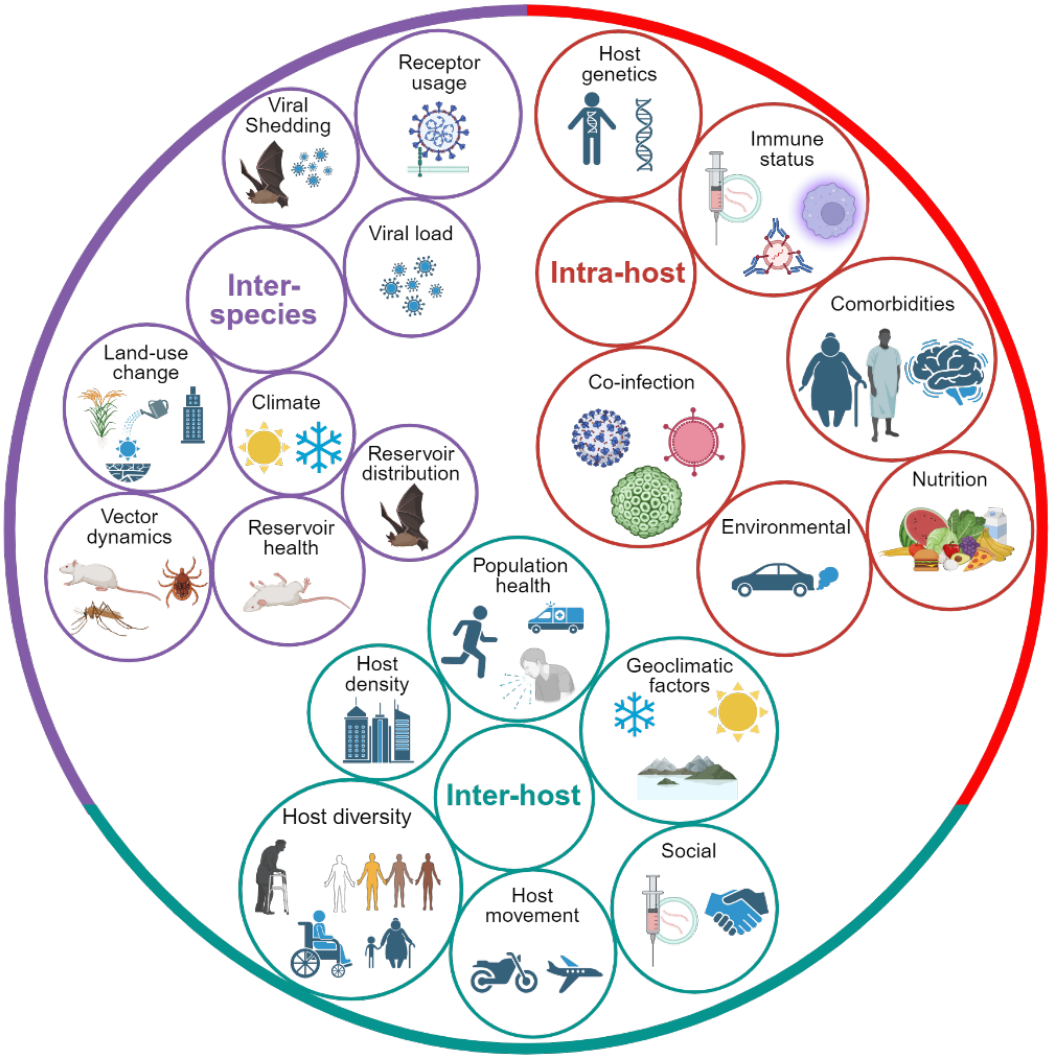


Figure 1-1. Zoonotic spillover.

A schematic to demonstrate factors influencing the outcome of a spillover event. Created with BioRender.com.

Multiple factors influence the resulting epidemiology of a spillover event (Figure 1-1). Assuming factors have aligned sufficiently to facilitate a spillover, there are two main outcomes: transmission throughout an immune naïve population and

causing infection (i.e. becoming endemic or pandemic), or dead end spillover where environmental or pathogen-related factors prevent maintenance in the population (Ellwanger & Chies, 2021).

One critical determinant is the type of pathogen, and specifically for viral spillover, the viral species involved. The route of transmission employed by the virus particle influences its epidemiological dynamics within the population. For instance, bloodborne viruses like CCHFv and Ebola virus (EBOV) typically exhibit comparative lower dissemination rates compared to respiratory viruses such as influenza virus. This discrepancy arises from the need for direct contact for transmission, which is characteristic of bloodborne viruses, contrasting with rapid transmission facilitated by respiratory droplet transmission (Burrell *et al*., 2017).

Other factors can be categorised into inter-species, inter-host, and intra-host factors (Figure 1-1). Inter-species factors refer to the interaction between the reservoir population and other populations which can be targeted for spillover prevention. For example, the introduction of rodent control measures in geographical locations with Lassa Virus (LASV) circulation reduces interactions between rodents and humans (Mari Saez *et al*., 2018). Inter-host refers to factors influencing interactions between individual members of a single population, either the reservoir or others. For example a high host density such as living in a metropolis such as New York City, increases potential transmission by proximity to susceptible hosts within the population (Sy *et al*., 2021). Intra-host refers to factors influencing the outcome of infection from a single individual. For example comorbidities such as obesity have been associated with negative outcomes upon infection with a virus (Badawi *et al*., 2018; Marimoutou *et al*., 2003; C. D. Russell *et al*., 2023).

1.6 Reverse zoonosis

Zoonotic spillover can also occur in the reverse direction, from a human into the known reservoir or naïve species (Ellwanger & Chies, 2021). Reverse spillover provides opportunity for rapid viral adaptation within the new host, allowing for the generation of pools harbouring heterogeneous viral genotypes. These genotypes can then spill back into the human or other population (Figure 1-2).

Influenza A virus is one of the major global pandemic threats, with the swine population acting as a mixing vessel for different variants and raising concerns of reassortment between avian and mammalian influenza strains. Most attention is directed towards spillover into humans after influenza virus mixing and reassortment within an intermediate swine host. However, there is significant evidence which supports the repeated introduction of avian influenza virus from humans into the swine population, (Cappuccio *et al*., 2011; Karasin *et al*., 2000; Nelson & Vincent, 2015; Ottis *et al*., 1982).

SARS-CoV-2 represents a unique risk for reverse zoonoses as a large majority of the global population has been infected between late 2019 and the present day. There have been many instances of reverse zoonosis of SARS-CoV-2 into animal populations, globally. Many reports have been into companion animals because of their proximity to humans infected with high viral titres with the first instance reported in a Pomeranian from Hong Kong (Sit *et al*., 2020). Since 2020, SARS-CoV-2 infections in animals resulting from transmission from humans has been reported in Europe, North and South America, Asia and Africa (Pramod *et al*., 2021). The importance of reverse zoonotic interactions were highlighted upon detection of extensive SARS-CoV-2 infection at mink farms in the Netherlands

with onward transmissions spilling back into the human population (Oude Munnink *et al*., 2020, 2021).

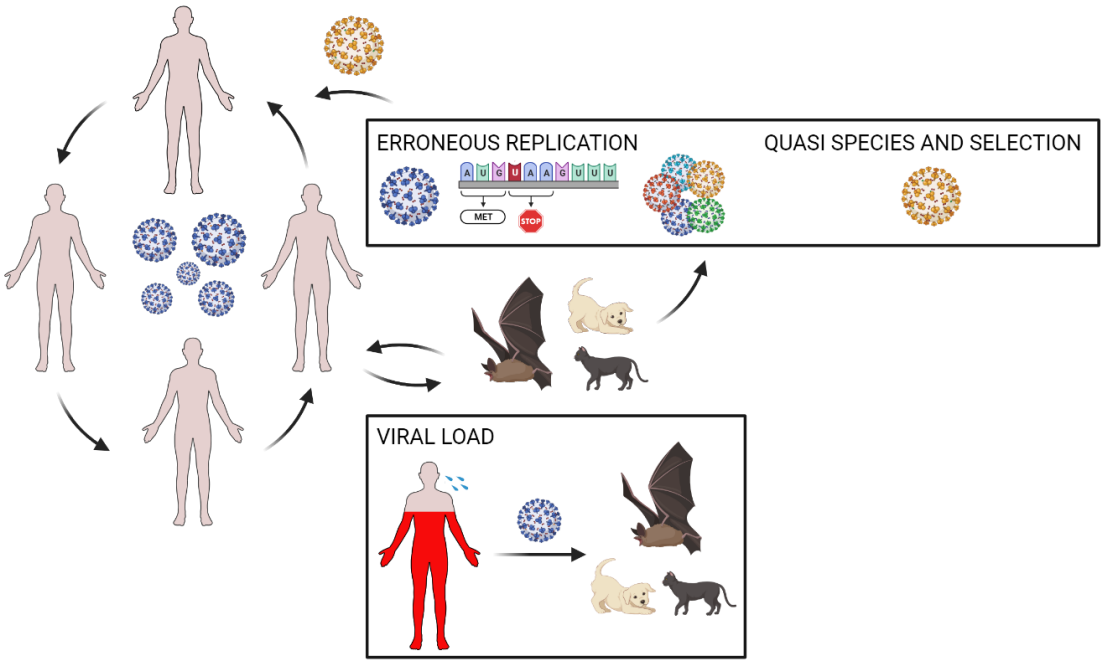


Figure 1-2. Reverse zoonosis.

A schematic to demonstrate a potential route of reverse zoonosis. Starting with a circulating virus in the human population. A human with high viral load encounters a susceptible animal species who then becomes infected (inter-species). Viral replication in a previously naïve host facilitates uncontrolled viral replication while the innate immune system mounts its initial response. The naïve immune system also exerts a different selection pressure upon this large viral population within the naïve animal and thus selects for different variants from a large pool of genotypes. The selected genotype can then undergo transmission to other animals of the same species, or back into the human population, possibly with altered pathogenicity. Created with BioRender.com.

1.7 A history of pandemics

Throughout history, EIDs have exerted a significant public health burden globally. The first recorded viral pandemic, “Russian flu”, was reported between 1889-94. This is termed “flu” but considering the transmissibility and low case fatality rates

observed during the COVID-19 pandemic, it has been hypothesised that a coronavirus was the etiological agent (Berche, 2022; Parsons, 1891).

Since the start of the 20th century, there have been three influenza pandemics (Al Hajjar & McIntosh, 2010; Cockburn *et al*., 1969; Honigsbaum, 2020; Kilbourne, 2006; Martini *et al*., 2019; Schäffr *et al*., 1993; Taubenberger & Morens, 2006), two coronavirus pandemics, SARS-CoV, SARS-CoV-2 (Stadler *et al*., 2003; Zhou *et al*., 2020) and an ongoing, *retroviridae* pandemic (HIV).

In 150 years, the global population have recorded and overcome/are still tackling six different viral pandemics all with likely animal origin. Given the increased connectivity in the global population in the 21st century, preparedness is a necessity with regards to containing further pandemics and minimising the potential economic and health burdens.

1.8 Bats as a reservoir of viral pathogens

Bats (order Chiroptera) are a monophyletic mammalian clade (Lei & Dong. 2016) of >1400 species (Fenton *et al*. 2015) which diversified from other mammalian lineages an estimated 65 million years ago (MYA) (O'Leary *et al*., 2013). This comprises >22% of extant mammalian species globally, superseded only by Rodentia (Burgin *et al*., 2018).

The divergence of Chiroptera into two suborders, Yinpterochiroptera and Yangochiroptera (previously Mega- and Microchiroptera), is hypothesised to have occurred >50 million years ago (MYA). Previously, higher order chiropteran taxonomic classification was based on echolocation ability with the Microchiroptera suborder containing true echolocators. Molecular data has

provided the basis for reclassification into Yinpterochiroptera and Yangochiroptera. (Jones & Teeling, 2006; Tsagkogeorga *et al*., 2013). Yinpterochiroptera is inclusive of megabats and five microbat families previous classified as Microchiroptera: *Rhinopomatidae*, *Rhinolophidae*, *Hipposdierae*, *Craseonycteridae* and *megadermatidae* (Tsagkogeorga *et al*., 2013) thus indicating some echolocating bat families are more evolutionary related to old world fruit bats than other echolocating members of the Yangochiroptera (Tsagkogeorga *et al*., 2013).

Bats play key roles for human populations such as ecosystem management. Roles such as predating on many agricultural pests having both production economic consequences (Boyles *et al*., 2011), and pollinating a variety of plants globally via seed dispersal (Fujita, 1991; Kasso & Balakrishnan, 2013). As a result of the link between bat populations and high consequence viruses, they can be targeted when aiming for reduced zoonotic spillover, disregarding these ecosystem roles.

However, bats can harbour viruses with pandemic potential as demonstrated by the SARS-CoV and SARS-CoV-2 and, as mentioned, an association with other high consequence viruses such as EBOV, Marburg, Nipah and Hendra viruses (Calisher *et al*., 2006; Epstein *et al*., 2020; Koch *et al*., 2020; O'Shea *et al*., 2014). Despite harbouring high consequence viral pathogens, there are limited reports of fatalities resulting from viral infection. The basis for the immunity to viral infection demonstrated by bat populations has been and is subject of continued research efforts in different fields e.g., host evolution, immunology, and ecology.

Evolutionarily, bats diverged from other mammalian lineages an estimated 65 MYA around the time of the Cretaceous-Paleogene (K-Pg) extinction event. The subsequent period was characterised by low species richness preventing inter-species viral dissemination and providing opportunity for a unique and extended temporal scale for virus-host evolution (P. Gupta *et al*., 2021).

Bats are also the sole mammalian lineage capable of extended flight and can cover large distances, providing means of transmission. This provides the basis for which viruses with mammalian tropisms can cover large distances, rapidly. This unique ability is likely the key factor in driving the observation of more zoonotic viruses in bats compared to rodents, despite greater species numbers for Rodentia (Luis *et al*., 2013). Bats have also undergone diverse speciation, providing a wide range of transmission opportunities (Guy *et al*., 2020) with only small numbers of extant species under surveillance (Phelps *et al*., 2019) resulting from economic drivers, inadequate provision globally, lack of academic motivation or an absence of global cooperation.

Multispecies roosts provide opportunity for widespread inter-species transmission. Simulated models predicted successful roost circulation resulted from uncontrolled heterogeneous species mixing. This resulted in more frequent spillover events from infected roosts into susceptible populations (Lunn *et al*., 2021). These environments provide opportunity for episodic accelerations in viral evolution associated with inter- and intra-species transmission events within a single roost. The resulting viral genomic population will harbour a diverse pangenome for the virus with the potential for spillover.

The immunological status and innate immune responses provide molecular mechanisms which promote resistance or tolerance to infection (Schneider & Ayres. 2008). The immune system of bats is of particular interest because of the number of human pathogens associated with circulation in bat populations. Early definitions of the bat innate and adaptive immune systems identified comparable immune cell populations to other mammalian species (M. L. Baker *et al*., 2013) suggesting the innate response is a main driver for the viral immunity observed in bats. The baseline expression of type 1 interferons (IFNα) in *Pteropus alecto* is increased compared to human type 1 interferon expression despite a reduced number of interferon genes (Zhou *et al*., 2016). Other observations such as different patterns of expression of IRF7 and enhanced IRF3-mediated antiviral responses may also impact susceptibility to viral infection (Irving *et al*., 2021).

1.9 Metagenomics for viral detection and characterisation

Metagenomic methods aim to capture an accurate representation of all nucleic acid contained within a sample. Advances in next generation sequencing (NGS) technologies and increases in computational power paired with metagenomic amplification methods have allowed researchers to targ*et all* potential pathogens. The first application of metagenomics to virus discovery was in 2002 (Breitbart *et al*., 2002), aiming to characterise the virome of marine samples. The nucleic acid was sheared and amplified following attachment of dsDNA linkers to fragment ends, termed linker amplified shotgun libraries. The same approach was then used to discover 1200 novel viral genotypes from human faeces, demonstrating the power of metagenomics for discovery of the human virome (Breitbart *et al*., 2003). Since 2002, the number of studies within which metagenomics were used

has increased massively, and the scope of application increased equally (Mokili *et al*., 2012).

There are many methodologies available for metagenomics. The choice of which is dependent on several factors. Two key considerations are the genome type

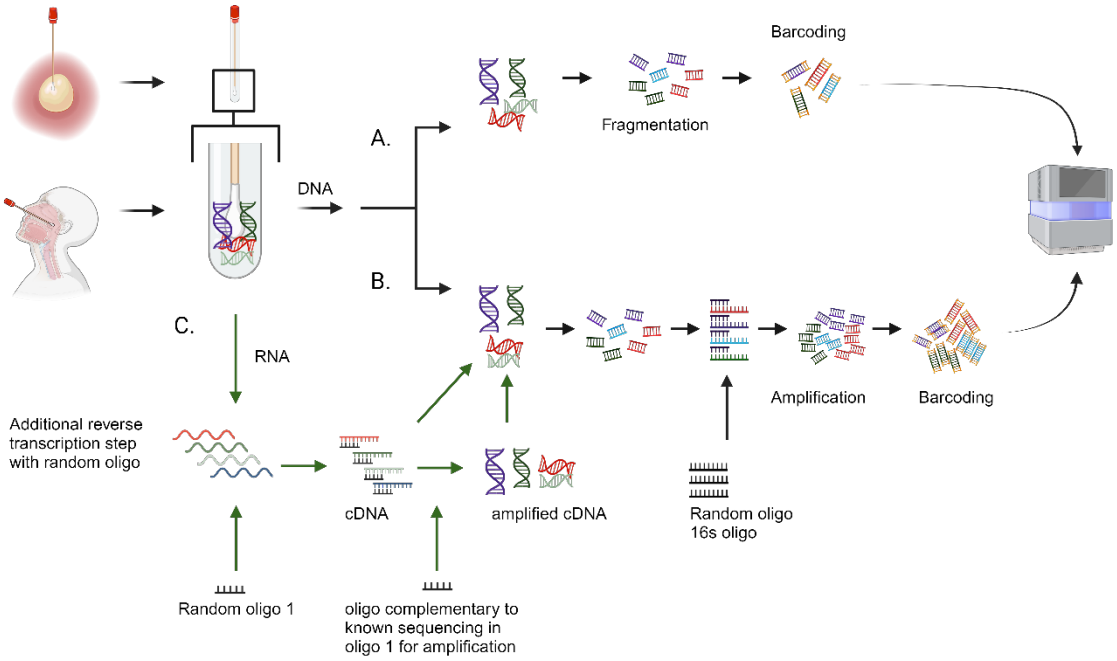


Figure 1-3. mNGS pathways.

A schematic demonstrating the three main avenues of processing nucleic acids for metagenomic NGS. This does not include prior steps of nucleic acid extraction and enrichment steps such as DNase treatment. The pathway labelled A represent the traditional shotgun sequencing approach to metagenomics. There is no amplification of target nucleic acid and includes fragmentation and barcoding steps before sequencing. The pathway A is also applicable to DNA pathogens where amplification is not desired and nucleic acid in the sample is barcoded and then sequenced. The pathway labelled B represents the metagenomic pipeline for DNA based pathogens when amplification of specific pan bacterial genes (16s) is desired. The pathway C represents the metagenomic pipeline for RNA viruses. There is an addition of a reverse transcription step with a random oligo with a known tag. This known tag is targeted in the addition of a second oligo if amplification is desired. cDNA from this pipeline can be used for mNGS either unamplified or amplified. Created with BioRender.com.

(RNA vs DNA) and whether amplification of the nucleic acid in the sample is desirable (Figure 1-3).

Direct detection of viruses can be achieved by either molecular or immunoassay-based techniques without the need for propagation. Molecular assays focus on detection, amplification and in some cases, quantification of viral nucleic acid e.g., qRT-PCR. Immuno-based assays include traditional serological techniques such as Enzyme-linked Immunosorbent Assays (ELISA) and provide viral detection during or after viral infection (Cassedy *et al*., 2021) (Table 1-2). The use of mNGS can offset some of the disadvantages of other viral detection methodologies in samples with a high viral load (Table 1-3).

Method	Type
Quantitative Real Time PCR	Molecular
Loop Mediation Isothermal Amplification	Molecular
Recombinase Polymerase Amplification	Molecular
Helicase-dependent Amplification	Molecular
Rolling Circle Amplification	Molecular
CRISPR-Cas9-Based Nucleic Acid Detection	Molecular
Enzyme-Linked Immunosorbent Assays (ELISA)	Immunoassay
Blotting techniques	Immunoassay
Lateral Flow Immunoassay	Immunoassay

Table 1-2. Assays for pathogen detection.

A boom in RNA virus discovery over recent years has resulted from improvements in NGS and surveillance in at risk reservoir species for potential zoonoses under the one health approach (Ergünay, 2021). The advancement of metagenomic methodologies have undoubtedly contributed to this massive expansion of knowledge around the global virome, especially the vertebrate RNA

virome (Edgar *et al*., 2022; Kawasaki *et al*., 2021; Nakagawa *et al*., 2023; Neri *et al*., 2022; Shi *et al*., 2023).

Advantages	Disadvantages
Detection of unknown pathogens	low viral load samples can be incorrectly negative because of reduced sensitivity
Removes pre-assay work e.g., designing of primers	High computational demands for data processing
Detection of co-infections	Background noise (host reads)
Provides details of minor variation present in quasispecies when paired with deep sequencing	Requires development and knowledge of tools for assembly of reads from metagenomic amplification
Application to any sequencing technology	Currently limited read length so perhaps not utilising the full power of long read sequencing technologies
An internal control can be spiked at a known concentration to provide quantification	Cost
	Cross-contamination

Table 1-3. mNGS for pathogen detection.

1.10 Molecular phylogenetics

The rapid improvement of sequencing technologies and storage capabilities for genomic data has facilitated the development of molecular phylogenetic methods as a crucial tool for evolutionary biology. Phylogenetic methods are now an important tool in a large number of biologically relevant analyses e.g., defining the history of different populations (Bowcock *et al*., 1994), genome comparisons (Sicheritz-Pontén & Andersson, 2001) and reconstruction of ancestral genomes (Dobzhansky & Sturtevant, 1938; Joy *et al*., 2016; Z. Yang & Rannala, 2012).

1.10.1 Constructing phylogenetic trees

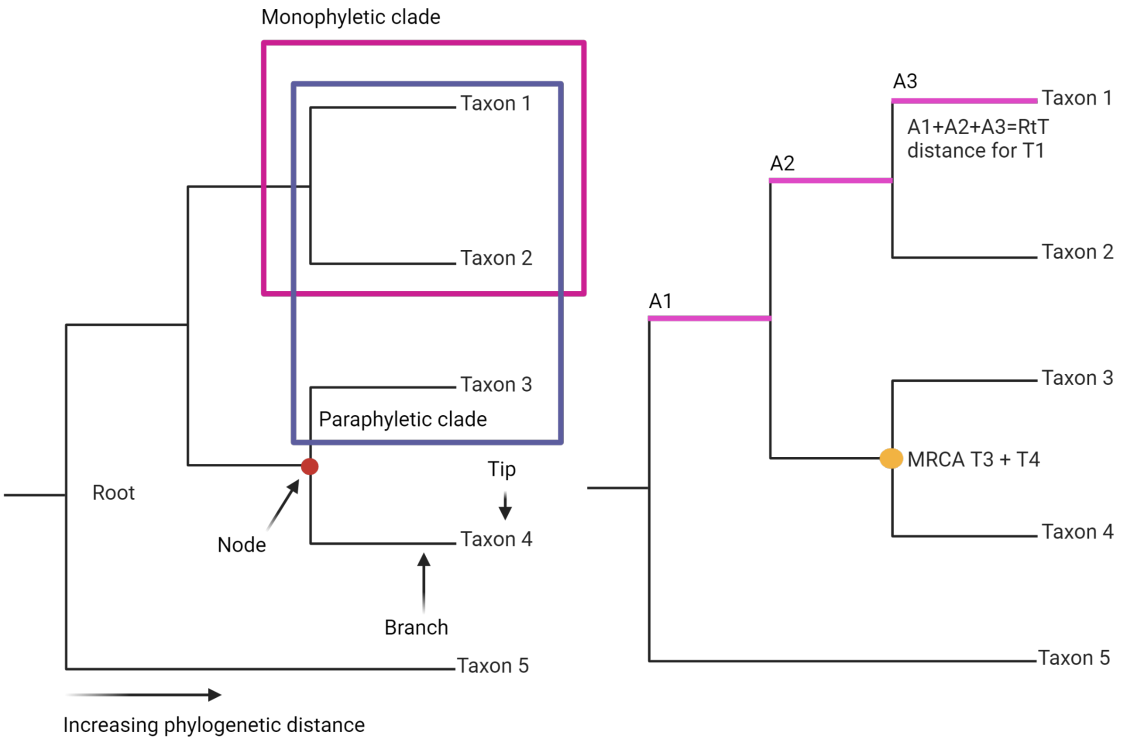


Figure 1-4. A hypothetical phylogenetic tree.

A schematic introducing key terms and concepts for molecular phylogenetics. Created with BioRender.com.

A phylogenetic tree is a method of representing relatedness between different organisms from molecular or other datatypes (Figure 1-4). A phylogenetic tree is a graph containing nodes, branches, and tips. The tips represent the data from which the tree is generated from. The branches depict changes over time (phylogenetic distance from the root) and depending on the methodology used to infer the phylogeny, can refer to genetic distance or time. Nodes represent coalescence events where the descendant tips coalesce into one ancestral state, termed the most recent common ancestor (MRCA). Tips can be grouped into either mono- or paraphyletic groups. Monophyletic groups contain clusters of tips

where the MRCA to all descending tips is included in the clade, whereas paraphyly refers to a group in absence of the MRCA node.

1.10.1.1 From alignment to topology

In molecular phylogenetics, trees and their topology are inferred from a set of nucleotide or amino acid sequences termed an alignment. The sequences in the alignment can be compared by either distance or character-based methods (Figure 1-5). In distance-based methods, the entire sequence is treated as "one" object and compared in this manner to other sequences in the alignment to group most similar sequences (e.g., neighbour joining). In character-based methods, sequences are compared in a character-wise approach (e.g., maximum parsimony, maximum likelihood, Bayesian inference). Sequence comparison aims to identify Single Nucleotide Polymorphisms (SNPs), regions of similarities and common insertions/deletions (indels), which are used to cluster the most closely related sequences.

Phylogenetic inference under character- and likelihood-based methodologies require a underlying substitution model (Kapli *et al*., 2020). The substitution model is used for modelling transitional probabilities between nucleotides in the alignment as a representation of the evolutionary process under which the alignment has been generated in nature (how all the sequences have evolved with regards to one another) (Arenas, 2015).

Many different substitution models have been proposed to describe the evolutionary process, allowing varying degrees of flexibility for parameters such as nucleotide exchange rates, the proportion of invariant sites, nucleotide

frequency and rate heterogeneity within a single sequence (Arenas, 2015; Z. Yang & Rannala, 2012).

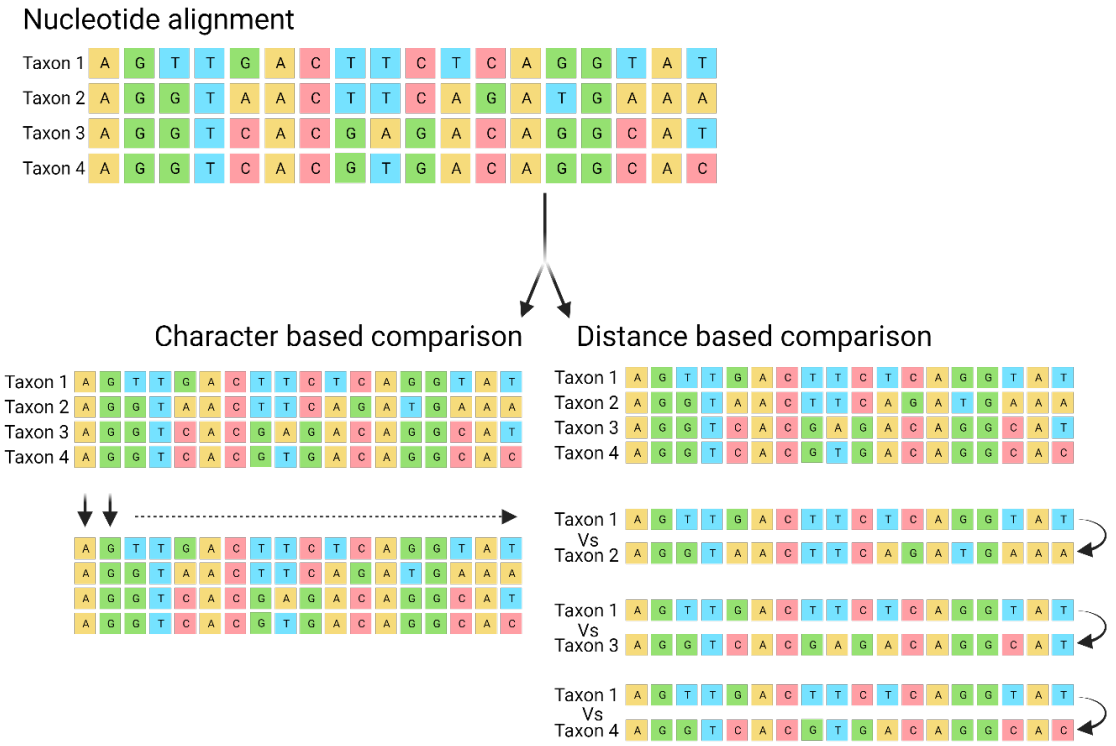


Figure 1-5. The two phylogenetic methods.

A schematic showing the two alignment-based methods at the start of phylogeny construction. Created with BioRender.com.

The alignment and substitution model are used to generate a distance matrix which is forms the basis of the phylogenetic topology (Agarwala *et al.*, 1998; Saitou & Nei, 1986). The distance matrix contains a numerical value which represents the genetic distance between all sequences, which is then represented by branch lengths in the topology.

Character-based methodologies are fundamental to maximum parsimony, maximum likelihood, and Bayesian inference. The most basic is parsimony-based inference as the criteria aims to reduce the total character state changes

required to account for the variation in the alignment, and this forms the phylogenetic topology (Edwards & SForza, 1963; Steel & Penny, 2000). This is not based on an underlying model and thus does not account for any parameters included in maximum likelihood and Bayesian inference.

Under a maximum-likelihood framework the calculated likelihood is the optimal likelihood of the DNA sequences being observed as a function of the unknown topology and branch lengths. Thus, the estimated branch lengths and substitution rate support the observed diversity in the nucleotide alignment at the optimal likelihood (Dhar & Minin, 2016).

The nucleotide substitution model is important in this process as the parameters provided are utilised during the likelihood calculation to constrain the range sampled for the final likelihood. However, early comparisons of the impact of nucleotide substitution models on phylogenetic estimation suggested topology may be robust despite incorrect model selection (Z. Yang *et al*., 1994) but can reduce accuracy of ancestrally reconstructed protein sequences (Del Amparo & Arenas, 2022).

1.10.2 Estimating evolutionary rates

1.10.2.1 Evolutionary models of increasing complexity capture biological systems with improved accuracy

The simplest nucleotide substitution model was proposed by Jukes and Cantor proposing uniform nucleotide exchange rates (Jukes & Cantor, 1968). The JC69 model represents a formal model of the hypothesis presented by Zuckerkandl and Pauling, who hypothesised a constant evolutionary rate, termed the strict

molecular clock (Dickerson, 1971; Kimura, 1965; Margoliash, 1963; Zuckerkandl & Pauling, 1962, 1965a, 1965b).

The JC69 model was extended by Kimura who proposed the addition of different exchange rates for transitions and transversions resulting from the chemical composition of purines and pyrimidine nucleotides (Kimura, 1980). Felsenstein then proposed that the frequency of nucleotides is variable for all four nucleotides and thus rates are frequency dependant (Felsenstein, 1981). The general time reversible model (GTR) allows each nucleotide to have its own transition probability independently and in accordance with its frequency (Lanave *et al*., 1984) providing researchers with high levels of flexibility.

1.10.2.2 Accounting for rate variation between lineages, genes, and sites

A limitation of early evolutionary models was the inability to vary evolutionary rates between lineages, genes, and sites within a gene. Inherently, at all scales, rates demonstrate heterogeneity. Viral lineages evolve at different rates because of ecological niches, geographic regions, or hosts (Spielman *et al*., 2019). Genes can evolve at different rates because of different functional impacts on viral fitness (Narain & Chakraborty, 1987). Sites within a gene can evolve at different rates because of domain specific functions within a protein e.g., the active site of an enzyme would be constrained and evolve more slowly than functionally irrelevant regions (Figure 1-6) (Philippe et al., 2003). As substitution models have increased in complexity, a gamma-distribution (Γ) is used to allowed transitional rate variation between sites of the same gene or genome. This variation is represented by a Γ-distribution to model a range of evolutionary rates (Figure 1-6) (Z. Yang, 1993). The shape parameter, α, provides the numerical value

determining the shape of the Γ-distribution. A high α suggests little variation in rate across the gene, whereas a low α suggests the inverse. Some nucleotide substitution models bin rates from across the gene into different categories, termed a discrete Γ-rate heterogeneity modelling and the number of categories can often be manipulated by the investigator.

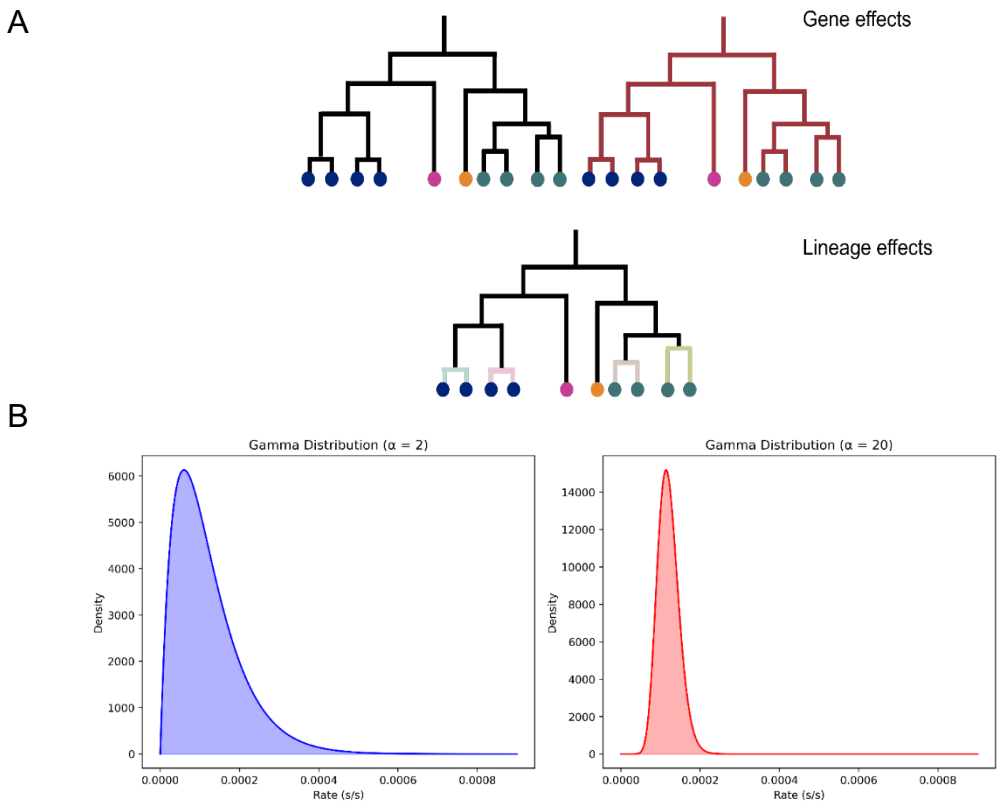


Figure 1-6. Types of heterogeneity within genomic sequences.

Rate variation across genes can be termed “gene effects” and rate variation across lineages can be termed “lineage effects” (A). The Γ-distribution is a continuous probability distribution which when implemented under evolutionary models permits different regions across a gene or genome to exhibit differences in evolutionary rates (B).

1.10.3 Relaxing the molecular clock to account for variation in evolutionary rates caused by many biological factors

The molecular clock was initially hypothesised to represent a strict evolutionary rate across all organisms (Zuckerkandl & Pauling, 1962, 1965a) which was met with resistance as others hypothesised rates of evolution would vary widely over time with changes in selection pressures e.g. environmental conditions (Bromham *et al*., 2018; Bromham & Penny, 2003). It is now known that molecular evolution fluctuates across multiple planes, time, genomic regions both in close and far proximity and between lineages when comparing genomes and/or genes. To account for this variation of rates across different planes, models now exist for relaxation of the assumption of rate constancy.

For the molecular clock to be utilised as a tool for defining evolutionary timescales and divergence dating, calibrations are required to place the observed diversity within a temporal scale (Duchêne *et al*., 2014). Within viral phylogenetics where samples tend to be collected and sequenced in relatively modern times (since 1900), this tends to be a collection date. Other calibration methods include the use of fossils, geological evidence e.g. the formation of large geographic structures (mountain ranges, islands), climatic evidence and previously published results (Ho *et al*., 2015).

There are two main categories under which modern models for relaxing the molecular clock are placed into: relaxed clock models where the rate can change at any node or branch across the phylogeny, and local clock models, where clades are assumed to possess local clocks which are conserved among closely related sequences (Figure 1-7).

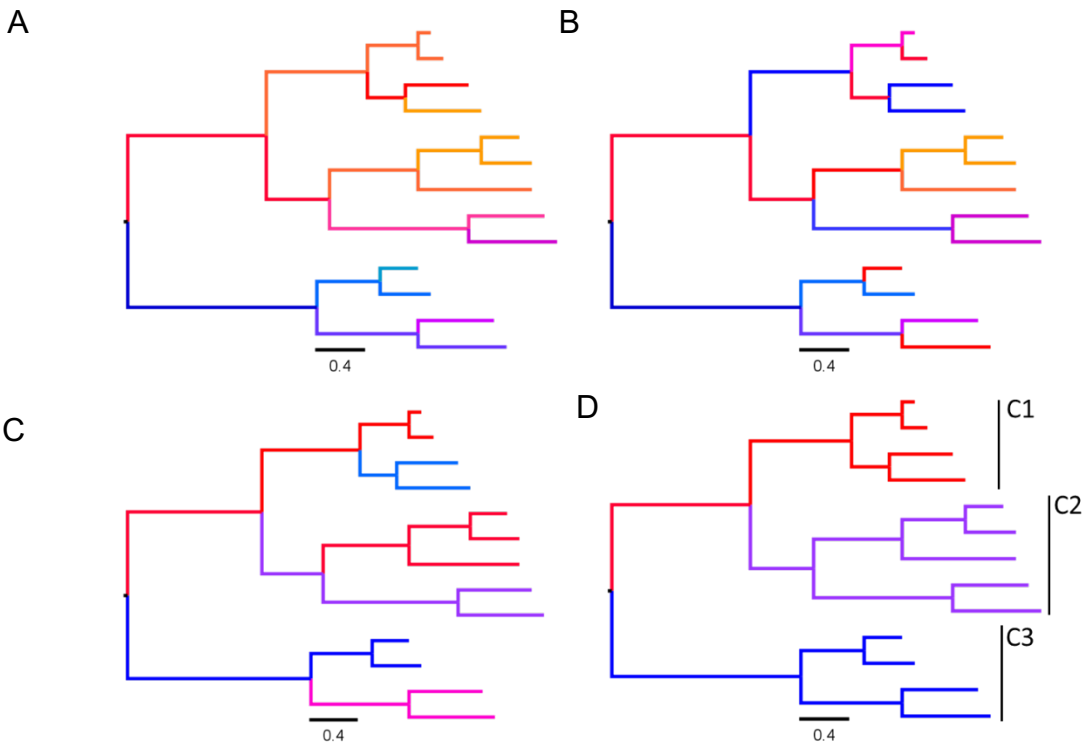


Figure 1-7. Methods for relaxing the clock.

A schematic demonstrating four commonly employed options which can be used for evolutionary rate estimation and divergence time estimating by relaxing the molecular clock. Rate is depicted as colour along the branch as a red-blue scale (red, fast) A strict clock is not shown, but one rate is applied across the tree. In a relaxed autocorrelated model (A), the descendant branches are limited in the rate variation which can occur. For a relaxed uncorrelated model, the descendant rate is not reliant on parental branches. Other options include a random local clock (C) and a fixed local clock (D). Local clock models look to identify clades which share evolutionary rates. In a fixed local clock model, the taxa ids for the clades are provided e.g., Clade 1 (C1), clade 2 (C2), and clade 3 (C3).

Many different models have been proposed for modelling the process of evolution across a phylogeny (Aris-Brosou & Yang, 2003; Drummon *et al*., 2003; Huelsenbeck *et al*., 2000; Kishino *et al*., 2001; Sanderson, 1997, 2002). These fall into two broad categories: nonparametric models which do not assume statistical distributions for parameters such as evolutionary rate, and parametric models which assume statistical distributions as a method of allowing uncertainty across different parameter estimates for example, a normal distribution (Figure

1-8). A complete explanation of these models and the statistical strengths and weaknesses fall outside the scope of this thesis but the benefits of allowing uncertainty in evolutionary models will be discussed.

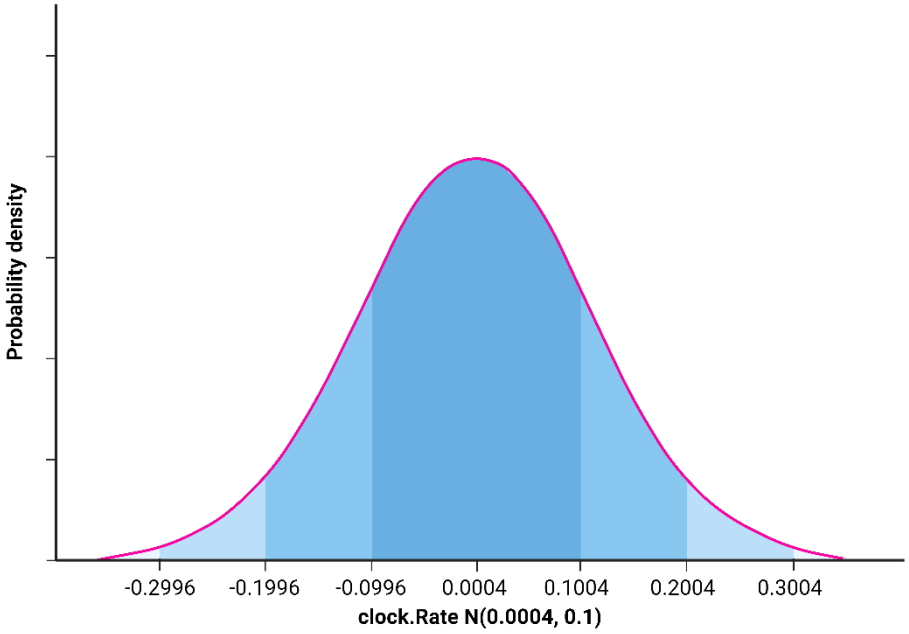


Figure 1-8. A normally distributed prior for the evolutionary rate.

The mean rate as is expected to be around 10-3 -10-4and is provided a standard deviation of 0.1. This model would allow rate estimates using the information stored within the nucleotide alignment to be within the bounds of the statistical distribution. Created with BioRender.com.

1.10.4 How does phylogenetic tree construction under a Bayesian framework add to this process?

In non-Bayesian phylogenetic tree construction, the topology is inferred using the evolutionary data stored within an alignment and the branching pattern represents relatedness based on the genetic distance stored in the matrix. These phylogenies can be useful for a variety of tasks without the complex and computationally extensive Bayesian process, including but not limited to testing

phylogenetic hypothesises e.g., testing monophyly, mapping discrete or continuous traits to phylogenetic topology and comparing genomes.

However, phylogenetic inference under a Bayesian framework offers different avenues of investigation and thus can be used to answer different evolutionary questions. It allows the researcher to provide prior knowledge about the biological system being modelled to improve estimates. For example, for ancient phylogenies such as that of Simian Immunodeficiency virus (SIV), there are certain historical events which can be used to provide calibration points in the phylogeny. The formation of Bioko Island isolated a population of the Bioko species from the mainland. Enforcing this separation event pushes the age of the SIV most recent common ancestor (MRCA) back (Worobey *et al*., 2010). Calibration dates can also be in the form of previous publications characterising features of the virus being studied e.g., evolutionary rate, which can then be used.

Under a Bayesian statistical framework, the user can allow uncertainty around specific parameters by providing a statistical distribution around suitable values for the biological system. The allocation of a statistical distribution facilitates handling parameters such as coalescence points (MRCA) and evolutionary rates across a phylogeny and can provide a range of values within a 95% confidence interval (CI) (Nascimento *et al*., 2017).

The features provided under a Bayesian framework with the ability to relax the molecular clock for accurate variation of rate across time provides the ability to generate time-scaled phylogenies using tools such as Bayesian Evolution Analysis by Samling Trees (BEAST) (Drummond & Rambaut, 2007).

1.11 Phylodynamics, phylogeography and tracking RNA viruses

Phylodynamics refers to the field investigating the behaviour of pathogens which cause infectious diseases from an evolutionary (pathogen) and ecological (epidemiological) perspective (Grenfell *et al*., 2004). As a rule, epidemic dynamics are reliant on a continuous source of susceptible individuals for ongoing infection which provides the basis for SIR model (Featherstone *et al*., 2022). The temporal scale over which the epidemic occurs is an important consideration when considering pathogen dynamics. Epidemics *tend* to be limited in duration and RNA viruses are excellent candidates because of the rapid evolutionary rate. This rapid evolution leaves a molecular footprint across a measurable temporal scale which represents the microevolutionary processes, which can be used to correlate specific genotypes to points in time (Drummond *et al*., 2003; Featherstone *et al*., 2022).

This genotypic variation has historically been qualitatively observed spatially across populations and this variation could be harnessed for hypothesises regarding gene flow and movement of populations (Templeton, 1998).

Phylogeographic methodologies now aid the public health response by uncovering epidemiological transmission and spread of pathogens across geographic regions (Faria *et al*., 2011). Phylogenetic reconstructions of these viral outbreaks can be linked to spatial processes such as viral diffusion between locations but are limited by quality of metadata attached to the sequence. This approach has been useful in highlighting the impact of the initial spread of epizootic rabies and how it has shaped the modern viral distributions (Biek *et al*., 2007), the origin of the 2009 H1N1 epidemic (G. J. D. Smith *et al*., 2009) and

extensively used for tracking the dissemination of SARS-CoV-2 (Attwood *et al*., 2022; du Plessis *et al*., 2021; Paredes *et al*., 2022; Tsui *et al*., 2023).

1.12 Project aims

This thesis is composed of two projects rooted in the prediction and prevention of zoonoses. Project one aimed to first generate novel genomes using mNGS from archival and contemporary clinical CCHF samples stored at Porton Down. We hypothesised that the addition of novel sequences to a curated CCHFv dataset containing all publicly available sequences and those generated in this study would facilitate the identification of geographic regions at risk of outbreaks with virulent strains and highlight common transmission routes. The additional novel genome sequences were used for determination of the molecular epidemiology of CCHFv globally with the aim of identifying geographic regions at risk of strains associated with increased pathogenicity. Bayesian phylodynamic and phylogeographic methods were then employed for inference of evolutionary rates across the CCHFv clades in the small segment, and prediction of dissemination of CCHFv across time. The novel dataset was also used for detecting regions under selection across the three genomic segments for identification of therapeutic targets.

More specifically, because of the bias towards CCHFv samples from one outbreak in Turkey, we hypothesised that the generation of genomes for the 2015 Turkey outbreak would allow inference regarding the viral population circulating in Turkey at that time.

Project two materialised because of the COVID-19 pandemic. We hypothesised that UK horseshoe bat populations likely harboured coronaviruses, particularly sarbecoviruses, as observed in eastern Asia, Africa, and to a lesser extent, Europe. Faecal samples were collected from a single *Rhinolophus hipposideros* (lesser horseshoe) roost across four years (2020-2023). mNGS and targeted amplification approaches were used to first, detect, and then characterise sarbecoviruses circulating within the roost. Other roosts in the south of England were included in the project as to further characterise lesser horseshoe bats as a reservoir and to detect other viruses in UK bat species.

2 Materials and Methodology

2.1 Crimean-Congo haemorrhagic fever virus sample pre-processing

2.1.1 CCHFv sample selection

At Porton Down historical and contemporary CCHFv samples from suspected and confirmed cases are stored at the BSL containment level 4 facility. 19 samples were retrieved from storage and propagated before sequencing. 41 samples were processed without propagation (Appendix 1, Appendix 2).

2.1.2 Inactivation, extraction, and DNase treatment of nucleic acids from CCHFv samples

Historical and contemporary samples from clinical samples or tissue culture were inactivated at BSL containment level 4. Nucleic acid was extracted using the manual extraction QIAamp viral RNA kit (Qiagen) as per manufacturer instructions. The eluate was DNase treated with TURBO DNase (ThermoFisher) and incubated at 37 ºC for 30 minutes (Table 2-1).

Reagent	Volume (µL)
10X TURBO DNase buffer	4
TURBO DNase enzyme	1
Sample	45
Total	50

Table 2-1. Turbo DNase reaction constituents.

DNase treated nucleic acid was cleaned with Zymo RNA clean and concentrator kit (Zymo Research). 100 µL RNA binding buffer and 150 µL molecular grade

ethanol was added to the sample and mixed by pipetting. This solution was added to a Zymo RNA spin column for centrifuging. After each step the flowthrough is discarded, with the eluate being retained. First the sample was centrifuged at 16000g x 30 seconds. 400 µL RNA preparation buffer was added and centrifuged at 16000g x 30 seconds. 400 µL RNA wash buffer was added to the column and centrifuged at 16000g x 30 seconds, discarding flowthrough, and centrifuged again at 16000g x 30 seconds to remove excess reagent. 50 µL molecular grade water was added to the column, incubated for 2 minutes, and eluate retained after being centrifuged for 30 seconds at 16000g.

2.2 UK bat dropping collection and pre-processing

2.2.1 Sample collection and selection

2.2.1.1 2020 Cinderford samples

Faecal samples were collected from *Rhinolophus hipposideros* (lesser horseshoe) bats during an annual population survey in Gloucestershire in June 2020 (Appendix 3). A sterile sheet was placed beneath a roost and left for 30 minutes. Pellets were then collected and stored in 2 mL RNALater, at -20°C.

2.2.1.2 2021 Cinderford samples

Faecal samples were collected from the Gloucestershire roost in June and July 2021 (Appendix 4). Pellets were collected as above but placed in 1 mL virus transport media SIGMA-VIROCULT (Medical Wire & Equipment Co Ltd) (VTM) to preserve viable virus for isolation attempts and stored at -20ºC.

2.2.1.3 2021 Somerset and Devon sample collection

Faecal samples were collected from various locations (appendix table) across South England. Faeces was collected by trapping a bat until pellets were released and stored in 1 mL RNAlater and stored at -20ºC for subsequent analysis.

2.2.1.4 2022 Cinderford samples

Faecal samples were collected again from the same roost in Gloucestershire. The storage agent for screening was altered from RNAlater to DNA/RNA shield (Zymo Research) to reduce reliance on CL3 cabinet access. The agent for preservation of viable virus was VTM.

Firstly, for maximum coverage and to obtain sufficient material for viral isolation, 10 pellets were combined and divided into two; into 1 mL RNA/DNA shield (Zymo Research) or 1 mL VTM to allow concurrent screening and isolation of positives. 200 samples were obtained. A follow up collection obtained 1 pellet per tube into 1 mL for screening and isolation. 100 samples were obtained. Samples were stored at 4ºC until processing and then frozen at -20ºC prior to analysis (Appendix 4).

2.2.1.5 2023 Cinderford samples

48 Faecal samples were collected from the Gloucestershire roost in August 2023. Pellets were collected as above but placed in 1 mL virus transport media SIGMA-VIROCULT (Medical Wire & Equipment Co Ltd) (VTM) to preserve viable virus for isolation attempts. Samples were stored at -20ºC prior to analysis (Appendix 4).

2.2.1.6 2022 Plymouth samples

40 faecal pellets were collected from different areas of the roost floor and stored in 1 mL RNA/DNA shield (Zymo Research) for initial screening. Samples were stored at -20ºC prior to analysis (Appendix 4).

2.2.1.7 2023 *Rhinolophus ferrumequinum* dropping collection

48 droppings were collected from a *Rhinolophus ferrumequinum* roost in Dorset. 1 dropping per tube was placed in 1 mL DNA/RNA shield (Zymo Research) for inactivation and preservation of nucleic acid. Droppings were stored at -20ºC before processing.

2.2.2 Homogenisation of faecal pellets

Faeces homogenisation was performed in either a Precelleys (Fisher Scientific UK), or Sarstedt tubes (VWR International Ltd) with ceramic or glass beads. Samples were homogenised by vortex until the pellet had disappeared.

2.2.3 Inactivation, extraction, and DNase treatment of nucleic acids from faecal droppings

For faeces collected in 2020 10^6 / mL^{-1} genome copies Hazara virus was added before extraction as an internal sequencing control. All faeces were added to binding buffer under CL3 conditions.

Nucleic acid extraction was performed using an automated extraction protocol using the MagMAX viral RNA extraction kit (ThermoFisher UK) on the Kingfisher

Flex machine. 460 µL sample (RNA/DNA shield, VTM or RNAlater) was added to 265 µL binding buffer (inactivating agent), 6 µL proteinase K, and 10 µL MagMAX beads. Two washes were performed, wash one containing guanidine thiocyanate, and wash two 80% EtOH and then eluted into 70 µL MagMAX elution buffer. This eluate was DNase treated with EzDNase (Table 2-2).

Reagent	Volume (µL)
EzDNase enzyme	1
Buffer	1
Sample	8
Total	10

Table 2-2. EzDNase treatment reaction.

2.2.4 Screening methodologies

2.2.4.1 qPCR screening faecal samples with the pan-Sarbecovirus WHO RdRp assay

Faecal samples were screened using WHO Pan *Sarbecovirus* assay with the RNA dependant RNA polymerase (RdRp) targeting primers (Corman et al., 2020). The cycling conditions for this reaction were as follows: 1 x (55 ºC, 10 minutes), 1 x (94 ºC, 3 minutes), 45 x (94 ºC for 15 seconds, 58 ºC for 30 seconds). Separate assays were conducted with probe 1 and probe 2 to delineate between pan *Sarbecovirus* and SARS-CoV-2 specific positivity.

Reagent	Volume (µL)
TaqMan Fast Virus 1X	5
Fwd primer (0.6 µM)	1.5 (10 µM)
Rv primer (0.9 µM)	2.0 (10 µM)
Probe (Pan/SARS-CoV-2 specific, 0.1 µM)	0.25 (10 µM)
Sample	5
Water	11.25
Total	25

Table 2-3. qRT-PCR reaction and conditions.

2.3 Sequence Independent Single Primer Amplification

RNA metagenomic reverse transcription and amplification was performed using SISPA (Table 2-4).

	Reagent	Volume (µL)
	Sol A (40 µM)	1
	Sample (DNase treated)	4
65 °C for 5 minutes, 23 °C for 5 minutes		
RT		
	5X first strand buffer	2
	Molecular grade water	1
	dNTP (10 µM)	1
	DTT (0.1 M)	0.5
	Superscript IV RT	0.5
	Sample	5
	Total	10
42 °C for 10 minutes		
2nd strand		
	Water	3.85
	5X sequenase reaction buffer	1
	Sequenase dilution buffer	0.45
	Sequenase enzyme	0.5
	Sample	10
	Total	15.85
37 °C for 8 minutes		
Amplification		
	Water	19
	Q5 hotstart mastermix (NEB) (X10)	25
	Sol B	1
	cDNA (step 4 output)	5
	Total	50
98 °C for 30 seconds, 35 x (98 °C for 10 seconds, 50 °C for 20 seconds, 72 °C for 2 mintes), 72 °C for 2 minutes		

Table 2-4. The SISPA protocol.

2.3.1 Quantification of SISPA product

cDNA was quantified using a Qubit fluorometer (ThermoFisher) using the high sensitivity dsDNA kit. For each sample, 199 µL buffer is added to 1 µL nucleic acid binding dye. 1 µL sample is added to this and quantified after a 2-minute

incubation. SISPA-cDNA concentrations for all samples are found in the appendix (Appendix 3, Appendix 4).

2.4 Targeted amplification

Whole genome tiling amplicons, amplified by 34 primer pairs were designed against a the *Sarbecovirus* identified from the Cinderford roost in 2022 using PrimalScheme (Quick *et al*., 2017). These primer pairs were divided into two pools to prevent overlapping amplicons in the same reaction (Appendix 5).

2.4.1 Working stocks and balancing

All primers were diluted to 10 µM working stock solution. 5 µL each working primer was pooled into respective pools to create two working primer pools. After an initial run, 1 µL 100 mM stock for primers 9, 11, 13, 15, 17, 23, 27 in pool 1 and 12, 14, 18, 22, 26 in pool 2 were additionally spiked into the working stocks to balance the primer concentrations.

2.4.2 Reactions conditions

For each pool a reverse transcription (RT) step was carried out. 4 µL LunaScript RT supermix (NEB) was added to 7.5 µL RNA and 8.3 µL molecular grade water (NEB). This was incubated at 25ºC for 2 minutes, 55ºC for 10 minutes and 95 ºC 1 minute. The resulting reaction mix, including cDNA, was then added directly (without clean up) into the amplification reaction (Table 2-5). The amplified product was then sequenced by ONT (Oxford Nanopore Technology).

Reagent	Volumes (µL)
Q5 hotstart mastermix	12.5
Pool 1 / Pool 2 primer pool	4
Nuclease free water	6
cDNA	2.5
Total	25

Table 2-5. The amplification reaction for the amplicon scheme PCR.

2.5 Next generation sequencing

2.5.1 Oxford Nanopore Technology

Long read sequencing was performed using the ligation sequencing kit (LSK109) with native barcode expansion kit (EXP-NBD 196) on R9.4 flow cells for 72-hours on a GridION with high-accuracy base calling via guppy. Multiplexed libraries were prepared with a total input of 1.5 µg nucleic acid per flow cell shared evenly across the samples. All ONT libraries were prepared as per manufacturer's instructions. Briefly, cDNA was end-prepped to generate an overhanging sticky end for barcode addition. An adapter was then added to facilitate pore transversal in the flow cell membrane.

2.5.2 Illumina - Nextera XT Library preparation

Short read sequencing was performed using Illumina technology. Illumina libraries were prepared according to the Nextera XT protocol with 2 x 150-bp paired ends. Briefly, cDNA was diluted to 0.33 ng/µL and used as input for the tagmentation reaction where cDNA is fragmented and tagged for analysis. Fragmented cDNA was then barcoded with a unique pair of indexes for multiplexing. The fragmented and barcoded cDNA was then quantified by Qubit

and the concentrations used for calculating the volumes used to provide the correct molarity for sequencing.

- nM = (concentration of product (ng/µl) / (500 (bp) x 660 (g/mol)) x 1000000

2.6 Receptor binding assay methodology

2.6.1 Generating replication deficient lentiviral-based pseudoparticles

The pcDNA3.1 vector was constructed by GeneArt (ThermoFisher) with the RhGB01 spike or RhGB01 RBD-SARS-CoV spike chimera. The construct was orientated as:

- HindIII - Kozak – ATG – Spike (codon-optimised) – stop – FLAG – NotL

2.6.2 Pseudoparticle production

Human Embryonic Kidney-293T (HEK293T) cells (ATCC: CRL-3216) were maintained for pseudoparticle production in 25 mL DMEM-10% (Merck) in a 75 cm^2 tissue culture. HEK293T cells were seeded at a concentration $7.5x10^5$ cells per well in a 6-well plate in 3 mL DMEM-10% and agitated to ensure equal distribution. Seeded cells were incubated overnight at 37ºC, 5% CO2.

When HEK293T cells were 60-80% confluent, the transfections were set up for pseudoparticle production. Three plasmids were transfected into the HEK293T cell line: pCSFLW which encoded the reporter gene and LTR. P8.91 which encoded the HIV gag/pol genes and pCDNA3.1 which encoded the RhGB01 spike or RhGB01-SARS-CoV Spike-RBD chimera.

In a 1.5 mL Eppendorf tube, 100 µL Opti-MEM (Thermofisher) was added with 0.6 µg CSFLW, P8.91 and 0.5 µg pCDNA3.1 or empty vector. The negative control was an empty pCDNA3.1 vector and positive control a VSV-G plasmid. All were incubated at room temperature for 5 minutes.

In a separate 1.5 mL Eppendorf tube, 100 mL Opti-MEM was added with 10 µL Polyethyleneimine (PEI) (Merck) (1 µg/mL) – setting up 1 per transfection and incubated at room temperature for 5 minutes.

For each 100 µL transfection mix with plasmid in Opti-MEM, 100 µL PEI in Opti-MEM was added and mixed vigorously and incubated at room temperature for 20 minutes. This transfection mix was then added to each well of the 6-well plate and incubated overnight at 37ºC, 5% CO_2.

After incubation, the media was removed from the 6-well plates containing the transfection mix and ensured to aspirate from the edge of the well by tilting the plate. The removed media was replaced with 3 mL DMEM-10% and incubated overnight at 37ºC, 5% CO_2.

For harvesting, 50 mL supernatant was transferred to falcon tubes and stored at 4ºC. The media was replaced again with 3 mL DMEM-10% and incubated overnight at 37ºC, 5% CO_2. The following day the harvest procedure was repeated and pooled with the corresponding well. To pellet the cells, falcon tubes were centrifuged at 2500 x g for 10 minutes at 4ºC.

2.6.3 ACE2 receptor binding screening assay

For the receptor screen, Baby Hamster Kidney-21 (BHK-21) cells (ATCC: CCL-10) were seeded in 24-well plates at $1x10^5$/well in DMEM-10% and incubated

overnight at 37ºC, 5% CO_2. At 60-80% confluence, transfection mixes were set up with 100 µL Opti-MEM and 500 ng different species of ACE2-expressing constructs (pDISPLAY, thermoFisher) and the control, a pDISPLAY vector. 3 µL Tran*IT*-X2 transfection reagent (Mirus) was added to the transfection mix for every 1 µg of DNA (1.5 µL per transfection mix). This was then incubated at room temperature for 20 minutes.

100 µL transfection mix was added to each well of plated BHK-21 cells and incubated overnight at 37 ºC, 5% CO_2. The media was removed and 0.5 mL EDTA-Trypsin (2 mM) (ThermoFisher) in PBS added to transfected BHK-21 cells to detach the monolayer, the detached cells were aliquoted into Bijou tubes and diluted to 1 x 10^5/mL with DMEM-10%. For these diluted cells, 100µL (1 x 10^4 per well) was seeded into flat, white-bottomed 96 well plates and incubated overnight at 37ºC, 5% CO_2.

The media was removed from BHK-21 seeded cells in the 96 well plate and infected with pseudoparticles at a concentration relative to 10^6-10^7 RLU or the control vector and incubated for 48 hours at 37ºC, 5% CO_2.

The media was removed from BHK-21 cells after incubation to allow ACE2 binding. 50 µL BrightGlo™ (Promega) was diluted 11 with serum free, phenol red DMEM (Merck) and added to wells for detection of ACE2 binding. After the addition of BrightGlo™ the plate was incubated in the dark for 5 minutes and then read of a GloMax Multi+ detection system (GloMax Discover Microplate Reader) using the luminescence protocol with 0.5s integration.

2.6.4 Confirming the expression of spike on the surface of the pseudoparticle and in the cytoplasm of HEK293T cells

The HEK293T cells or transfected pseudoparticles were mixed 1:1 with Laemmli buffer (BioRad) and reduced at 95 ºC for 5 minutes for denaturation. Reduced samples were loaded onto 7.5% acrylamide gels SDS-PAGE and ran. The separated proteins were transferred from the gel using a semi-dry transfer onto a nitrocellulose membrane. The pseudoparticle fraction was probed with mouse anti-FLAG and rabbit anti-P24. The HEK293T fraction was probed with mouse anti-FLAG and mouse anti-Tubulin for the cellular control.

2.7 General bioinformatic tools

2.7.1 Taxonomic classification

In both projects Kraken2 was used as a method of initial investigation (Wood *et al.*, 2019). Raw reads from either sequencing technology were used as input. The database used was downloaded on 13/07/2023.

2.7.2 Mapping

Raw reads from ONT and Illumina were mapped to reference genomes using two different tools. Short, paired end reads generated by Illumina technology are mapped using BWA-MEM (H. Li, 2013) and longer ONT reads mapped using minimap2 (H. Li, 2018). The commands used are as follows:

- Bwa mem reference.fa raw_fastq.R1.fq raw_fastq.R2.fq > output.sam
- Minimap2 -ax map-ont reference.fa raw.fastq > output.sam

2.7.3 Samtools

Samtools view and samtools sort were used to convert from a SAM file type to a BAM file type (H. Li *et al*., 2009). Samtools flagstat was used to extract mapping statistics. samtools depth was used to investigating mapping positions of raw reads across reference genomes with the -aa flag to prevent masking of sites with mapped reads.

- samtools view -bh input.sam | samtools sort -o output_sorted.bam
- samtools flagstat output_sorted.bam
- samtools depth -aa output_sorted.bam

2.7.4 Quasibam

Consensus sequences for reference guided assemblies for both projects were generated using Quasibam (Penedos *et al*., 2015). A tool which generates a consensus sequence and minor variant output file. The command differs slightly depending on if the raw data is from ONT or Illumina because of the lower accuracy of ONT reads using R9.4 flow cells. All consensus genomes called from Illumina data were called at a minimum depth 10 and 80% agreement for a consensus base. For consensus sequences called from ONT data (amplicon data) a minimum depth 20 was used, and the consensus agreement required for a base to be called dropped to 60% to capture more sequence.

2.7.5 De novo and hybrid assemblies

De novo and hybrid assemblies were performed using SPAdes genome assembler (Bankevich et al., 2012). Hybrid assemblies were performed with

spades.py with additional –nanopore flag to provide long reads in addition to shorter reads.

2.7.6 Alignments

Nucleotide and amino acid datasets were aligned using the MAFFT aligner (Katoh & Standley, 2013). Partial sequences were added to complete alignments using the “--addfragment” and “--keeplength” flags.

- mafft --auto in.fasta > out_aln.fasta
- mafft --auto --addfragments partials.fa --keeplength full_aln.fasta > full_partial_aln.fasta

2.7.7 Maximum likelihood phylogenies

Maximum likelihood phylogenies were inferred using IQTREE (Nguyen *et al*., 2015). For larger alignments the nucleotide substitution model HKY+G4+I was selected to reduce computational time and its widespread applicability to phylogenetic inferences without the risk of over parameterisation provided by GTR nucleotide substitution models.

- iqtree -s input_aln.fasta -m MFP -B 1000
- iqtree -s input_aln.fasta -m HKY+G4+I

2.8 Computational analysis of amplicon scheme raw data

Amplicons which failed/underperformed were identified using a bespoke python tool. Briefly, this command line tool requires the raw reads from each pool to be

mapped to the reference genome used for generating the primer scheme. It also requires the samtools depth command to be used for generation of a tab separated text file containing three columns (reference id, base, depth). A bed file containing the start and end coordinates in columns 2 and 3 is also needed. The tool will use the coordinates in the bed file to rapidly extract the lowest depth present in the region covered by the primer coordinates. A depth is provided for the forward and reverse primer and the amplicon itself and user-customisable plots are outputted. A similar tool is also available for extracting the mean depth across the primers and amplicon. It is also possible to use these tools to look at specific primer numbers in the scheme and doesn't require all primer pairs to be included in the analysis. Links to tools provided below.

- https://github.com/jackmicr0/GenomicsTools/blob/main/Primer_scheme_gap_finder_minimum_depth.py
- https://github.com/jackmicr0/GenomicsTools/blob/main/Primer_scheme_gap_finder_mean.py

2.9 Computational analysis for the CCHFv dataset

2.9.1 Generating CCHFv consensus sequences

To generate consensus sequences, raw data was mapped to a representative CCHFv sequences from relevant geographic locations and across temporal space. Raw reads were mapped to reference sequences (UG3010 [DQ211650, DQ211637, DQ211624], Min [KX013482, KX013481, KX013480], Borno [KX238958, KX238957, KX238965], NIV1311122 [MH396645, MH396643, MH396644]) using bwa mem, and mapping success was assessed via samtools

flagstat and samtools depth with flag -aa. Quasi-bam was used for generating consensus sequences.

For each segment the reference sequence which generated the most complete sequence was selected to generate sequences for analysis (Appendix 6). First, the consensus sequences from an in-house pile up script were used. The first five “ATG” occurrences in the sequence identified and evaluated by the success of translating the sequence from this point to find the most appropriate start codon.

Secondly, the quality of reading frames was assessed. Quality control metrics included the number of unknown bases in the nucleotide sequence. Each potential reading frame (5 frames) are assessed for suitability. Full sequences are identified by confirming the presence of a start codon, and an end codon matching that of currently available CCHFv sequences, and ensuring the sequence is in-frame.

Accession	Strain	Location	Date	Continent	Segment
DQ211650, DQ211637 DQ211624	UG3010	Republic of the Congo	1956	Africa	S,M,L
KX013482, KX013481, KX013480	Min	Russia	1968	Europe	S,M,L
KX238958, KX238957, KX238956	Borno	Nigeria	2012	Africa	S,M,L
MH396645, MH396643 MH396644	NIV1311122	India	2013	Asia	S,M,L

Table 2-6. Metadata associated with the reference sequences.

2.10 Retrieving NCBI sequences for dataset generation

Sequences from the public domain were retrieved from NCBI utilising the E-fetch tools via a bespoke command line python tool. To retrieve sequences, the search terms used were “Crimean-Congo”, “CCHF”, “Crimean-Congo haemorrhagic fever”, and “Crimean-Congo haemorrhagic fever orthonairovirus”. Sequences were then filtered using terms, “UNVERIFIED”, “Jingmen”, “Bos”, “mitochondrial”, “Mitochondrial”, “Bacterial”, “modified”, “Modified”, “Patent”, “Construct”, “Plasmid”, “plasmid”, “construct”, “microbial”, “Microbial”, and “Compositions”.

- https://github.com/jackmicr0/GenomicsTools/blob/main/Retrieve_with_EaZeEeeee.py

The tool extracts the metadata associated with the NCBI entry which was then improved manually by literature searches. Datasets were then separated based on completeness – partial and complete and the complete sequences taken forward.

2.11 Phylogenetic analyses

The three sets of nucleotide sequences were aligned using MAFFT aligner and visually inspected post alignment to ensure quality. For generating maximum likelihood trees, IQTREE was used with parameter -B 1000 for bootstrapping, and nucleotide substitution model HKY+G4+I used.

2.12 Time calibrated phylogenetic and phylogeographic analyses

For evolutionary rate and TMRCA estimates, models were implemented in BEAST v10.4 (Drummond & Rambaut, 2007). Prior to setting up BEAST models, exploratory analysis of temporal signal was carried out on each clade using an in-house R script and TempEst (Rambaut *et al*., 2016). The R script extracts the patristic distance (distance from tip to root by sum of intermediate branches) from a maximum likelihood phylogeny and carries out a linear regression between the patristic distance and the collection date attached to the sequence. This is an informal measure of temporal signal which assesses the correlation between sequence diversity and time. It is only an informal measure as all sequences have a shared evolutionary history and thus are not independent points suitable for statistical analysis.

For the small segment, the clade with the strongest signal as assessed by R^2 value, an initial BEAST model was set up to estimate a rate which could be used for rate calibration. To assess the suitability of the calibration rate to each clade, different models were used with different statistical distributions for the clock rate prior. Normal and lognormal distributions were tested with standard deviations 0.1, 10 and 100. All models used an HKY+G4+I substitution model and a constant population tree prior. As the standard deviation is decreased, there is less parametric space allowed to be explored during rate estimation using the nucleotide sequence data and thus the data has less influence on the parameter estimate.

For the full small segment phylogeny, a lognormally distributed uncorrelated relaxed clock and fixed local clock models were used, with a lognormal distribution of rates among branches with a rate prior calibrated with evolutionary

rate estimated in Tempest, $1.2x10^{-4}$ with a standard deviation 100. Both models used an HKY+G4+I substitution model and a constant population tree prior.

MCMC chains were run for either 50 million iterations with samples taken every 5000 iterations or for larger clades, for 100 million iterations with samples taken every 10000 iterations. For full small segment models, duplicate 200 million MCMC runs were completed and combined, removing a 10% burn-in. All sampling methods allow for at least 10000 individual samples to be taken from the posterior for each parameter being estimated. Chain convergence was assumed if the effective sample size (ESS) was >200 for all parameters summarised by Tracer (Rambaut *et al*., 2018).

2.13 Model selection

Comparisons of evolutionary models were conducted using path sampling. Path sampling calculates a log marginal likelihood which is then used to compare models. The model was selected using Bayes factors (BF) which considers which of two models is best supported given the observed data, for example model 1 (M1) vs model 2 (M2). The Bayes factor was calculated by exponentiating the difference between the log marginal likelihoods of the two models (ML1 and ML2). A BF < 1 suggests evidence for M2, 1 < BF < 3 suggests weak evidence for M1, 3 < BF < 20 suggests strong evidence for M1, and 20 < BF < 150 suggests very strong evidence for M1.

2.14 Discrete trait phylogeographic diffusion

A discrete trait analysis (DTA) was implemented using BEAST v10.4. Discrete location traits were assigned to phylogenetic tips via metadata. All locations were restricted to country level because of the absence of higher resolution geographic locations across the CCHFv small segment phylogeny. To assess viral diffusion at local and global scales, individual clades and entire phylogenies were modelled independently.

Spatial dissemination across a phylogeny is modelled as a continuous-time Markov chain (CTMC) which allows freedom for state transitions to occur continuously across a phylogeny with transition rates between different states. An asymmetric transition matrix was used for transition rates which facilitates different rates for two-way transitions between any two selected locations.

Briefly, to assess the significance of transitions between regions, all transitions are assigned a prior probability of zero, to represent no discrete trait transition. The prior probability is compared to a posterior distribution of observed transition rates calculated from the phylogeny to generate a Bayes factor. This is implemented using a Bayesian Stochastic Search Variable Selection (BSSVS) approach. A Bayes factor >3 was considered as significant for diffusion between discrete traits. A posterior probability >0.5 and >0.7 were both considered for different levels of support for transition rates between discrete traits.

For the discrete trait analysis, MCMC chains for smaller clades of 100 million iterations sampled every 10000 were used. For the larger phylogenies, MCMC chains of 200 million iterations sampled every 20000 were used. Convergence was assessed using Tracer.

2.14.1 DTA visualisation

DTA runs were summarised using SPREAD4 (Nahata *et al*., 2022), a tool for visualisation of spatiotemporal reconstructions inferred from the genomic data. SPREAD4 output was then plotted as edge network graphs onto geographic maps using python modules, NetworkX (Hagberg et al., 2008) and GeoPandas (Jordahl, 2014).

2.15 Testing for selection

Evidence for selection was inferred using nucleotide alignments with codon-based methods, as implemented in HyPhy using the DataMonkey server. Command line HyPhy was used for large alignments with prolonged running times. Coordinates for functional annotations were taken from UniProt.

2.15.1 Site-specific selection

Site-specific selection was tested using Fast, Unconstrained Bayesian AppRoximation (FUBAR) (Murrell *et al.,* 2013). FUBAR also assumes constant rates for each site across the entire phylogeny. Under the Bayesian framework, FUBAR reports a posterior probability rather than a p-value. FUBAR can also perform favourably to another method, fixed effects likelihood (FEL) with weak positive selection detection.

Another method for detecting episodic positive selection is Mixed Effects Models of Evolution (MEME) (Murrell *et al.,* 2012). MEME will report the presence of positive selection at an amino acid site and calculate the number of branches in the phylogeny demonstrating this selection. Each site in a phylogeny is provided

two rate (ω) classes and provides a weight to each class, where the weight represents the probability that the site evolves under either rate. For calculating the rate classes, a single dS (synonymous) value is provided (α) and two different dN values (β^-, β^+) to represent negative (disadvantageous mutations) and positive (advantageous mutations) selection, respectively. If $\beta^+ > \alpha$ positive selection is inferred. Significance of these estimates is assessed using a likelihood ratio test. The LRT score is calculated as the difference between the log likelihood scores of the null model (positive selection cannot happen) and the alternative model (positive selection is unrestricted).

2.16 Testing for recombination

Recombinant sequences were detected using Recombination Detection Program 4 (RDP4) (Martin *et al*., 2015). This tool combines multiple different recombination detection tools to validate predictions across different tools. Tools used are RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiSscan, PhylPro, LARD, 3Seq. RDP provides a minor and major parental sequence where possible, and the recombinant sequence. Where there is uncertainty around which sequence in the trio is the recombinant, this is provided. If a recombinant is detected in ≥3 methods at <0.01 P-value, it will be removed from further analysis. If there is uncertainty around a statically supported triplet of sequences, the triplet is removed.

2.17 Testing for reassortment

To identify reassortment within full CCHFv genomes two methods were employed. Graph-Incompatibility-based Reassortment Finder (GiRaF) (Nagarajan & Kingsford, 2011) utilises phylogenies generated using MrBayes (Ronquist *et al*., 2012) to identify incompatible divergence events in the phylogenies for the different segments. This results in either single or multi taxa candidate sets of reassorted taxa where multiple taxa candidate sets indicate a potential ancient reassortment event. The candidate taxa identified by phylogenetic incompatibility are also assessed by pairwise distance to surrounding sequences. Reassortment networks were visualised using the Baltic python package (https://pypi.org/project/baltic/).

In addition to this, segments were concatenated and inputted into RDP4. Pairwise distances were also calculated for each strain for the small, medium, and large segments. Viruses were labelled as possible reassortant by having highest similarity to a virus collected outside of its collection location. Reassortant viruses were then inspected phylogenetically for differential clustering.

2.18 UK Bat bioinformatic analyses

2.18.1 Genome annotation

Consensus sequences were annotated using PROKKA (Seemann, 2014). Genomic annotations were then checked visually and by translation to confirm predicted genes were in-frame.

- prokka –kingdom Viruses

2.18.2 Pairwise comparisons

Pairwise comparisons were conducted using Biopython (Cock *et al*., 2009) module “Bio.pairwise2”. Raw nucleotide and amino acid differences were calculated using pre-aligned sequences and a bespoke python script counting raw differences between sequences.

2.18.3 In silico protein structure prediction

2.18.3.1 AlphaFold2

Protein structures were predicted using AlphaFold2 (Jumper *et al*., 2021) via the Galaxy EU server (Afgan *et al*., 2018). The full database was used with the monomer model. The input target sequence for RBD modelling was residues 324–515 in the spike protein sequence.

2.18.3.2 Homology modelling

Homology modelling was implemented using the SWISS-MODEL server (Schwede *et al*., 2003). The input target sequence for RBD modelling was residues 324–515 in the spike protein sequence. SWISS-MODEL performs a template search of BLAST and HHblits databases for pre-existing characterised structures with homology. The most suitable template was selected according to the highest Global Model Quality Estimate (GMQE) and lowest Quaternary Structure Quality Estimate (QMEAN). GMQE expresses the accuracy of the model built as a number between 0 and 1. QMEAN uses statistical potentials of mean force to generate global and per residue estimates.

2.18.3.2.1 Assessment of protein model quality

To assess the quality of predicted structures two methods were selected in addition to visual inspection via structural superimposition to similar experimentally characterised structures. ProCheck and ERRAT2 (Colovos & Yeates, 1993; Laskowski *et al*., 1993) were implemented using the Structural Analysis and Verification Server (SAVES). ProCheck generates a Ramachandran plot which classifies the torsion angles of each residue in the predicted 3D structure into 3 groups to assess its validity. ERRAT2 uses a database of 96 high-resolution protein structures to classify the 6 types of atomic interactions (carbon–carbon, carbon–nitrogen, carbon–oxygen, nitrogen-nitrogen, nitrogen–oxygen, oxygen–oxygen) into correct or incorrectly determined protein regions based on the atomic interactions. This provides an overall quality factor for non-bonded atomic interactions expressed as the percentage of a protein for which the calculated values sit below the set rejection limits, with a score > 50 indicating a high-quality model.

2.18.3.3 ChimeraX

ChimeraX was used for all protein visualisations (Pettersen *et al*., 2021). RMSD calculations were performed using the matchmaker command. The Cα atom of every amino acid in the sequence alignment serves to align the chosen structures, with the distance between their respective Cα atoms then measured. Electrophoretic calculations were completed using the coulombic command in ChimeraX, which calculated the coulombic electrostatic potential.

3 Molecular sequence analysis of Crimean-Congo haemorrhagic fever virus

3.1 Introduction

3.1.1 Crimean-Congo haemorrhagic fever virus and its discovery

Crimean-Congo haemorrhagic fever virus (CCHFv) is one of the most widespread tick-borne zoonotic viruses within the *Hareavirales* order and the etiological agent of Crimean-Congo haemorrhagic fever (CCHF). Upon the implementation of a comparable system to the Linnaean taxonomic system used for the classification of cellular organisms, CCHFv has been reclassified as *Orthonairovirus haemorrhagiae*. The most recent reclassification divided the *Bunyavirales* order (Briese et al., 2016) into two orders, *Elliovirales* and *Hareavirales* under the *Bunyavircetes* class (*Riboviria*; *Orthornavirae*; *Negarnaviricota*; *Polyploviricotina*; *Bunyaviricetes*; *Hareavirales*; *Nairoviridae*; *Orthonairovirus*; *Orthonairovirus haemorrhagiae ICTV*).

Human infection results in disease of varying severities which range from asymptomatic (Bodur *et al*., 2012), to mild and self-limiting disease, to haemorrhagic fever symptoms and death or convalescence (Shahhosseini *et al*., 2021a). It was discovered concurrently in Crimea and Zaire (now the Democratic Republic of Congo) (Casals, 1969). The first retrospectively confirmed cases were recorded in 1944 when more than 200 soviet military personnel presented with haemorrhagic fever after the liberation of Crimea from Nazi occupation (Butenko & Karganova, 2007; Hoogstraal, 1979a).

Successful isolation was not reported until the introduction of the intracerebral inoculation of newborn mouse model for arbovirus isolation in 1967. In 1968, the

Drozdov strain from the Astrakhan region of Russia was identified as antigenically indistinguishable to a virus isolated from a patient in Zaire in 1956; highlighting the global prevalence of CCHFv upon characterisation (Butenko & Karganova, 2007; Casals, 1969; Causey *et al*., 1970; Hoogstraal, 1979a). Now, CCHFv is endemic in many regions within Asia, Africa and Europe (Hoogstraal, 1979a; Shahhosseini *et al*., 2021a).

3.1.2 CCHFv and the enzootic cycle

CCHFv is silently maintained within an enzootic cycle of hard bodied ticks (*Ixodidae*) and vertebrate hosts (Figure 3-1). Ticks can be categorised by the number of feeding hosts required during their lifecycle. Single host ticks e.g., *Rhipicephalus* species which attach to one host and complete their lifecycle on that host. Two host ticks e.g., *Hyalomma* species in which larvae attach to a smaller mammal, and quest as an adult tick prior to final blood meal on a larger mammalian host. Three host ticks which reattach to a novel host for each lifecycle stage e.g., *Ixodes, Ricinus, Haemophysalis, Punctate* and *Dermacentor* species.

Primary human exposure is generally attributed to tick bites or exposure to the contaminated blood of infected animals. Nosocomial and secondary transmission can occur because of nonspecific febrile symptoms during the pre-haemorrhagic phase or atypical clinical presentation, leading to an outbreak. Tertiary transmission would indicate a poor containment of primary and secondary cases and a more severe outbreak. Despite knowledge of the overall structure of the maintenance cycle, vector-vector, vector-host and vector-pathogen interactions

are non-linear and intra-vector variation prevent a simple relationship being observed between the cycle, and viral population dynamics (Cator *et al*., 2020).

The dynamics of transmission between the three life stages are not clearly defined (Estrada-Peña, Ottowell, *et al*., 2012; Hoogstraal, 1979a), but the overwintering of tick populations is a risk factor for endemicity (Gargili *et al*., 2017). Despite a lack of clarity for transmission within the cycle, CCHFv seroprevalence and viral RNA (vRNA) has been demonstrated in all three stages in disparate geographic locations.

Globally, the northwards spread of tick species, and vectors generally, is predicted to have dire consequences for the incidence of vector-borne diseases (Liu-Helmersson *et al*., 2014; Rocklöv & Dubrow, 2020). A host of factors have been identified to impact distribution: these include climate change, social conditions, environmental conditions and local alterations to the environment and economic fluctuations (England *et al*., 2016; Gray *et al*., 2009; Lindgren *et al*., 2000; Randolph, 2001; Sumilo *et al*., 2007).

CCHFv is traditionally associated with hard bodied ticks (*Ixodidae*) and vertebrates, mostly domestic livestock near human populations. However, this pattern in domestic livestock could be artificially inflated by sampling bias. This interaction between tick populations and domestic livestock is complex because of the variation in life cycle dynamics of the various tick species and a high number of possible intermediate hosts (Figure 3.1).

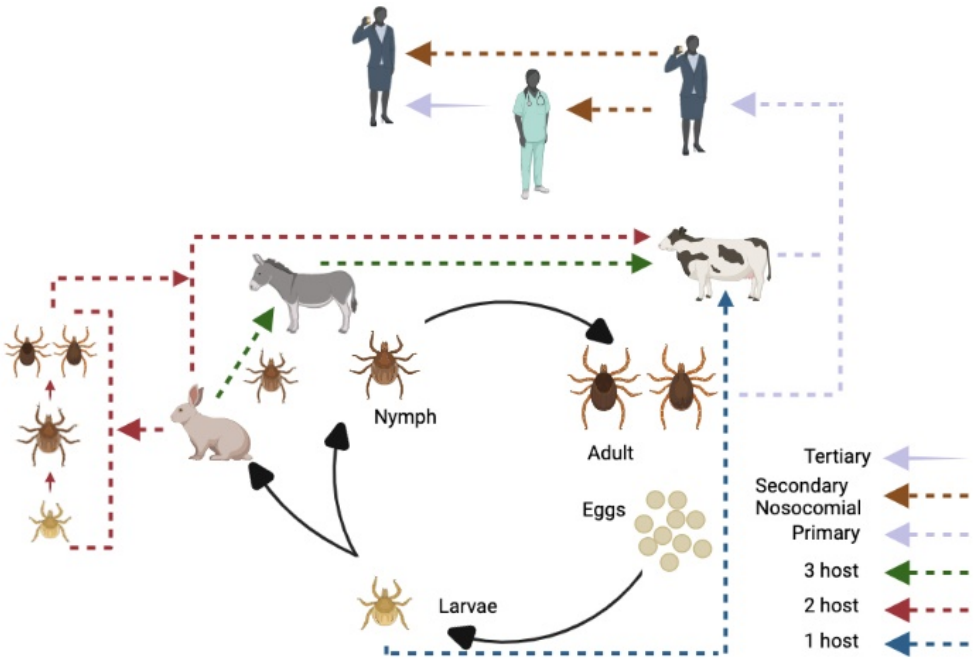


Figure 3-1.The enzootic maintenance cycle.

The transmission cycle in nature which maintains CCHFv between mammalian hosts and arthropod ticks. Created with BioRender.com.

3.1.3 Genomic organisation and functional proteins

CCHFv has a tri-segmented single stranded negative sense RNA genome, with a small (S), medium (M) and large (L) segment (Figure 3-2). At the 5' and 3' termini of the S, M and L segments are complementary non-coding regions (NCR) which are important for initiation of transcription and thus replication of the virion by serving as a promoter region of the RNA dependent RNA polymerase (Ferron *et al*., 2017; Zivcec *et al*., 2016a).

3.1.3.1 The small segment

The S segment is ambisense and encodes the nucleocapsid protein (NP) in the negative sense and the NSs protein in the positive sense direction (Figure 3-2) (Sanches *et al*. 2002; Altamura *et al*. 2007; Zivcec *et al*. 2016). The structure of

the NP was elucidated in 2012 highlighting multiple conformations, one as a monomer and one as an oligomerised dimer which suggests different stable conformations for the NP, and the potential for different functions within these conformations (Carter *et al*., 2012; Guo *et al*., 2012; Y. Wang *et al*., 2012). The NP is comprised of two domains: the head and stalk (Figure 3-3).

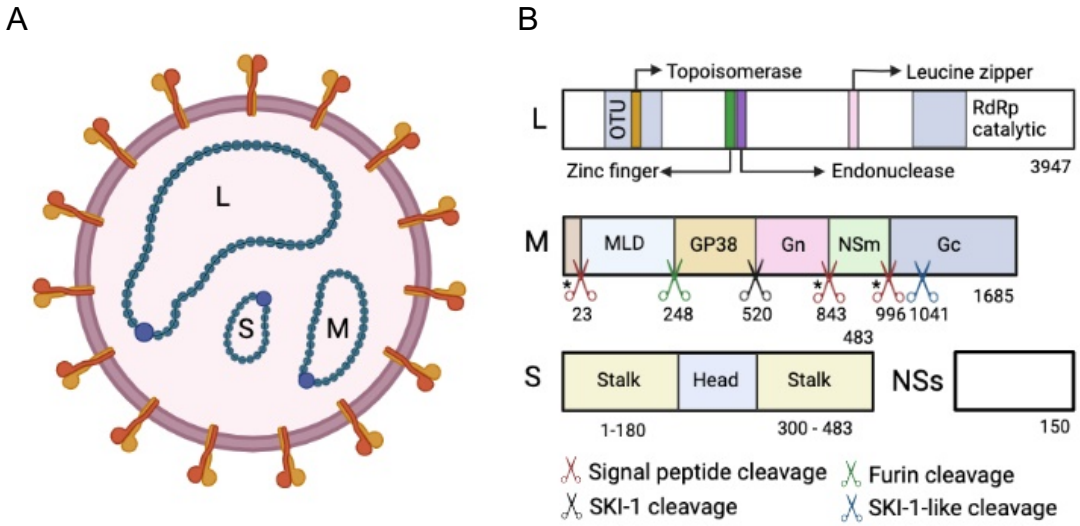


Figure 3-2. The virion and genome.

The tri-segmented negative sense genome is represented by three circular segments as genomic RNA with the RdRp attached (A). The large (L) segment is translated into the mature RNA dependent RNA polymerase (RdRp) with additional domains, such as a proteolytic Ovarian Tumour domain (OTU) with a topoisomerase region, a zinc finger, endonuclease, leucine zipper and RdRp catalytic domain. The medium (M) segment is translated into a polyprotein which goes through two rounds of cleavage (* indicates round 1) to liberate mature glycoproteins Gn and Gc, and soluble proteins, GP38, GP85, GP160 and NSm. GP38, GP85 and GP160 are soluble proteins including the mucin like domain (MLD) (B). Created with BioRender.com.

The main function of the NP is the encapsidation of copy RNA (cRNA) and vRNA into ribonucleoprotein complexes (RNC) for replication. The NP interacts with both vRNA and the L segment protein to initiate transcription (Bergeron *et al*.,

2010). vRNA binding in NP is mediated by positively charged regions (Y. Wang *et al*., 2012) (Figure 3-3).

At the apex of the stalk domain, a caspase-3 proteolytic cleavage site is present (Guo *et al*., 2012). Caspase-3 mediated cleavage represents a host defence mechanism against viral infection (D. Wu *et al*., 2022). Caspase-3 induces cellular apoptosis for clearance of infected cells but also reduces the cellular concentration of NP limiting the generation of nascent virions by reduction of RNCs (Karlberg *et al*., 2011). NP dimerisation masks the caspase-3 cleavage site and increases polymerase activity by increasing RNC availability. Homotypic NP oligomerisation was found to be dependent on the presence of resides 240-482, inclusive of the DEVD motif (Macleod *et al*., 2015).

Upon incubation with ssRNA, the NP dimer undergoes conformational change releasing the NP monomer which forms the RNC for replication with the caveat of increased caspase-3 mediated cleavage. As the vRNA concentration increases within the cell, more monomeric forms of NP will become available for RNC formation, but caspase-3 mediated cleavage of NP will also increase (Karlberg *et al*., 2011).

The NP also functions in divalent cation-dependent endonuclease cleaving single and double stranded DNA. The endonuclease function was also observed to be differentially regulated by different divalent cations. Mutations of these sites abrogated DNA endonuclease activity R384A, E387A, K411A, H453A, Q457A, whereas Y374A only marginally impacted function (Guo *et al*., 2012)

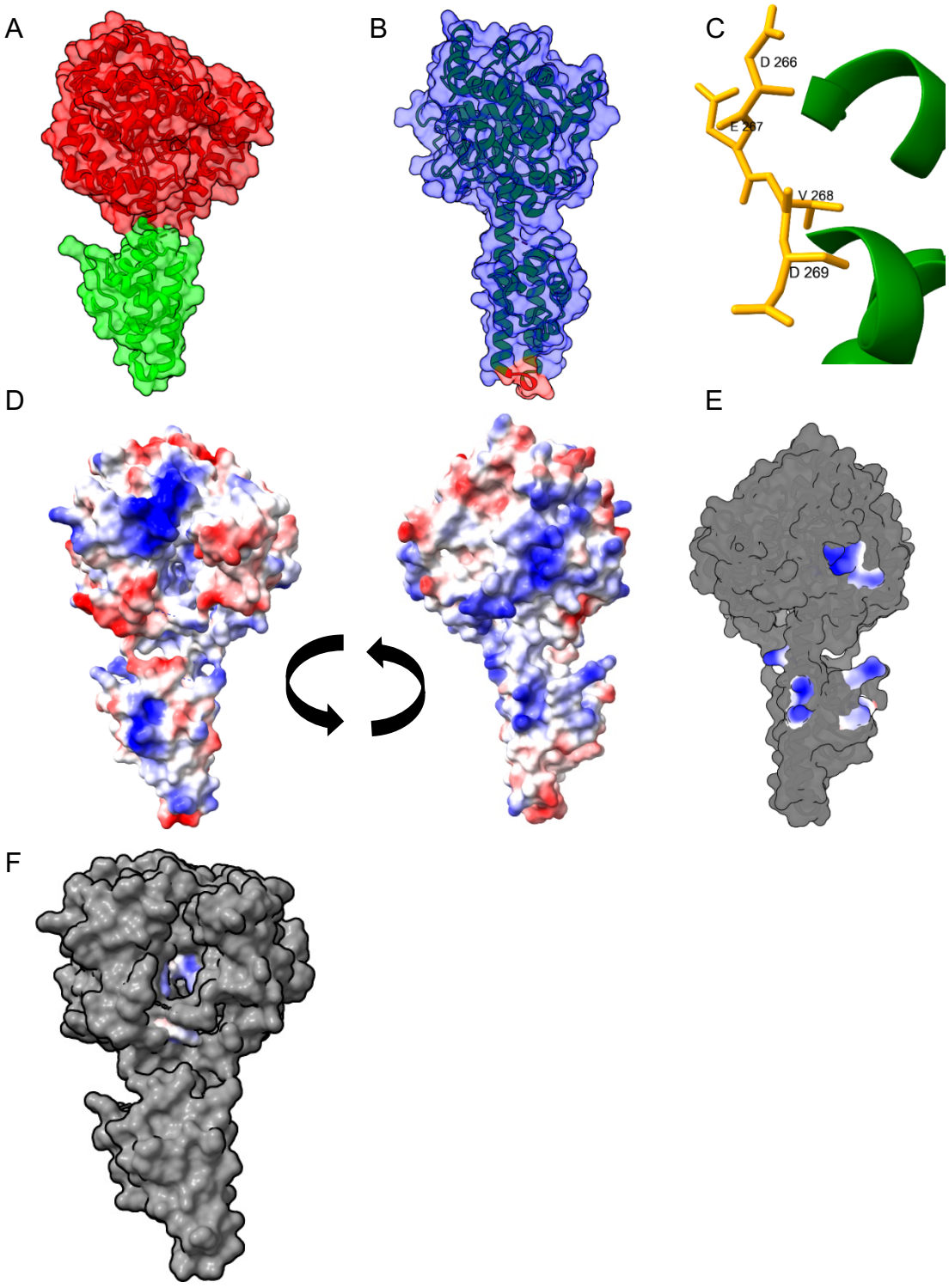


Figure 3-3. The nucleoprotein.

The nucleoprotein with domains and sites of interest. The CCHFv nucleoprotein with the stalk (181-299) and globular head (1-180,300-482) domain (A) and the region containing the caspase-3 DEVD motif (266-269; Asp-Glu-Val-Asp) (B, C). The nucleoprotein coloured by charge (D). Predicted RNA binding sites reported by Wang *et al*. 2012. Region 1 (K132, R134, K135, N468) is found in the globular head domain, and region 2 (H197, K222, R225, K282, K286, K292) in the stalk domain (E). The endonuclease domains functioning in ss and dsDNA cleavage (F).

3.1.3.2 The medium segment

The M segment is a single open reading frame encoding the glycoprotein precursor (GPC). The GPC contains signal peptides at the PreGn NTD (MLD/GP38/Gn), NSm and Gc which directs synthesis and processing to the secretory pathway where a complex cascade of cleavage and maturation into mature proteins will take place in a co-translational process. The GPC is ordered as the following: a highly variable mucin-like domain (MLD), GP38, PreGn and PreGc. Between the different coding regions are tetrapeptide cleavage sites required for cleavage and release of mature proteins (Hulswit *et al*., 2021; Zivcec *et al*., 2016a) (Figure 3-2).

3.1.3.2.1 Non-structural proteins

Four non-structural proteins are encoded within the medium segment, namely, GP38, NSm, GP85 and GP160 (Leventhal *et al*. 2021). GP38 is a soluble protein encoded downstream of the MLD and does not show significant homology to known protein structures/domains preventing function prediction (Mishra *et al*. 2020). However, GP38-specific monoclonal antibodies (mAbs) have been characterised and are protective in a mouse model suggesting a potential role in pathogenesis (Bertolotti-Ciarlet *et al*., 2005; Durie *et al*., 2022; Golden *et al*., 2019) (Figure 3-4).

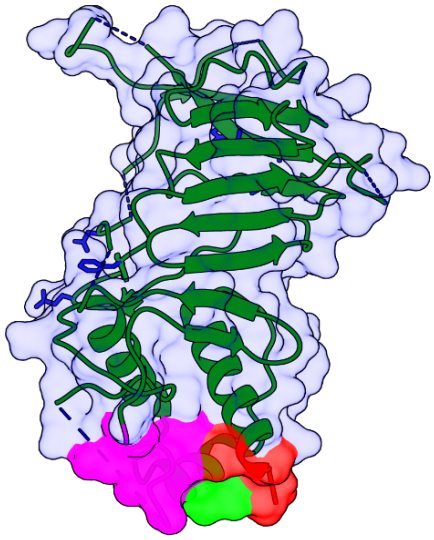


Figure 3-4.The GP38 protein.

The GP38 protein from CCHFv strain Hoti from Kosovo depicted with the three mAb binding epitopes visualised. Region 1 amino acids 253-261 (red), region 2 amino acids 289-294 (green) and region 3 amino acids 343-349 (magenta). The mAb binding sites were predicted from mAbs isolated from human survivors and demonstrated increased cross-protection between CCHFv clades compared to Anti-Gc mAbs (Durie *et al*., 2022).

NSm is a double membrane spanning protein with studies showing that mutant strains (10200ΔNSm) demonstrate wild-type replication but delayed pathogenesis in mice, improper Gc processing, defective viral particle formation and reduced secretion (Welch *et al*., 2020). NSm expression was identified to be important for N-glycosylation of glycoproteins during transport through the golgi-network and proper secretion of infectious particles (Freitas *et al*., 2020). The NSm is also reported to be under positive selection in a tick host and potentially associated with increased viral fitness in the tick host (H. Xia *et al*., 2016).

The GP38 and MLD can be cleaved from the GPC together forming two other non-structural proteins, GP160/85, which have been detected in the supernatant of infected cells (Hawman & Feldmann, 2023; Sanchez *et al*., 2006). The MLD in

CCHFv could function as observed in EBOV and prevent neutralisation by masking epitope binding sites (Fusco *et al*. 2015).

3.1.3.2.2 Structural proteins

The mature Gc and Gn glycoproteins have roles in viral tropism, generation of RNCs, and localisation of mature glycoproteins in the host cell (Zivcec *et al*., 2016a). The virion has spike-like projections from the envelope (Gn and Gc) which initiate infection by interaction with host cell receptors. It was recently identified that the low density lipoprotein receptor (LDLR) is a potential entry receptor and binds Gc (Z.-S. Xu *et al*., 2024). Entry was also identified to be via clathrin-mediated entry using early endosomes by use of small molecule inhibitor pitstop2 preventing CCHFv infection (Garrison *et al*., 2013). This evidence paired with the observation that antibodies raised against Gc, and not Gn, are protective suggest Gc is the main binding protein (Bertolotti-Ciarlet *et al*., 2005). Gc has recently been structurally characterised as a class II viral fusion protein with three domains (N. Li *et al*., 2022) and protective mAb sites have been predicted from survivors (Fels *et al*., 2021) (Figure 3-5).

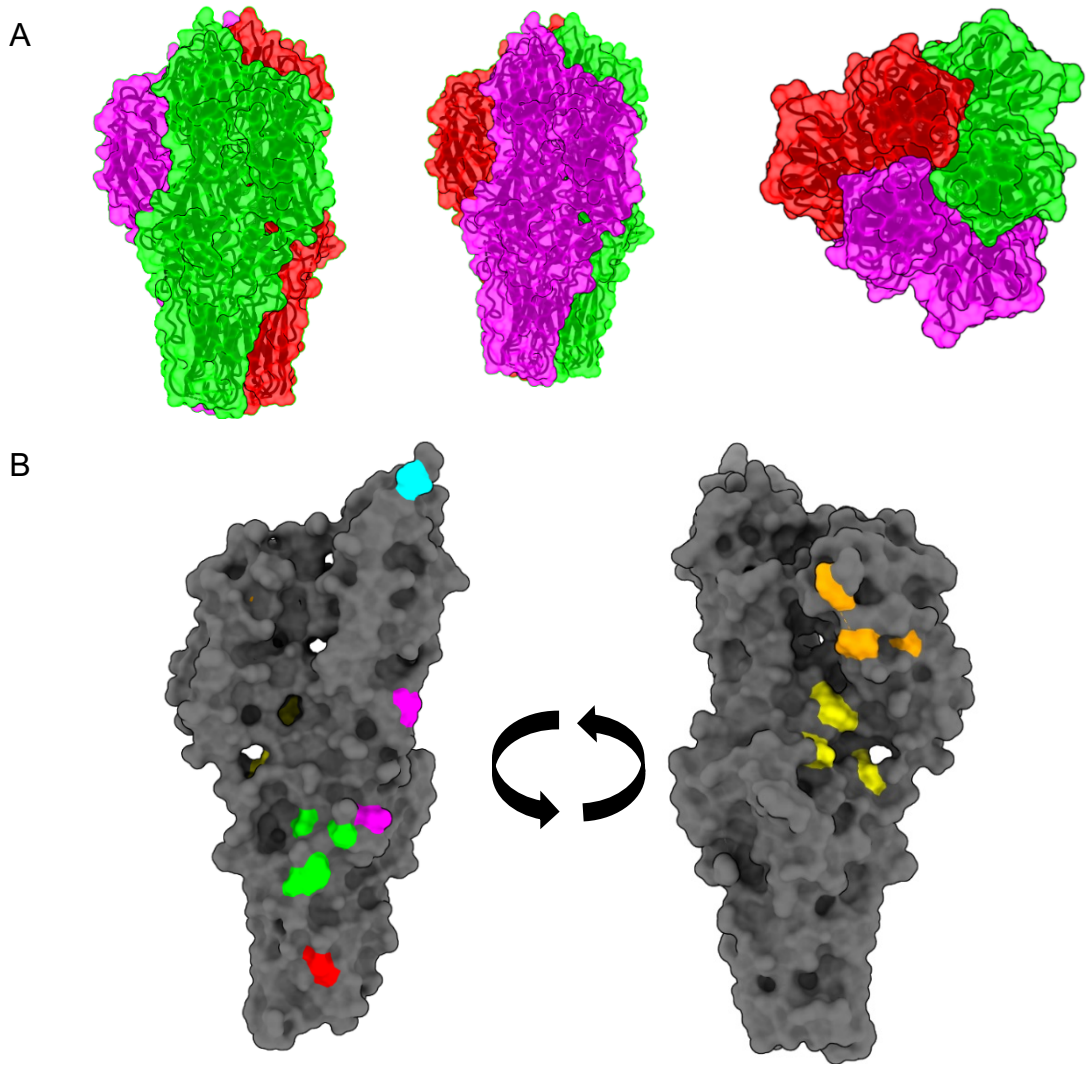


Figure 3-5. The Gc protein.

The trimeric Gc structure for CCHFv strain 10200 from front, back and top (A). Six antigenic sites predicted. Amino acid sites as follows: site 1 1185, 1199 (red), site 2 1309, 1314, 1352, 1353 (green), site 3 1307, 1229 (pink), site 4 1297, 1386,1393 (yellow), site 5 1124, 1128, 1093 (blue) and site 6 1500, 1504, 1450, 1522, 1518 and 1520 (orange) (B). from front, back and top (A). Six antigenic sites predicted. Amino acid sites as follows: site 1 1185, 1199 (red), site 2 1309, 1314, 1352, 1353 (green), site 3 1307, 1229 (pink), site 4 1297, 1386,1393 (yellow), site 5 1124, 1128, 1093 (blue) and site 6 1500, 1504, 1450, 1522, 1518 and 1520 (orange) (B).

The mature Gn protein contains two predicted N-linked glycosylation sites (557N and 755N) but N-linked glycans have only been observed at site 557N. Abrogation of glycosylation at 557N results in mis localisation in the secretory pathway and retention of GPC encoded proteins in the endoplasmic reticulum (Erickson *et al*., 2007). This suggests proper Gn maturation is required for transport of mature proteins and formation of functional virions. N-linked glycosylation is also seen to occur in the Gc protein, but abrogation of these sites did not impact localisation of the proteins. Gn expression in the absence of Gc demonstrated Gn processing and localisation is independent of Gc, but glycoprotein localisation, folding and transport are dependent on Gn N-glycosylation (Erickson *et al*. 2007).

The CTD of Gn is predicted to function in host pathogen interactions and viral assembly. The electrostatics in Gn provide a binding surface for vRNA, confirmed by EMSA using residues 729-805. The full Gn is yet to be fully structurally characterised, with structures limited to the Gn tail (Fernando Estrada & De Guzman. 2011).

3.1.3.3 The large segment

The L segment encodes the RdRp but also encodes a CCHFv-specific (within *Bunyaviricetes*) deubiquitinase of the Ovarian Tumour Proteases (OTU) family in the NTD (Figure 3-2) (Bergeron *et al*., 2010; Zivcec *et al*., 2016b). The main function of the L segment is to encode the RdRp for transcription of vRNA. The catalytic domain is residues 2342-2551, but has been reported to extend a wider region outside the catalytic domain 2043-2714 (Kaushal & Baranwal, 2023) and 2043-2714 (Mirza *et al*., 2019). Mutation of amino acids 2518-2519 (∆DD)

abrogated polymerase activity by removing the aspartic acids of the SDD polymerase motif (Bergeron *et al*., 2010). For transcription, the L segment must interact with the NP in the RNC. The 3' and 5' regions of the L protein were found to interact with the NP, but not the central region (1-1325, 2590-3945) (Macleod *et al*., 2015).

The endonuclease encoded by the L segment also has a crucial function in optimal viral replication. The endonuclease cleaves host mRNA and uses the cleaved 5' capped m7G oligonucleotides as a primer to initiate viral transcription for RdRp attachment. This prevents recognition by the RIG-1 pattern recognition receptor (PRR) of uncapped vRNA and promotes immune evasion (Habjan *et al*., 2008; Spengler *et al*., 2015). The D693 residue is predicted to coordinate Mn^{2+}, which is critical for cap snatching activity and mutating D693 abolishes transcription but doesn't impair the ability to replicate CCHFv genome analogues (Devingnot *et al*. 2015; Bergeron *et al*. 2010).

OTU domains are present in eukaryotic, viral, and bacterial proteins and function as a cysteine protease capable of deconjugating a broad spectrum of proteins modified with ubiquitin or ubiquitin-like proteins to manipulate host ubiquitin signalling (A. F. Schubert *et al*., 2020; Tchesnokov *et al*., 2020). The OTU domain has been demonstrated to hydrolyse Ubiquitin and ISG15 from cellular target proteins. There is an N-terminal extension in viral OTU (vOTU) not seen in eukaryotic OTU (eOTU) which can bind a specific hydrophobic patch of ubiquitin and results in a different ubiquitin orientation compared to an eOTU-Ub complex (Akutsu *et al*. 2011). The OTU domain has been specifically defined to inactivate cellular proteins involved in RIG-like receptors (RLR) mediated immune signalling through deubiquitinating the proteins (Van Kasteren *et al*. 2012). Additionally,

nuclear factor κ-light-chain-enhancer of activated B cells (NF-κβ) dependent signalling (activated by TNF-α via ubiquitination), an important innate and adaptive immune signalling pathway is negatively influenced by OTU domain containing proteases (Frias-Staheli *et al*. 2007).

3.1.4 CCHFv genotypes

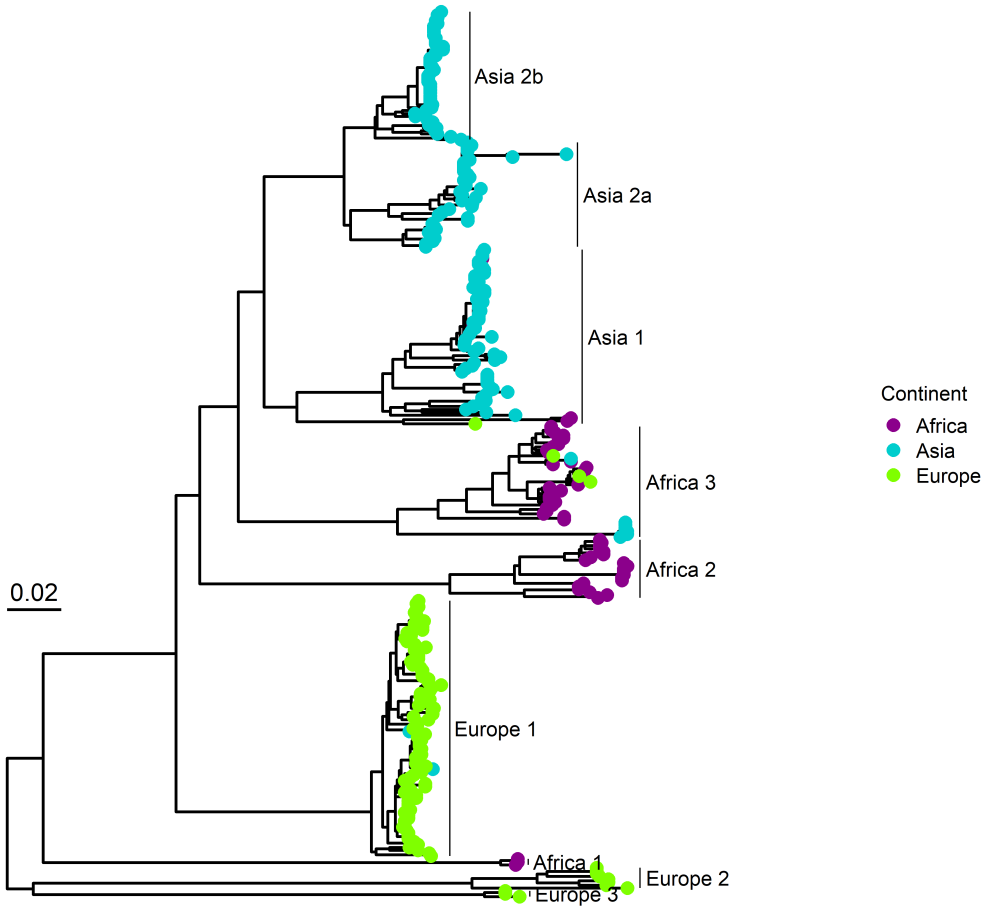


Figure 3-6. The small segment phylogeny.

The small segment phylogeny for publicly available CCHFv sequences before reclassification of Europe 2 into Aigai virus. Tips are coloured by continent.

Genotyping is based on the S segment because of its widespread application as a PCR target for diagnosis. The S phylogeny clusters CCHFv into 6 structured monophyletic clades which align with geographic location. Africa 1-3 consists of viruses isolated from Africa. Asia 1 and Asia 2 consists of viruses from Asia. Asia 2 is split into two subgroups: Asia 2a and Asia 2b. The final clade is a clade containing CCHFv from Europe (Deyde *et al*., 2006) (Figure 3-6). Until recently, there was a second Europe clade but this has been reclassified as Aigai virus (Papa (Άννα Παπά) *et al*., 2022).

3.1.5 Genomic representation of CCHFv in the public domain is an underrepresentation of the global distribution

This section will present the molecular epidemiology from the perspective of publicly available complete small segment sequences and the distribution of CCHFv across Africa, Asia, and Europe. Additionally, it will discuss other evidence indicating the presence of CCHFv, including molecular and serological detection in animal, human, and tick populations documented in the literature. Regions where genomic representation is inadequate in each continent will be highlighted. The metadata extracted from the literature will also highlight significant discrepancies in serological surveillance capabilities, diagnostic accuracy, case reporting, and tick detection across all three continents where CCHFv is reported.

For concision, and because of the reliance on the small segment for genotyping, only the small segment has been used for the comparison between genomic and

other evidence of CCHFC prevalence. A comparison of small, medium, and large clustering is in Phylogenetic congruency between segments.

3.1.5.1 CCHFv clades in Africa

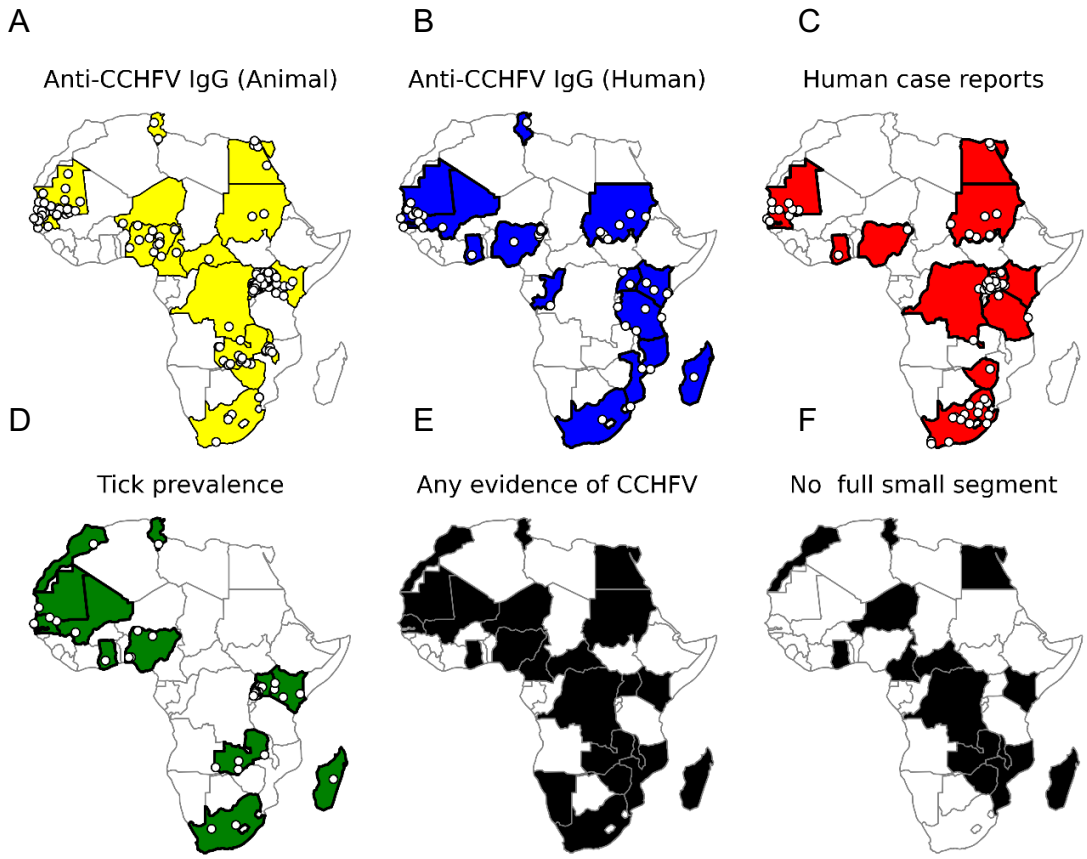


Figure 3-7. Evidence for CCHFv in Africa.

Maps showing Anti-CCHFv IgG detection in animal (A) and human (B) populations, human cases reported in the literature (C) and tick prevalence (D). Another map is used to show the total estimated distribution in Africa as a sum of all available evidence (E) and the countries with no small segment sequence data but other forms of evidence (F). In A, B, C and D, dots mean evidence for CCHFv occurrence in that region.

There are three genotypes present in Africa (Figure 3-7). Africa 1 consists of sequences collected from Senegal. Africa 2 consists of sequences collected from

Uganda, Democratic Republic of the Congo (DRC) and Republic of the Congo (RC) (Central Africa). Africa 3 consists of sequences from a much larger collection of countries from south, central, west, and north Africa (Deyde *et al.*, 2006) (Figure 3-8).

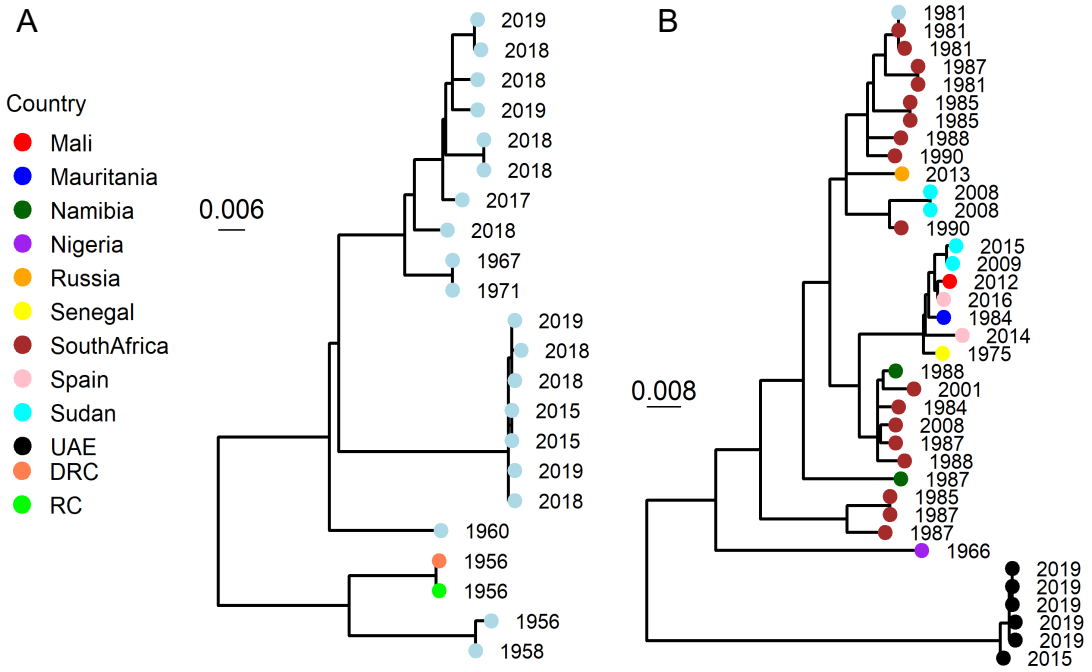


Figure 3-8. The Africa 2 and Africa 3 clades.

Maximum likelihood phylogenies for the small segment inferred under an HKY+G4+I nucleotide substitution model for Africa 2 (A) and Africa 3 (B) clades.

In Africa 2 and Africa 3, sampling bias is evident within the phylogenies. Uganda is the most represented location in Africa 2 (Figure 3-8) with most sequences originating from recent outbreaks. This increased representation is attributed to heightened surveillance and reporting efforts since 2013, following a span of 50 years without cases. (Atim *et al.*, 2023; Balinandi *et al.*, 2018, 2021, 2022; Kizito *et al.*, 2018; Nyakarahuka *et al.*, 2022). Seroprevalence studies in human and animal populations and knowledge surrounding the tick vectors are limited in

Uganda compared to other endemic locations in Asia, or Europe (Atim *et al*., 2023; Balinandi *et al*., 2021; Fisher-Hoch *et al*., 1992; Lule *et al*., 2022; Nyakarahuka *et al*., 2023). There is a lack of historical serological data in Uganda with the first report of animal seroprevalence in cattle in 2017 (Nyakarahuka *et al*., 2023) and human seroprevalence in 2020 (Lule *et al*., 2022). The Ugandan sequences cluster into three subclades (Figure 3-8).

The three historical sequences from the 1950s cluster with other historical sequences from the DRC and RC. The contemporary Ugandan sequences cluster into two separate clades indicating at least two genotypes circulating in Uganda. Despite the earliest sequences originating in the DRC and RC (previously Zaire), additional evidence of circulation is limited. However, animal seroprevalence was detected in the Katanga province of DRC in 2013 (Sas, Mertens, Kadiat, *et al*., 2017) and a second case was reported in 2008 in Berwe (Grard *et al*., 2011). No additional case reports are available for the Republic of the Congo and no human seroprevalence or positive ticks have been reported.

South Africa is the most represented country in the Africa 3 clade (Figure 3-8). The first human case was reported in 1981 with subsequent cases reported throughout the 1980s (Burt *et al*., 1996; Swanepoel *et al*., 1983, 1985, 1987; van Eeden, 1985; Vawda *et al*., 2018). It was first hypothesised that introduction into South Africa was because of the action of migratory birds prior to this index case. However widespread seroprevalence in animal populations including livestock and wild animals was discovered soon after which contradicted this hypothesis and suggested long-term circulation (Adam *et al*., 2013; Burt *et al*., 1996; Fisher-Hoch *et al*., 1992; Guilherme *et al*., 1996; Kajihara *et al*., 2021; Khan *et al*., 1997; Morrill *et al*., 1990; Msimang *et al*., 2021; Sas, Mertens, Kadiat, *et al*., 2017;

Shepherd *et al*., 1987; Zeller *et al*., 1994). Positive ticks of the *Hyalomma* and *Rhipicephalus* genus were reported at the location of the index case (Lombard reserve) (Swanepoel *et al*., 1983), which has since expanded to ticks in Free State and Northern Cape (Msimang *et al*., 2021).

The historical sequences from South Africa cluster into two distinct smaller clades. In subclade one, in addition to the South African sequences, there is one sequence each from Iran, Russia, Sudan, and Uganda indicating circulation of Africa 3 in Europe, Asia, and Southern and Northern Africa. The second subclade contains two smaller subclades: one with two sequences from Spain and Sudan and one from Mali, Senegal, Mauritania, and Uganda. The sequences in this subclade are from disparate geographic regions but are more homologous than the second smaller subclade. This second subclade contains sequences from South Africa and Uganda highlighting circulation of closely related viruses across a large geographic region.

3.1.5.1.1 Underrepresented countries in Africa with other reported evidence of circulation

Within Africa 3, nine countries are represented by 33 sequences, 18 of which are from South Africa. Although there is significant evidence in South Africa outside of genomic data, this is not representative of the evidence in the literature of circulation within Africa (Figure 3-7). Mauritania and Senegal will be used as examples of countries underrepresented by genomic data in the public domain.

The first human case in Mauritania was reported in 1983 with additional cases since (Bob *et al*., 2017; Boushab *et al*., 2020; J. P. Gonzalez *et al*., 1990; Kleib *et al*., 2016; Nabeth *et al*., 2004; Saluzzo *et al*., 1985; Sow *et al*., 2014). There is

also evidence for circulation in animal, human and tick populations (J. P. Gonzalez *et al*., 1990; Nabeth *et al*., 2004; Saluzzo *et al*., 1985; Sas, Mertens, Isselmou, *et al*., 2017; Schulz *et al*., 2021) (Figure 3-9).

Despite this, Mauritania is represented by one sequence from 1984 which clusters with sequences from Spain in 2014 and 2016 (Figure 3-8). This provides evidence for the action of migratory birds in the transfer of CCHFv from West Africa to Spain (Figure 3-9).

Human cases were first reported in Senegal in 2003/04 and more recently in 2017-19 (Dieng *et al*., 2020; Jauréguiberry *et al*., 2005; Tall *et al*., 2009). Animal and human seroprevalence were first detected in Senegal in 1969 (J.-P. Gonzalez *et al*., 1990; Mangombi *et al*., 2020; Wilson *et al*., 1990; Zeller *et al*., 1994) and 1987, respectively (Chapman *et al*., 1991; Mhamadi *et al*., 2022; Wilson *et al*., 1990). CCHFv has also been detected in *Hyalomma*, *Boophilus*, *Amblyomma*, and *Rhipicephalus* ticks (Mhamadi *et al*., 2022; Zeller *et al*., 1997) (Figure 3-9).

Senegal is the sole location for Africa 1 with two sequences from ticks and one from a goat. The location of Senegal in Western Africa suggests it is unlikely that the Africa 1 is limited to those foci and the absence of representation is likely an artefact of sampling and surveillance efforts (Figure 3-8).

Across Africa there are many countries which are underrepresented in the public domain. Of particular interest are regions in the north and west because of proposed action of migratory birds. Additionally, countries North of Uganda in Eastern Africa e.g., Sudan and Egypt have both contemporary and historical evidence of CCHFv while being part of a flyway into central Asia and being near

land borders into the middle east which have been associated with importation of CCHFv into Saudi Arabia (Hassanein *et al.*, 1997).

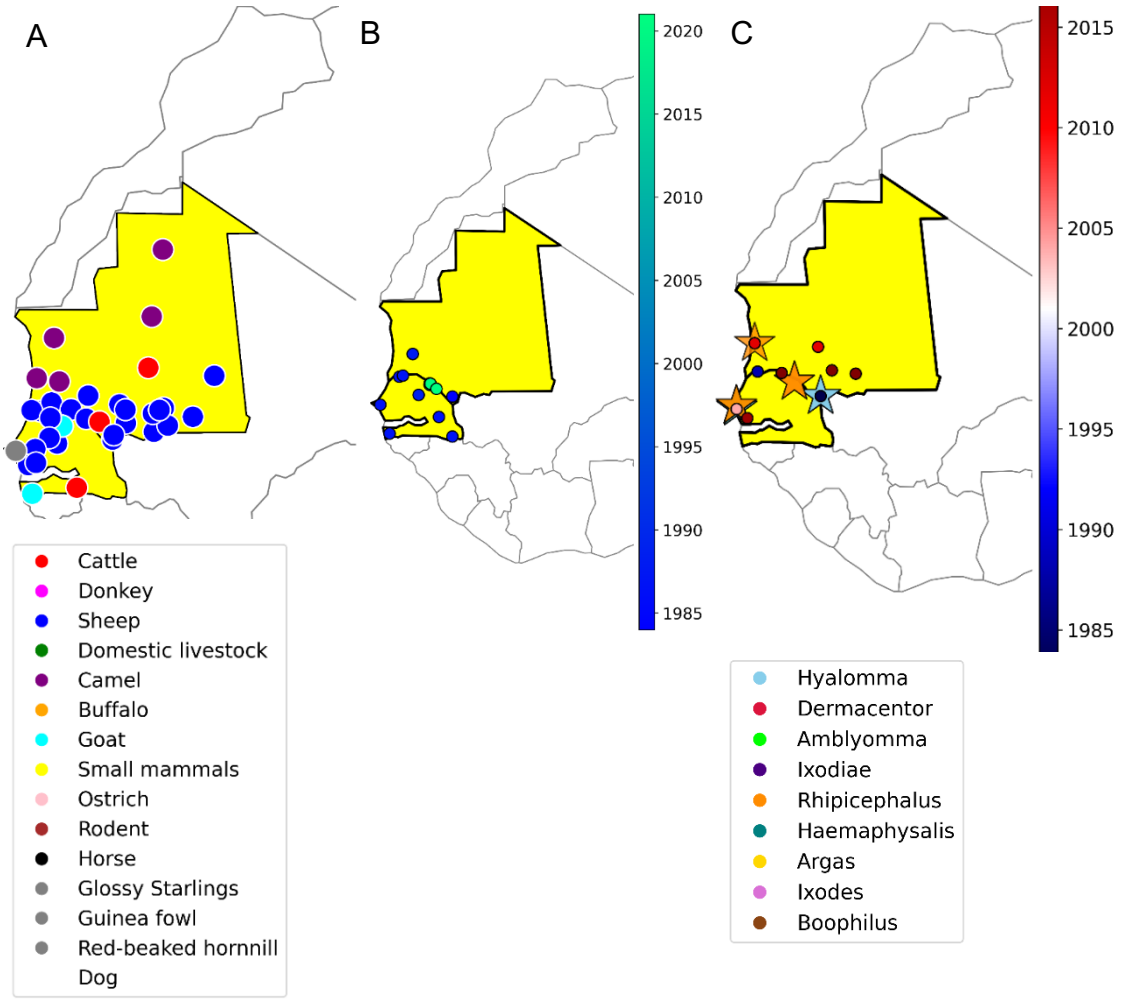


Figure 3-9. Evidence for CCHFv in Senegal and Mauritania.

Maps showing animal seroprevalence (A), human seroprevalence (B) and human cases overtime with tick prevalence (C) for Senegal and Mauritania in West Africa. Dots represent evidence of CCHFv in that area.

3.1.5.1.2 African countries with evidence of CCHFv without representation

12 other African countries have reported evidence for CCHFv in the last 70 years but are without genomic representation. Animal seroprevalence has been

detected in Egypt (Darwish *et al*., 1977; Mohamed *et al*., 2008), Niger (Mariner *et al*., 1995), Kenya (Obanda *et al*., 2021), Cameroon (González Gordon *et al*., 2022), Malawi (Phonera *et al*., 2021), Tunisia (Bouaicha *et al*., 2021; Khamassi Khbou *et al*., 2021; Zouaghi *et al*., 2021) Zambia (Kajihara *et al*., 2021; Lysholm *et al*., 2022), and Zimbabwe (J.-P. Gonzalez *et al*., 1990). Human seroprevalence has been reported in Kenya (Lwande *et al*., 2012; Tigoi *et al*., 2015), Ghana (Akuffo *et al*., 2016), Tunisia (Wasfi *et al*., 2016), Tanzania (Rugarabamu *et al*., 2021), Mozambique (Muianga *et al*., 2017) and Madagascar (Shahhosseini *et al*., 2021b).

CCHFv positive ticks have been reported in Kenya in *Rhipicephalus* and *Hyalomma* genus (Chiuya *et al*., 2021; Sang *et al*., 2011; Shahhosseini *et al*., 2021b), in Ghana in *Amblyomma*, *Hyalomma* and *Boophilus* (Akuffo *et al*., 2016), in Tunisia in *Hyalomma* (Bouaicha *et al*., 2021), in Zambia in *Hyalomma* (Kajihara *et al*., 2021) and in Morocco from *Hyalomma* collected from migratory birds (Palomar *et al*., 2013).

3.1.6 CCHFv clades in Asia

The Asian clades are more heterogeneous with regards to geographic location. Surveillance and sequencing efforts have also been more extensive across more countries compared to the African clades (Figure 3-10). Asia 1 is the largest Asian clade with 61 tips. The Asia 1 clade has an outgroup of 3 sequences from Nigeria,

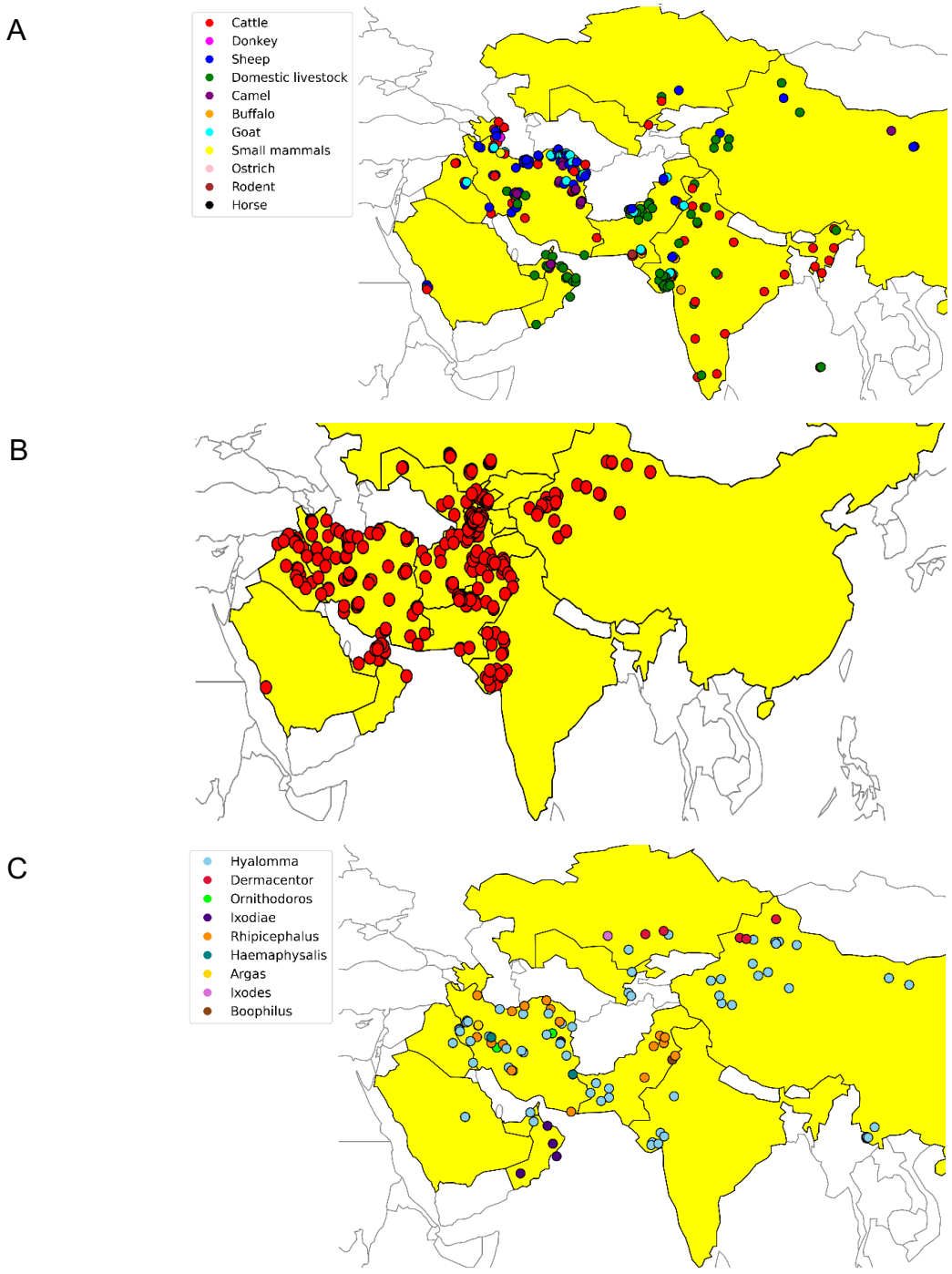


Figure 3-10. Evidence for CCHFv in Asia.

Maps showing different types of evidence for CCHFv prevalence in Asia. Animal seroprevalence (A), human cases (B) and tick prevalence (C). Each dot represents a report of CCHFv prevalence in that area. Dots are not scaled for prevalence reported in each report for

Central Africa Republic (CAR) and Spain, “Africa 4”. This suggests that an Asia 1-like genotype is circulating on all continents with endemic foci. The main Asia 1 clade contains three subclades. One subclade at the base of Asia 1 contains genetically diverse sequences compared to the rest of Asia 1, a second subclade contains mostly sequences from Indian ticks, and third subclade containing a more homogeneous viral population from countries in the Middle East and India (Figure 3-11).

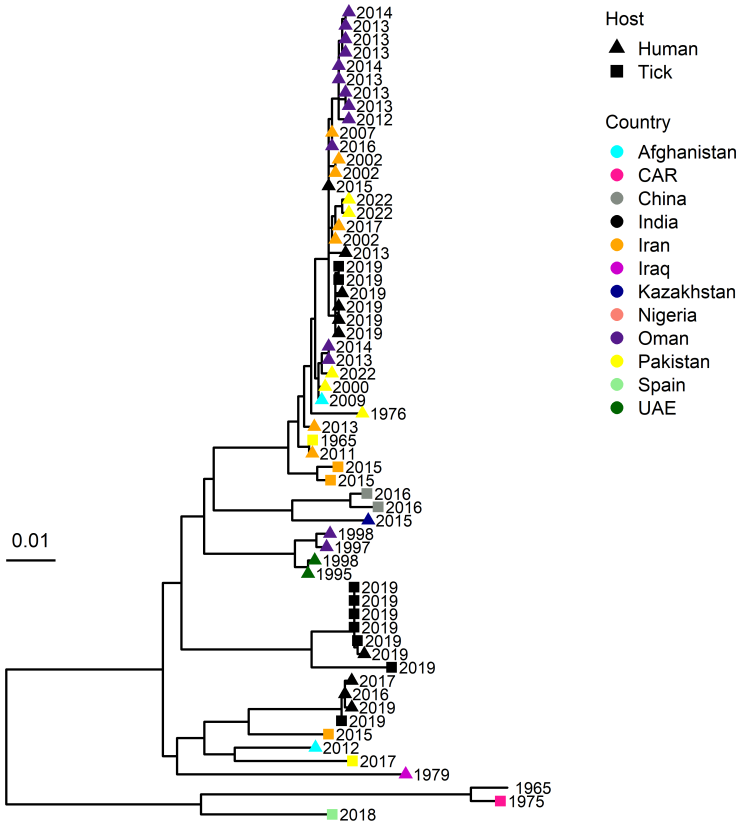


Figure 3-11. The Asia 1 clade.

A maximum likelihood phylogeny for Asia 1 small segment clade with tips coloured by geographic location and shaped by host. The empty node in the protruding outgroup is the Nigerian goat sequence novel to this study.

The Asia 2a and Asia 2b clades have a comparable topology with regards to geographic location. In both, a monophyletic clade containing sequences from China sit at the base of the clade with sporadic sequences from central Asian countries before a comparatively large cluster of homogeneous viruses from India (Figure 3-12).

In all three clades, India is over-represented by contemporary sequences resulting from extensive sequencing of outbreaks since 2011. All Indian CCHFv sequences in the public domain are from two Northern states, Gujarat and Rajasthan, potentially because of land borders to other endemic regions (Appannanavar & Mishra, 2011; Gandhi *et al*., 2011; Makwana *et al*., 2015; Mishra *et al*., 2011; Mourya *et al*., 2012, 2017; Sahay *et al*., 2021; P. Singh *et al*., 2016; Tripathi *et al*., 2020; P. D. Yadav, Cherian, *et al*., 2013; P. D. Yadav, Raut, *et al*., 2013; P. D. Yadav, Gurav, *et al*., 2014; P. D. Yadav, Raut, *et al*., 2014; P. D. Yadav *et al*., 2016).

Despite this, animal and human seroprevalence were first reported in southern states Kerala and Pondicherry in 1976 (Shanmugam *et al*., 1976), where cases still haven't been reported. Since 2011, livestock and human serosurveillance studies have highlighted seropositivity within different India states (Mourya *et al*., 2012, 2014, 2015; Sahay *et al*., 2021) and circulation is known in many tick genus (Kaul *et al*., 1990; Mourya *et al*., 2012; Sahay *et al*., 2020, 2021; P. D. Yadav, Cherian, *et al*., 2013) (Figure 3-13).

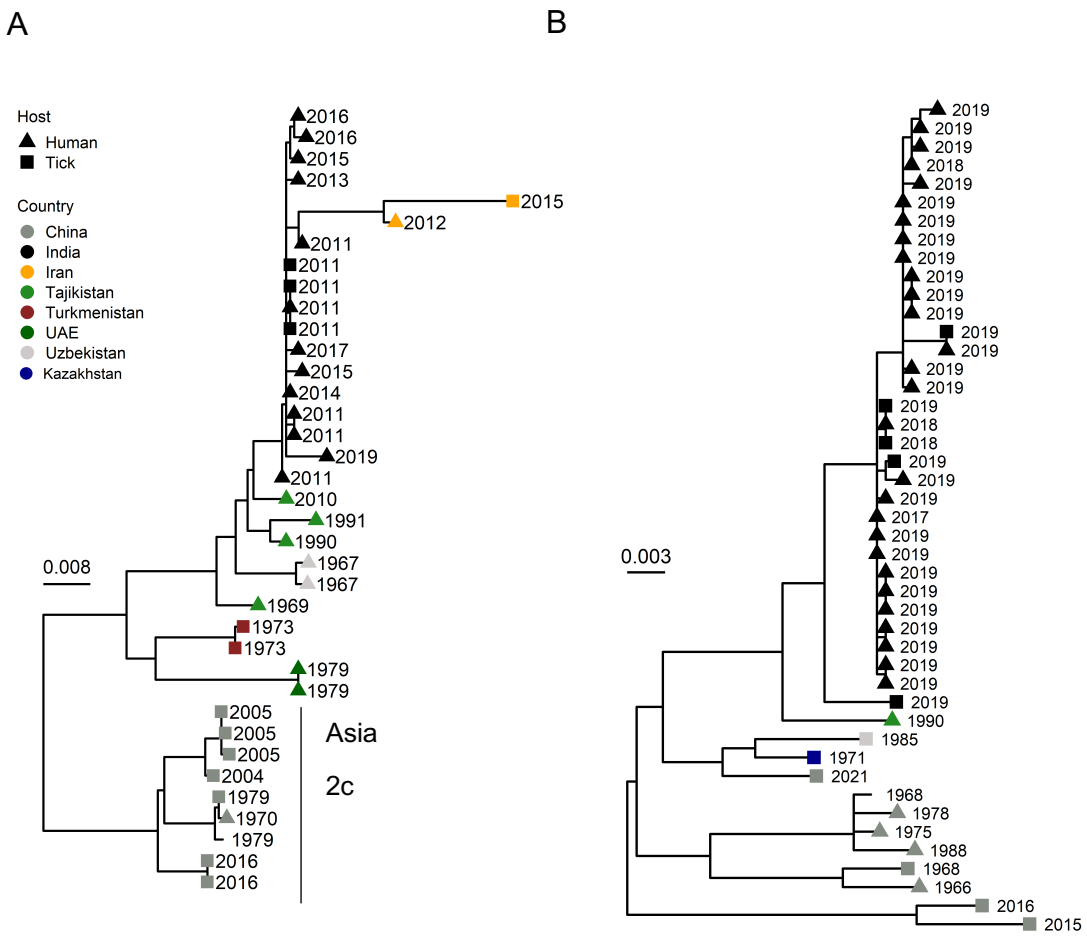


Figure 3-12. The Asia 2 clades.

Maximum likelihood phylogenies for Asia 2a (with tentatively named Asia 2c containing sequences from China) (A) and Asia 2b (B).

Outbreaks in India are unique compared to other endemic regions as genomic representation is available for each reported outbreak (2011, 2013-2019). Phylogenies indicate a mixed viral population circulating within the region. Viruses from Gujarat and Rajasthan collected between 2011 and 2019 cluster in both Asian clades (Figure 3-11; Figure 3-12). The metadata attached to these sequences highlight highly similar viruses belonging to both Asian clades isolated from human and tick hosts in Gujarat and Rajasthan highlighting the importance of tick bites as a risk factor for transmission.

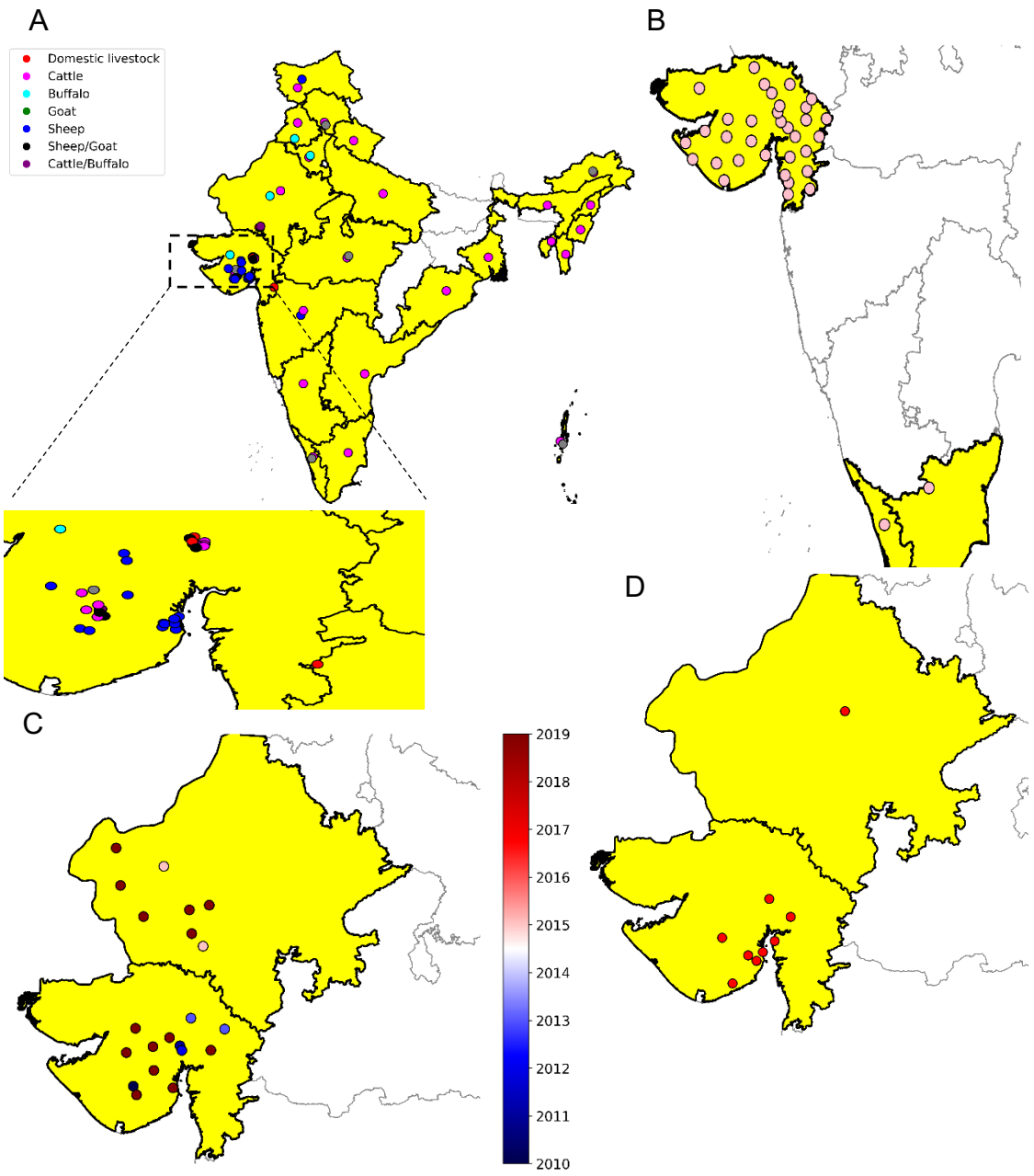


Figure 3-13. Evidence for CCHFv in India.

Maps showing animal seroprevalence in India in animal populations (A), human seroprevalence (B), human case numbers (C) and tick prevalence (D). Each dot represents a report of CCHFv prevalence in that area. Dots are not scaled for prevalence reported in each report for clarity. In Panel C and D only Gujarat and Rajasthan are shown due to lack of reports for human cases and tick prevalence in the other regions.

China is comparable to India with regards to representation in the public domain but has sequences from both historical and contemporary timepoints. The North-

Western region of China, Xinjiang, is a hotspot with most of the evidence for CCHFv found in the region (Dai *et al*., 2006; X. Gao *et al*., 2010; R. Guo *et al*., 2017; Han *et al*., 2002; Kong *et al*., 2022; Y. Li *et al*., 2020; Lv *et al*., 2006; Moming *et al*., 2018; Morikawa *et al*., 2002; Papa *et al*., 2002; Qing *et al*., 2003; Saijo *et al*., 2004; Spengler *et al*., 2016; Sun *et al*., 2009; Q. Tang *et al*., 2006; H. Xia *et al*., 2011; Yen *et al*., 1985, 1985; Y. Zhang *et al*., 2018) (Figure 3-14).

On the other hand, some central Asian and middle eastern countries are underrepresented in the public domain despite evidence of circulation in animal, human and tick populations. There is a major language barrier to assessing historical records and publication regarding the circulation in Central Asian countries but this has been well summarised by Fereidouni *et al* (2023) which was used as a reference for this section for the distribution at earlier dates prior to publication in English (Fereidouni *et al*., 2023).

The southern part of Central Asia, the Middle East and western part of China demonstrate circulation in many animal and tick species and human cases (Figure 3-14). From the current predicted phylogenetic topology, it can be hypothesised that the central Asian region may act as a conduit for dissemination of CCHFv from China to the Middle East and vice versa.

This could be attributed to the geographic barrier posed by the Himalayas (Couret *et al*., 2022), preventing direct introduction from China into India. There is also a potential pathway for introduction from Eastern Africa into the Middle East or vice versa via land borders and sea ports, specifically because of importation of livestock during Eid celebrations (el-Azazy & Scrimgeour, 1997; Hassanein *et al*., 1997; Rodriguez *et al*., 1997). However, the sampling bias present in the current

dataset e.g., overrepresentation of China and India and limited representation for some Middle Eastern and Central Asian countries, prevent conclusions with certainty as to the evolutionary pathway that CCHFv has undergone throughout Asia.

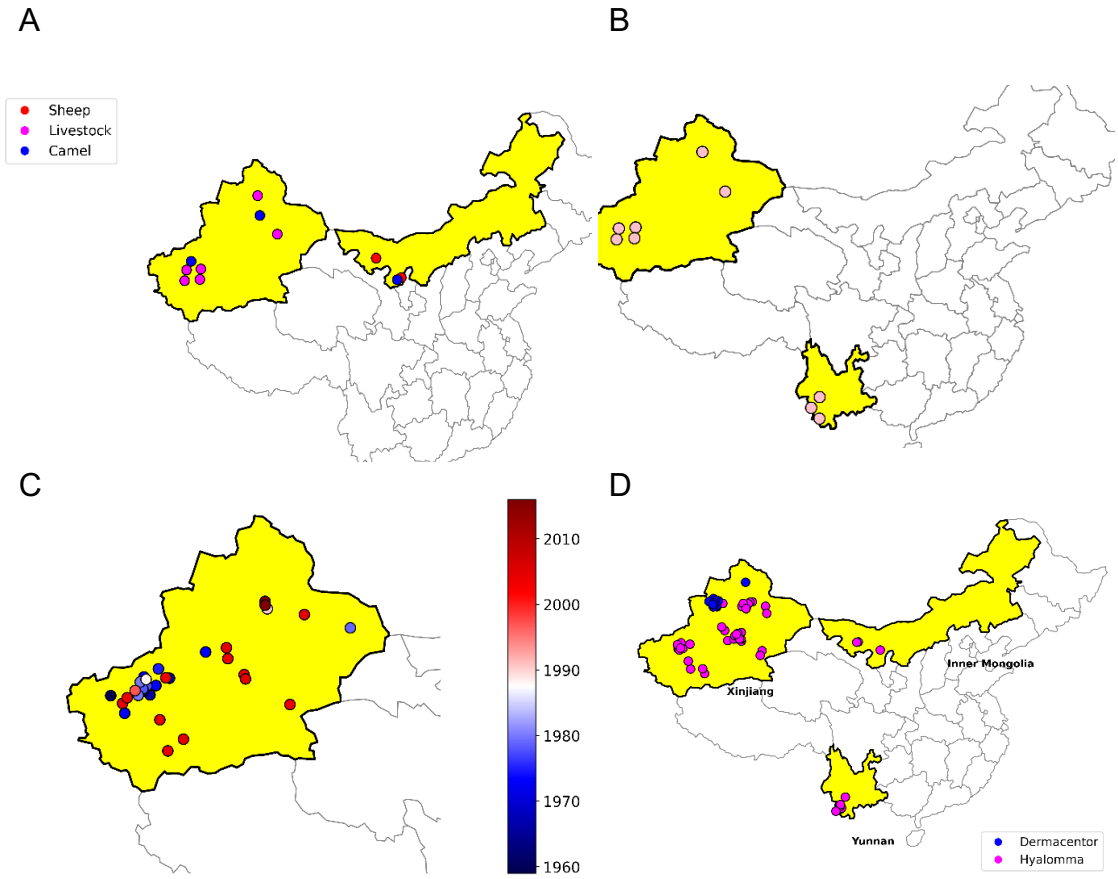


Figure 3-14. Evidence for CCHFv in China.

Maps showing animal seroprevalence (A), human seroprevalence (B), human cases (C) and tick prevalence in regions of China. Each dot represents a report of CCHFv prevalence in that area. Dots are not scaled for prevalence reported in each report for clarity.

3.1.7 CCHFv clades in Europe

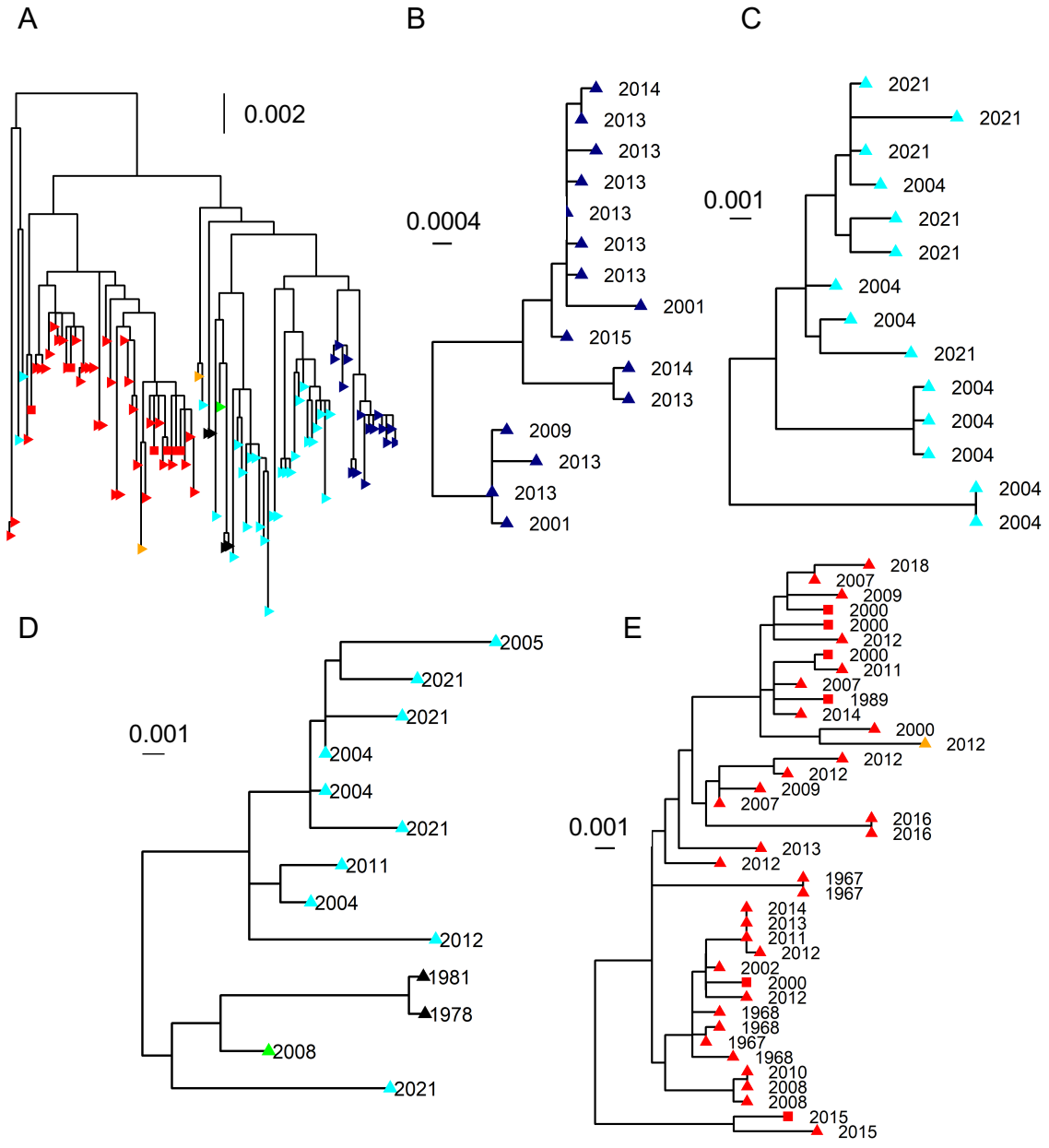


Figure 3-15. The European clade.

Individual maximum likelihood phylogenies inferred under an HKY+G4+I nucleotide substitution model using the small segment sequences for the entire European clade (A), Kosovo clade (B), Turkey 1 (C), Turkey 2 (D) and Russian clade (E).

The European clade has sequence representatives from Russia, Turkey, Kosovo, Bulgaria, Greece, and Iran (Figure 3-15). The three main foci where CCHFv is reported in Europe are Russia, Turkey, and Kosovo. All three form monophyletic clades representing limited exchange within the regions. As with the Asian and African clades, some European countries are also underrepresented by the genomic data (Figure 3-16). but there is a more even distribution of sequences across the main regions. Additionally, two sequences from Bulgaria are located at the tip of a clade opposed to the base because of accelerated evolution associated with cell culture (Figure 3-15).

3.1.7.1 CCHFv in Spain

Spain is one of the most recent countries to have reported the onset of autochthonous cases. In 2010, a positive *Hyalomma lusicatium* tick collected from Red Deer was reported from Caceres, Western Spain (Estrada-Peña, Palomar, *et al*., 2012). The first autochthonous case was recorded in Avila in 2013 (Negredo *et al*., 2017). Cases have since been recorded in 2016 (Avila (1), Madrid (2)), 2018 (Badajoz (1), Salamanca (2)), and 2020 (Avila (1), Salamanca (3)) (Monsalve Arteaga *et al*., 2021; Negredo *et al*., 2021).

A seroprevalence study in North-western Castile-y-leon demonstrated low seropositivity, comparable to that observed in Southern Portugal and Hungary in the 1980s (Filipe *et al*., 1985; Horváth, 1976). This suggests that asymptomatic circulation in the human population is still limited but should be expected after identification in tick populations.

Several tick survey studies have been conducted since positivity in *Hyalomma* populations in Spain in 2010, finding positive ticks from Red Deer, Wild Boar, Fallow Deer and Roe Deer (Cajimat *et al*., 2017; Moraga-Fernández *et al*., 2021; Sánchez-Seco *et al*., 2022).

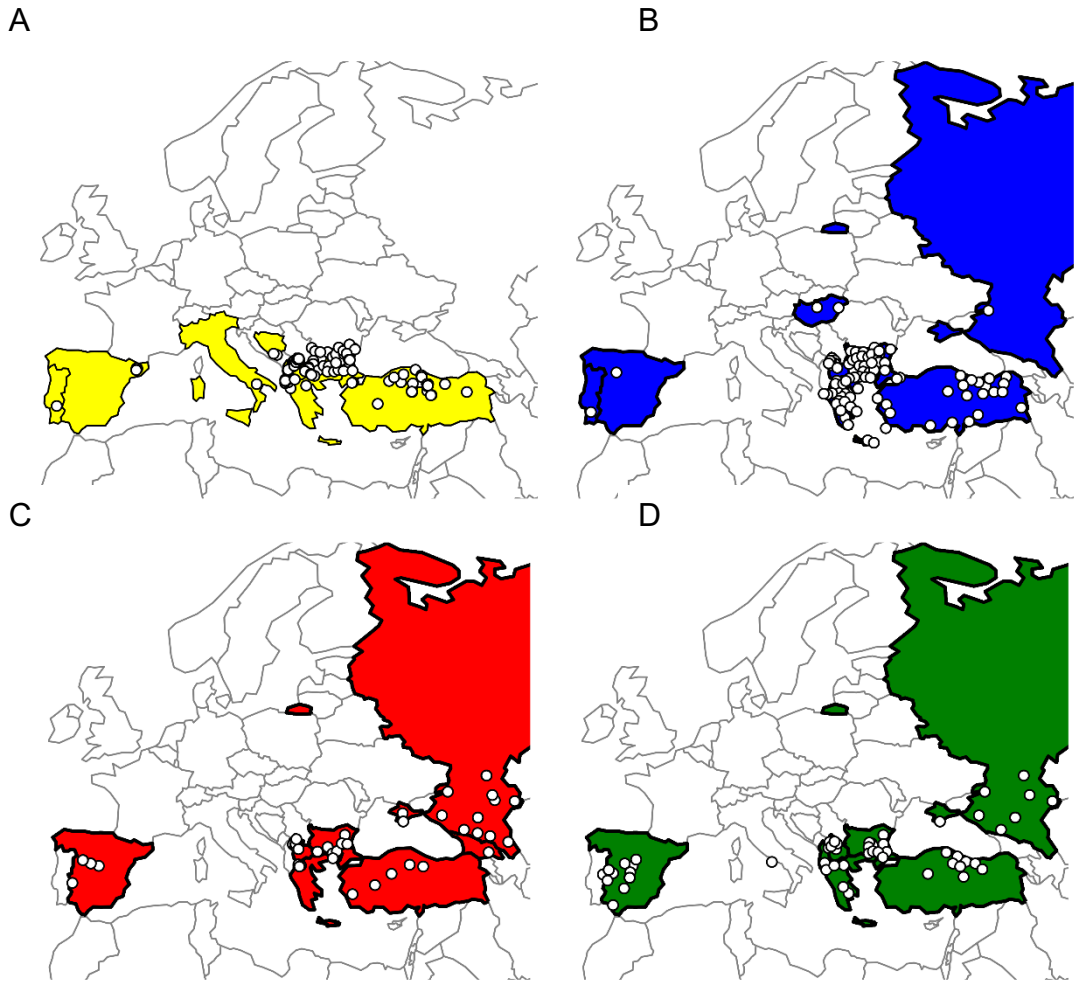


Figure 3-16. Evidence for CCHFv in Europe.

Maps showing animal seroprevalence (A), human seroprevalence (B), human case reports (C) and tick prevalence (D) in Europe.

Three different genotypes have been reported in Spain (Lorenzo Juanes *et al.*, 2023). In the first cases, the Africa 3 genotype was detected (Figure 3-8). After this, the Africa 4 (Asia 1 outgroup) was detected (Figure 3-11). The European clade has also been detected but with full segment representative not available. The presence of Africa 3, Asia 1 outgroup (Africa 4) and Europe 1 highlights circulation of diverse genotypes in Spain.

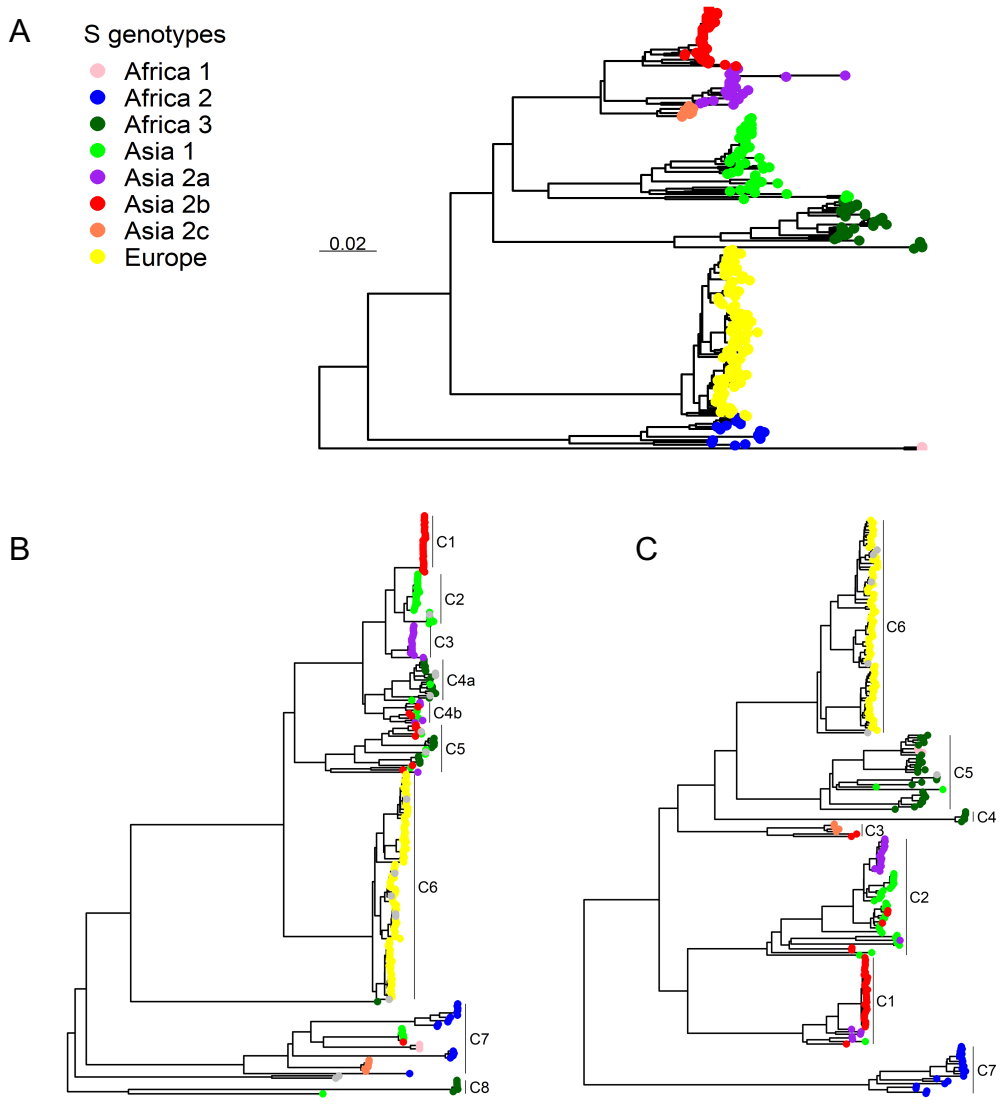


Figure 3-17. S, M and L phylogenies.

Maximum likelihood phylogenies for the S (A), M (B) and L (C) segments inferred under an HKY+G4+I nucleotide substitution model coloured by small clade genotyping.

3.1.8 Phylogenetic congruency between segments

The geographic clustering observed in the S segment phylogeny is conserved within some M clades (Figure 3-17). The European clade and most of the Asia 1, Asia 2a, Asia 2b, and Africa 2 clades are conserved. The Africa 3 clade is divided between two smaller clades, C4a and C5 with strains from Asia. M segment C4 (MC4) contains two subclades, MC4a one with Africa 3 and Asia 1 and MC4b with CCHFv from Asia 1, Asia 2a and 2b. Another more divergent clade, MC5, contains sequences from Africa 3 and the three Asian clades. MC7 contains Africa 2 and Africa 1 in addition to sequences from Asia 1 and the Asia 2c clade. The large segment clustering more closely aligns with the small segment clades, with the formation of larger clades for sequences from Asia and Africa suggesting more conservation in the L segment in these clades/regions (Figure 3-17).

3.2 Results

3.2.1 Novel CCHFv sequences were generated by metagenomic NGS

Forty-three, thirty-four and thirty-three novel sequences were generated for the S, M and L segments, respectively (Appendix 7). Many novel sequences are from an outbreak in Turkey in 2015. Novel sequences were generated from samples originating from Pakistan, Kosovo, Democratic Republic of Congo, Nigeria, Uganda, and Tajikistan (Table 3-1). In total, 29 full genomes were generated from metagenomic sequencing: one from Kosovo, Pakistan, and the DRC, and 26 from Turkey.

Sample	Location	Date	Host	segments
D1126	Pakistan	1998/99	Human	S, M, L
Kos2009	Kosovo	2009	Human	S, L
S3409	Kosovo	2009	Human	S, M, L
3011	DRC	1956	Human	S, M, L
AR7620	Nigeria	1965	Goat	S
SE7239	Uganda	1971	Human	S, L
SE7300	Uganda	1967	Human	S, L
Khodzha	Tajikistan	1975	Human	S
Kos.2475	Kosovo	2009	Human	M, L

Table 3-1. Metadata for non-turkish novel sequences.

In the final dataset there are 333, 259 and 233 sequences from the S, M, and L segments (Appendix 8). Sampling spanned from 1956 to 2022, with an increasing density of samples towards the present day (Appendix 9). Most sequences were derived from human cases, but all segments have >45 sequences from ticks (Appendix 10). Both Asia and Europe were comparably sampled but Africa demonstrates >50% less sequences, representing a large sampling bias within the dataset.

3.2.2 Novel sequences cluster within Asia 2a, the Asia 1 outgroup and the European clade

A

Sample	S	M	L
Khodzha	Asia 2a	NA	NA
D1126	Asia 2a	Asia 2a (clade 2)	Asia 2a
Tur.1009	Europe (Russia)	Europe (Russia)	Europe (Russia)
AR7620	Africa 4 (Asia 1 out)	NA	NA
SE7300	Africa 2	NA	Africa 2
SE7329	Africa 2	NA	Africa 2
3011	Africa 2	Africa 2 (clade 7)	Africa 2
Kos2009	Europe (Kosovo)	NA	Europe (Kosovo)
S3409	Europe (Kosovo	Europe (Kosovo)	Europe (Kosovo)
Kos.2475	Europe (Kosovo)	Europe (Kosovo)	Europe (Kosovo)
Other Turkey	Europe (Turkey)	Europe (Turkey)	Europe (Turkey)

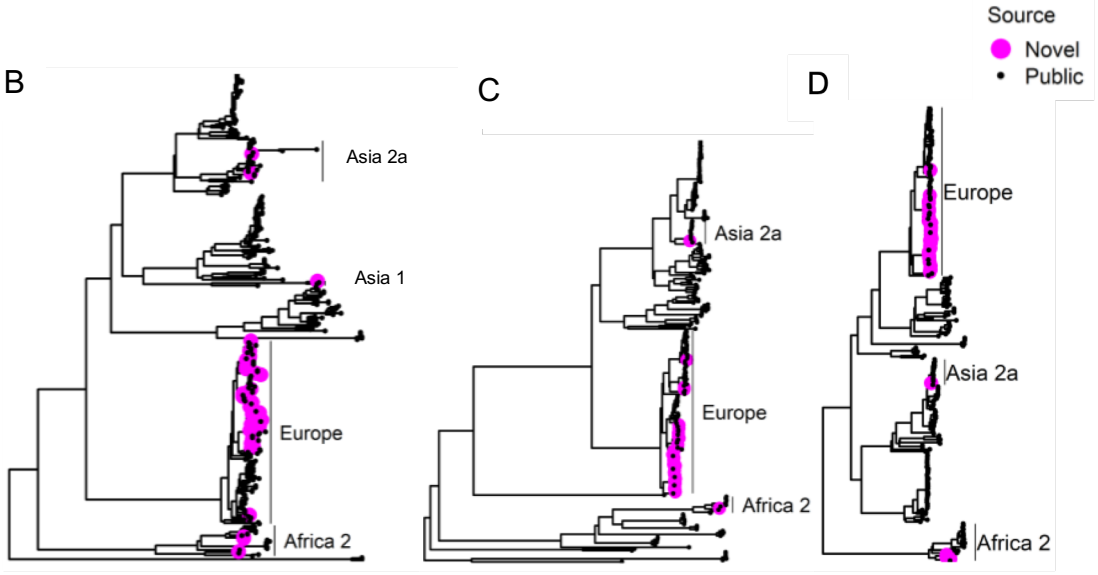


Figure 3-18. The metadata and clustering of novel sequences.

A table containing clade assignments for the novel sequences generated in the study. NA represents segments with no sequence (A). Maximum likelihood phylogenies inferred under an HKY+G4+I nucleotide substitution model highlighting the placement of novel sequences in the S (B), M (C) and L (D) segments.

For the small segment, 37 sequences were added to the European clade, three into Africa 2, two into Asia 2a and one into Asia 1. For the medium segment, 32 sequences were added to the European clade with the same clustering pattern observed in the small segment. One sequence (D1126) was added to clade 3 and one sequence to Africa 2. Clade 3 corresponds to the Asia 2a clade. For the L segment 29 sequences have been added to the European clade following the same clustering pattern as for the S and M segments.

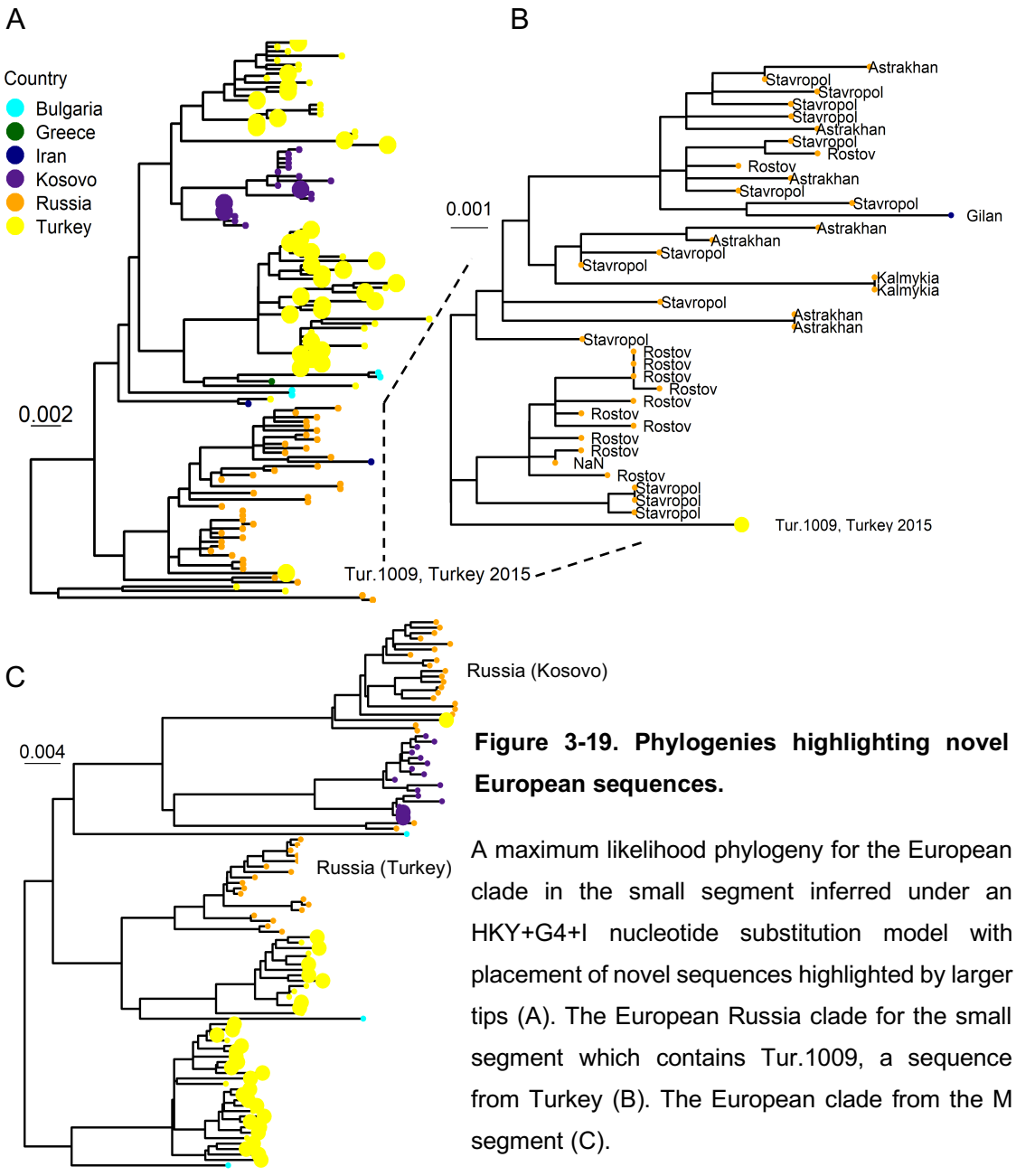


Figure 3-19. Phylogenies highlighting novel European sequences.

A maximum likelihood phylogeny for the European clade in the small segment inferred under an HKY+G4+I nucleotide substitution model with placement of novel sequences highlighted by larger tips (A). The European Russia clade for the small segment which contains Tur.1009, a sequence from Turkey (B). The European clade from the M segment (C).

Specifically, D1126 has been added to Asia 2a, and 3011, SE7239, and SE300 added to Africa 2 (Figure 3-18). Within the European clade all but one of the novel Turkish sequences' clusters in previously characterised Turkish subclades. Tur.1009 clusters with the Russian subclade in all three segments (). In the M segment phylogeny, Tur.1009 clusters with the Russian subclade most related to sequences from Kosovo than Turkey. The three added Kosovan sequences cluster in the Kosovo subclade (Figure 3-18).

Three historical samples from Porton Down have been incorporated into the Africa 2 containing clade. Among these, sample 3011, retrieved from the Democratic Republic of Congo (DRC) in 1956, cluster closely with UG3010,

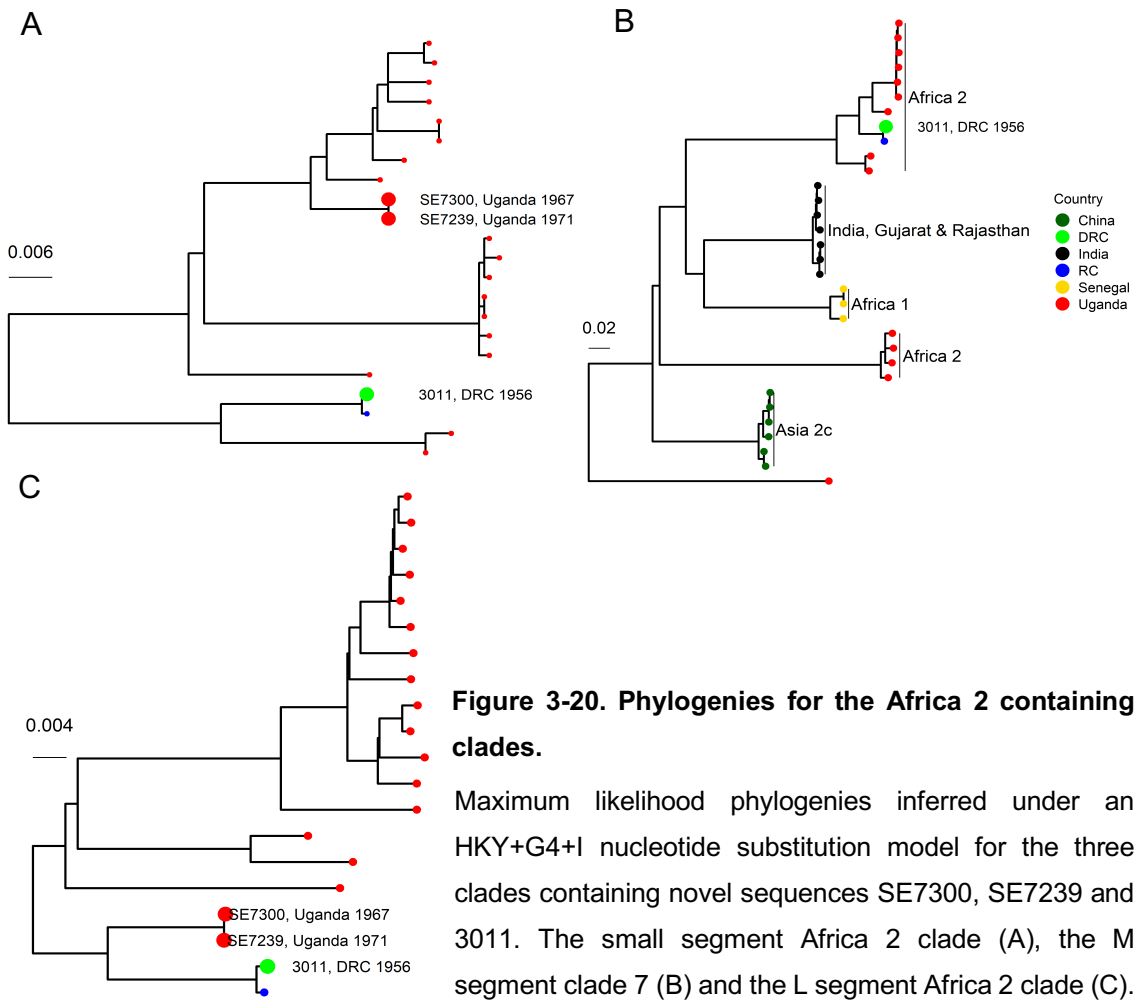


Figure 3-20. Phylogenies for the Africa 2 containing clades.

Maximum likelihood phylogenies inferred under an HKY+G4+I nucleotide substitution model for the three clades containing novel sequences SE7300, SE7239 and 3011. The small segment Africa 2 clade (A), the M segment clade 7 (B) and the L segment Africa 2 clade (C).

obtained from the Republic of the Congo in the same year. Additionally, SE7300 and SE7239 have also been added into this clade, both collected from Semunya in Uganda in 1967 and 1971 respectively. These sequences cluster at the base of one of the Ugandan clades alongside contemporary Ugandan samples. In terms of the L segment, these sequences cluster with other historical counterparts within the clade (Figure 3-20).

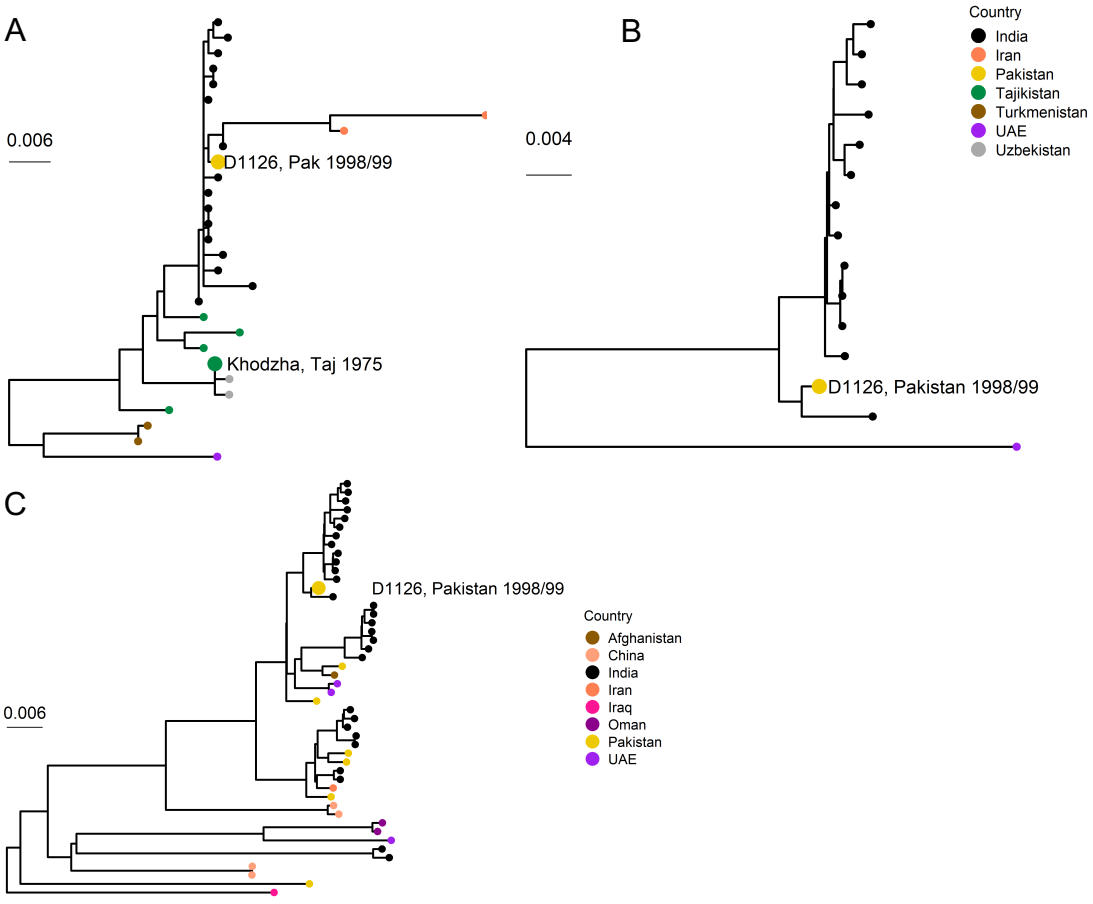


Figure 3-21. Phylogenies for the Asia 2a containing clades.

Maximum likelihood phylogenies inferred under an HKY+G4+I nucleotide substitution model highlighting the position of D1126 in the Asia 2a clades for the small (A), medium (B) and large (C) segments.

Within the Asia 2a subclade D1126 clusters closely with a sequence from a patient sample collected in India (Gujarat) in 2011 in all segments. Khodzha clusters closely with Hodzha, a sequence from a fatal infection in Uzbekistan in 1967. This provides genomic evidence of related CCHFv circulating in Pakistan and India and Tajikistan and Uzbekistan (Figure 3-21). In the M segment, D1126 is positioned at the base of MC3 providing a historical date for that clade (Figure 3-21).

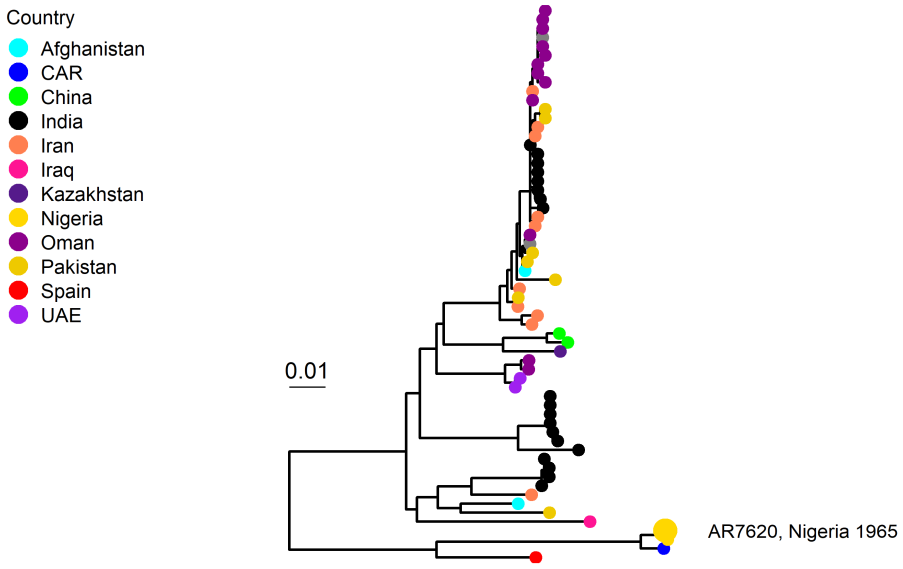


Figure 3-22. A Phylogeny for the Asia 1 clade.

Maximum likelihood phylogenies for the Asia 1 clade inferred under an HKY+G4+I nucleotide substitution model with the novel Nigeria sequence in its outgroup.

MC3 otherwise contains sequences from human samples collected from Gujarat and Rajasthan and ticks from Gujarat and thus provides evidence for the introduction of CCHFv from Pakistan into Northern India. Also, at the base of MC3 is a sequence from a patient sample collected from Dubai in 1979. Although part of the clade, this sequence is contained within its own terminal branch, indicating its disparity from the rest Asia 2a, specifically in the medium segment (Figure

3-21). Within the Asia 1 outgroup (also called “Africa 4”) which contains sequences from Nigeria, CAR, and Spain, one new sequence is added, AR7620. This was generated from a serum sample taken from a goat in Nigeria in 1965. AR7620 has no M and L sequence (Figure 3-22).

3.2.3 Detection of recombinant CCHFv strains

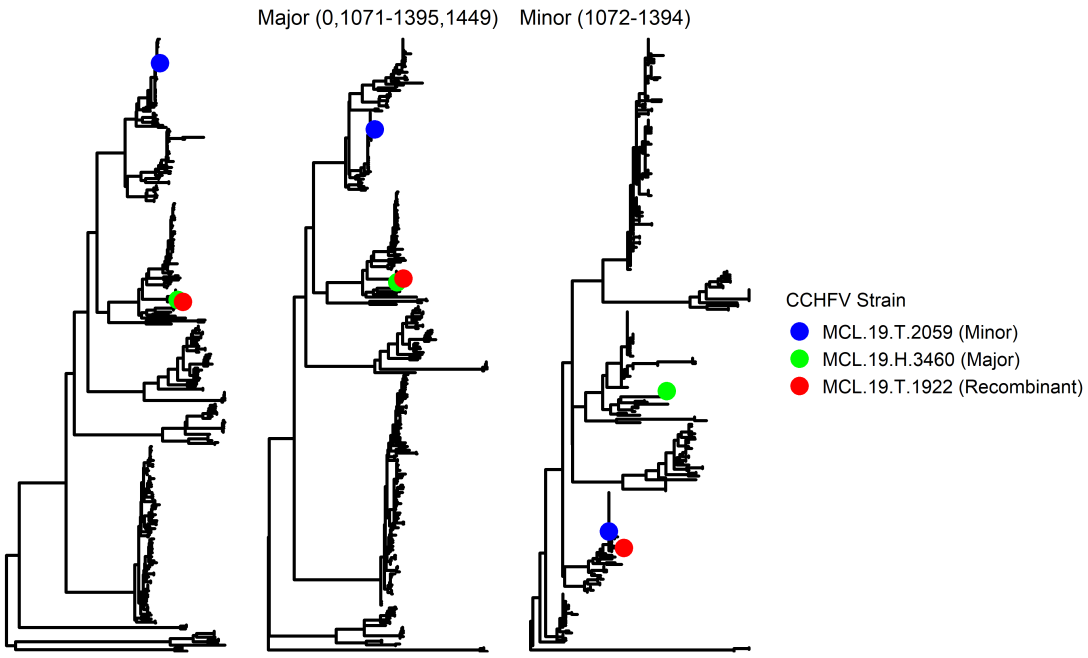


Figure 3-23. Small segment recombination candidates.

Phylogenetic representation of the small segment recombinant, MCL.19.T.1922 (MN866211). The minor parental strain is predicted to be MCL.19.T.2059 (MN930422) and major parental strain MCL.19.H.3460 (MW298552).

No recombination events were detected in the novel sequences generated in this study. In the small segment, RDP4 detected one recombinant sequence: MCL.19.T.1922. This sequence was derived from a tick collected in Rajasthan

from India in 2019 and clusters in Asia 1. The minor parental strain is predicted to be from a tick from Gujarat in India in 2019 which clusters in Asia 2b. The major parental lineage is predicted to be from a human patient in Gujarat collected in 2019 which clusters in Asia 1 (Figure 3-23).

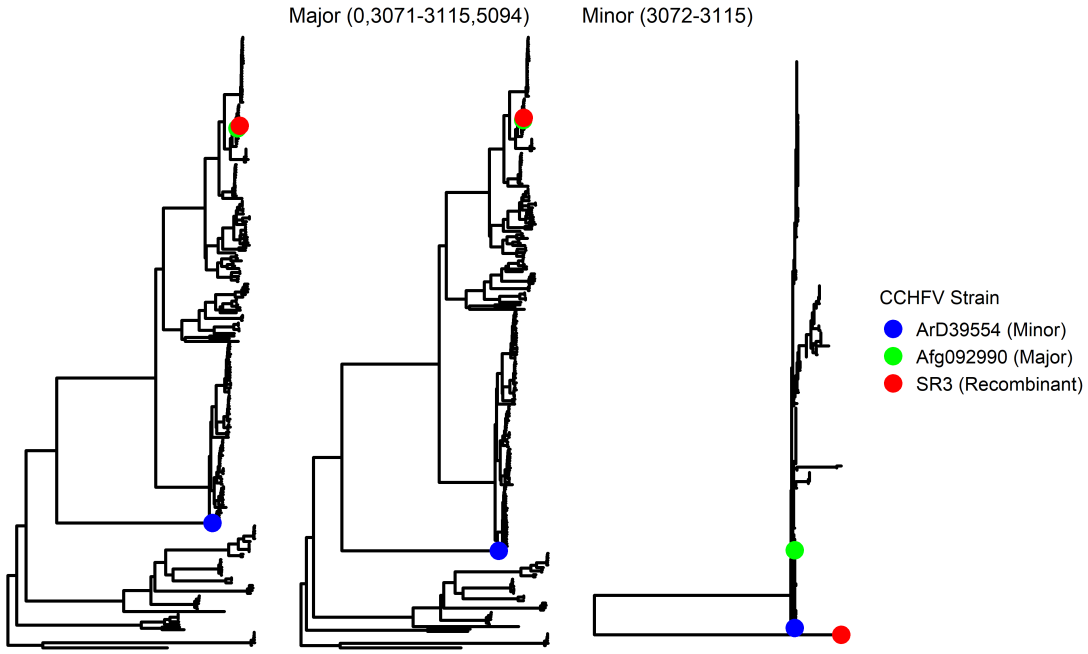


Figure 3-24. Medium segment recombination candidates.

Phylogenetic representation of the potential medium segment recombinant, SR3 (AJ538199). The minor parent is predicted to be ArD39954 (DQ211628) and major parent Arg092990 (HM452306).

In the medium segment, RDP4 detects three potential recombinant events, but only one with detection in ≥ 3 methods. Strain SR3 was collected from a human from Karachi in Pakistan in 2000. Further investigation demonstrated SR3 is likely an erroneous sequence and will be removed (Figure 3-24).

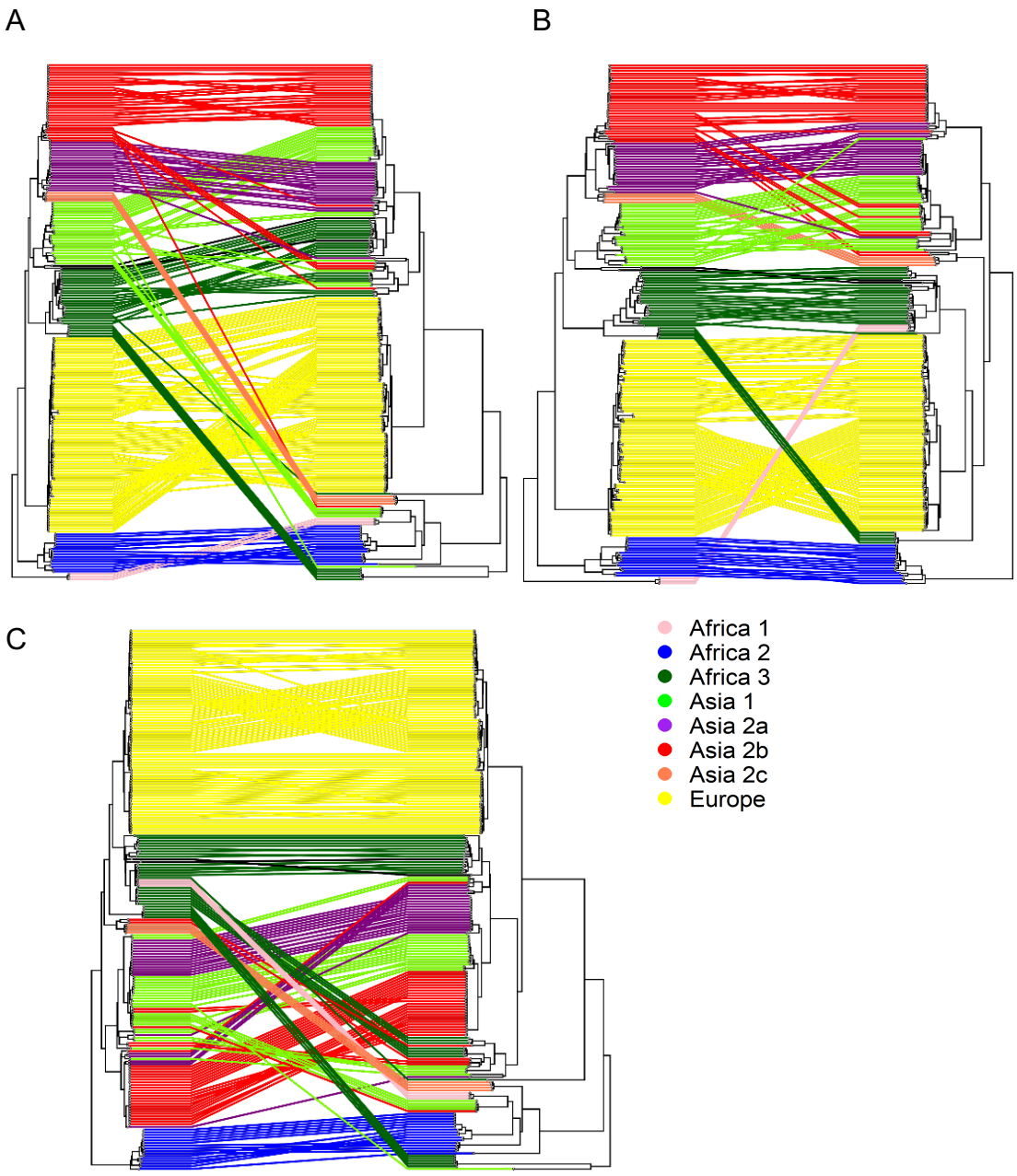


Figure 3-25. Phylogenetic incongruence between segments.

Phylogenetic tanglegrams for the S vs M (A), S vs L (B) and M vs L (C) segments. Joining lines are coloured by small segment clade.

In the large segment, three different breakpoints were identified in the same triplet of sequences (Semunya (DQ076412), UG3010.RC (DQ211624) and Nakiwogo

(KX013483)). From this recombination analysis, five strains were removed from the dataset for each segment used for further analysis.

3.2.4 CCHFv demonstrates evidence of extensive reassortment

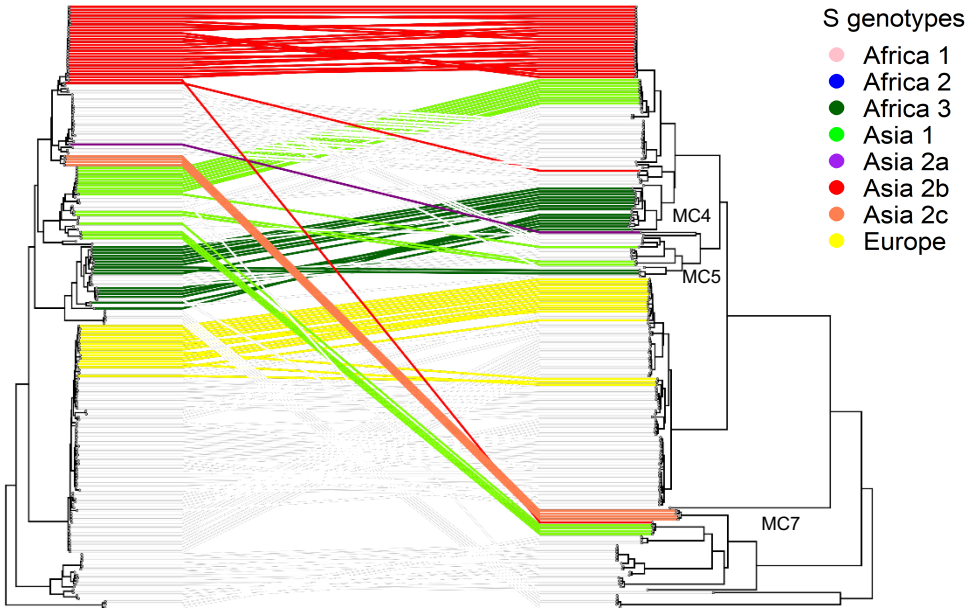


Figure 3-26. Medium segment reassortment candidates.

A tanglegram of the S and M segment phylogenies with candidate M segment reassorted taxa as identified by GiRaF highlighted. The S segment phylogeny is provided on the left and M segment phylogeny on the right. Candidate taxa are joined by lines coloured by small segment clade.

Reassortment events can account for some of the observed phylogenetic incongruence between the three segments. Two methods were used for the detection of reassortment events in the curated dataset. As is known for CCHFv, the majority of reassortment events predicted involve exchange of the M segment

demonstrated by increased incongruence between the S and M and M and L phylogenetic topologies (Figure 3-25).

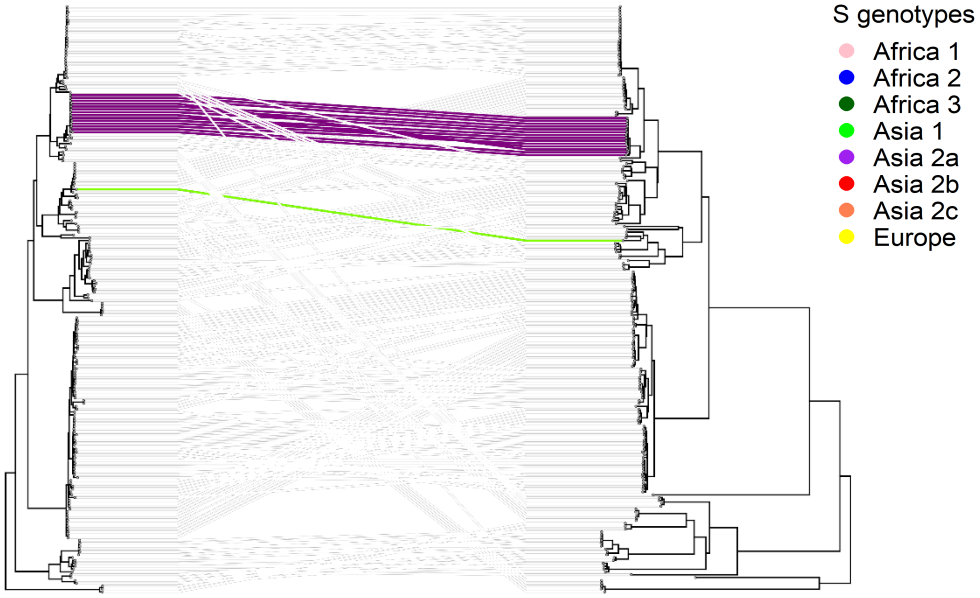


Figure 3-27. Small segment reassortment candidates.

A tanglegram of the S and M segment phylogenies with small segment candidate reassorted taxa highlighted.

M segment reassortment was detected in 80 unique strains, categorised into 14 different sets. The three heterogeneous M segment clades, MC4, MC5 and MC7 contain reassorted taxa from Asia 1, Asia 2a, Asia 2b and Asia 2c. Outside of theses clades, most of Asia 2b, a subsection of Asia 1 and taxa within the Russian subclade are also predicted to demonstrate incongruence in line with reassortment events (Figure 3-26). S segment reassortment was detected in a section of Asia 2a suggesting acquisition of a central Asian-like S segment at an ancestral timepoint. A lone strain, Matin, from Pakistan was also identified as a potential S segment reassorted strain (Figure 3-27).

L segment reassortment was detected in 76 strains but many are duplicates of M segment reassorted strains. This is supported by the topological congruence between the S and L. The Africa 1 clade is identified as an L segment reassortment event, clustering closely to other Northern African strains for the L segment (Figure 3-28).

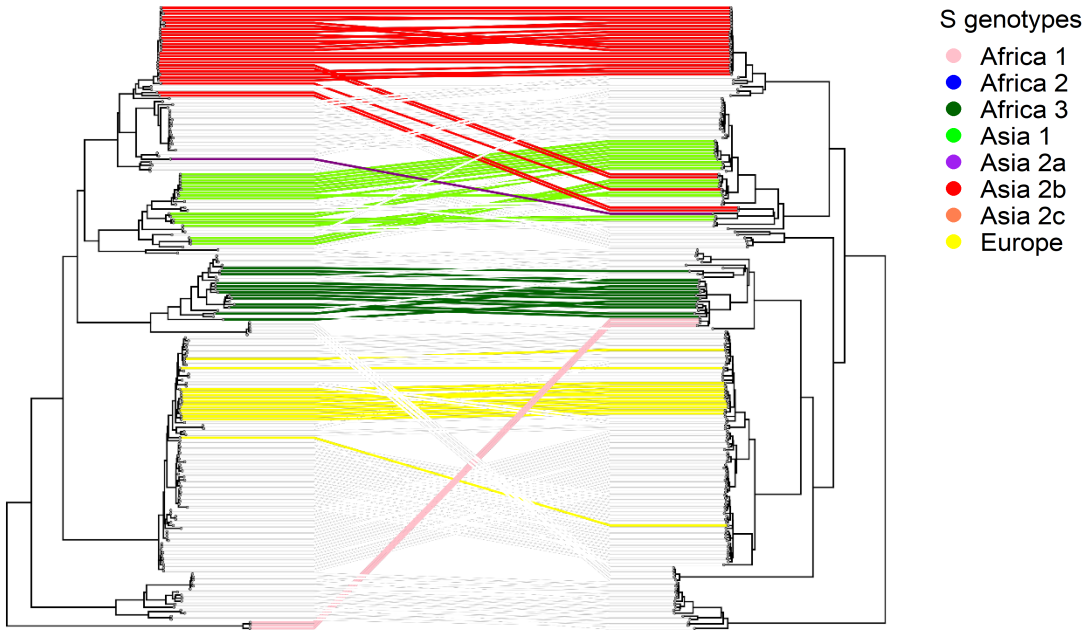


Figure 3-28. Large segment reassortment candidates.

A tanglegram of the S and L segment phylogenies with the L segment reassorted taxa highlighted. To identify which are incorrectly identified as L segment reassortment events and represent M segment reassortment events the L segment candidates are visualised with the S segment. Congruence between the S and L phylogenies is indication of M segment reassortment.

A

	Clade	Segment	Start	End	Minor	Major
DAK8194	Africa 1	L	6636	18473*	ArD39554	Nakiwogo
Tur.2050	Europe	M	1456	6626*	Tur.2066	S3409
3809ST2007	Europe	M	1436	6717	Unknown	Ast137
81R2014	Europe	M	1305*	6638	46ST2012	3756ST2007

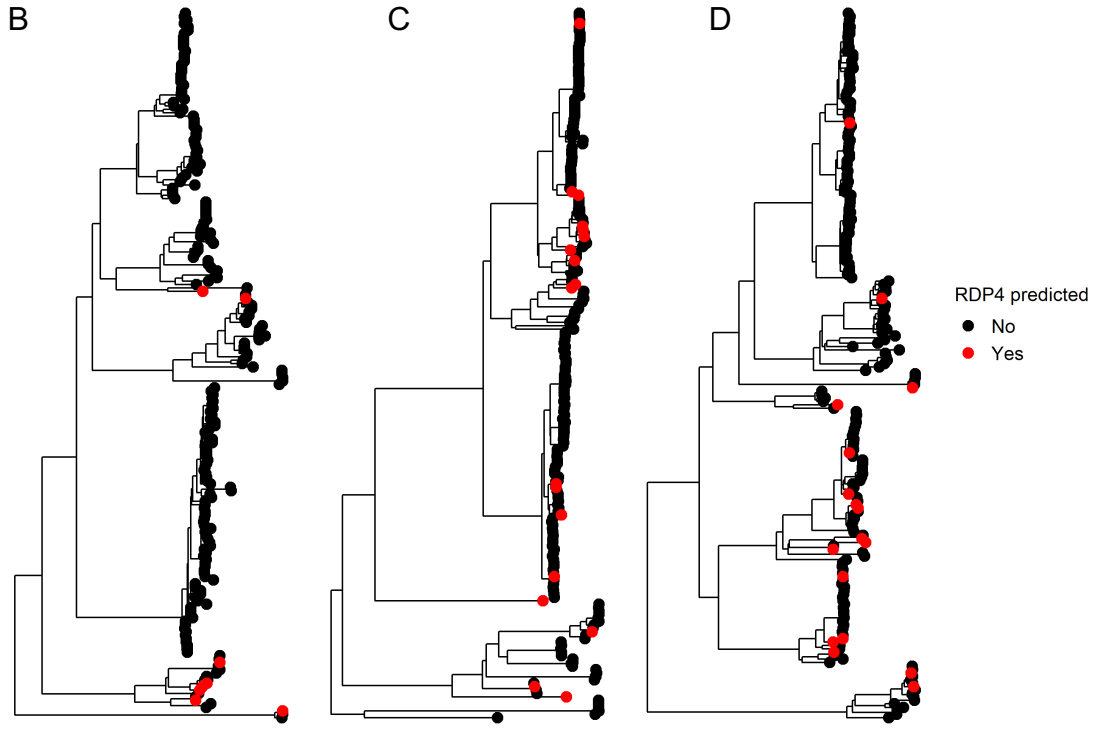


Figure 3-29. Candidate reassorted taxa identified by RDP4.

A table describing reassortment events in which the reassorted taxa is confidently predicted (A). Phylogenies for the S, M and L segments with predicted reassorted taxa predicted by RDP4 highlighted (B, C, D). If a reassortment event was predicted but the recombinant was potentially misidentified, the minor and major parental virus are considered as potential reassortments and highlighted. "*" indicates uncertainty for the breakpoints identified by RDP4.

RDP4 predicts 8, 19 and 20 reassortment events for the S, M and L segments, respectively. Four reassortment events were identified with high confidence detection of the correct reassorted taxa. The remaining events represent predicted reassortment events with uncertainty surrounding which sequence in

the triplet has resulted from reassortment. Novel genome Tur.2050 is identified as a M segment reassortment (Figure 3-29).

3.2.5 Testing for selection across the three segments

3.2.5.1 Detecting sites under selection in the three segments using codon-based models

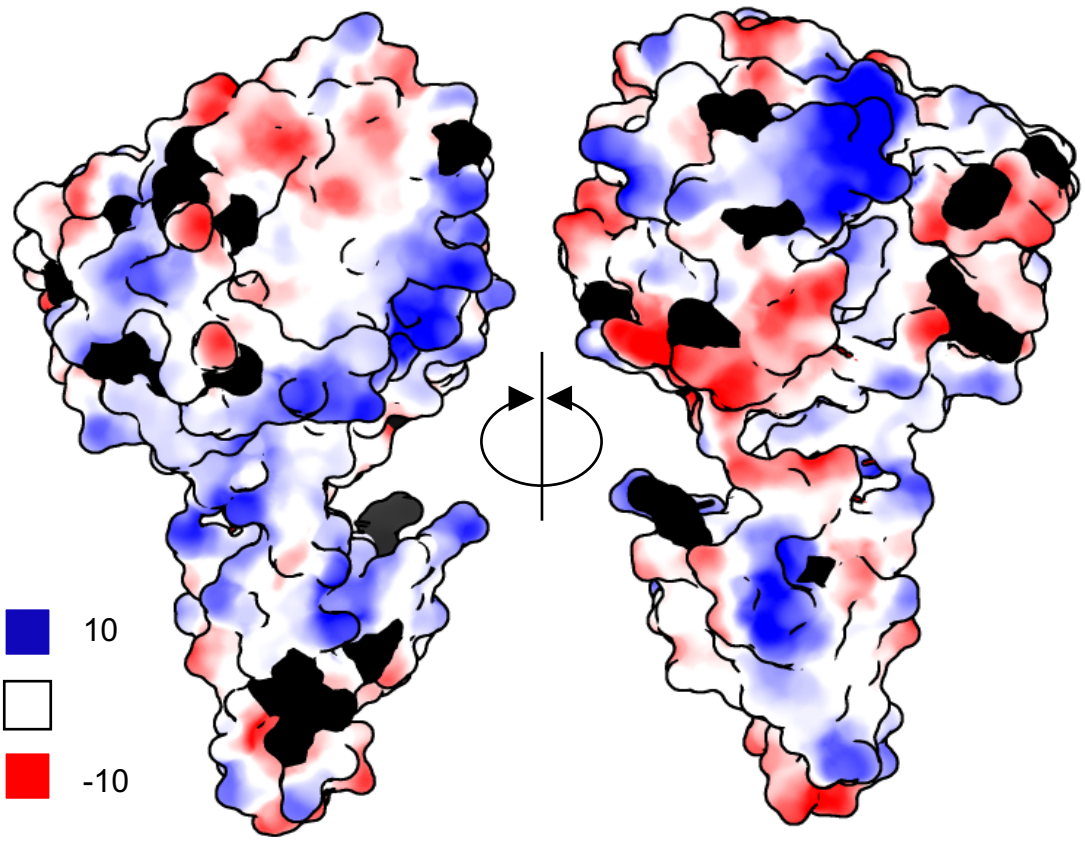


Figure 3-30. The nucleoprotein sites under negative selection.

The CCHFv nucleoprotein coloured by electrophoretic charge. Blacked out residues represent residues not under negative purifying selection as identified by FUBAR.

FUBAR was used to test for pervasive selection at the codon level. No sites were predicted to be under positive selection (posterior=>0.9). Sites 30 (ω=2.67, BF=39, posterior=0.881) and 125 (ω=2.03, BF=19.5, posterior=0.78) are under

positive selection at a reduced stringency. 447 sites are observed to be under negative, purifying selection (w=<1, posterior=>0.9) (Figure 3-30).

A limitation of FUBAR is that a period of episodic positive selection can be masked by a period of conservation, whereas MEME can identify episodic positive selection. MEME identifies six codon sites are under positive selection (LRT p >0.05), sites 11, 208, 256, 302, 319, 350 (Figure 3-31).

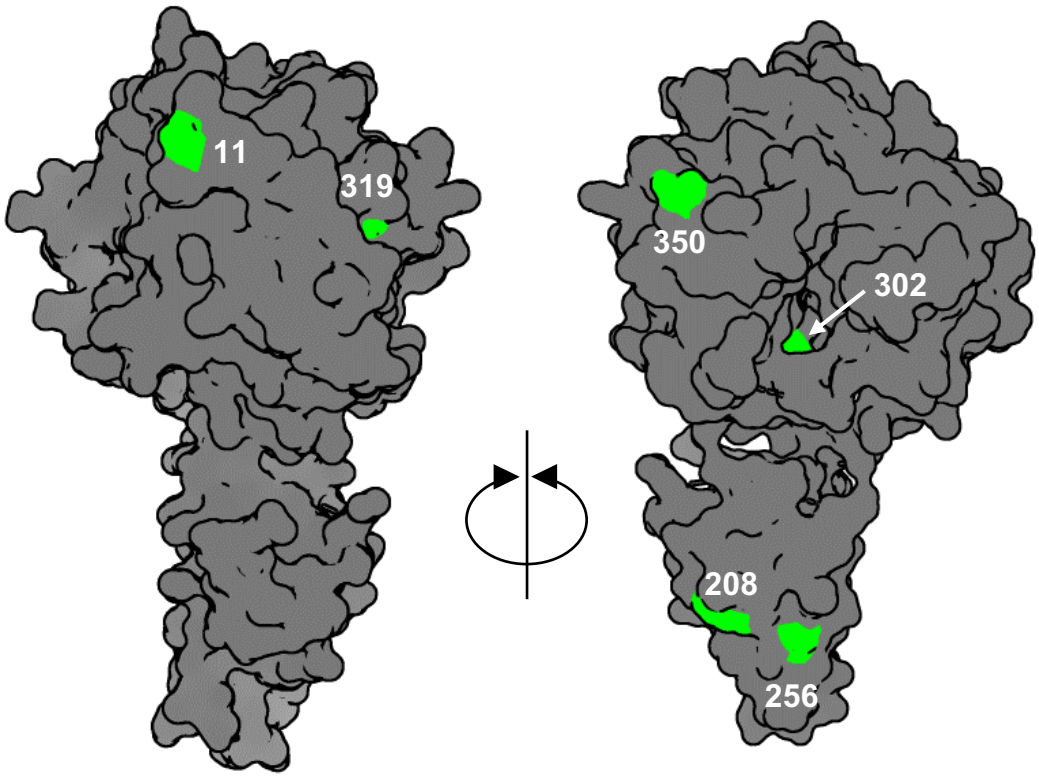


Figure 3-31. The nucleoprotein sites under episodic positive selection.

Nucleoprotein sites predicted to be under positive selection by MEME. The two models represent the front and back of the nucleoprotein.

FUBAR detects positive selection at 25 sites in the M segment, seven sites in Gc and five in GP38 (Figure 3-32). An additional 10 sites in mucin-like-domain, two

in Gn, and two in NSm were detected as under positive selection but resolved structures for these are not available (Appendix 11).

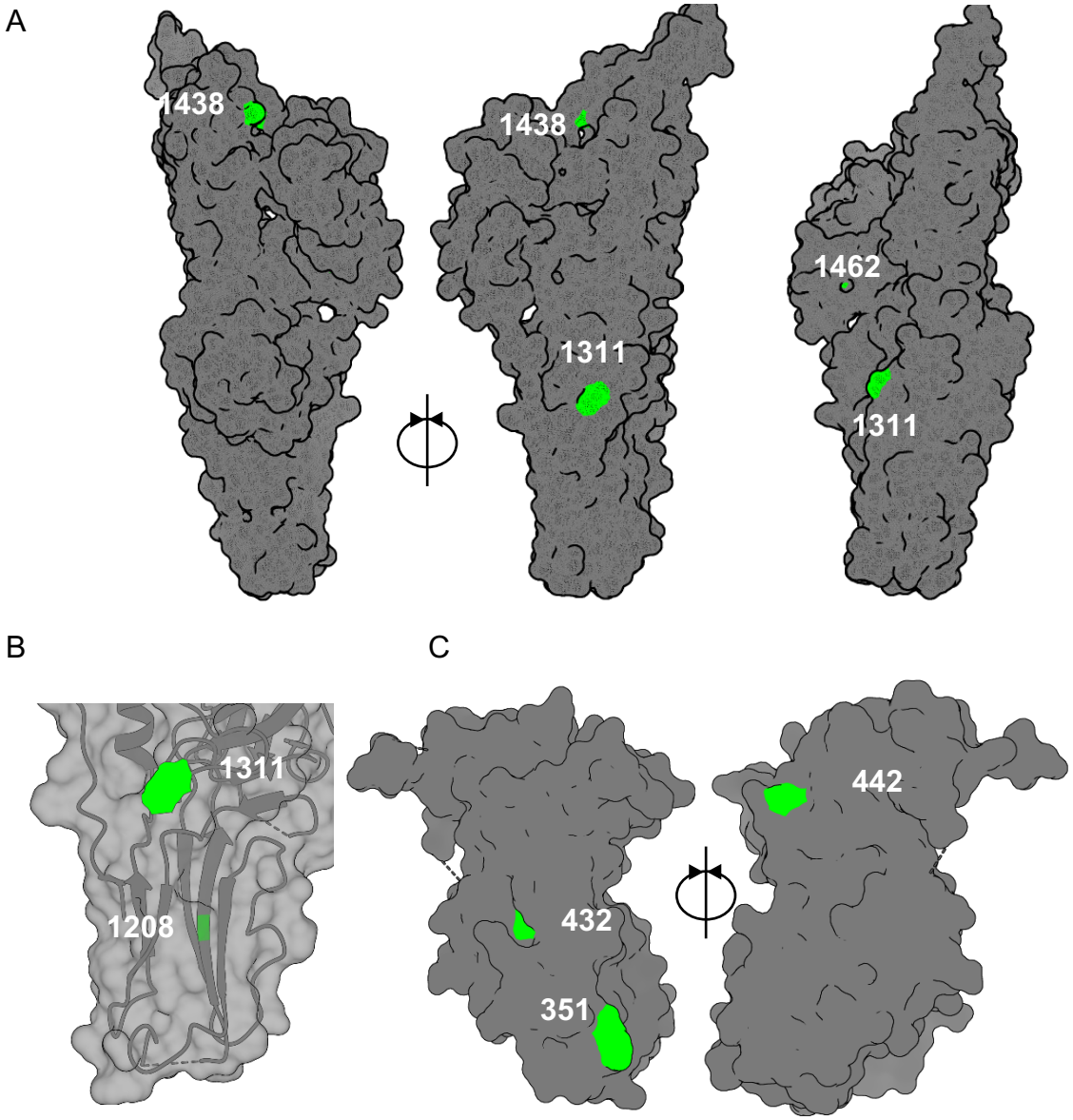


Figure 3-32. Sites in the Gc and GP38 protein under positive selection

Highlighted sites were predicted by FUBAR. The Gc protein (PDB 8JKD) with sites under positive selection which are exposed on the surface of the protein highlighted. Sites 1605 and 1607 aren't included in the model (A). Residue 1208 in the Gc is a cysteine residue in center of the protein (B). The GP38 (PDB 8DC5) with sites 432, 442 and 351 highlighted (C). Site 341 was not resolved in available GP38 models.

FUBAR detects 1313 sites under negative selection (ω=0.02–0.54) suggesting strong negative selection in a large proportion of the GPC: 561 Gc, 237 Gn, 207 GP38, 190 NSm, 100 MLD. MEME identifies 30 sites across the GPC under episodic positive selection, 17 in Gc (Figure 3-33), 10 in MLD, one in Gn, one in NSm and one in GP38 (Appendix 12). Site 341 in the GP38 is located near sites 343-349 which were predicted to be important binding sites for mAb isolated from CCHFv survivors.

In the L segment, site 784 was predicted to be under positive selection by FUBAR (ω=2.78, p=0.004). This site is in the endonuclease domain. Using MEME, 17 sites are predicted to be under positive selection ($p < 0.01$). Site 769 is within the endonuclease domain, and sites 83 and 157 are in the OTU domain. The remaining 14 are not within functionally assigned regions.

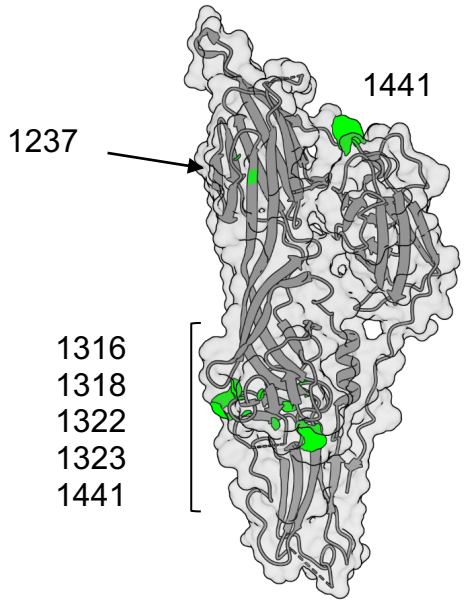


Figure 3-33. Sites in the Gc protein under episodic positive selection.

Residues detected in Gc as under episodic positive selection based on MEME. Sites 1662, 1664, 1668, 1670, 1671, 1673, 1676, 1682, 1683 and 1684 were also detected as under selection but are not resolved in the model (PDB 8JKD).

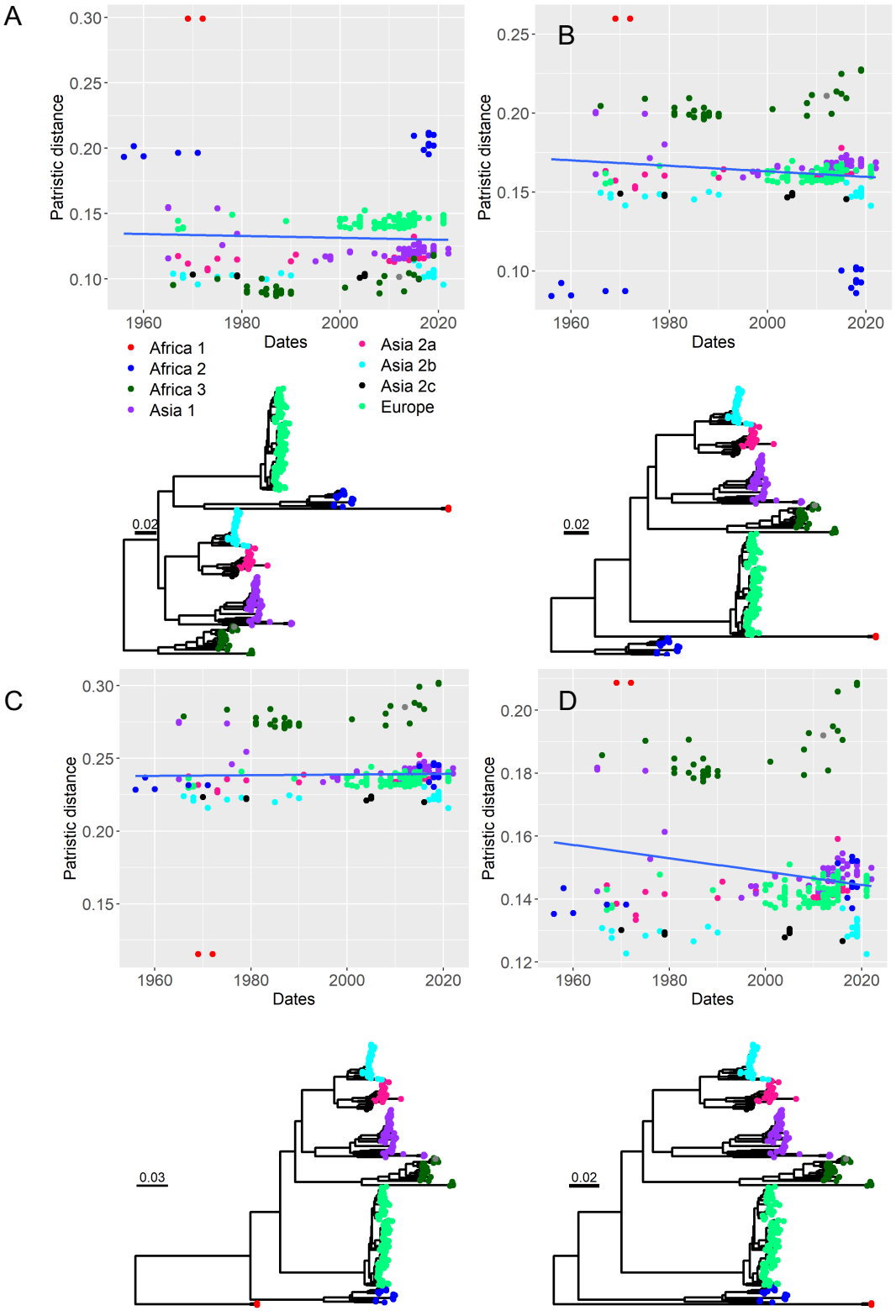


Figure 3-34. Rooting hypotheses and temporal signal investigation.

The three African clades were used as root clades (Africa 3 (A), Africa 2 (B) and Africa 1 (C)) as a comparison against the midpoint rooting (D) in attempts to identify the most temporally favored rooting configuration for the small segment. Tips are coloured by small segment clades.

3.2.6 The evolutionary dynamics of the CCHFv inferred from the small segment

3.2.6.1 Temporal signal and root location

No correlation between root-to-tip distance and collection date was observed utilising a midpoint rooted phylogeny or when rooting was manually placed at any of the African clades (Figure 3-34). The African clades were selected as candidate root clades because of the observation of increased root-to-tip distance in the midpoint rooted phylogeny and previous publications (Carroll *et al*., 2010; Zehender *et al*., 2013). Midpoint rooting places the root at the centre of the root-to-tip distance and doesn't necessarily select the most biological relevant rooting locations. As little difference was observed between rooting hypothesises the rooting taken forward was selected by the optimising residual mean squared error using TempEst, which mirrored the midpoint rooting with Africa 1 as the root clade.

The long ancestral branches prior each clade paired with the absence of temporal signal through all six clades using a single molecular clock suggests the presence of multiple clocks, and thus each clade was interrogated individually for temporal signal. The strongest temporal signal was observed in the Africa 2 and Africa 3 clades (R^2=0.47, R^2=0.53) (Figure 3-35). The rate estimated for the Africa 3 clade was used for a calibration rate for all small segment clades ($1.2x10^{-4}$ s/s/y^{-1}).

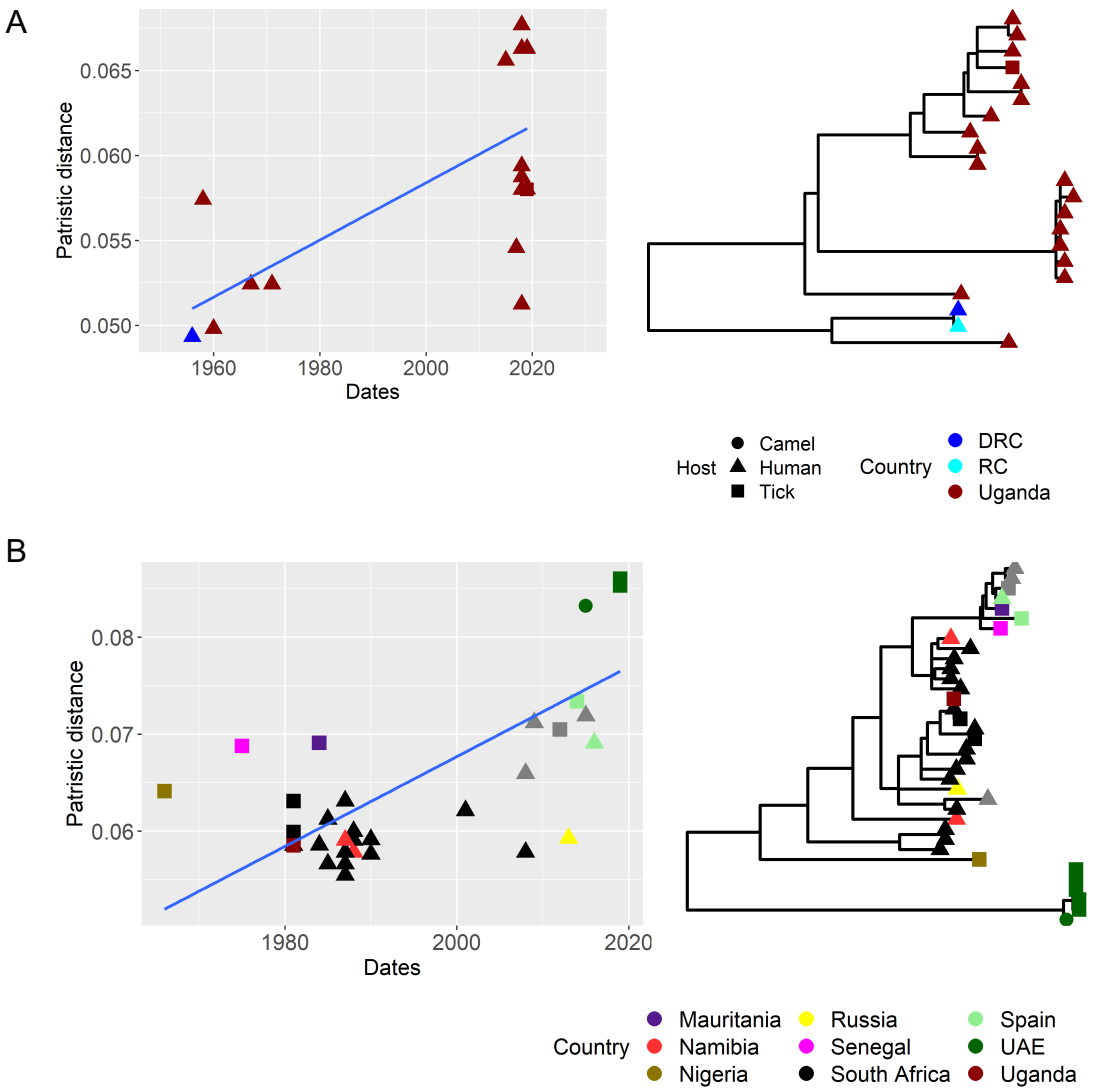


Figure 3-35. Clades with temporal signal.

Linear regressions for the Africa 2 (A) and Africa 3 (B) clades as the clades with the strongest correlation between collection date and root to tip distance. Points are coloured by collection location and shaped by host. The corresponding phylogeny is provided,

The performance of normal and lognormal priors of varying stringency were compared and the best model selected according to the Bayes factor. This stringency is referring to the parameter spaces provided for rate estimation by the statistical distribution used for the clock rate prior. The clock rate prior impacts evolutionary rate estimates and this TMRCA for the clades. All models for all clades converged with >200 ESS for each parameter.

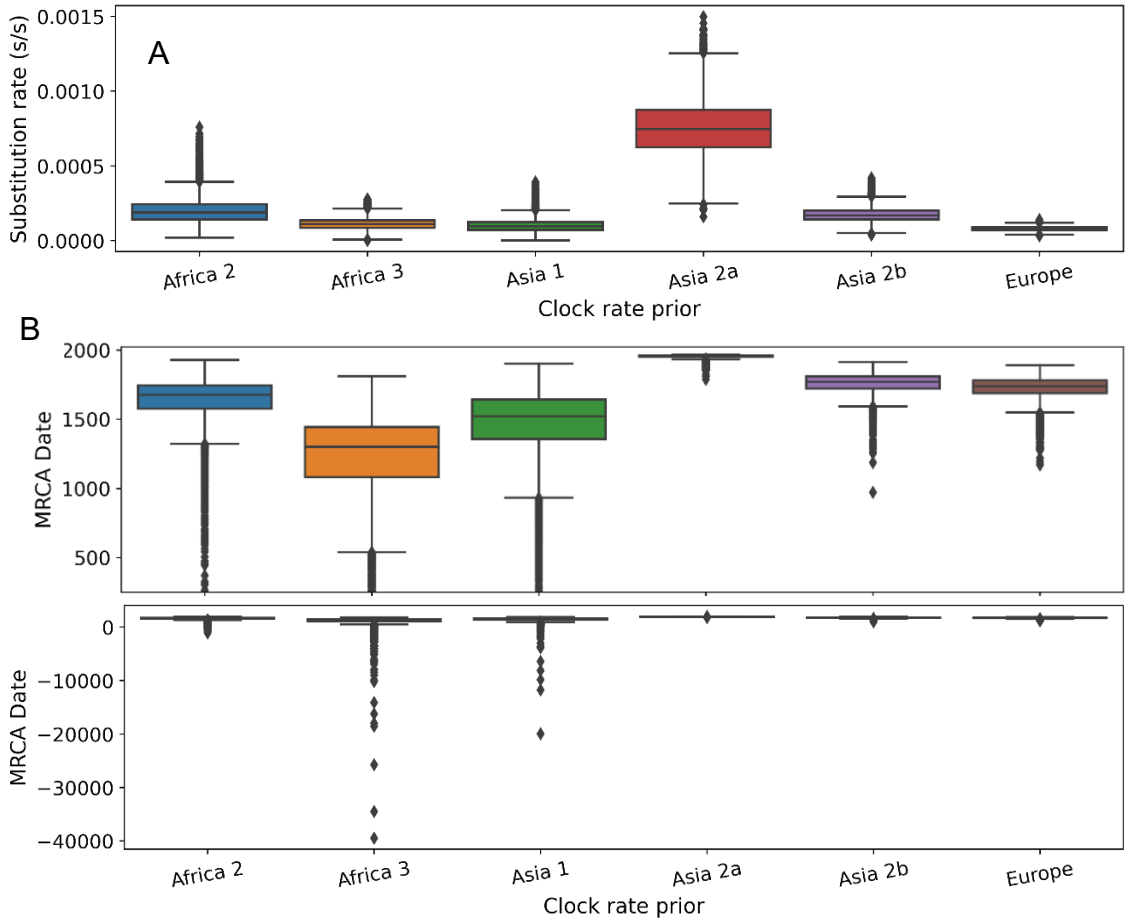


Figure 3-36. Small segment clade estimates.

Estimates of evolutionary rate (A) and time to most recent common ancestor (TMRCA) (B) for the Africa 2, Africa 3, Asia 1, Asia 2a, Asia 2b and Europe clades. Models were selected in this order, logN(1.2E-4, 10), logN(1.2E-4, 0.1), logN(1.2E-4,10), N(1.2E-4,0.1), logN(1.2E-4,0.1) and N(1.2E-4,100). Where N or logN refers to a normal or lognormal distribution, 1.2E-4 is the rate ($1.2x10^{-4}$) and 10 is the standard deviation of the distribution.

Africa 3 is estimated to have the most ancient root age, followed by Asia 1, Africa 2, Asia 2b, Europe, and Asia 2a. The European clade is estimated to have the slowest evolutionary rate with $7.9x10^{-5}$ substitutions/site/yr^{-1} (s/s/y^{-1}) and Asia 2a the fastest evolutionary rate with a rate of $7.5x10^{-4}$ s/s/y^{-1} (Table 3-2, Table 3-3). These rates and TMRCA need to be estimated in context of the entire phylogeny to confirm estimates (Figure 3-36).

As stringency was manipulated for clade-specific clock rate and TMRCA estimates the impact of the prior distribution on estimates was investigated but repeating evolutionary models without genomic data. Prior only models were not comparable to those using genomic data suggesting little influence on estimates.

Clade	Clock.rate prior	Mean	Lower	Upper
Africa 2	logN(1.2E-4,10)	1634.11	1343.93	1893.61
Africa 3	logN(1.2E-4,0.1)	1149.54	491.87	1718.26
Asia 1	logN(1.2E-4,10)	1448.52	967.4	1845.4
Asia 2a	N(1.2E-4,0.1)	1953.97	1934.95	1966.83
Asia 2b	logN(1.2E-4,0.1)	1758.46	1625.71	1877.04
Europe	N(1.2E-4,100)	1727.14	1579.78	1849.42

Table 3-2. TMRCA for the small segment clades.

TMRCA for the small segment clades from Bayes factor selected models. Coalescence date (expressed as years before 2022) estimates for the root of each clade from the model selected by BF test with the mean and lower and upper 95% CI.

Clade	Clock.rate prior	Mean	Lower	Upper
Africa 2	logN(1.2E-4,10)	0.0002001	5.77E-05	0.0003538
Africa 3	logN(1.2E-4,0.1)	0.000112	3.90E-05	0.000188
Asia 1	logN(1.2E-4,10)	0.000102	2.95E-05	0.000184
Asia 2a	N(1.2E-4,0.1)	0.000753	0.000426	0.001108
Asia 2b	logN(1.2E-4,0.1)	0.000174	8.92E-05	0.000262
Europe	N(1.2E-4,100)	7.97E-05	5.26E-05	0.000108

Table 3-3. Evolutionary rates for the small segment clades.

Evolutionary rate estimates for the small segment clades from Bayes factor selected models. Rate estimates for each clade with the mean and lower an upper 95% CI.

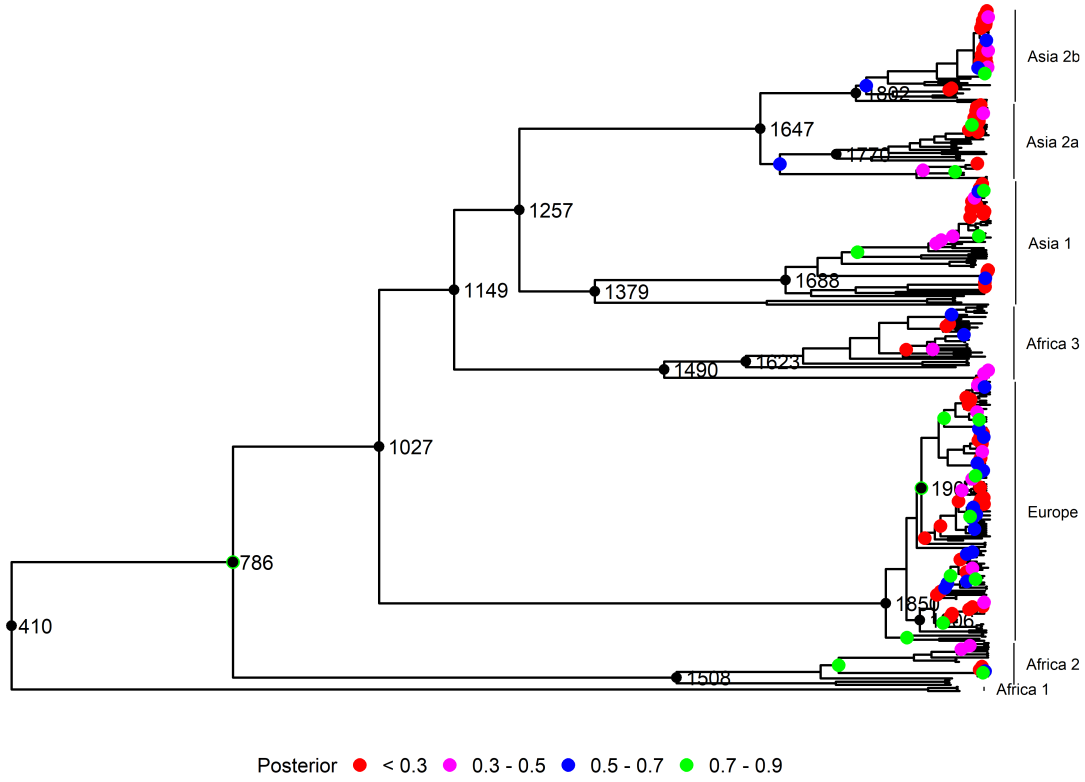


Figure 3-37. Small segment time scaled phylogeny.

A maximum clade credibility (MCC) time scaled phylogeny with median node heights for the small segment inferred under a relaxed clock model with a constant population demographic parameter. Internal nodes are labelled with mean TMRCA calendar date. Coloured nodes represent posterior support for the node. Nodes with >0.9 aren't shown. Black nodes demonstrate those annotated with dates.

To accommodate for the highly structured topology of the CCHFv small segment, an uncorrelated relaxed clock was used with clades enforced by partitioning into monophyletic taxon groups. Under this model the evolutionary rate is estimated to be $1.2x10^{-4}$ (Highest Posterior Density (HPD) $9.4x10^{-5}$–$1.4x10^{-4}$). The root of the phylogeny and the divergence of the Africa 1 clade is estimated to have occurred in 410 CE (HPD 266 BCE–998 CE). Of particular interest are the TRMCAs for the deep ancestral nodes which depict divergence into the modern-day clades. The Africa 2 clade diverged in 786 CE (HPD 454-1106) and the

European clade diverged from a likely ancestral African lineage in 1027 CE (HPD 763–1271). The Africa 3 and the Asian lineages diverged from each other in 1149 CE (HPD 918-1360) before further divergence into Asia 1 in 1379 CE (HPD 1185-1553), Asia 2a in 1770 CE (HPD 1697-1835) and Asia 2b in 1802 CE (HPD 1740-1860) (Figure 3-37).

3.2.7 The CCHFv small segment under a fixed local clock model

As the CCHFv phylogeny demonstrates a structured topology, we hypothesised each clade could evolve under separate molecular clocks. To test this, evolutionary models were implemented constraining each clade to monophyly and assigned its own prior for clock rate and root age estimations in accordance with clade specific rate inference (Figure 3-36).

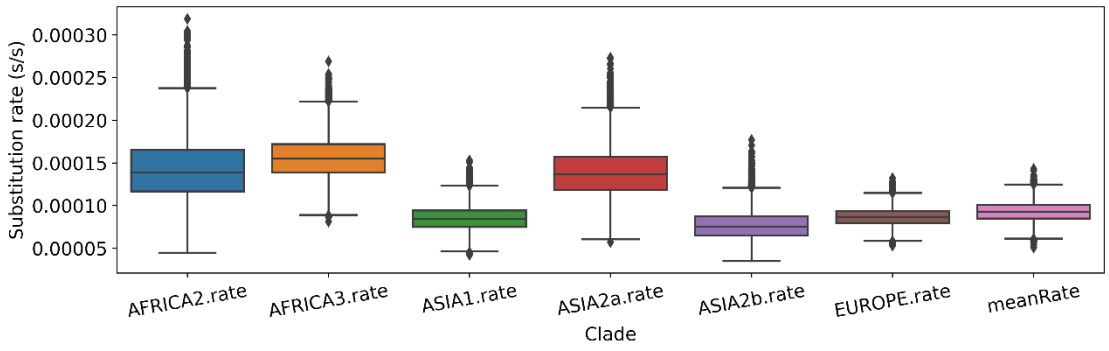


Figure 3-38. Rate estimates under fixed local clocks.

Rate estimates for each clade under a fixed local clock model where clades with >3 sequences were constrained to monophyly and have a unique rate estimated. meanRate demonstrates the rate estimated across all branches in the phylogeny.

Small variations in rate were observed across the clades when heterogeneous evolutionary rates are permitted (Figure 3-38). The quickest evolutionary rate was

observed in the Africa 3 clade at $1.5x10^{-4}$ s/s/y^{-1} (HPD $1.1x10^{-4}$ – $2.0x10^{-4}$), which varies from the rate estimated when cladistic evolutionary rates were estimated independently (Table 3-3). The slowest rate among clades was observed in the European clade at $8.6x10^{-5}$ s/s/y^{-1} (HPD $6.7x10^{-5}$ – $1x10^{-4}$). The rate estimated for the entire phylogeny is reduced to $7.6x10^{-5}$ s/s/y^{-1} (HPD $4.9x10^{-5}$ – $1.05x10^{-4}$) from $1.2x10^{-4}$ s/s/y^{-1} estimated without local clocks. The Asia 1 and Asia 2b clades are suggested to evolve at rates of $8.9x10^{-5}$ s/s/y^{-1} (HPD $4.2x10^{-5}$ – $1.4x10^{-4}$) and $7.7x10^{-5}$ s/s/y^{-1} (HPD $3.5x10^{-5}$ – $1.1x10^{-4}$), respectively. The model selected by Bayes factor for each clade was used for these estimates.

When estimated under a fixed local clock, the root of the phylogeny is reduced from 410 CE (HPD 266 BCE – 998 CE) to 275 BCE (HPD 523 BCE – 1173 CE). The TMRCAs for Africa 2, Africa 3, and Asia 2a are brought closer to the current day. Under a fixed local clock model Africa 2 diverged in 1570 CE (HPD 1358-1760), Africa 3 diverged from Asian clades in 840 CE (HPD 454–1195) and Asia 2a diverged in 1778 CE (HPD 1687-1865). The remaining clades TMRCAs are estimated to be more historical. Asia 1 is predicted to have diverged in 950 CE (HPD 591-1265) and Asia 2b in 1692 CE (HPD 1557-1815) (Figure 3-39).

Three different models were utilised to model the small segment phylogeny. Uncorrelated relaxed clock models with and without clades monophyletic constraints, and a fixed local clock model with a rate prior per clade. A Bayes Factor test overwhelmingly selected the relaxed clock model with monophyletic clades and one rate prior as the model (BF vs no monophyly = $7.9x10^{198}$, BF vs local clocks = $9.5x10^{14}$).

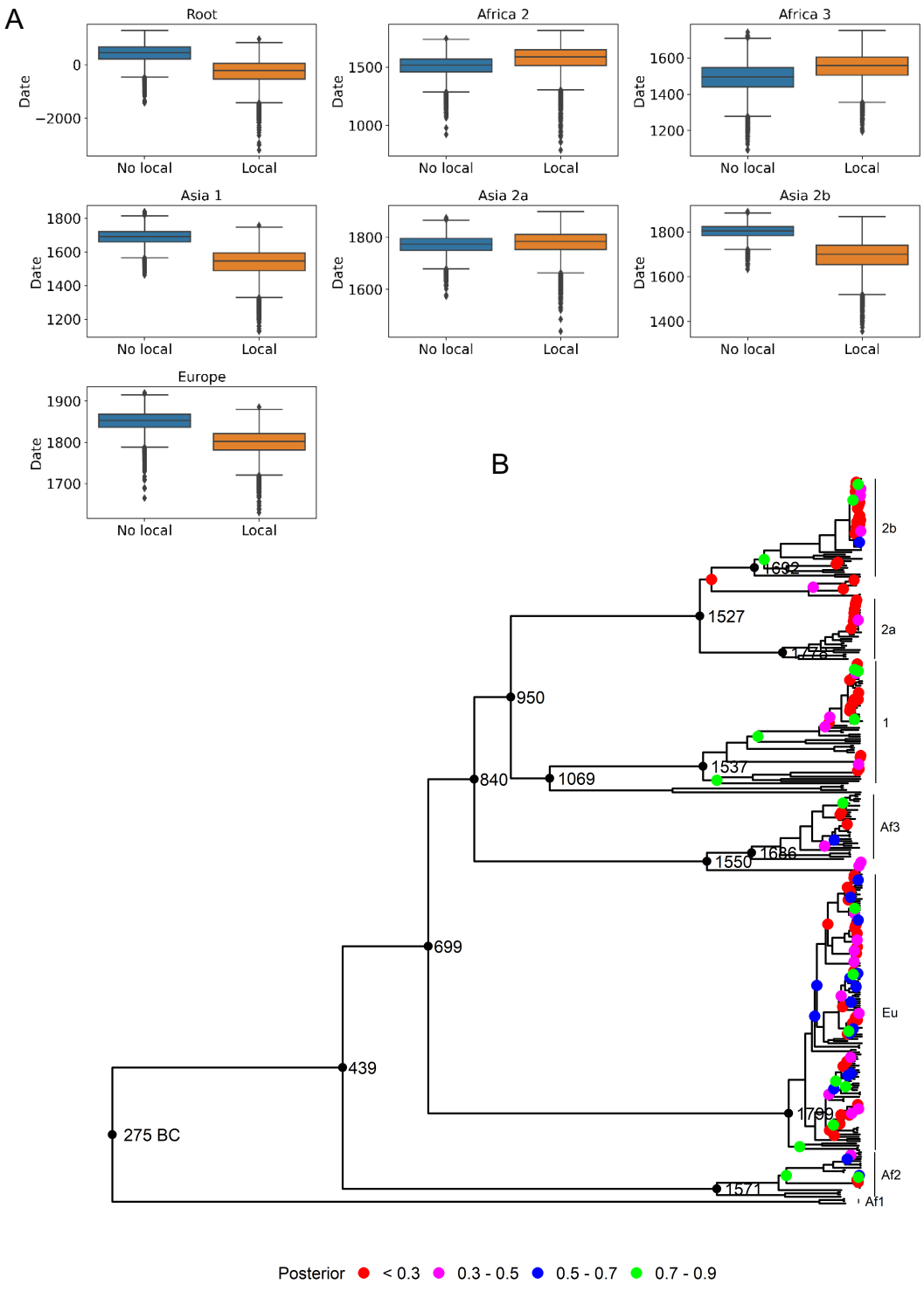


Figure 3-39. Differential coalescence and a time scaled phylogeny.

Differential TMRCA dates for each clade comparing estimates made under a relaxed clock model and a fixed local clock model. No local refers to estimates from the uncorrelated log normal relaxed clock model (A). The MCC time-scaled small segment phylogeny with median node heights as inferred under a fixed local clock model. Nodes annotated with mean heights represent divergence dates. Node colours denote posterior support (>0.9 nodes omitted) (B).

3.2.8 A BEAST reconstruction of CCHFv dissemination at local scales

Using a time-calibrated phylogeny, hypotheses can be made regarding dissemination across both local and global scales. However, a formal assessment can also be made by reconstruction of ancestral locations by discrete trait analysis (DTA).

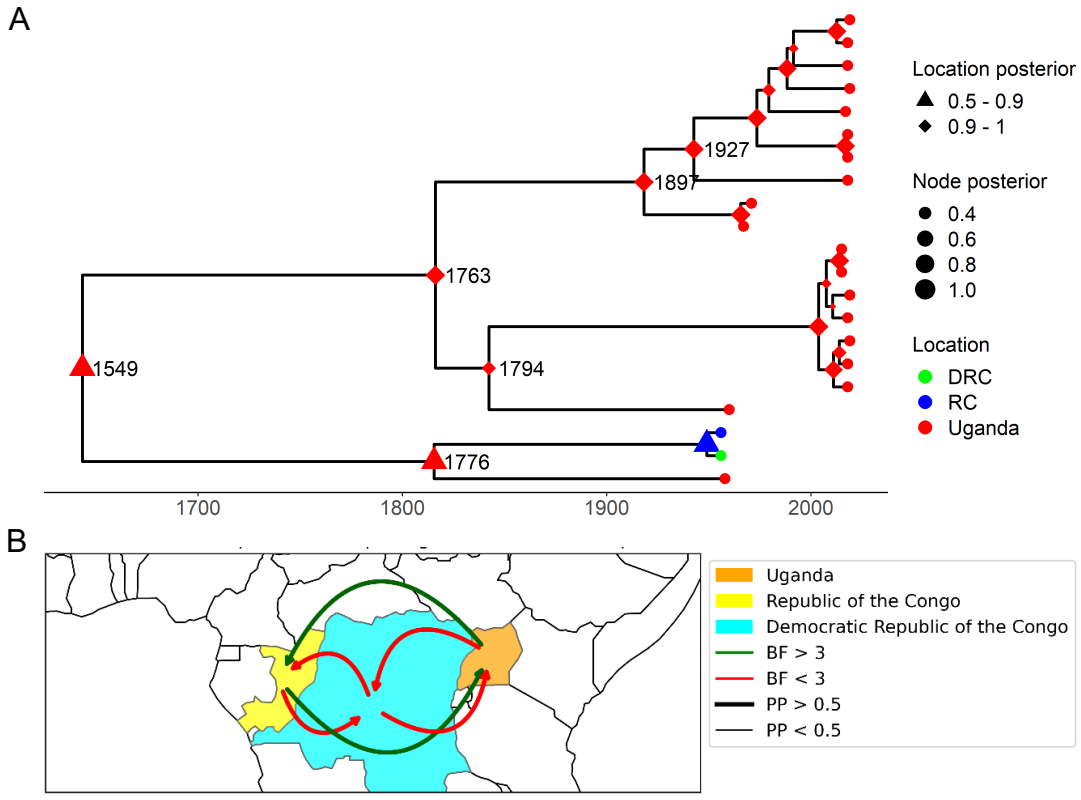


Figure 3-40. Discrete trait reconstruction for the Africa 2 clade.

Time-scaled MCC phylogenies for the Africa 2 with ancestral nodes coloured by most highly supported discrete location. Node labels show mean heights as calendar dates coalescent time. The shape of the icon at each node depicts the probability of the ancestral location. The size of the node depicts the posterior support for the node (A). Edge network graphs plotted with geographical coordinates for visualisation of CCHFv clade Africa 2. The directionality of movement is assessed via Bayes Factor, and support for the significance of the movement is assessed via the posterior probability (B).

Ancestral state reconstruction predicts Uganda as the ancestral location for the Africa 2 clade (posterior=0.72) with the progenitor virus estimated to be circulating prior to 1549 CE (HPD 1137-1903). Node transitions rates suggest mixing between the DRC and RC and DRC/RC to Uganda, and vice versa are possible (posterior=>0.5) (Figure 3-40).

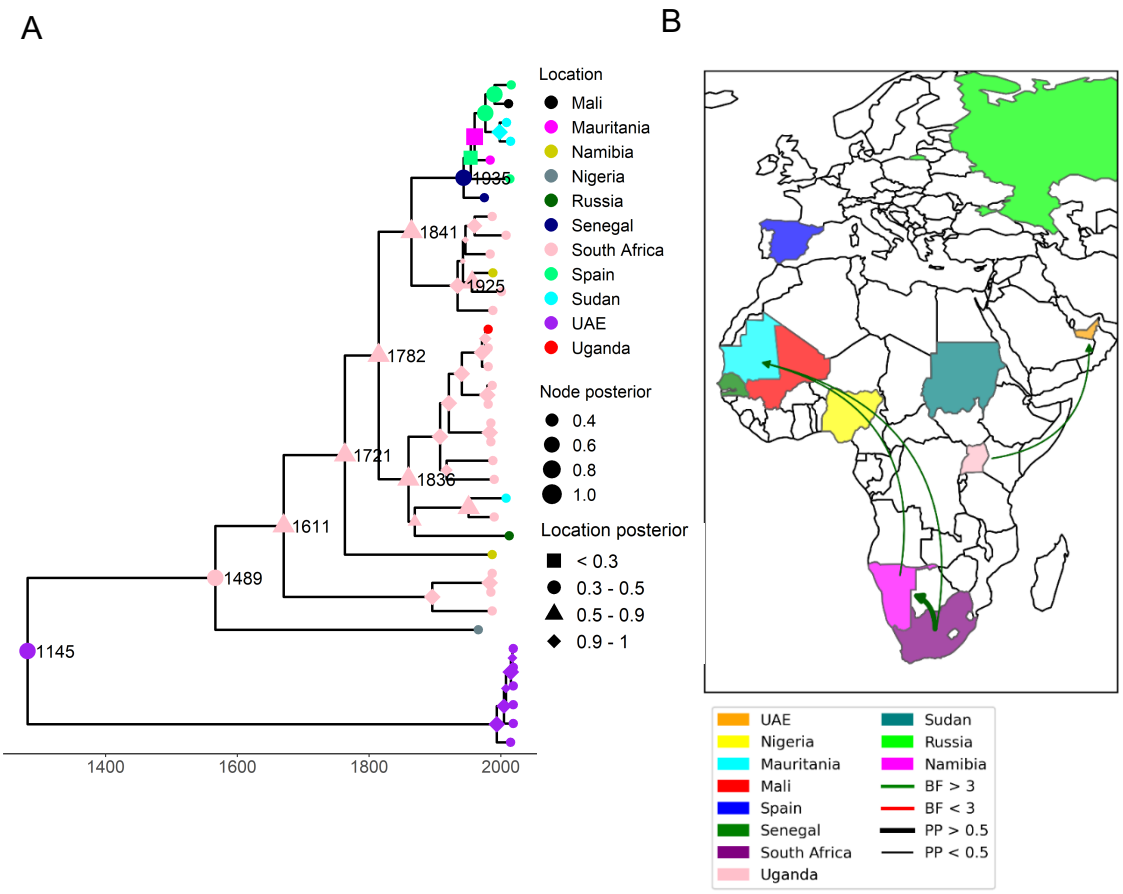


Figure 3-41. Discrete trait reconstruction for the Africa 3 clade.

Time-scaled MCC phylogenies for the Africa 3 with ancestral nodes coloured by most highly supported discrete location and node labels showing mean heights as calendar dates coalescent time. The shape of the icon at each node depicts the probability of the ancestral reconstruction. The size of the node depicts the posterior support for the node positioning (A). Edge network graphs plotted with geographical coordinates for visualisation of CCHFv clade Africa 2. The directionality of movement is assessed via Bayes factor, and support for the significance of the movement is assessed via the posterior probability (B).

The ancestral location of the Africa 3 clade with the UAE outgroup was predicted to be the UAE (posterior=0.3) in 1145 CE (HPD 289-1707). South Africa is reconstructed as the ancestral location for the main Africa 3 clade at the two descending nodes for the root (posterior=0.33, 0.61). The most supported node reconstruction predicts South Africa as the ancestral location in 1611 CE (HPD 1245-1847). The reconstructed location at the parental node to the clade containing sequences from Europe, Western and Eastern African is South Africa (posterior=0.76) in 1841 CE (HPD 1698-1939) (Figure 3-41). No transition rates with >0.7 posterior support was detected in the Africa 3 clade. Dissemination from South Africa to Namibia is supported with lower statistical support (posterior >0.5, BF=13.8).

The reconstructed location at the root of the Asia 1 clade is Iraq (posterior=0.3) in 1593 CE (HPD 1158-1906). The reconstructed location at the most historical but supported node is Pakistan (posterior=0.58) in 1948 CE (HPD 1918-1970). Other ancestral locations are predicted at deeper nodes but with low probability. Transition rates from Oman to Iraq (BF=15.2, posterior=0.64), Iraq to Oman (BF=8.5, posterior=0.50), and UAE to China (BF=66.6, posterior=0.88) are statistically supported (Figure 3-42).

Tajikistan is predicted to be the ancestral location (posterior=0.49) at the root of the Asia 2a clade in 1951 CE (HPD 1925-1967). Other ancestral nodes are also predicted to be from Tajikistan but is with strongest support in 1978 CE (HPD 1966-1989, posterior=0.93). Pakistan is predicted as the proximal origin prior to dissemination into India in 1995 CE (HPD 1984-1999, posterior=0.62) and the presence of Asia 2a in India is most strongly supported in 2009 CE (HPD 2003-2012, posterior=0.98) (Figure 3-43).

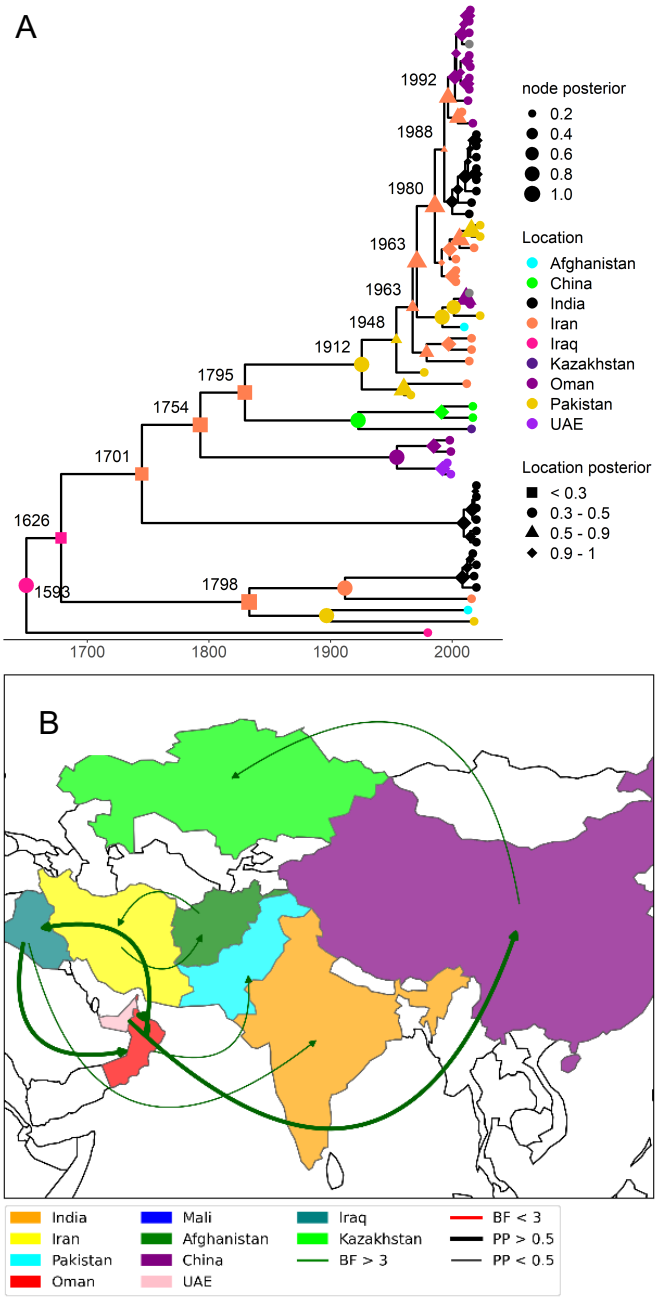


Figure 3-42. Discrete trait reconstruction for the Asia 1 clade.

Time-scaled MCC phylogenies for the Asia 1 clade with ancestral nodes coloured by most highly supported discrete location. Node labels show mean heights as calendar dates coalescent time. The shape of the icon at each node depicts the probability of the ancestral location. The size of the node depicts the posterior support for the node positioning (A). Edge network graphs for discrete phylogeographic results for the Asia 1 clade (B).

No transition rates were supported with a BF >3 and a posterior probability >0.7. Transitions between India into Tajikistan (BF=6.9, posterior=0.68), Tajikistan into Uzbekistan (BF=5.2, posterior=0.61), Pakistan into Iran (BF=5.87, posterior=0.64), and Iran into India (BF=5.6, posterior=0.63) are suggested.

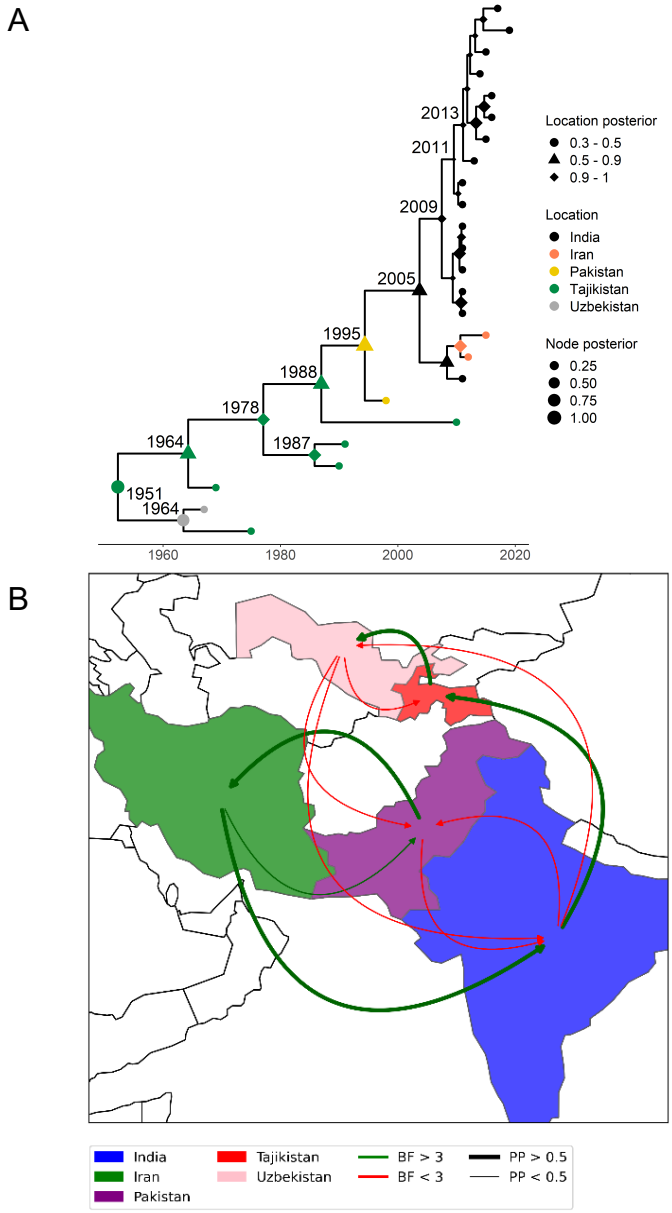


Figure 3-43. Discrete trait reconstruction for the Asia 2a clade.

Time-scaled MCC phylogenies for the Asia 2a clade with ancestral nodes coloured by most highly supported discrete location. The shape of the icon at each node depicts the probability of the ancestral location. The size of the node depicts the posterior support for the node. Node labels show mean heights as calendar dates coalescent time (A). Edge network graphs for discrete phylogeographic results for the Asia 2a clade (B).

Turkmenistan is reconstructed as the ancestral location for the Asia 2b clade (posterior=0.24) in 1747 CE (HPD 1596-1875). China is the most supported ancestral reconstruction at internal nodes in 1896 CE (HPD 1840-1944, posterior=0.84) and 1927 CE (HPD 1885-1959, posterior=0.96), as potentially different Asia 2b subclades. The Asia 2b clade is predicted to have been circulating in India in 1979 CE (HPD 1953-2000, posterior=0.9). At a posterior probability >0.5, transition from China into Kazakhstan (BF=8.2, posterior=0.61) and UAE into Turkmenistan (BF=7.16, posterior=0.57) are supported (Figure 3-44).

As the clade with the most genomic information and reduced sampling bias towards single well represented locations, well supported ancestral reconstructions are present at deeper nodes in the European clade. Russia is predicted as the ancestral location for the European clade in 1722 CE (HPD 1557-1845, posterior=0.87). Turkey is predicted to have had CCHFv circulating in 1822 CE (HPD 1737-1895, posterior=0.87), and Kosovo in 1926 CE (HPD 1875-1966, posterior=0.97). Transition rates support Russia as the source of movement into the other European regions (posterior=>0.7, BF=>3). Transition from Bulgaria to Russia is also a highly supported transition rate exchange (BF=59, posterior=0.93). Transition rate support also implicates Russia to Iran as an important pathway for movement (BF=13, posterior=0.75) (Figure 3-45).

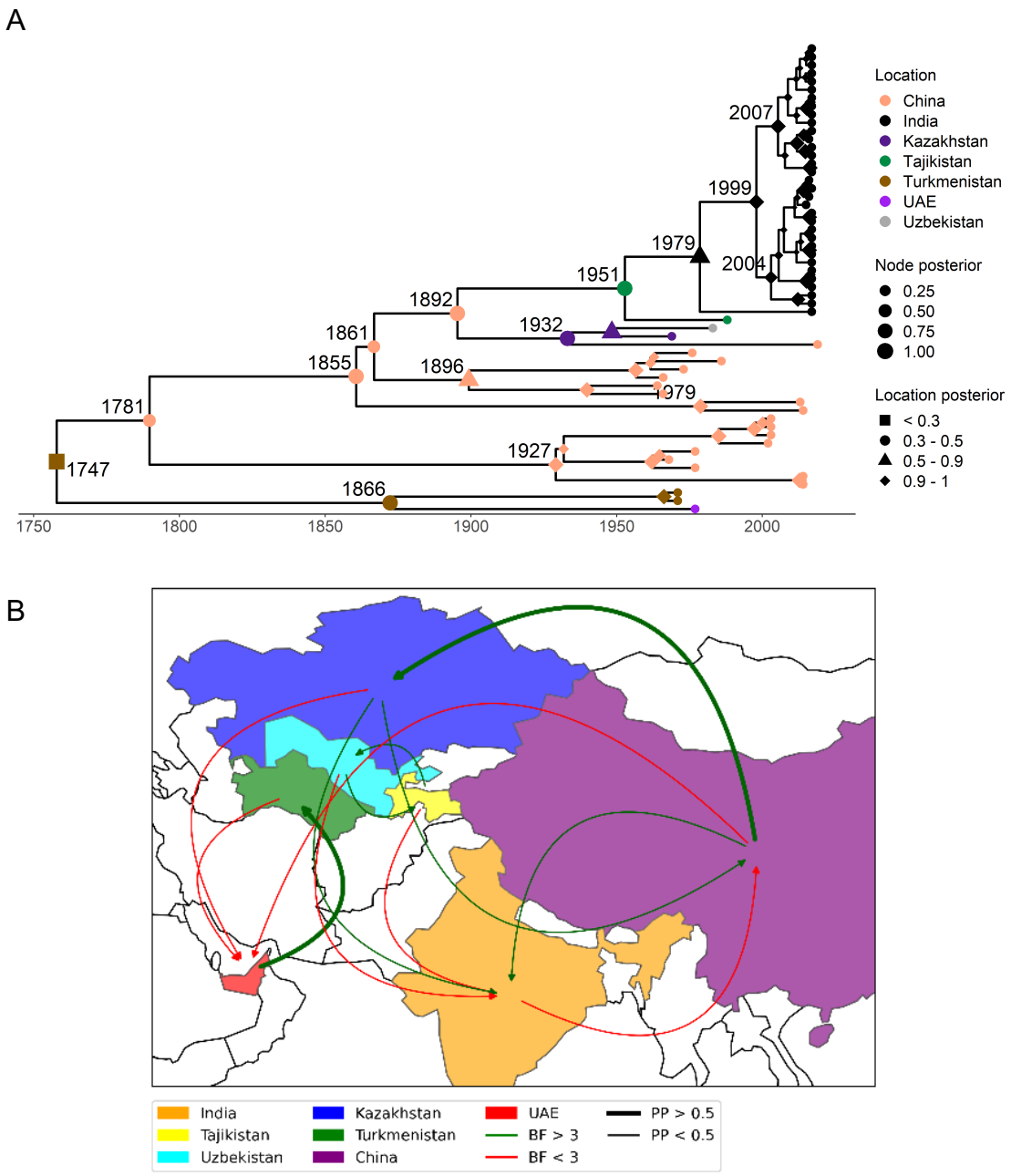
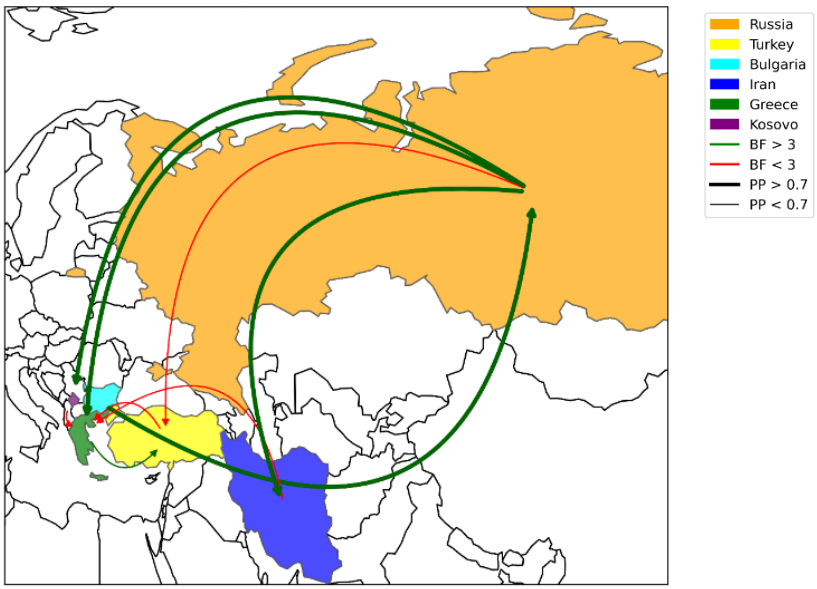


Figure 3-44. Discrete trait reconstruction for the Asia 2b clade.

Time-scaled MCC phylogenies for the Asia 2b clade with ancestral nodes coloured by most highly supported discrete location. The shape of the icon at each node depicts the probability of the ancestral location. The size of the node depicts the posterior support for the node positioning. Node labels show mean heights as calendar dates coalescent time (A). Edge network graphs for discrete phylogeographic results for the Asia 2b clade (B).

A



B

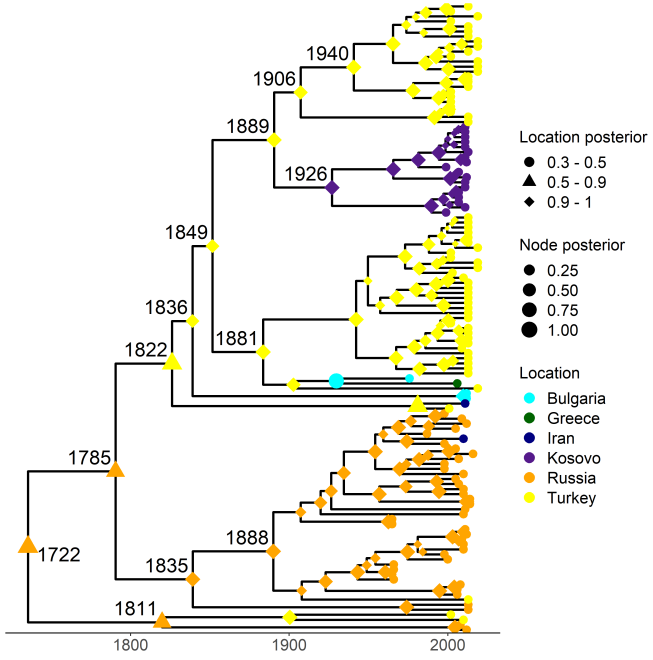


Figure 3-45. Discrete trait reconstruction for the European clade.

Time scaled MCC phylogeny inferred under a discrete phylogeographic model. Nodes are coloured by ancestral location, shaped by posterior support for the ancestral reconstruction and sized by support for the node positioning. Node labels show mean heights as calendar dates coalescent time (A). Edge network graph for discrete location geographic reconstruction of regions in linked by the CCHFv S European clade (B).

3.2.9 BEAST reconstruction of CCHFv dissemination on a global scale

The above section specifically investigated clade dissemination within their respective geographic regions. Using the small segment phylogeny, we investigate the global movement of CCHFv across the three continents within which it is reported. All reconstructed locations at the deep nodes have <0.3 posterior probability support for South Africa as the ancestral location (Figure 3-46). There are two well supported ancestral reconstructions. China is reconstructed as the ancestral location for the Asia 2a and Asia 2b clade in 1666 CE (HPD 1571-1758, posterior=0.82). Russia is reconstructed as the ancestral location for the European clade in 1855 CE (HPD 1809-1899, posterior=0.96). There are eight transition rates supported for CCHFv dissemination on a global scale. These highly supported transitions can be split into four categories: Africa into Asia, Asia into Africa, Asia into Europe, and intra-Asian (Table 3-4).

Location 1	Location 2	Bayes factor	Posterior probability
Mauritania	Pakistan	31.42	0.53
Mauritania	Oman	122.6	0.88
Turkmenistan	CAR	43.1	0.61
Oman	Uganda	28.2	0.51
Afghanistan	Nigeria	30.4	0.52
Pakistan	Namibia	531.09	0.95
China	Greece	48.1	0.63
Kazakhstan	India	98.4	0.78

Table 3-4. Supported discrete trait transitions for the small segment.

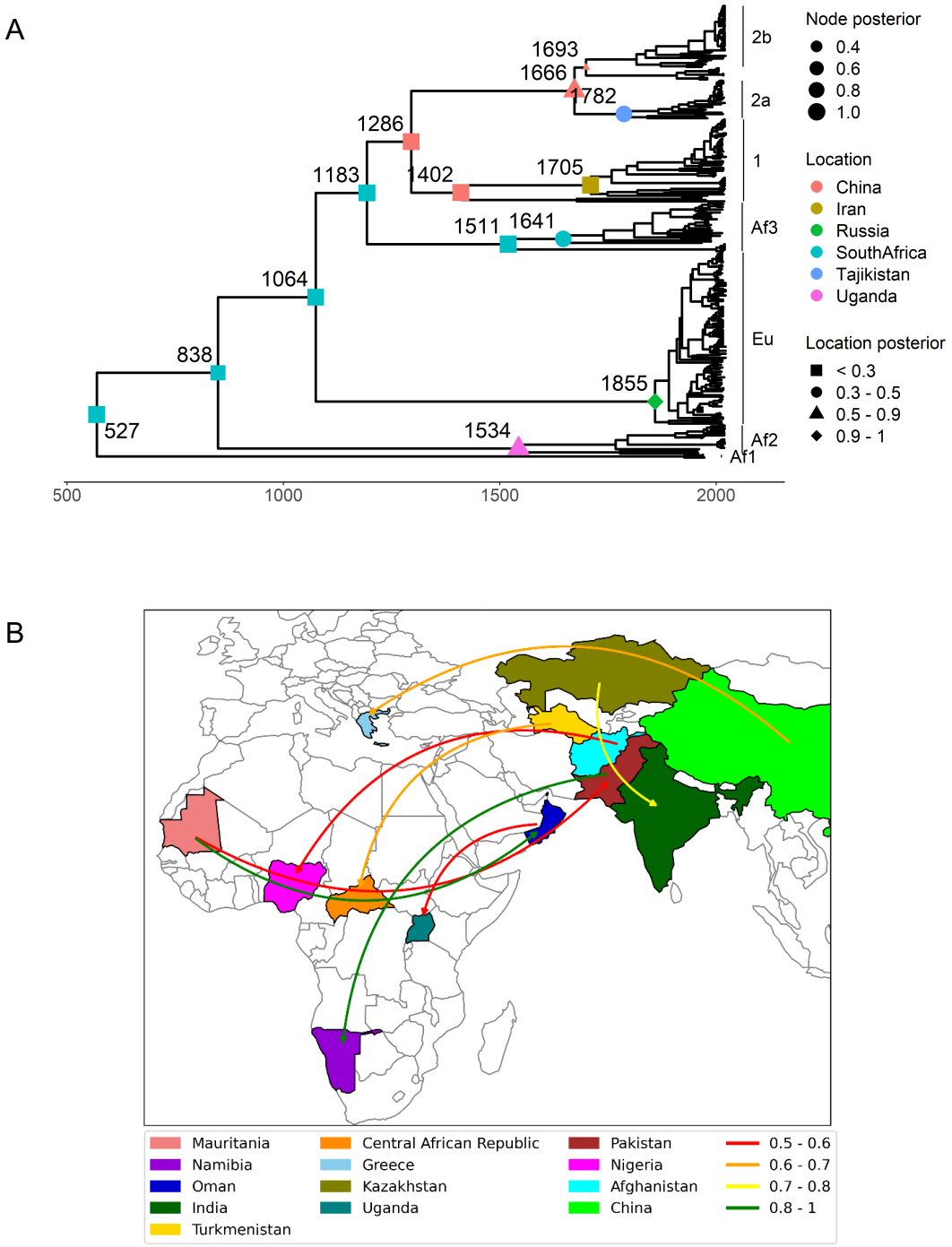


Figure 3-46. Discrete trait reconstruction for the full small segment.

Time scaled MCC phylogeny inferred under a discrete phylogeographic model. Ancestral state reconstruction for the discrete location trait was estimated. Nodes are coloured by ancestral location, shaped by location support and sized by posterior support for node positioning. Node labels show mean heights as calendar dates coalescent time (A). Edge network graph for discrete location geographic reconstruction of regions. Transitions supported are represented by arrows which depict the direction of movement coloured by a bayes factor. Only transitions supported by posterior probability >0.5 are shown (B).

3.3 Discussion

3.3.1 Novel CCHFv sequences added to the Europe, Asia 2a and Asia 1 clade provide insight into potential dissemination routes

This project generated 43, 34 and 33 novel sequences for the small, medium, and large segments, respectively. Most of these sequences were recovered from clinical samples collected in 2015 during an outbreak in Turkey. Within the European clade all but one of the Turkish sequences clusters in the previously known Turkish subclades. Here, we present a complete genome from Turkey which clusters within the Russian subclade in all three segments, indicating this is not a reassortant virus but the introduction of a Russian subclade-like virus into Turkey. This is the first evidence of mixing between the two European regions contained within the Russian and Turkish clades. This provides an additional avenue with potential for generation of recombinant or reassortant European lineages.

All Russian sequences in the public domain are from southern districts which border Georgia, Azerbaijan, and Kazakhstan, all of which report circulation in human, animal and/or tick populations (Abdiyeva *et al*., 2019; Bryant-Genevier *et al*., 2022; Chumakov *et al*., 1970; Fereidouni *et al*., 2023; Greiner *et al*., 2016; Head *et al*., 2018, 2020; Lotfollahzadeh *et al*., 2011; Nurmakhanov *et al*., 2015; Saidi *et al*., 1975; Sultankulova *et al*., 2022; Yashina *et al*., 2003; Zakhashvili *et al*., 2010). One hypothesis is that a Russia-like variant has circulated through these regions into Turkey. An alternative hypothesis is one that is conserved across CCHFv epidemiology, through the action of migratory birds (Leblebicioglu *et al*., 2014). It is also possible that Tur.1009 clusters in the Russian subclade because of natural variation, as the sequence is present on its own terminal

branch and not part of a small cluster of Russian sequences, which would support a recent introduction event into Turkey. This clustering, and the observation that the remaining recovered genomes cluster in both Turkish subclades provides evidence that the outbreak in 2015 was caused by a heterogeneous viral population.

Within the Asia 2a subclade the sequence from a clinical case in Pakistan clusters closely with a sequence from a clinical sample collected in Gujarat in 2011. D1126 is from a previously unrepresented outbreak in Pakistan in 1998/99 and provides evidence of a similar virus circulating and causing clinical infections in 2011. In the M segment, D1126 is positioned at the base of clade 3 (Asia 2a), providing a historical date for that clade. This is not included in the presented analysis, but this historical timepoint provided by D1126 provides a reference point for temporal signal in this clade highlighting the importance of retrospective sequencing for phylodynamic inference.

3.3.2 A recombinant sequence from a tick in Rajasthan resulting from cocirculation near a village with a human case report in Gujarat

Five potential recombinant viruses were detected by RDP4. In the small segment a recombinant sequence was detected from a tick collected in Rajasthan in 2019. The minor parent was collected from a tick from Gujarat in 2019 and the major parent from a human in Gujarat in 2019. In the study reporting MCL-19-T-1922 it is not defined as a recombinant (Sahay *et al*., 2020) with only reassorted strains reported (MCL.19.T.1929, -T.1916 and -T.1812). The study reporting the Gujarati tick and human sequences was structured as to sample *Hyalomma* populations

surrounding the patient houses. This suggests the *Hyalomma* population surrounding the human case location harboured both the CCHFv which caused the fatal human infection and a second CCHFv population, resulting in the recombinant virus detected here which was sampled from Rajasthan. As the major parent in this recombination was collected from a patient, this indicates circulation of CCHFv with clinical significance potential circulating in Rajasthani *Hyalomma* populations. To the best of my knowledge this is the first report of this recombinant strain in the literature.

3.3.3 Ancestral and contemporary reassortment events highlight the complex evolutionary history of CCHFv

Both GiRaF and RDP4 offer insights into the causal evolutionary processes of phylogenetic incongruence observed within the segment phylogenies. The ability of GiRaF to predict sets of taxa by detection of phylogenetic incongruencies at deep nodes opposed to tips facilitates hypotheses to ancestral reassortment events (Klempa, 2018). These events have likely superseded the expansion of clades across diverse geographical regions, highlighting the complex interplay between genetic exchange and geographic dispersion in evolutionary dynamics.

An ancestral virus to a subset of strains in the Asia 2a clade in the S segment is predicted to have acquired an S segment from a strain related to contemporary Central Asian strains. The subset contains sequences from India and Pakistan, including D1126, presented in this study. The S segment phylogeny supports this by placing this subset evolutionarily closer to Central Asian strains in the S compared to the M and L segments, where the reassorted Asia 2a subset strains

are monophyletic. This provides supporting evidence for circulation of ancestral strains in Central Asia before introduction into India and the Middle East.

Ancestral M segment reassortment events are supported by the formation of heterogeneous clades, namely, MC4, MC5 and MC7. These clades in the M segment contain sequences from different clades and disparate geographical locations, indicating a possible ancestral reassortment event where an ancestral strain obtained a disparate M segment before dissemination or reintroduction into both African and Asian countries. Sequences in the MC4 clade are more closely related to the Asia 1, Asia 2a and Asia 2b clades than MC5, yet both contain sequences from Asia and Africa. It is possible that two ancestral reassortment events have resulted in the formation of these two clades. MC7 contains all Africa 2 and Africa 1 strains, in addition to a subset of Asia 1 and Asia 2c. This supports a reassortment event in which Asian strains have acquired an African-like M segment. In the L segment, the three sequences which comprise the Africa 1 clade are reassorted into LC5, closely related to the other sequences from Senegal which is in the Africa 3 clade.

3.3.4 Codon-based models for detecting sites under selection detect episodic pervasive selection at six sites in the nucleoprotein

The nucleoprotein plays numerous roles in enhancing the fitness of CCHFv supported by the identification of most codons within NP are subject to purifying selection. MEME identified episodic positive selection at six codons sites. Site 302 is located deep within a positively charged groove predicted to cleave single and double stranded DNA. The amino acid observed at site 302 is alanine, a

neutral amino acid, important for the structure of peptides and unlikely to directly interact with nucleic acids but potentially involved in the structure of the groove (Kubyshkin & Budisa, 2019). There are several predicted nucleic acid binding grooves predicted by positively charged pockets on the nucleoprotein. Site 350 is also located close to a predicted binding groove where lysine residues are observed at sites 342, 343, 345, and 346.

Six sites on the Gc protein are known epitopes for protective mAbs (Fels *et al*., 2021) which have been characterised from survivors. No sites within these epitopes are predicted to be under positive selection but sites in proximity are. Sites 1311 and 1341 are located near mAb binding sites at 1309, 1314, 1352, 1253 and 343-349, respectively. Alterations in proximity to binding sites could induce conformational change and disrupt binding (Carpentier & Chomilier, 2020).

3.3.5 A relaxed molecular clock with enforced monophyletic clades was selected as the best evolutionary model for the small segment opposed to a fixed local clock model

When considering the historical setting within which CCHFv has emerged from a single progenitor virus into a highly divergent population, there are many divergence events which are important to estimate a coalescence date for.

Under a lognormal relaxed clock model with monophyletic taxa the root is estimated to be approximately the year 449 BCE when the Africa 1 clade diverged. This clade contains three sequences, two derived from a tick (1969) and one from a goat (1972). Previous studies investigating the origins of the

CCHFv have included the previously included Europe 2 clade and thus the root calibration has differed, estimated as 1128 BCE (Carroll *et al*., 2010) and more recently, 1110 BCE (Zehender *et al*., 2013). The difference in root MRCA between these studies can be explained by the different evolutionary rates estimated for the small segments in these studies: $1.09x10^{-4}$ (HPD $0.17x10^{-4}$ – $2.09x10^{-4}$) and $2.96x10^{-4}$ ($1.67x10^{-4}$– $4.7x10^{-4}$). In this study, we estimate the evolutionary rate of the S segment to be $1.2x10^{-4}$ (HPD $9.4x10^{-5}$ – $1.4x10^{-4}$), between the two previous estimates. The study estimating the inflated evolutionary rate and more recent TMRCA used a Bayesian Skyline Plot (BSP) (Drummond *et al*., 2005) as the demographic population model compared to a constant population prior. A BSP allows the population size to vary across time which may influence rate estimates at different timepoints in the phylogeny.

The diverging order of clades in the small segment is conserved in this and previous studies (Carroll *et al*., 2010; Zehender *et al*., 2013). First, Africa 1 diverges from an ancestral lineage potentially from Aigai virus (Papa (Άννα Παπά) *et al*., 2022). Then Africa 2, Europe, Asia 1, Asia 2a and Asia 2b. We estimate a TMRCA for an Asia-Africa progenitor lineage is estimated to have been introduced in approximately 1149-1183 BCE, potentially into China from Africa. This ancestral node which diverges into the Asian lineages, also diverges into the Africa 3 clade – the clade with the largest geographic range.

3.3.6 The global spread of the Africa 3 clade

The Africa 3 clade has been detected in Southern, Northern and Western Africa, and most recently, Spain. A popular hypothesis for this distribution is proposed to be because of the action of migratory birds. This is supported by the detection of CCHFv in migratory birds in Greece, Morocco and Italy (Lindeborg *et al*., 2012; Mancuso *et al*., 2019; Palomar *et al*., 2013). In addition to this, the Ebro Delta wetland in South Catalonia in the Northeast of Spain is a key stop over for migratory birds where very high rates of seroprevalence in the Iberian Ibex supports circulation of CCHFv (Espunyes *et al*., 2021).

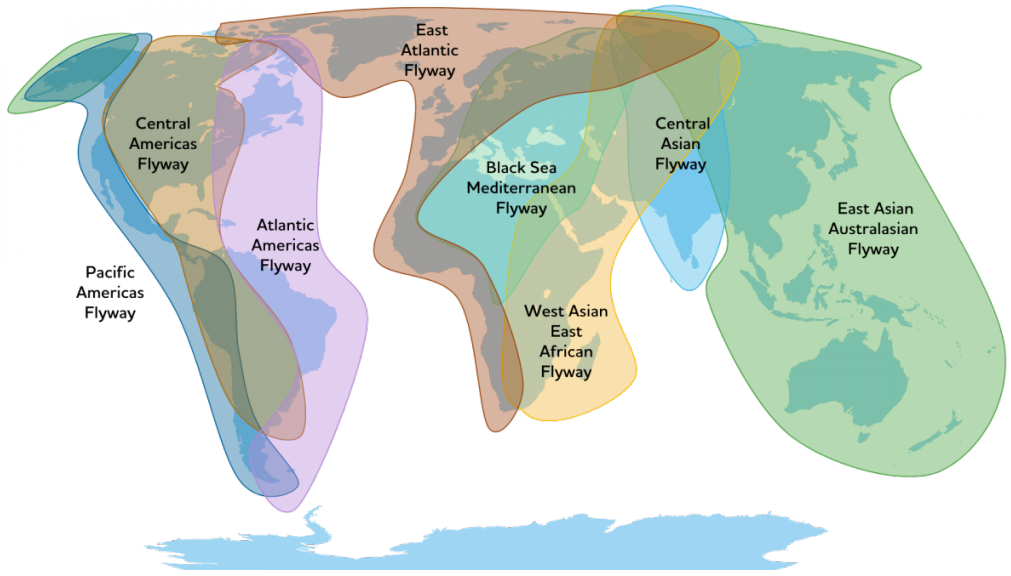


Figure 3-47. Major global bird flyways.
Global flyways with potential for transmission of CCHFv globally. Provided by UNEP/GRID-Arendal.

The introduction of *Hyalomma* into Europe was initially hypothesised to be resulting from annual migrations of bird populations from Western Africa to Europe (Hoogstraal, 1979b). There are eight global multi-species flyways, five of

which span the old world (Figure 3-47) (Boere & Stroud, 2006). One flyway is the East Atlantic which could support dissemination from South Africa into Western Asia and to Europe. This is one of the main hypotheses for the distribution of the Africa 3 clade into Europe. The migratory bird hypothesis can also be applied to dissemination between Africa and Asian using different flyways (Figure 3-47).

3.3.7 The impact of inadequate sampling phylogeographic inference

CCHFv is unevenly represented in the public domain with regards to geographic region. This is especially apparent in endemic regions where case reports are highest globally. Thus, it follows that large amount of genomic variation which informs epidemiology is likely absent from inferred phylogenies. For CCHFv, other hazard group 4 viruses and many other viruses of clinical importance, this is likely an artefact of an inability to investigate such viruses because of a lack of resource in endemic or risk regions (Vaiente & Scotch, 2020). For CCHFv, as demonstrated by differences in sequence availability globally, it is also likely because of differing viral sequencing policies and opinions on public health response priorities (P. Liu *et al*., 2022).

As phylogeography directly utilises genomic sequences in relation to the geographic location provided in the metadata, the lack of sequence availability compared to known prevalence can result in biased ancestral reconstructions (Kalkauskas *et al*., 2021). This is a limitation of phylogeography as a statistical method and is likely influencing the ancestral reconstructions observed within the BEAST reconstructions of clade-specific and S segment evolution across time and space.

For CCHFv, metadata surrounding the collection of sequences is, for the most part, limited to country resolution labels reducing its applicability to continuous phylogeographic models which can reduce the sampling bias which influence the result of discrete trait analysis.

Continuous phylogeographic methods require longitude and latitude coordinates as metadata, which facilitates smaller movements between sequence foci compared to discrete traits (Dellicour *et al*., 2021). With discrete trait analysis where geographic traits are binned into groups called “demes”, there is no delineation of migration between geographically separate demes where true migration events are less likely than local migration. A dataset with sampling bias can result in large migrations being predicted and indicate linkage between unlinked regions (Guindon *et al*., 2016). Despite this, as with the action of migratory birds, it is important to consider such estimates in accordance with other forms of evidence. However, an improved and standardised metadata requirement for viral sequences uploaded to the public domain will enhance the application of phylogeographic methods to a larger group of viruses and prevent reliance on models of discrete trait evolution.

3.3.8 Discrete trait analysis and phylogenetic topology indicate potential transmission routes

The metadata attached to sequences (collection date and a location) can be utilised to predict the most probable ancestral locations at internal nodes within a phylogeny. Ancestral state reconstruction is used to generate intermediate sequence states at otherwise unobservable internal nodes (Cunningham *et al*.,

1998), and identify the statistically most probable geographic locations harbouring ancestral genotypes.

All three Asian clades have comparable structure with regards to geographic location and the distribution of the locations across each clade. At the base of the clades a higher number of sequences from ticks in Central Asian/China and Middle Eastern regions are observed prior to human cases mainly from India (Figure 3-11, Figure 3-12). This suggests potential circulation in “source” locations in Central Asia/China/Middle East before India acts a “sink” for all three Asian genotypes. Intra-asia dissemination between Asian regions is supported by DTA with many, often reciprocal, linkages predicted (Figure 3-42, Figure 3-43, Figure 3-44).

Both Africa 2 and Africa 3 clades exhibit sampling bias towards specific countries, Uganda, and South Africa, respectively, with both over-represented locations predicted as the ancestral location for these clades. This is potentially caused by sampling bias within the dataset. An improvement in sequencing policies and access to resource in African regions will provide greater resolution of the molecular epidemiology and dynamics of CCHFv in Africa.

The phylogeographic reconstructions using the entire small segment are more liable to biases in the estimates from sampling inequalities using discrete trait analysis. This is largely because of the equal likelihood of transfer between geographically proximal and disparate locations. However, supported transitions highlight a broad linkage between Africa and Asia. This linkage is also supported by the extensive reassortment events identified.

3.3.9 The introduction of the three Asian clades into India from three different locations

In all three Asian clades there have been introduction events into India since the mid-1800s. To contextualise the region at this historical timepoint it is important to recognise Pakistan and Bangladesh as part of the British Empire in Colonial India (G. Smith, 1882). In the three clades, ancestral dissemination is estimated into India/British India 97 (Asia 1), 40 (Asia 2b) and 15 years ago (Asia 2a), representing the three temporal scales under which the clades have been introduced into Southern Asia. As this pattern was observed three times, we can hypothesise what underlying processes or factors are driving the observed pattern.

One hypothesis is that the sequencing procedure used in India has generated an over representation of sequences from India. While this is true, the observed pattern across the clades requires different hypotheses as in Asia 1, the over representation of India is less apparent, but the same pattern observed. An alternative hypothesis is the switch from local subsistence-based agricultural practices to large-scale agricultural practices during colonial rule between 1858-1947 (Dutt, 1960) provided favourable conditions for propagation of tick-borne viruses (Sylvain *et al*., 2021) and a niche for Asia 1 to circulate within Northern Indian states.

It is an interesting observation that there is no direct introduction indicated from the phylogenetic topology from China into India despite tick prevalence and human cases of “Xinjiang haemorrhagic fever”, a previously used term for CCHFv in China (Yen *et al*., 1985). One hypothesis is the Himalayan Mountain range provides physical dispersal barrier (Caplat *et al*., 2016) for introduction of China-

like strains into North Eastern Indian states, and dissemination through Central Asia and the Middle East is required for movement from China into India (Mills *et al.*, 2010).

3.3.10 Russia is predicted as the source of introduction into multiple European countries

Russia is predicted to have been the source of introduction of CCHFv into Greece, Kosovo, and Iran. Two flyways cover these regions – the east Atlantic flyway and black sea Mediterranean flyway. Modern day Russia is also only separated from Iran and Turkey by Georgia and Azerbaijan. Azerbaijan is where some of the first modern day Iranian cases were reported. The absence of the *Hyalomma* tick in some Northern European countries means that the route from Russia to Turkey and the Balkans is perhaps through Georgia and Azerbaijan.

3.4 Conclusion

Crimean-Congo haemorrhagic fever virus is the most widespread tick-borne bunyavirus globally. In this study, we generated novel S, M and L sequences from contemporary and archival samples at Porton Down increasing the respective datasets by 12.9, 13.1 and 14.2%, with most novel sequences attributed to an outbreak in Turkey in 2015. The sequences generated from this outbreak suggest a heterogeneous viral population was in circulation at the time, with both Turkish subclades and a Russia-like CCHFv strain detected. In addition to this, one isolate is predicted to be resulting from a reassortment event where a Kosovo-like strain has acquired an M segment from a Turkish strain.

Evolutionary rate estimates inferred from BEAST indicate all clades evolve at standard RNA virus rates, 10^{-3}-10^{-4} s/s/y^{-1}. Rate heterogeneity between clades was observed, but potentially as an artefact of sampling bias between clades. Phylogeographic ancestral reconstruction for the S segment utilising the largest number of sequences available supports the order of divergence into modern day clades previously reported.

Significant bias was observed within publicly available genomic sequences, especially compared to evidence provided by other molecular and serological assays. To improve the accuracy and resolution of ancestrally reconstructed geographic locations under discrete trait analysis, increased sampling is required in poorly sampled regions. This is especially apparent in African countries outside of Uganda and South Africa, with Africa predicted to be the historical reservoir for dissemination. Additionally, improvements in location metadata e.g., latitude and longitude coordinates, attached to sequences could remove the necessity for even sampling between regions by removing the reliance on discrete trait reconstruction.

4 The discovery and characterisation of novel sarbecoviruses

4.1 Introduction

This section serves as an introduction to results chapters 2 and 3, owing to their thematic overlap. The focus will be placed on the *Coronavirinae* subfamily, specifically the *betacoronavirus* genus and the *Sarbecovirus* subgenus. A comprehensive overview of both structural and non-structural proteins is provided, elucidating their respective functions. Furthermore, the global distribution of sarbecoviruses is explored, drawing from extensive sampling in the literature and the availability of complete genome sequences in the public domain. Lastly, a concise summary of current findings concerning bat-associated sarbecoviruses and their interactions with ACE2 receptors is presented.

4.1.1 Taxonomy

Several viruses that are pathogenic in numerous mammals and birds are contained within the *Coronaviridae* family (Order Nidovirales). Within the *Coronavirinae* subfamily are four genus, *Alpha*-, *Beta*-, *Delta*-, and *Gammacoronavirus* (Lauber *et al*., 2013). Within the *betacoronavirus* genus are five subgenus: *Embecovirus* (lineage A), *Sarbecovirus* (lineage B), *Merbecovirus* (lineage C), *Nobecovirus* (lineage D) and recently added, *Hibecovirus* (lineage E) (Decaro & Lorusso, 2020; Z. Wu *et al*., 2016). Known human pathogens are contained within the *betacoronavirus* genus. Namely, embecoviruses HCoV-OC43 and HKU-1 (Woo *et al*., 2005), sarbecoviruses SARS-CoV (Guan, 2003;

Peiris *et al*., 2003) and SARS-CoV-2 (Zhou *et al*., 2020) and *Merbecovirus* MERS-CoV (Zumla *et al*., 2015).

4.1.2 Coronaviridae genome organisation

The *Coronaviridae* family contains viruses with positive sense single stranded RNA genomes. With genomes >26kb, they represent the largest genomes within the RNA virus space (Lauber *et al*., 2013) and can encode between 7-10 open reading frames. The 5' 2/3rd of the genome (ORF1ab) encodes several non-structural proteins required for transcription, replication, and maintenance of the large genome (Table 4-1). It is the functions exerted by these non-structural proteins (nsp) which allow maintenance of larger genomes (Kakavandi *et al*., 2023; R. Yadav *et al*., 2021). After ORF1ab, there are 4 structural proteins; the spike glycoprotein (S), envelope protein (E), membrane protein (M) and nucleocapsid protein (N) (Payne, 2017). Other accessory ORFs can be found 3' of ORF1ab along with these structural proteins and often provide the virus with fitness benefits compared with genotypes without the accessory ORFs (Hurtado-Tamayo *et al*., 2023; D. X. Liu *et al*., 2014; Y.-J. Tan *et al*., 2005).

Non structural protein		Function	Reference
nsp1		Block host mRNA translation by forming complex with 40S ribosomal subunit – "Host shutoff factor" CTD (148-180) insert into mRNA tunnel of 40S ribosomal subunit Trigger cleavage of host mRNA Inhibit innate immune response by dampening expression of genes such as type 1 interferon	(Jahirul Islam *et al*., 2023; Jin *et al*., 2022; K. Schubert *et al*., 2020; Y. Wang *et al*., 2023)
nsp2		Potential RNA binding protein Potential regulation of host intracellular signaling	(Gupta *et al*., 2021; Jin et al., 2022)
nsp3	Ubl 1	ssRNA binding (potential AUA target) Interacts with nucleoprotein for RTC formation Competes with TRS for N complex formation Dampen innate immune response by removing ubiquitin or ISG15 from ubiquitinated or ISGylated proteins and inhibition of these complexes forming	(Hurst *et al*., 2013; Keane & Giedroc, 2013; T. Peng *et al*., 2008; Serrano *et al*., 2007)
	HVR (Glu-rich)	Hyper variable region with unknown function	
	PL1Pro	Cleavage of nsp1, nsp2 and nsp3 from NTD of pp1a/1ab Absent from Sarbecoviruses	(Harcourt *et al*., 2004; Lei *et al*., 2018)
	Mac1	Binds ADP-ribose or poly(ADP-ribose) and losing this function could lead to lower pathogenicity De-ribosylation of ADP-ribosylated proteins (de-MARylation (mono) and de-PARylation (poly)) for interference with host cellular functioning Potential Inhibition of expression of innate immunity genes and/or upregulation of proinflammatory cytokines Potential interacting with the RdRp	(Eriksson *et al*., 2008; Fehr *et al*., 2016; Geilhausen *et al*., 1973; Imbert *et al*., 2008; Jankevicius *et al*., 2013; Piotrowski *et al*., 2009; Putics *et al*., 2006)
	Mac 2	Mac3 Essential for replication of SARS-CoV but not conserved in all coronaviruses RNA binding Mac2/3 binds guanidine preferentially – abrogation of binding complex abolishes SARS-CoV replication Mac3 potential binding to poly(A) aswell as poly(G) Mac2/3 Interacting with host proteins e.g. RCHY1 which results in downregulation of antiviral, p53.	(M. A. Johnson *et al*., 2010; Kusov *et al*., 2015; Ma-Lauer *et al*., 2016; J. Tan *et al*., 2009)
	Mac 3		
	DPUP	∆dpup results in depletion of RNA synthesis in SARS-CoV Mac2-Mac3-DPUP can bind cobalt ions, whereas removal of DPUP removes this function Mac2-3-DPUP complex can interact with nsp3, nsp9 and nsp12	(W. Hu *et al*., 2012; Neuman *et al*., 2008)

	Ubl2	Potential to antagonize host innate responses by blocking IRF3 or NF-kB Potential to influence stability of PL2pro	(Chen *et al*., 2015; Clementz *et al*., 2010; Frieman *et al*., 2009)
	PL2pro	Deubiquitylation and deISGylation of host proteins Direct interaction with IRF3 to block functionality nsp3|nsp4 cleavage Interaction with nsp2, ORF3a and ORF8b (id via Y2H assay)	(Devaraj *et al*., 2007; Lei *et al*., 2018; von Brunn *et al*., 2007)
	NAB	Only within betacoronaviruses Bind ssRNA, preference for G repeat regions Unwind dsRNA	(Lei *et al*., 2018; Serrano *et al*., 2007)
	βsM	Only within betacoronaviruses Unknown function	(Lei *et al*., 2018)
	TM1	First pass through ER membrane Transmembrane domains may bring PL2pro close to the cleavage site	(Harcourt *et al*., 2004; Kanjanahaluethai *et al*., 2007)
	3Ecto	Lumen localization required for cleavage between nsp3 and nsp4 by PL2pro	(Harcourt *et al*., 2004; Kanjanahaluethai *et al*., 2007)
	TM2	Second pass through ER membrane	
	Y1 CoVY	Cytosolic Y1 conserved in nidovirales, CoV-Y in coronaviruses	(Neuman, 2016)
nsp4		Anchor RTC with nsp3 and nsp6 via three TM domains traversing the ER membrane Involved in ER membrane rearrangement for viral replication	(Angelini *et al*., 2013; Jin *et al*., 2022; Santerre *et al*., 2021)
nsp5		Main protease (MPro) / 3C-like protease (3CLpro) required for processing of pp1a/pp1ab Cleavage at 11 sites and release of nsp4 to nsp16 for formation of RTC	(Jahirul Islam *et al*., 2023)
nsp6		Multiple spanning TM protein involved in packaging viral components into compartments for lysosomal degradation	(Cottam *et al*., 2014; Jahirul Islam *et al*., 2023)
nsp7, nsp8		nsp7/nsp8 form a hexadecameric structure with a hollow lumen which is positively charged for nucleic acid traversion nsp7/nsp8 complex attaches to RdRp to stabilize the polymerase domain	(Biswal *et al*., 2021; Gao *et al*., 2020)
nsp9		Homodimer ssRNA binding protein	(de O. Araújo *et al*., 2021; Frieman *et al*., 2012)
nsp10		Small protein acting to scaffold nsp14 and nsp16 for exonuclease and 5' mRNA cap methylation functions, respectively	(Jin *et al*., 2022)

nsp11		Product from cleavage at nsp10/nsp11 site	(Jin *et al*., 2022)
nsp12		RNA dependent RNA polymerase N terminus contains a NiRAN domain functioning to catalyze transfer of nucleotide monophosphates NiRAN also functions in capping viral RNA at the 5' end to mimic host mRNA	(Jahirul Islam *et al*., 2023)
nsp13		RNA helicase unwinding RNA secondary structures, duplexes which may have formed nsp12/nsp13 complex improves unwinding – two nsp13 monoers interact with the nsp7-nsp8-nsp12 complex on opposing sides of the RNA binding site.	(Jahirul Islam *et al*., 2023; Jia *et al*., 2019)
nsp14		3'-5' Exonuclease activity N7-MTase by methyltransferase domain Critical for proof-reading, correcting errors and thus preventing mutation in the larger CoV genomes by exision of 3'-end mismatch Requires complex with nsp10 for optimal exonuclease function	(Bouvet *et al*., 2012; Ma *et al*., 2015; Ziebuhr, 2004)
nsp15		Uridine specific endoribonuclease for cleavage of RNA with long uridine homopolymers Process vRNA and prevent host innate immune activation	(Pillon *et al*., 2021)
nsp16		2'-O-methyltransferase for 5' methyl capping of viral mRNA to prevent degradation of vRNA	(Jahirul Islam *et al*., 2023; Rosas-Lemus *et al*., 2020)

Table 4-1. The functions of the 16 non-structural proteins.

4.1.3 Structural proteins

4.1.3.1 Spike glycoprotein

The spike glycoprotein (S) is classified as a class I viral membrane fusion protein which mediates entry into host cells (Bosch *et al*., 2003). This group includes functional proteins from other pathogenic viruses including hemagglutinin used for cellular entry by Influenza viruses (C. J. Russell, 2014). Class I viral membrane fusion proteins contain both the receptor binding subunit and a membrane fusion subunit and are the only viral proteins required for entry into a host cell (White *et al*., 2008). Class II fusion proteins require non-specific proteolytic cleavage, and class III proteins are composed of a canonical five domains structure (Backovic & Jardetzky, 2009).

The S protein can be split into two functional subunits: S1 and S2. The S1 subunit mediates receptor binding to the cell surface via the receptor binding domain (RBD) and S2 fuses with the host membrane (F. Li, 2016a; M.-Y. Wang *et al*., 2020) . The location of the RBD in the S1 domain is dependent on the host cell component utilised for entry. As a rule, if the RBD is found in the S1-NTD tropism tends to be mediated by sugar/carbohydrate recognition, whereas as S1-CTD RBD localisation indicates usage of a proteinaceous receptor (Figure 4-1).

The S2 subunit is composed of five domains: the fusion protein, heptapeptide repeats 1 and 2 (HR), the transmembrane domain and the cytoplasmic fusion domain (CT). The fusion peptide is rich in hydrophobic glycine and alanine amino acids required for insertion into the target membrane (Millet & Whittaker, 2018). The two HR regions form a helical bundle, the fusion core, which is indispensable for viral entry and conserved such that pan-coronavirus antiviral EK1 targeting a site in HR1 was designed as a fusion inhibitor (S. Xia *et al*., 2019, 2020). The

transmembrane domain anchors the spike inside the viral membrane (Y. Huang *et al*., 2020).

The S protein forms a trimer and requires activation by cleavage to undergo conformational change for host membrane fusion. The first cleavage event occurs at the S1/S2 site by host proteases. The host protease acting at this site is determined by the presence of a poly-, bi- or monobasic cleavage site. In SARS-CoV-2, IBV (Bestle *et al*., 2020) and lab adapted OC43 (de Haan *et al*., 2008) a polybasic R-R-x-R motif is cleaved by host furin (Örd *et al*., 2020). Interestingly, SARS-CoV possesses a monobasic cleavage site which is cleaved by host trypsin (Belouzard *et al*., 2009). Some bat sarbecoviruses possess a dibasic motif at the S1/S2 cleavage site but mechanisms of cleavage have not been investigated. After cleavage at S1/S2, conformational change exposes the S2' TMPRSS site for cleavage by type II transmembrane serine protease (TMPRSS2) and cellular entry occurs via direction fusion of the viral envelope with the cell membrane (Bestle *et al*., 2020). The omicron variant of SARS-CoV-2 demonstrates less reliance on TMPRSS2 cleavage for cellular attachment and entry owning to a greater reliance on endocytic pathway (Zhao et al., 2022).

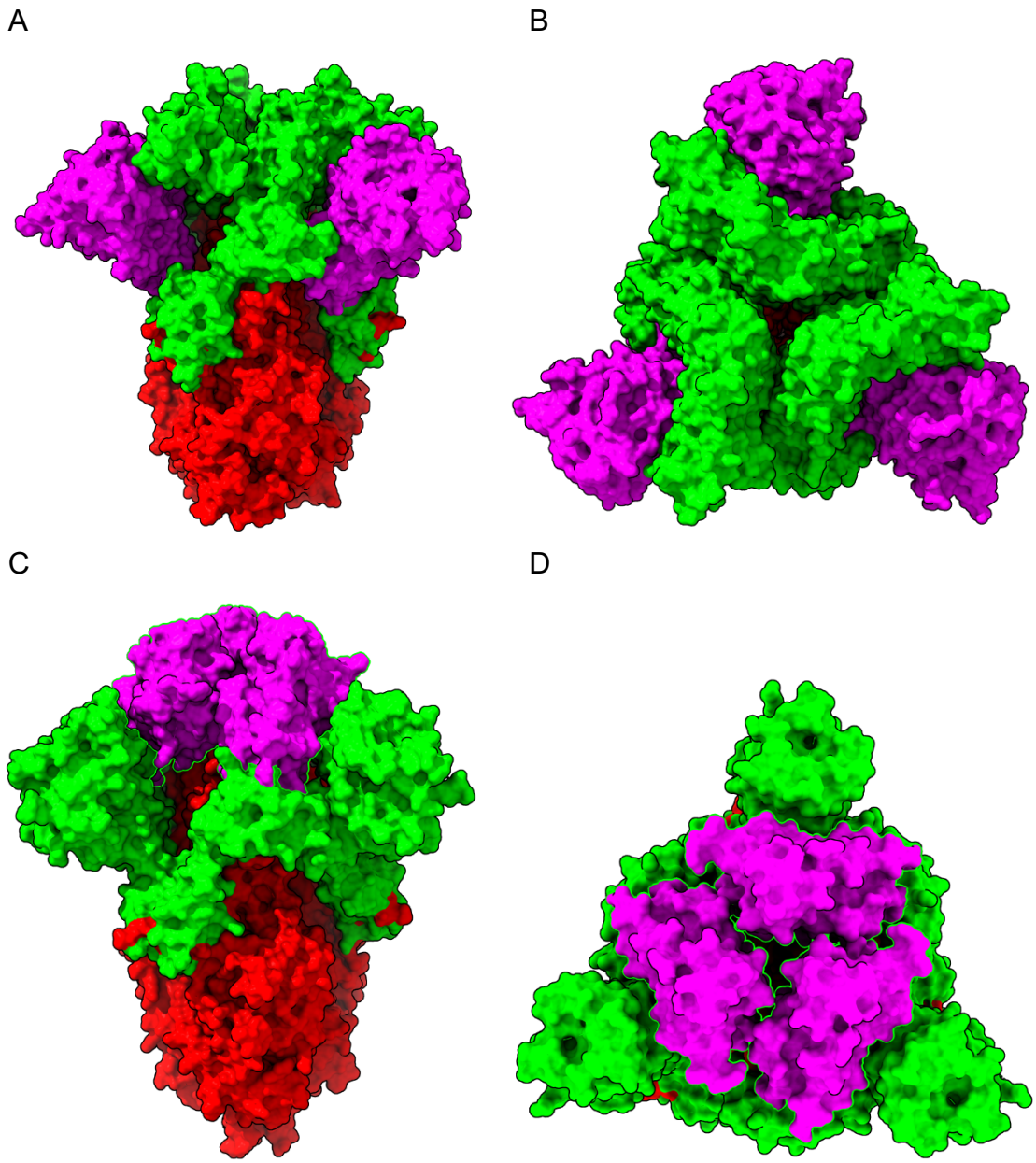


Figure 4-1. Spikes with the S1, S2 and RBD highlighted.

Trimeric spike proteins from SARS-CoV-2 (PDB 8D55) (A, B) and OC43 (PDB 7SB3) (C, D). The S1 subunit is shown in magenta, the RBD location in green and S2 in red. OC43 and SARS-CoV-2 depict the opposing receptor binding domain locations for the spike glycoprotein, NTD and CTD of S1, respectively. The S2 domain is shown in red. Visualised using ChimeraX.

4.1.3.2 Receptor binding domain in S1-NTD

Within the *Betacoronavirus* genus, Murine Hepatitis Virus (MHV), Bovine Coronavirus (BCoV) and related human pathogen HCoV-OC34 are examples of viruses with S1-NTD located RBDs.

4.1.3.2.1 MHV

MHV is an exception to the rule suggested above with an S1-CTD RBD that targets a proteinaceous receptor, the carcinoembryonic antigen related cell adhesion molecule (CEACAM1) (F. Li, 2016a; G. Peng *et al*., 2011; Williams *et al*., 1991). This is the only known exception to a NTD-S1 binding a protein receptor for entry (Kubo *et al*., 1994). Within this NTD, seven MHV strains were used to identified conserved residues absent in other *betacoronaviruses*. S33, L79, T82, Y162 and K183. Y162 were identified as important for binding (Thackray *et al*., 2005).

4.1.3.2.2 BCoV and HCoV-OC43

Bovine CoV (BCoV) and MHV possess domains in the S1-NTD which mimic the structure of eukaryotic galectin which bind to sugars exposed in the host cell surface. However, only the BCoV galectin domain recognises a sugar, 5-N-acetyl-9-O-acetylneuraminic acid, for entry (G. Peng *et al*., 2012). Human pathogen OC43, is closely related to BCoV and potentially occurred because of a zoonotic spillover (Vijgen *et al*., 2006). The 5-N-acetyl-9-O-acetylneuraminic acid ligand is expressed widely across mammalian tissues and thus the sugar

receptor is an potential risk factor in the transmission of BCoV from cattle to humans (F. Li, 2016b).

4.1.3.3 Receptor binding domain in S1-CTD

4.1.3.3.1 MERS

Middle Eastern Respiratory Syndrome (MERS) emerged as a novel coronavirus in 2012 after spillover and drew global attention as a potential pandemic agent with high case fatality rates (CFR). At this time, Angiotensin Converting Enzyme 2 (ACE2) was known as the entry receptor for SARS-CoV, but this was eliminated as the target receptor for MERS-CoV EMC12, the outbreak strain. The host cell receptor for MERS-CoV entry was identified as DPP4 (CD26) which is expressed in the lower respiratory tract (Raj *et al*., 2013). The MERS-CoV RBD was found to be in the S1-CTD at amino acids 367-606 (Lu *et al*., 2013). Interestingly, an extensive receptor screening of merbecoviruses highlighted DPP4 binding is a trait conserved to clade 1 merbecoviruses (Letko, 2024).

4.1.3.3.2 SARS-CoV

The SARS-CoV spike binds human ACE2 (hACE2). The S1-CTD of SARS-CoV contains a receptor binding domain (RBD) with two subdomains: a core and receptor binding motif (RBM). The RBM is a concave surface which binds to hACE2. The RBM core structure is a five-stranded anti-parallel beta-sheet with two ridges formed by loops (F. Li, 2016b). A 193 amino acid fragment (318-510) was demonstrated to bind hACE2 more effectively than full length S1, with smaller fragments 327-510 and 318-490 not binding hACE2 at a detectable level.

One mutation of aspartic acid at amino acid 454 abolished the association of full S1 with ACE2 (F. Li *et al*., 2005a; Wong *et al*., 2004).

4.1.3.3.3 SARS-CoV-2

SARS-CoV-2 is comparable to SARS-CoV in that it utilises hACE2 for cellular entry via binding of the S1-CTD. The RBD has very similar structure to SARS-CoV. The contact between ACE2 and SARS-CoV-2 RBM can be split into three clusters, The two ends of the RBM bridge interact with the N and CTD of the hACE2 α1 helix and small areas on the α2 helix. The N terminus of α1 in hACE2 (Tyr41, Gln42, Lys353 and Arg357) binds residues Gln498, Thr500, Asn501 in the RBM by a network of hydrogen bonds. In the middle of the RBM bridge, Lys417 and Tyr453 interact with hACE2 Asp30 and His34. At the C terminus of the α1 in ACE2, Gln474 in RBM is hydrogen bonded to Gln24 of hACE2 and Phe486 interacts with Met82 in ACE2 via Van der Waals forces (Yan *et al*., 2020). The *Coronaviridae* family contains many different viruses with a variety of cellular entry tropisms. Specifically, within the *Sarbecovirus* genus, it is hypothesised that the main receptor used for host cell entry is ACE2, as seen for SARS-CoV and SARS-CoV-2.

4.1.4 Other Coronavirus structural proteins

4.1.4.1 Envelope

The envelope protein (E) is the smallest coronavirus structural protein, measuring between 76 and 109 residues long (Mukherjee *et al*., 2020). It is important in the functional interface of the virus particle with the host and has functions in virion assembly, budding (CoVs obtain lipid bilayer from cis-golgi network), morphogenesis and regulation of other cellular functions (Liao *et al*., 2006).

4.1.4.2 Envelope protein structure and topology

The E protein can be split into the short hydrophilic NTD (MYSFVSEETG), a hydrophobic transmembrane domain with an aromatic belt (FXXFXXF), a hydrophobic CTD which contains the oligomerisation unit (SFYVYSRVK; SARS) (Liao *et al*., 2006) and a PDZ (postsynaptic density protein 95/*Drosphila* disc large tumour suppressor/Zonula occludens-1 protein (zo-1)) (Schoeman & Fielding, 2019).

The topology of the E protein in coronaviruses has undergone scrutiny in the field with experimental studies utilising antibodies targeting the CTD and NTD demonstrating opposing topologies for the E protein in different coronaviruses (Godet *et al*., 1992; Maeda *et al*., 2001; Q. Yuan *et al*., 2006). In sarbecoviruses, the SARS-CoV E is predicted to have two different topologies from experimental (Q. Yuan *et al*., 2006) and in silico predictions (Mukherjee *et al*., 2020). Experimentally, the structure of the SARS-CoV E protein was demonstrated to have an atypical short transmembrane helical hairpin with Phe23 the centre of the hairpin inversion within the lipid bilayer (Figure 4-2). Whereas, in silico methods do not predict the atypical helical hairpin structure and instead predict a typical transmembrane helical structure (Liao *et al*., 2004; Torres *et al*., 2005). Such structural differences would suggest different functions. The experimentally supported monomeric structure is predicted to form a pentameric pore in the lipid bilayer, and is supported by bacterial studies which demonstrate membrane permeability (Arbely *et al*., 2004).

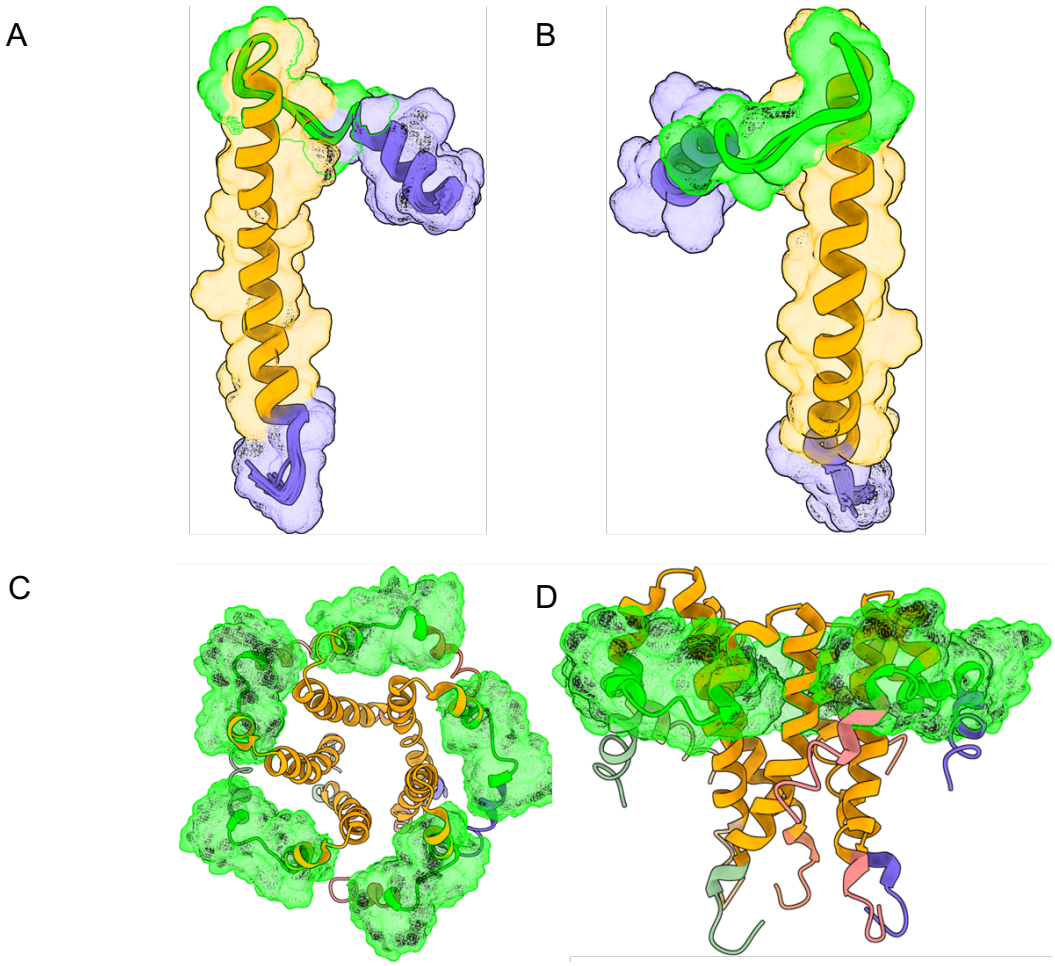


Figure 4-2. The coronavirus envelope protein.

One of the possible topologies of the coronavirus envelope protein. In its monomeric form (A, B (PDB 2MM4)) there is an alpha helices transmembrane domain (orange) and a flexible linker domain (green) which causes the kink in the membrane. In its pentameric pore-forming structure, the flexible linker domains form a pore in the membrane (C, D (PDB 5X29)).

4.1.4.3 Envelope functions

The E protein has a multitude of functions related to coronavirus replication, morphogenesis, virion assembly and virulence. This is exemplified by high levels of E expression during viral replication (DeDiego *et al.*, 2007; Venkatagopalan *et al.*, 2015).

Viral particle formation can be E-dependent, but this observation is not conserved between coronaviruses, and nucleoprotein- (N), membrane- (M) and E-

dependent particle formation mechanisms have also been reported (Baudoux *et al*., 1998; Corse & Machamer, 2000; Hsieh *et al*., 2005; Q. Huang *et al*., 2004; Vennema *et al*., 1996). ΔE mutants have been observed to have reduced virion quality (DeDiego *et al*., 2007; Kuo & Masters, 2003; Ortego *et al*., 2007), altered morphology and reduced plaque size Fischer *et al*., 1998), pathogenesis attenuation in in vivo infection models (DeDiego *et al*., 2007) and interestingly, increased apoptosis (DeDiego *et al*., 2011; Ruch & Machamer, 2012). However, E has also been associated with pro-apoptotic effects on the host cell by ionic disturbance and removing a host cell from homeostasis (Mukherjee *et al*., 2020). E is also hypothesised to interact with other structural proteins, such as M, ORF2a, and nsp3 (Álvarez *et al*., 2010; Corse & Machamer, 2003; de Haan *et al*., 2000).

The oligomerisation unit in the CTD is proposed to function in the amyloidogenic propensity of E, potentially forming helical structures (Ghosh *et al*., 2015; Mukherjee *et al*., 2020). The pentameric structure of E in the host cell membrane forms viroporins which have been demonstrated to increase membrane permeability e.g. Influenza virus M2 (M. E. Gonzalez & Carrasco, 2003; Hong & DeGrado, 2008; Nieto-Torres *et al*., 2011; Pervushin *et al*., 2009). In experimental studies investigating membrane permeability of *Escherichia coli,* expression of SARS-CoV E induces increased membrane permeability (Madan *et al*., 2005).

The SARS-CoV E protein has a PDZ-binding motif at the terminus of the CTD which binds to PALS1 PDZ domain. PALS1 is a tight junction associated protein and thus is directly implicated in the membranous structure in mammalian cells. This interaction with a protein directly involved in cellular structure implicates the envelope protein as a virulence factor in disruption of epithelial structure in SARS-

CoV patients (Teoh *et al*., 2010). This PDZ domain is a novel feature of SARS-CoV and thus may be a product of recent adaptation for increased fitness to human hosts.

To summarise, the E protein of coronaviruses performs diverse functions because of its association with other structural proteins, localisation at the Golgi, ER and ERGIC structures, roles in virion assembly and potentially, pathogenesis.

4.1.5 Membrane

The M protein has a triple helix bundle which spans the membrane three times and presents a NH2-terminal domain on the virion surface and a long COOH cytoplasmic domain inside the virus. The M protein is the most abundant structural protein in the virion of SARS-CoV (Bianchi *et al*., 2020). The M protein interacts with all other structural proteins, including the E and S in the ERGIC (de Haan *et al*., 1999; Kumar *et al*., 2023) and then N for virion assembly. M stabilises the interaction between N and vRNA, aiding in assembly completion and vRNA packaging (Thomas, 2020; Z. Zhang *et al*., 2022).

The M protein also interacts with spike protein, demonstrated by coimmunoprecipitation and co-localisation experiments (de Haan *et al*., 1999). Alternation in the transmembrane domain of the triple spanning membrane protein was found to impact complex formation with the spike protein; but complete abrogation was not observed (de Haan *et al*., 1999).

4.1.6 Nucleoprotein

The N protein forms a dimer where each monomer has three distinct domains which are highly conserved across all coronaviruses. At both termini there are intrinsically disordered regions near RNA binding domains required for the main functionality of N. Spanning the middle between either RNA binding domain is a serine/arginine rich linker region. A nuclear localisation signal for cellular localisation of the protein is at the CTD RNA binding domain (Caruso *et al*., 2022; McBride *et al*., 2014). The CTD also contains the domain which forms a hydrophobic helix-rich region required for homodimerisation (Caruso et al., 2022; Chang et al., 2005). The homodimerisation feature of N is conserved within *alpha*, *beta* and *gammacoronaviruses* (Yu *et al*., 2006).

4.1.6.1 Formation of the ribonucleoprotein complex and viral replication

The most important role of N in an infected host cell is the packaging of genomic RNA into ribonucleoprotein (RNP) complexes called nucleocapsids. The NTD associates with the 3' end of vRNA when forming the RNP structure, at least in part, by electrostatic interactions (Spencer & Hiscox, 2006). More specific mapping of RNA binding sites for vRNA in the SARS and IBV N protein is mediated by positively charged lysine and arginine residues which bind a 32-nucleotide stem loop (Q. Huang *et al*., 2004). Also important to RNP formation is the homodimerisation of N.

4.1.6.2 Coronavirus replication

Coronavirus replication occurs in two different processes: replication of the genome for packaging into virions, and the generation of sub genomic mRNA (sgmRNA) for translation into functional proteins. The genomic RNA (gRNA) released by viral uncoating undergoes immediate translation into the two primary nsp precursors, ORF1a and ORF1b. These are co- and post-translationally processed into the 16 mature non-structural proteins which form the replication-transcription complex (RTC) needed for replication of the large coronavirus genome. During this process, organelles for viral replication are generated at the ER e.g., double membrane vesicles (DMV), convoluted membranes (CM), and small open double membrane spherules (DMS), to generate a protective cellular environment for replication and transcription of sgmRNA.

For the generation of mRNA transcripts, a negative sense genome copy is produced to act as a template for transcription. This transcription process is termed a discontinuous transcription process. ORFs in the 3' region of the genome possesses transcription regulatory sequences (TRS) which the RTC can recognise during transcription. The discontinuous transcription process results in set a negative strand sgRNA which act as templates for synthesising mRNA for translation (Caruso *et al*., 2022; V'kovski *et al*., 2021; Zúñiga *et al*., 2010).

4.1.7 Global distribution of sarbecoviruses

A systematic review summarising coronavirus surveillance in bats between 1996 and 2019 was published in 2023 (Cohen *et al*., 2023). This dataset was combined with additional records collected from literature searches as to included studies up to 2023. The meta-analysis identified a disproportionate number of studies

had been conducted per country in Asia compared to Europe or Africa and highlighted regions surveillance is absent (Cohen *et al*., 2023). This sampling bias highlighted in the number of *Rhinolophus* species within which sarbecoviruses have been discovered (Figure 4-3).

Of particular interest to this project is the detection of sarbecoviruses in Europe and the expansion of knowledge since 2020. Before the discovery of RhGB01 (Crook et al., 2021), only one fully resolved genome was in the public domain for Europe, a *Sarbecovirus* from Bulgaria, BM48-31/2008. This *Sarbecovirus* was discovered In *Rhinolophus blasii* in the Strandja National Park in 2008 when four *Rhinolophus* species were positive by PCR for SARS-related sarbecoviruses (Drexler *et al*., 2010).

Since RhGB01 was discovered in 2020, seven additional sarbecoviruses have been characterised from *Rhinolophus* species in Europe. Two sarbecoviruses were detected and characterised in Russia. Khosta-1 identified in *Rhinolophus ferrumequinum* and Khosta-2 *in Rhinolophus hipposideros* during sampling near Sochi in 2020. The prevalence of Khosta-1 was much higher than Khosta-2 in the sampling locations (62.5 vs 1.75%) and experimental investigation has identified different host ACE2 tropisms. A second study in the UK identified five new sarbecoviruses closely related to RhGB01 after sampling via a conservationist network, highlighting prevalence in *Rhinolophus hipposideros* and, for the first time in the UK, *Rhinolophus ferrumequinum*.

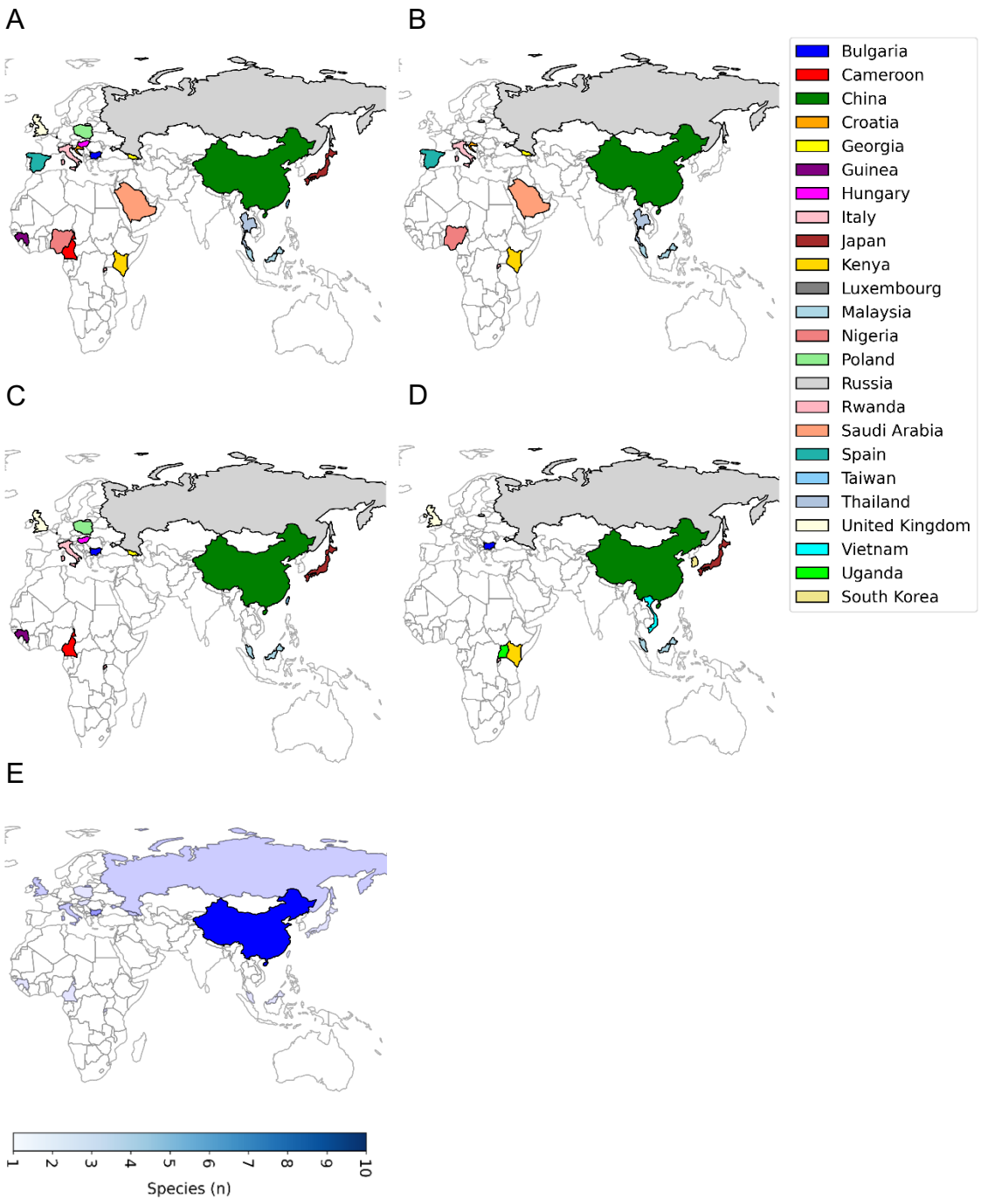


Figure 4-3. The global distribution of sarbecoviruses.

The global distribution of sarbecoviruses based on sampling records from the literature and available full genomes. Here *Sarbecovirus* distribution is categorised by country. Specifically, in horseshoes and other species (A), non-horseshoe species (B), and exclusively horseshoe species (C). Additionally, the availability of full genomes (D). Horseshoe species' distribution is coloured based on the number of species sarbecoviruses have been detected in (E).

4.1.8 Variation in ACE2 usage for sarbecoviruses from bats

The major human cellular receptor for host cell entry for SARS-CoV and SARS-CoV-2 is hACE2. 14 contact residues were identified as important for the recognition of hACE2 by the SARS-CoV spike protein: T402, R406, Y436, Y440, Y442, L472, N473, Y475, N479, Y484, T486, T487, G488, Y491 with Y442, L472, N479, T487 and Y491 being critical for the interaction (F. Li *et al*., 2005a). The emergence of SARS-CoV-2 highlighted an extended set of contact residues for interaction with hACE2, namely, V417, T446, Y449, Y453, Y455, L456, P475, L486, N 487, Y489, N493, G496, Y498, T500, T501, G502, Y505 (Lan *et al*., 2020).

Prior to the COVID-19 pandemic, ACE2 binding had only been assessed in SARS-CoV and closely related sarbecoviruses sampled from Yunnan in China. After the emergence of SARS-CoV, surveillance studies identified sarbecoviruses from a diverse range of *Rhinolophus* bats – termed SARS-like CoVs (SL-CoVs) which phylogenetically cluster with SARS-CoV (Ge *et al*., 2013; B. Hu *et al*., 2017; D. Hu *et al*., 2018; Lau *et al*., 2005, 2007, 2010; W. Li, 2005; Lin *et al*., 2017; Poon *et al*., 2005; X. C. Tang *et al*., 2006; L. Wang *et al*., 2017; Woo *et al*., 2006; Z. Wu *et al*., 2016; L. Xu *et al*., 2016; J. Yuan *et al*., 2010). Despite the observed clustering, it was demonstrates in vitro that RBDs from SL-CoVs possess different ACE2 binding profiles and thus not mirroring SARS-CoV tropism (Ren *et al*., 2008).

Rp3 was identified in *Rhinolophus pearsonii* in China in 2004 (W. Li, 2005). The Rp3 spike does not bind hACE2 and requires the replacement of its RBD with that of BJ01 (SARS-CoV) to restore the hACE2 binding phenotype. This was the first indication that sarbecoviruses from *Rhinolophus* bats could demonstrate

different ACE2 binding profiles (Ren *et al*., 2008). Additional SL-CoVs, WIV1 and WIV16, were identified to have a broad ACE2 tropism, binding human, Chinese horseshoe and civet ACE2s (Ge *et al*., 2013; X.-L. Yang *et al*., 2016) and highlighting potential transmission pathways for zoonotic spillover.

RsSHC014-CoV (SCH014) was isolated from Chinese horseshoe bats in 2011-12 (Ge *et al*., 2013). SCH014 demonstrates the ability to replicate in vitro but despite being closely related to WIV1, with reduced viral replication in mouse models compared to SARS-CoV strain Urbani (Menachery *et al*., 2015). WIV1 contains three amino acid differences compared to SARS-CoV Urbani outside of the hACE2 contact residues. Whereas SCH014 has five amino acid differences within the hACE2 contact residues (Y442, L472, N479, T487 and Y491) (Menachery *et al*., 2015), potentially explaining the difference in ACE2 binding and replication in vivo between SCH014 and WIV1.

A longitudinal study in Yunnan between 2011-15 identified 11 novel SARSr-CoV strains. Three strains were identified to bind hACE2 (Rs4874, Rs4241S, Rs7327S) (B. Hu *et al*., 2017). Over four years, this study identified multiple strains from the *same* cave, highlighting the sheer prevalence of sarbecoviruses with pandemic potential in the region.

In 2020, a large scale *Sarbecovirus* ACE2 binding study utilising pseudoparticles highlighted hACE2 usage in only *Sarbecovirus* clade 1 (SARS-CoV and SL-CoVs), supporting previous characterisation studies. hACE2 binding was absent in *Sarbecovirus* clade 2 (SARSr-CoV from Asia) and clade 3 (BM48-31-like/African-sarbecoviruses) viruses, suggesting clade specific receptor usage. Clade 2 sarbecoviruses consist of SARSr-CoVs clustering outside of SL-CoVs.

The clade assignments were based on spike and RBD phylogenies. Interestingly, treatment with host protease enhanced hACE2 binding in clade 2 sarbecoviruses identified in China. Representative RBD sequences from clades 2 and 3 demonstrate significant variation from the defined contact residues for SARS-CoV hACE2 interaction and only replacing these 14 residues rescued hACE2 binding in clade 2 and 3 RBDs. It is hypothesised that the deletions observed in the RBD are also critical for the structure of the ACE2 binding interface leading to reduced ACE2 binding (Letko *et al*., 2020).

The USAID-PREDICT project (Gilardi & Mazet, 2019) has identified multiple sarbecoviruses from Uganda and Africa with limited nucleotide identity to SARS-CoV and SARS-CoV-2 (Wells *et al*., 2021), in addition to BM48-31 in Europe (Drexler *et al*., 2010). A *Sarbecovirus* from Kenya was characterised in 2010 (Tao & Tong, 2019). These non-asian sarbecoviruses are contained within lineage 4 (clade 3). It was identified that the European/African sarbecoviruses differ from ACE2 tropic sarbecoviruses in the receptor binding interface, including small deletions within the RBD. This was corroborated by the absence of ACE2 binding in SARS-CoV chimeric pseudoparticles (Wells *et al*., 2021).

A high throughput assay investigating binding to ACE2 orthologues for a range of sarbecoviruses identified binding as a trait which had been lost in specific clades, and reappeared in some sarbecoviruses outside of the Asian clades (Starr *et al*., 2022). This high throughput assay was the first investigation into bat ACE2 binding profiles for different sarbecoviruses, providing inference for transmission networks between *Rhinolophus* species. Clade 1 SL-CoVs SCH014 and SARS-CoV Urbani successfully bind to all ACE2 alleles from *Rhinolophus sinicus* included in the screening whereas this was absent in clade 2

representative strains, YN2013 and HKU3. RBDs from sarbecoviruses in Clade 3 were identified to have affinity for ACE2s from *Rhinolophus affinis*. Khosta-2 (clade 3) was also identified to bind hACE2 at comparable levels to clade 1 sarbecoviruses (Starr *et al*., 2022).

Another study using chimeric bat *Sarbecovirus* RBD in a SARS-CoV backbone expanded the ACE2 binding profile of specific clade 3 sarbecoviruses. The Khosta-1 and BtKY72 ACE2 binding range was expanded to include ACE2 from *Rhinolophus ferrumequinum*, *Rhinolophus alycone*, *Rhinolophus landeri* and *Rhinolophus sinicus* at reduced affinity compared to SARS-CoV hACE2 binding. Khosta-1 binding to *Rhinolophus alcyone* is comparable to that of SARS-CoV hACE2, suggesting a strong affinity. All other clade 3 sarbecoviruses tested were identified to bind *Rhinolophus alycone* (Roelle *et al*., 2021).

Studies into *Sarbecovirus* tropisms have identified clades with variation in ACE2 tropism. In sarbecoviruses where ACE2 binding is currently unknown or hypothesised to be absent, hypothesises as to the evolutionary or methodological reason for this is required. It is possible some sarbecoviruses have lost the ability to bind ACE2 and utilise a different host cell receptor, or that the ACE2 range for these sarbecoviruses is narrow, possibly single species, and thus hard to identify in vitro, or ACE2 binding affinity is below the limit of detection in experimental studies conducted.

4.2 Results

A set of 53 faecal droppings from *Rhinolophus hipposideros* found in a roost Gloucestershire were collected during an annual population survey and screened using mNGS for the presence of RNA viruses.

4.2.1 Taxonomic classification reveals betacoronavirus RNA within a subset of faecal samples

An initial screening was conducted by taxonomic classification via Kraken2. Nucleic acid classified as either *Coronaviridae* (family) or *Betacoronavirus* (genus) were detected in five and four samples, respectively (Figure 4-4). The internal sequencing control (Hazara virus, *Orthonairovirus*) was spiked at 10^6/mL^{-1}, was observed in 12 samples. The percentage total reads classified in each sample ranged from 28-96%.

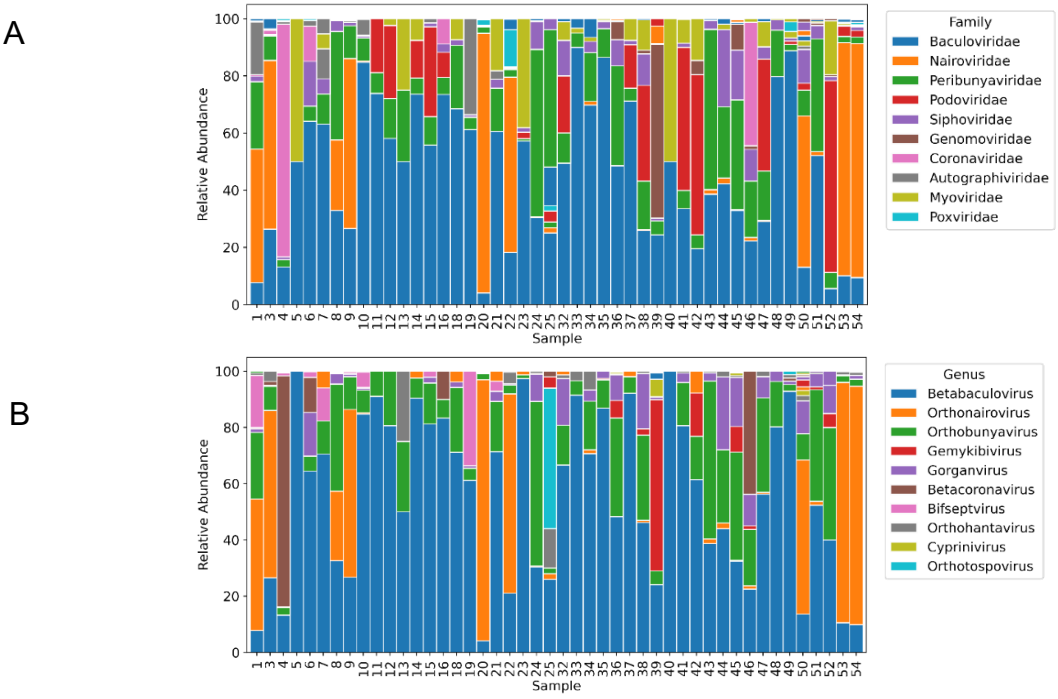


Figure 4-4. Read classification at family and genus level.

Taxonomic classification for the relative abundance at family (A) and genus (B) level for bat faecal samples from the 2020 collection at the Cinderford roost.

4.2.2 Identifying positive faecal samples

Four faecal droppings were detected to have >100 reads classified at family level to *Coronaviridae* (Figure 4-5). Sample four, six and 16 possess only reads classified as SARS-CoV-2. Sample four has 5193 reads classified to SARS-CoV-2 (2.54%). Sample six has 1595 reads classified to SARS-CoV-2 (1.35%). Sample 16 has 108 reads classified as SARS-CoV-2 (0.03%). Sample 46 has 1571 (0.26%), 269 (0.04%) and 212 (0.03%) reads classified as BM48-31, SARS-CoV and SARS-CoV-2, respectively.

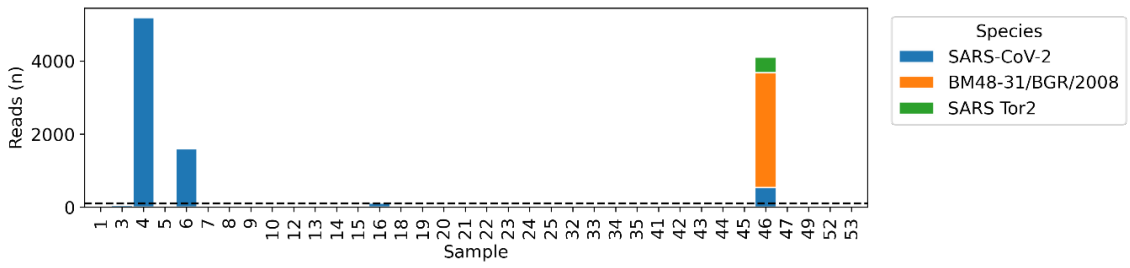


Figure 4-5. Read classification at species level.

The number of reads taxonomically classified to three different *Sarbecovirus* species for the four samples with >100 reads classified at family level to *Coronaviridae*.

Mapping raw reads for samples four, six and 46 suggests only sample 46 contains reads which are distributed across the betacoronavirus genome. Sample 46 has 5022 reads (0.45%) mapping to the BM48-31 reference genome, compared to 1571 classified by Kraken2. In samples four and six, reads are mapping in specific locations in the ORF1ab at the 5' of the genome, and region at the genomic 3' region.

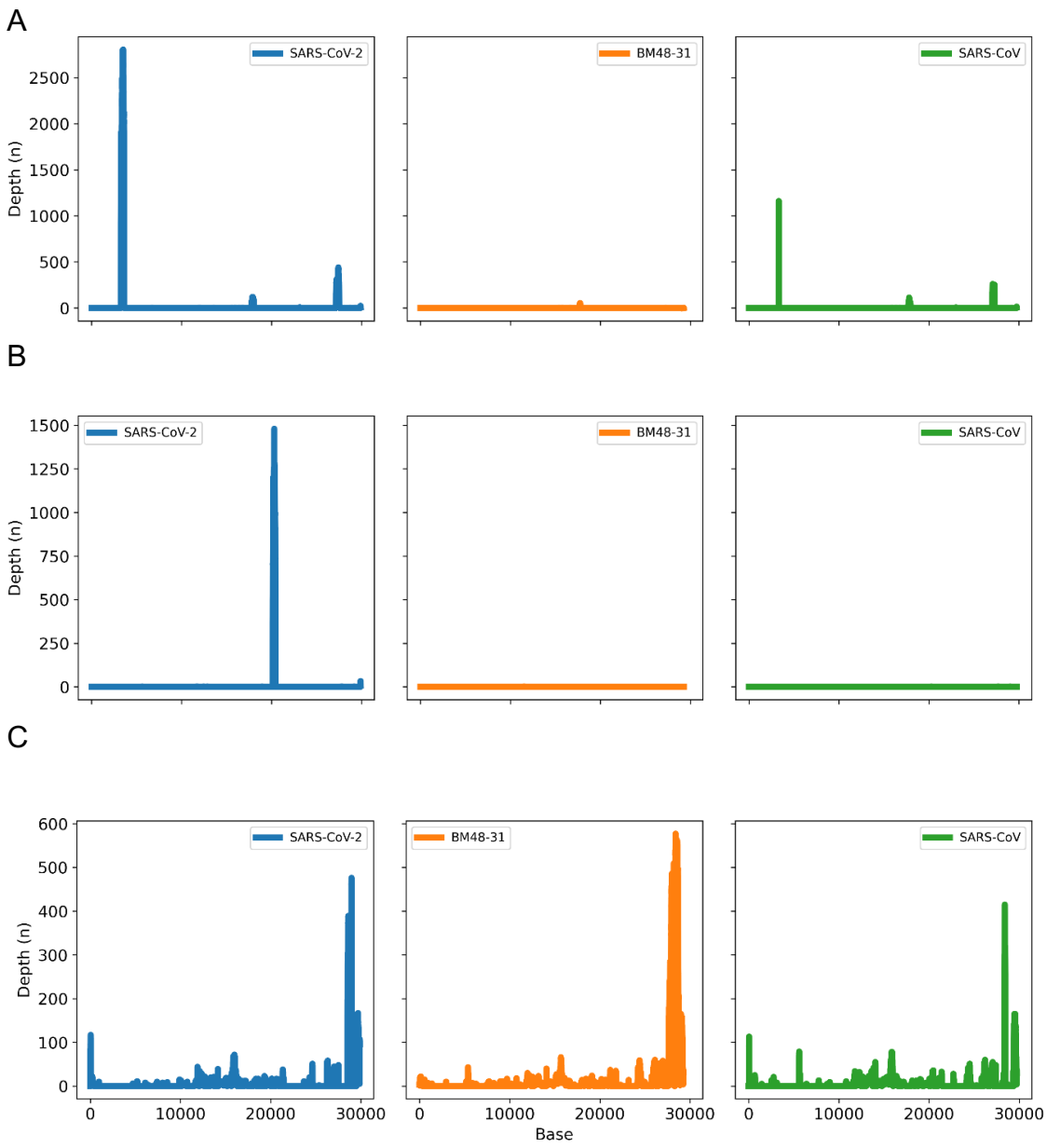


Figure 4-6. Reads mapped to representative *Sarbecovirus* genomes.

Raw reads mapped to three *Sarbecovirus* reference genomes identified by Kraken2 taxonomic classification for sample 4 (A), 6 (B) and 46 (C). Sample four has 10430, 1884 and 66 reads mapped to SARS-CoV-2, SARS-CoV and BM48-31. Sample six has 3248, 6 and 2 reads mapped to SARS-CoV-2, SARS-CoV and BM48-31. Sample 46 has 3177, 2699 and 5502 reads mapped to SARS-CoV-2, SARS-CoV and BM48-31.

4.2.3 Generating a *Sarbecovirus* genome from sample 46 using a de novo hybrid assembly

A de novo assembly of Illumina short sequencing reads (library 1) generated multiple contigs ranging from 500 bp to 6 kb with high levels of homology to BM48-31 by blastn search (Appendix 13). De novo assembly with additional reads from a second round (library 1 and 2) of short read sequencing increased the maximum contig length to 21kb, approaching the expected length for a coronavirus genome. Long reads (ONT) were obtained and used in conjunction with short reads to generate a *Sarbecovirus* genome using a de novo hybrid assembly (Table 4-2).

	Total reads (N)	Reads classified		Taxon (classification)
		N	%	
Illumina library 1	612,859	2084	0.34	Coronaviridae (F)
		1668	0.27	BM48-31 (S)
		359	0.06	SARS-CoV (S)
Illumina library 2	1,021,966	3817	0.37	Coronaviridae (F)
		3119	0.31	BM48-31 (S)
		628	0.06	SARS-CoV (S)
ONT library	562,461	5004	0.89	Coronaviridae (F)
		3681	0.65	BM48-31 (S)
		1148	0.20	SARS-CoV (S)

Table 4-2. Sequencing metrics for the 2020 faecal samples.

Sequencing metrics for the three libraries used for generating the RhGB01 genome assembly. The number of reads classified to each of the three representative *Sarbecovirus* genomes is provided.

The hybrid assembly generated a contig 29324 bp in length with 80.7, and 79.7% identity to BM48-31 and SARS-CoV-2, assessed by blastn (Appendix 14). To assess the coverage across the genome assembly, raw reads were mapped to the contig. Raw/trimmed nanopore reads mapped across the contig has >50X depth across the genome assembly. When combined, Illumina and ONT raw

reads demonstrate >100X depth across the genome assembly, providing high confidence for the genome assembly.

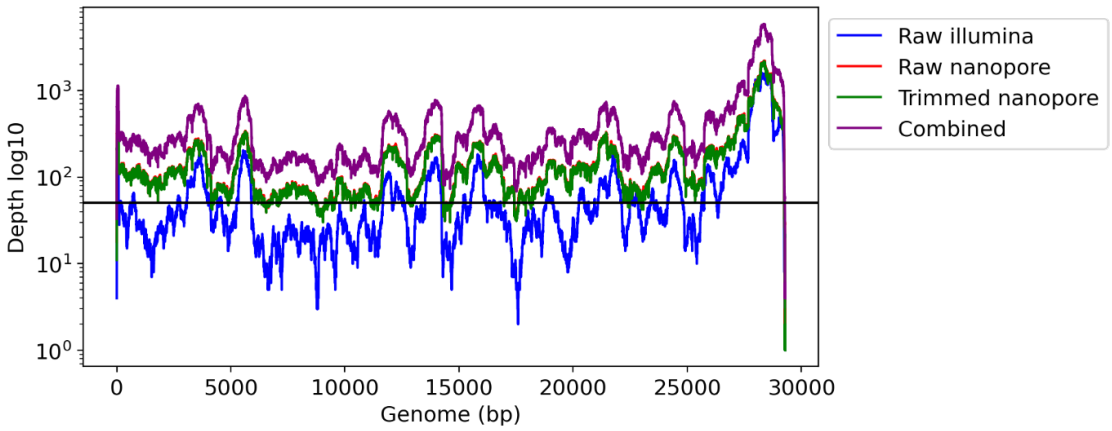


Figure 4-7. Sample 46 mapped to the genome assembly.

Read depths for both sequencing technologies across the 29.3kb contig generated by the hybrid de novo assembly. The horizontal line depicts 50X depth.

4.2.4 Characterising RhGB01

4.2.4.1 The genomic organisation of RhGB01

The genome assembly was named RhGB01, to represent its host, *Rhinolophus hipposideros* and the country of origin, Great Britain. This is the first identification and characterisation of a *Sarbecovirus* in lesser horseshoe bats in the UK and Europe. Annotation of the RhGB01 genome predicted 9 open reading frames (ORF). The genomic structure of RhGB01 resembles that of other sarbecoviruses. At the 5' end, a large open reading frame spanning approximately 21 kb encodes 16 non-structural proteins and possesses the characteristic -1 frameshifting ribosomal slippage site at the ORF1a and ORF1b intersection. At the 3' end are ORFs encoding structural proteins, and accessory

ORFs. In RhGB01, accessory proteins ORF3a, ORF6 and ORF7a are present. These annotations were validated manually by alignment with Wuhan Hu 1 (Figure 4-8). Sarbecoviruses often possess an ORF7b at the 5' of ORF7a. In RhGB01, ORF7b is truncated with 20 bases at the 5' region absent. There is also a small potential ORF upstream of the nucleoprotein, "ORF10/ORFX" but with no homology to known proteins.



Figure 4-8. The genome organisation of RhGB01.

Structural genes are highlighted by a bold border.

4.2.4.2 RhGB01 clusters monophyletically with European and African sarbecoviruses

Maximum likelihood phylogenies were inferred using the nucleotide and amino acid sequences for the nine predicted ORFs. Nucleotide phylogenies for ORF1ab, Spike (Figure 4-9), ORF3a, Membrane, ORF6 and Nucleocapsid all cluster RhGB01 monophyletically with two sarbecoviruses from Russia (Khosta-1 and Khosta-2), BM48-31 from Bulgaria, and a *Sarbecovirus* from Uganda. For the envelope nucleotide sequence, RhGB01 clusters with BM48-31, while other European and African sarbecoviruses cluster with sarbecoviruses form China. In the ORF7a nucleotide phylogeny, RhGB01 is located at the end of a long terminal branch extending from the root of the tree, possibly suggesting the presence of

an undiscovered genotype for ORF7a or a truncated\non-functional protein. The other sequences from Europe/Africa maintain the monophyletic clade separate from RhGB01.

For the spike glycoprotein, the amino acid phylogeny also clusters RhGB01 with the European and African sarbecoviruses. For ORF3a, the amino acid phylogeny clusters all European and African sarbecoviruses outside of the Asian *Sarbecovirus* clades, but also remove the monophyly of the Europe/Africa clade suggesting some differences within the ORF3a protein sequence. For the ORF3a, RhGB01 is more related to sarbecoviruses from China, Thailand, and Laos. RhGB01 clusters with a *Sarbecovirus* from China for the envelope protein sequence, suggesting the possibility for recombination.

For the membrane protein, RhGB01 clusters outside the other European/Asian sarbecoviruses, and is most closely related to sarbecoviruses from China. The ORF6, Nucleoprotein and ORF7a amino acid phylogenies agree with the nucleotide phylogeny, clustering European/African sarbecoviruses together and at the end of a terminal branch, respectively (Appendix 15).

4.2.4.3 RhGB01 is most similar at the nucleotide and amino acid level to Khosta-2

Nucleotide and amino acid alignments between *Sarbecovirus* sequences isolated from *Rhinolophus* bats facilitates identification of the number of differences between RhGB01 and other sarbecoviruses. At a nucleotide level RhGB01 shares closest identity to Khosta-2 for ORF1ab, spike, ORF3a, ORF6, ORF7a, and the nucleoprotein (Figure 4-9). For the envelope gene, RhGB01 shares equal

identity to BM48-31 and Khosta-1, demonstrating 18 nucleotide differences, compared to 20 nucleotide differences to Khosta-2. For the membrane gene, RhGB01 demonstrates a 90-nucleotide differences to Khosta-1 compared to 122 nucleotide differences to Khosta-2 (Appendix 16).

At the amino acid level RhGB01 shares closest identity with Khosta-2 for spike, ORF3a, ORF6, and the nucleoprotein. For the envelope protein, RhGB01 demonstrates 4 amino acid differences when compared to a selection of sequences from Rhinolophus bats from China, and 7 amino acid differences to other sarbecoviruses from bats in Europe. For the membrane protein, RhGB01 demonstrates 15 amino acid differences to Khosta-1 (Appendix 17).

4.2.4.4 Homology modelling for the structural proteins for RhGB01

The RhGB01 spike, envelope, membrane and nucleocapsid amino acid sequences were used for homology modelling with templates structures selected by highest GMQE score and percentage identity. The ERRAT2 scores range from 78.1-92.6 and the percentage residues in most favourable positions range from 81.5 to 92.6 suggesting models are of moderate quality (Appendix 18). Alphafold2 predictions were generated for structural proteins but exhibited large poorly supported regions or were outperformed by homology models in the chosen scoring parameters. All structural proteins for RhGB01 are structurally comparable to known structures from other coronaviruses (Figure 4-10).

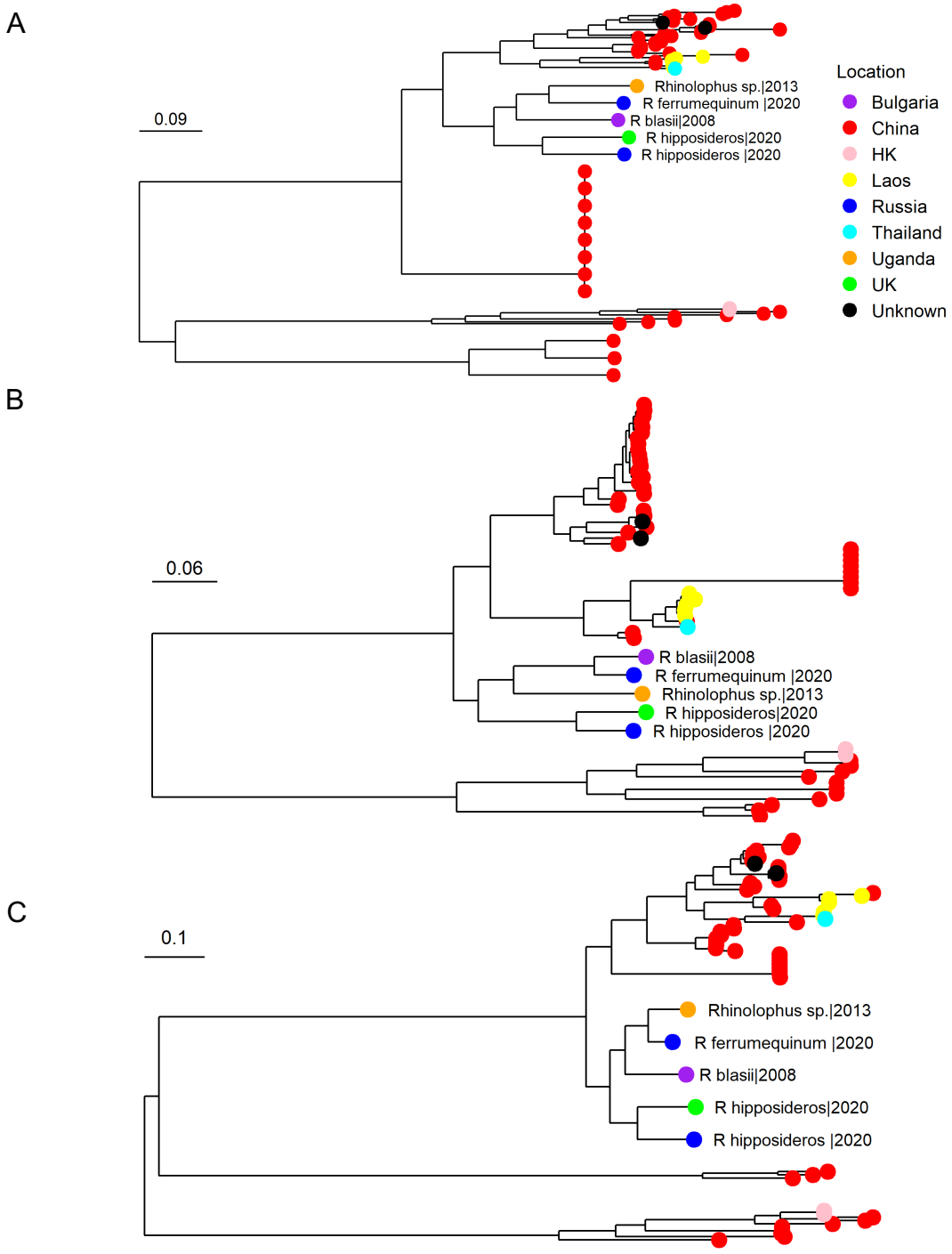


Figure 4-9. Phylogenies for *Rhinolophus* associated sarbecoviruses.

Maximum likelihood phylogenies inferred under an HKY+G4+I nucleotide substitution model for the spike (A) and ORF1ab (B) nucleotide alignments and spike glycoprotein amino acid sequence (C). Tips are coloured by sampling location. Phylogenies include all *Rhinolophus* associated sarbecoviruses in the public domain pre-2021.

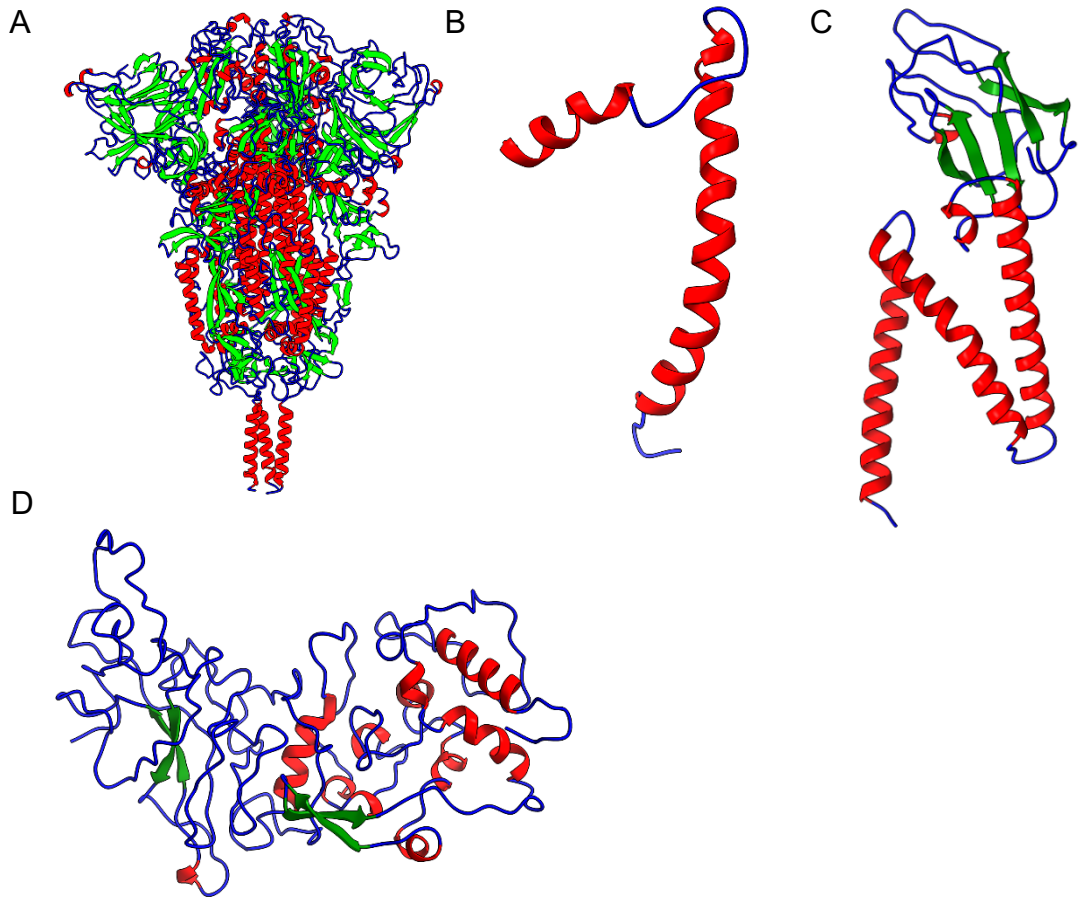


Figure 4-10. RhGB01 structural proteins.

Homology modelling for the RhGB01 structural proteins. The predicted structures for Spike (A), Envelope (B), Membrane (C) and Nucleocapsid proteins (D). Structures are coloured by secondary structure: helices (red), coil (blue), and strand (green).

4.2.4.5 RhGB01 as a public health risk

SARS-CoV-2 has demonstrated increased efficiency in utilising hACE2 for tropism, in part, because of a polybasic furin cleavage site (FCS) at the S1/S2 intersection resulting in increased cleavage and activation of spike (B. A. Johnson *et al*., 2021; Shang *et al*., 2020). The predicted RhGB01 spike protein demonstrates high structural similarity to SARS-CoV-2 spike in the S2 domain, but variation was observed in the N- and CTD of the S1 domain (entire spike, 1.33 RMSD). The spike protein of the European sarbecoviruses RhGB01 and

Khosta-2 possess a bibasic potential FCS precursor at the S1/S2 cleavage site (669-672 and 666-669). The R-A-K-Q motif observed here is one SNP from a polybasic R-A-K-R (R-X-K/R-R) canonical cleavage site (Q/CAA->R/CGA). Closer investigation into the minor nucleotide variation at the four codons for the RAKQ motif in RhGB01 demonstrated 100% conservation for all bases in this motif.

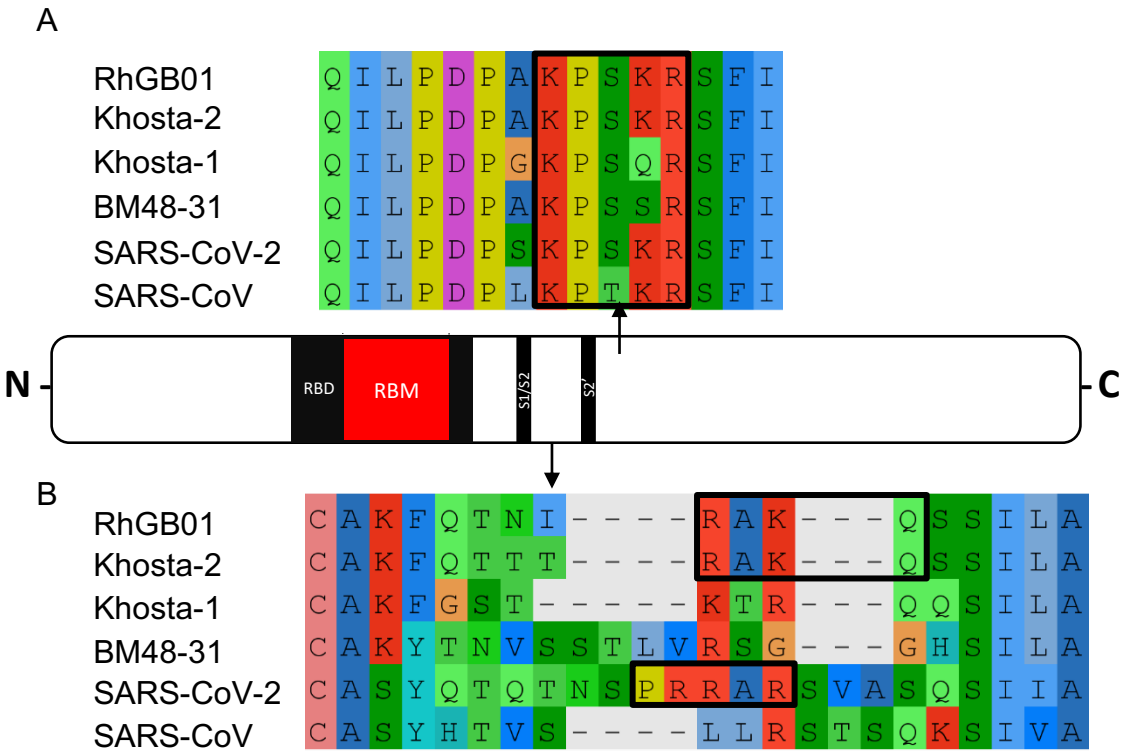


Figure 4-11. Cleavage sites present in the Spike.

A schematic of the two-cleavage sites in the Spike glycoprotein. The S2' cleavage site recognised by TMRPSS2 (A) with the cleavage motif highlighted. The furin cleavage site in SARS-CoV-2, with the bibasic alternative site in European sarbecoviruses (B).

The SARS-CoV-2 spike possess insertions up and downstream of the FCS within the disordered flexible loop which impacts accessibility for host furin. This increases the loop extension of the cleavage site from the core spike trimer,

allowing access for the host furin (Figure 4-12). RhGB01 also has the TMPRSS2 cleavage site observed in SARS-CoV-2 at the S2' site (Figure 4-11) required for conformational change of spike.

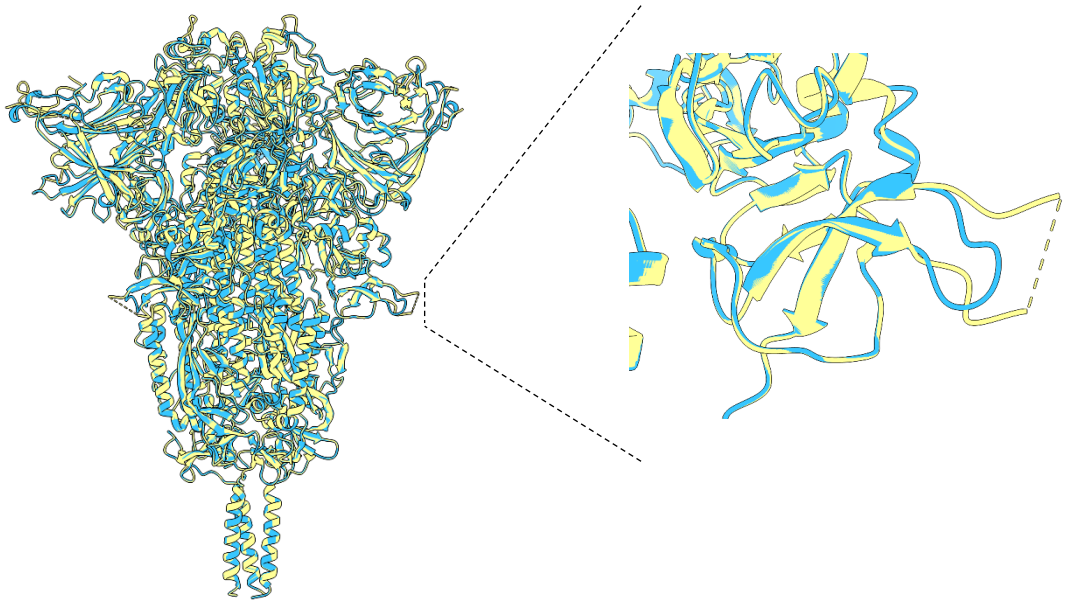


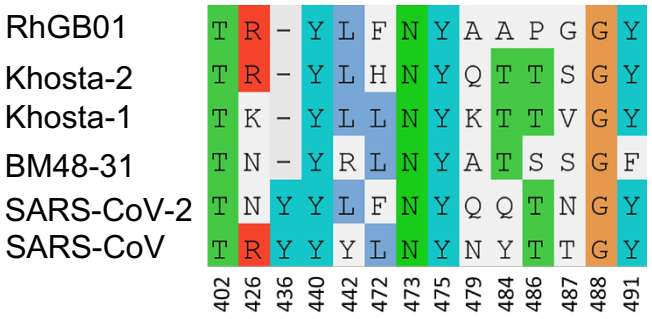
Figure 4-12. The RhGB01 and SARS-CoV-2 Spikes.

The RhGB01 spike protein (blue) superimposed to SARS-CoV-2 BA.2 Spike (PDB 8D55) (yellow) with the furin cleavage site on the extended loop shown. The "PRRAR" motif is not resolved in the crystal structure, but the base of the extended loop is visible.

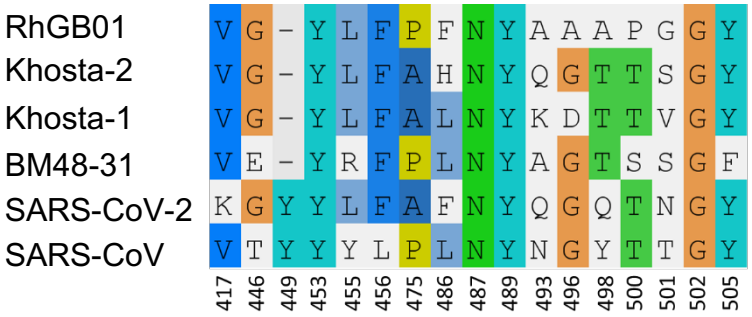
The RhGB01 RBD structure was predicted by homology modelling and superimposed onto other experimentally resolved RBD structures (Figure 4-13, Appendix 18). RhGB01 shares 7/14 and 9/17 contact residues with SARS-CoV and SARS-CoV-2. All sarbecoviruses included in the comparison have conserved contact residues at T402, Y440, N473, Y475, G488 for SARS-CoV and Y453, N487, and G502 for SARS-CoV-2. When superimposed onto SARS-CoV-2 in complex with hACE2, RhGB01 exhibits structural divergence in eight known

contact residues: V417K, Y449X, A475P, Q493A, G496A, Q498A, T500P, N501G (Figure 4-13).

A



B



C

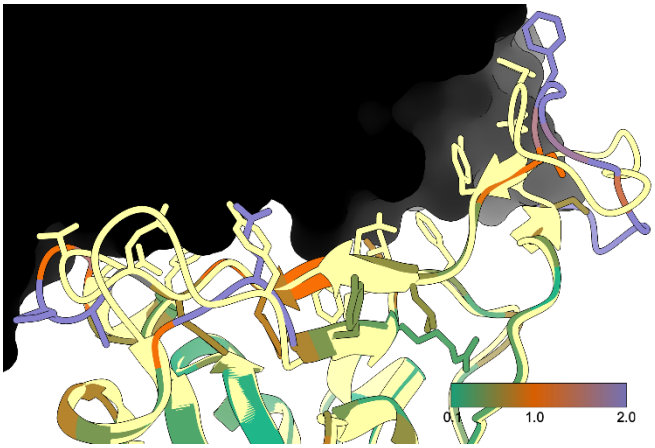


Figure 4-13. hACE2 contact residues in sarbecoviruses from Europe.

An alignment of SARS-CoV (A) and SARS-CoV-2 (B) hACE2 contact residues and their conservation in selected European sarbecoviruses discovered pre-2021. The structural superimposition of homology modelling predicted RhGB01 RBD (coloured by RMSD) to resolved SARS-CoV-2 in complex with hACE2 (PDB 6M0J) (C).

4.2.4.6 A mammalian ACE2 screening suggests a possible narrow host range for RhGB01 ACE2 usage

To assess the binding capability of the RhGB01 RBD, lentiviral-based pseudoviral particles expressing the RhGB01 spike or a chimera of RhGB01-RBD and SARS-CoV spike were produced. The expression of the RhGB01 spike and chimera are confirmed by bands at 250kDa and 130kDa. The 250kDa band indicates RhGB01 spike is full length as seen with SARS-CoV-2 D164 control. Comparably, the chimera construct presents a band at 130kDa band, suggesting cleavage of the S2 has occurred. All spike constructs are tagged at the S2 CTD and therefore the S1 subunit is not detectable after cleavage (Figure 4-14).

In the HEKT93 producer cells, the 230kDa band in the chimera confirms full length spike is expressed, supporting total cleavage of the chimera and confirms that the 130kDa band is not because of fragmented peptide during spike expression (Figure 4-14). A receptor screen of 22 ACE2s, human Aminopeptidase N (hAPN) (SARS tropism) and human Dipeptidyl Peptidase 4 (hDPP4) (MERS tropism) demonstrates no binding of RhGB01 spike compared to SARS-CoV-2 spike positive controls, suggesting RhGB01 tropism is not facilitate by an ACE2 included in the assay.

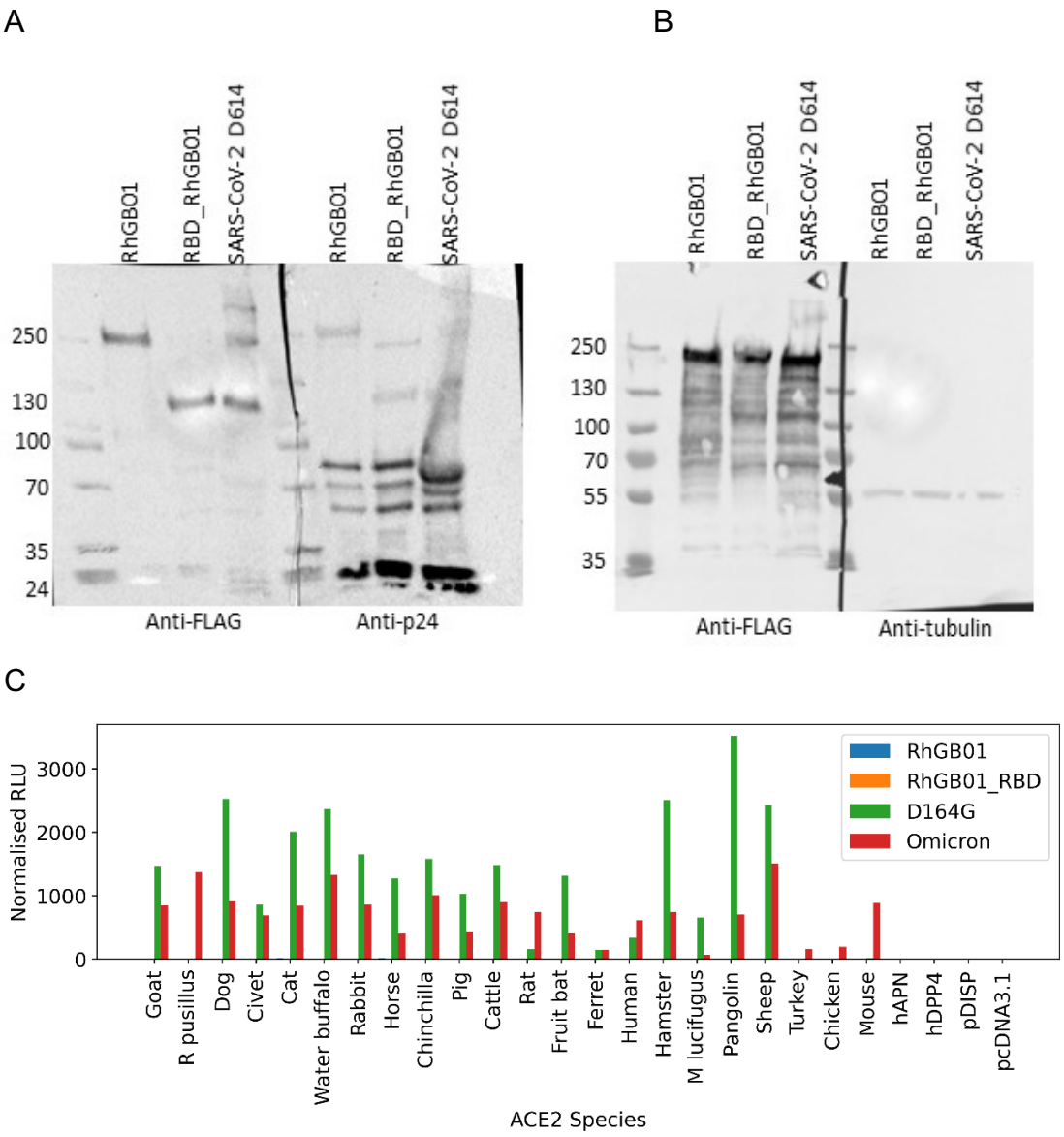


Figure 4-14. Mammalian ACE2 binding assay.

Western blots demonstrating RhGB01 spike protein expression in the pseudoparticle lysate (A) and producer cell (HEK293) lysate (B). Spike proteins are FLAG tagged at the S2 domain C terminal region and detected by an anti-FLAG antibody. For the receptor screen, ACE2 binding is measured by luciferase signal and expressed as relative light units, with values normalised by pCDNA3.1+ signal (C). The control receptors used are hAPN and hDPP4.

4.3 Discussion

4.3.1 RhGB01 is the first *Sarbecovirus* to be characterised from lesser horseshoe in Europe and the first in the UK

Here we discovered a novel *Sarbecovirus* (RhGB01) from faecal droppings collected at a lesser horseshoe roost in Cinderford, UK. This was the first *Sarbecovirus* genome to be fully recovered from horseshoe bats in the UK and was identified after sampling just 53 lesser horseshoe bats at 1 roost.

Outside of the UK, sarbecoviruses have been identified in *Rhinolophus* bats especially in China and Hong Kong after the initial spillover of SARS-CoV into the human population. Pre-2020, other European countries had reported the presence of coronaviruses by PCR and partial gene sequence analyses (Alkhovsky *et al*., 2022; Annan *et al*., 2013; Ar Gouilh *et al*., 2018; August *et al*., 2012; Balboni, Gallina, *et al*., 2011; Balboni, Palladini, *et al*., 2011; Brnić *et al*., 2023; Cholleti *et al*., 2022; Crook *et al*., 2021; De Benedictis *et al*., 2014; Drexler *et al*., 2010, 2011; Falcón *et al*., 2011; Fischer *et al*., 2016; Gloza-Rausch *et al*., 2008, p.; Goffard *et al*., 2015; Hardmeier *et al*., 2021; Hemnani *et al*., 2023; Kemenesi *et al*., 2014; Kivistö *et al*., 2020; Lazov *et al*., 2018; Lecis *et al*., 2019; Lelli *et al*., 2013; Lwande *et al*., 2022; Monchatre-Leroy *et al*., 2017; Moraga-Fernández *et al*., n.d.; Orłowska *et al*., 2022; Pauly *et al*., 2017; Popov *et al*., 2023; Reusken *et al*., 2010; Rihtaric *et al*., 2010; Rizzo *et al*., 2017; Speranskaya *et al*., 2023; C. C. S. Tan *et al*., 2023; Urushadze *et al*., 2021).

RhGB01 was the second *Sarbecovirus* from horseshoe bats in Europe to be fully sequenced, and the first from a lesser horseshoe bat. The other *Sarbecovirus* with a fully resolved consensus genome was detected in *Rhinolophus blasii*, BM48-31/BGR/2008. The detection of RhGB01 extends the geographic and

species range of SARSr-CoVs and suggest that sarbecoviruses are likely to be present throughout the range of the *Rhinolophus* bats, which are distributed throughout Africa, Eurasia, Oceania and Australia (Chornelia & Hughes, 2022).

Three faecal samples demonstrated >100 reads classified at the family taxonomic level to *Coronaviridae*. However, mapping of the raw reads to representative *Sarbecovirus* genomes highlighted these reads specifically mapping to small regions of the genome and provide no resolution across the genome. This indicates that the classifications by Kraken2 were not indicative of true positive samples. This mapping pattern demonstrated for SARS-CoV, and SARS-CoV-2 are caused by homopolymer regions. Mapping the raw reads from sample 46 demonstrated read distribution across the SARS-CoV, SARS-CoV-2 and BM48-31 genomes. This indicated the presence of a *Sarbecovirus* genome within this sample.

4.3.1.1 A de novo hybrid assembly was used to generate a genome assembly with nine predicted open reading frames

The RhGB01 genome was assembled by a de novo hybrid assembly, using both short and long read genomic data. The addition of long reads provides the scaffolding around which the shorter reads are assembled to complete the genome assembly. Mapping the raw reads to the genome assembly demonstrated the highest read depth at the 3' end of the genome where the nucleocapsid ORF sits, with >1000 reads mapping in this region. This suggests transcripts of the nucleocapsid are more abundant within the faecal samples potentially because of its importance for viral particle assembly and replication.

Genome annotation of the assembly predicted the presence of 9 open reading frames. Interestingly, RhGB01 encodes a truncated ORF7b with 20 nucleotides absent at the 5' region of the ORF. The SARS-CoV-2 ORF7b is predicted to cause dysregulation of the immune system by inducing the expression of TNF-α, IFN-β and IL-6 and activating TNF- α induced apoptosis (R. Yang *et al*., 2021).

The ORF8 is also absent from the RhGB01 genome. In SARS-CoV-2, ORF8 has been characterised as a secreted protein with the largest observed interactome of all the SARS-CoV-2 accessory proteins, interacting with host factors functioning in ER protein processing, extracellular matrix organisation, and post translational modifications e.g. glycosylation and glycosaminoglycan synthesis (Gordon *et al*., 2020). The presence of ORF8 is associated with alteration of the cellular environment for the enhancement of viral replication. Thus, the absence of this accessory protein suggests in the event of evolutionary modification required for hACE2 or entry via a different receptor, it is not primed for efficient human infection by ORF8-dependent modification of the cellular environment.

Other accessory factors which influence the severity of infection are observed in RhGB01. ORF3a is a viral protein which functions as a viroporin interfering with ion channel activity in host cell and endosomal membranes tot promotes viral entry and release by the lysosomal exocytosis pathway. ORF3a is also associated with ORF3a-induced apoptosis and necrosis, involvement in viral particle formation by association with M and E and influencing viral replication dynamics (J. Zhang *et al*., 2022). ORF6 is predicted to induce IFN activation but the function of ORF6 in immune modulation is not as well characterised (Li *et al*., 2023). In summary, the accessory proteins present in RhGB01 suggest

suboptimal immune regulation upon potential human infection if limitations regarding receptor usage are circumvented.

4.3.1.2 The RhGB01 spike demonstrates key differences compared to SARS-CoV-2 and the polybasic furin cleavage site is replaced with a bibasic motif

Across the entire spike protein RhB01 demonstrates high levels of structural similarity to SARS-CoV-2 spike, particularly within the S2 domain. However, variation was observed in SARS-CoV-2 spike regions associated with increased efficiency of infection in humans.

This variation is also observed in SARS-CoV with only one basic arginine residue at the S1/S2 cleavage site, representing a minimally basic site cleaved by host trypsin. The polybasic furin cleavage site observed at the S1/S2 intersection in SARS-CoV-2 is absent, but a potential precursor FCS is present (R-A-K-Q). This bibasic motif is 1 amino acid from the canonical polybasic cleavage site (R-X-K/R-R) and the potential for host furin to cleave at this site (Zabiegala *et al*., 2023). RhGB01 encodes more basic residues at the S1/S2 site than SARS-CoV but not the polybasic site observed in SARS-CoV-2. In models of the SARS-CoV-2 spike proteins publicly available, the furin cleavage (R-R-X-R) motif is unresolved which prevents molecular distance measurements between two motifs on the respective loops. However, insertions up and downstream of the furin cleavage motif in SARS-CoV-2 increase the protrusion of the cleavage motif outside of the S2 domain structure and allow binding of host furin to the site (B. A. Johnson *et al*., 2020). Thus, future mutation of Q into K or R wouldn't necessarily result in furin cleavage without extension of the loop. Additionally, investigations into the minor variant populations within the raw data at positions 2007-2016 (RhGB01

spike nucleotide coordinates) suggests no variation at these nucleotides. This is limited by read depth X25 at these sites and deeper sequencing could elucidate more variation at these sites.

4.3.1.3 hACE2 contact residues are absent in RhGB01

The receptor binding domain of the spike protein contains a receptor binding interface which contacts the ACE2 receptor. Specific residues have been identified as important for binding to hACE2 for SARS-CoV (F. Li *et al*., 2005b) and SARS-CoV-2 (Lan *et al*., 2020). RhGB01 has significant variation within these contact residues which we hypothesise prevent hACE2 binding.

4.3.1.4 An ACE2 binding assay suggests a narrow host range for RhGB01 ACE2 binding

To investigate the ACE2 binding profile of RhGB01, lentiviral-based pseudoparticles were generated expressing full length RhGB01 and a chimeric SARS-CoV spike backbone and RhGB01 RBD. The ACE2 receptors screened in this assay were representative of different mammalian ACE2 receptors, but bat ACE2s were not well represented with a notable absence of the *Rhinolophus hipposideros* ACE2. The only horseshoe bat ACE2 receptor present in the assay is that of *Rhinolophus pusillus* and the host range of this horseshoe bat is limited to Asia. The identification of a 250kDa protein in the RhGB01 WT condition suggests an absence of cleavage, potentially preventing conformational change required for RBD-ACE2 interaction. However, in the SARS-CoV chimera cleavage was observed by detection of 250kDa and 130kDa proteins. In both

conditions, no ACE2 binding is indicated by fluorescence which indicates a potentially narrow ACE2 binding range for the RhGB01 RBD. Future work on the RhGB01 spike should include a more extensive bat ACE2 screening as recently sampled sarbecoviruses demonstrate binding to ACE2 receptors not included in this screening.

4.3.1.5 Khosta-1 and -2 were concurrently characterised in Russia with Khosta-2 demonstrating hACE2 binding in vitro

With increased sampling since the COVID-19 pandemic (post-2020) a monophyletic European specific *Rhinolophus* associated clade has become apparent for most structural and accessory proteins. At this time, viruses within this clade had either undefined or predicted ACE2 host ranges with experimental efforts focused on SL-CoVs and RBD-inferred clade 1 viruses (Roelle *et al*., 2021; Seifert *et al*., 2022; Starr *et al*., 2022; Wells *et al*., 2021). This area has since developed rapidly. One publication tentatively named the clade containing European and central African sarbecoviruses as “lineage 4” (Wells *et al*., 2021). Khosta-2 and Khosta-1 are clustered within the “lineage 4” clade and were sampled from Russian lesser and greater horseshoe bats, respectively. Khosta-2, the most closely related non-UK European *Sarbecovirus* to RhGB01 has been shown to use hACE2 for entry in vitro when the RBD is expressed in a chimera with the SARS-CoV spike. Khosta-1, also closely related to RhGB01, demonstrates more diverse cellular entry than Khosta-2 for Huh-7 cells when the RBD is expressed in a chimera on pseudoparticles. Also contained within “Lineage 4” are *Rhinolophus* associated sarbecoviruses from Uganda and Rwanda, which demonstrate no cellular entry using an RBD-SARS-CoV chimera.

The *Sarbecovirus* from Bulgaria (BM48-31) full length spike also demonstrated no evidence of cellular entry (Seifert *et al*., 2022). Comparable results were seen in Starr *et al* (2022) who investigated the binding capacity of a set of recently sampled sarbecoviruses, including RhGB01. RhGB01 was identified to have reduced ACE2 binding compared to Khosta-1, Khosta-2, and selected African (BtKY72, PRD-0038, RDF-2370) and Asian sarbecoviruses (RsYN04). These African and Asian sarbecoviruses demonstrate broad usage of receptors from *R sinicus* and *R affinis*. As with the previous study, Khosta-2 was associated with hACE2 binding (Starr *et al*., 2022). Interestingly, RhGB01 demonstrated a non-significant peak of fluorescence for the hACE2 but was not significant enough to investigate further but suggests more research into hACE2-RhGB01 binding is required.

4.4 Conclusion

In this study, we hypothesised UK *Rhinolophus hipposideros* populations would harbour sarbecoviruses. From one faecal sample, RhGB01 was characterised, highlighting the presence of sarbecoviruses with comparable genomic organisation to other *Rhinolophus* associated sarbecoviruses detected in Asia and Africa. The determination of the ACE2 binding capability of novel sarbecoviruses has become the gold standard for prediction spillover risk to the human population. Experimental determination of ACE2 binding utilising a pseudoviral based assay determined a potentially narrow ACE2 host range for RhGB01, with the limitation of limited bat ACE2 receptors being utilised in the assay. In silico structural analysis of the RhGB01 spike highlighted the absence of a polybasic FCS observed in SARS-CoV-2, but potential FCS precursor in the

form of a bibasic motif (R-X-K-X). Additionally, the FCS loop in SARS-CoV-2 is present on a much longer loop which provides space for host furin binding, and cleavage. RhGB01 extends the western extreme for *Sarbecovirus* distribution in *Rhinolophus* bats and highlights the need for improvements in surveillance in the UK.

5 Longitudinal sampling of a lesser horseshoe roost in the UK

5.1 Results

5.1.1 Longitudinal qRT-PCR screening to assess *Sarbecovirus* prevalence in a single roost in Gloucestershire over a 4-year period

Year	Location	Month	Positive (total)	Positive (%)	mNGS	Ct range
2020	Cinderford	June	7 (53)	13.2	53	29.1-32.4
2021	Cinderford	June	40 (64)	62.5	16	27.2-35.2
2021	Cinderford	July	8 (18)	44	1	28.6-33.87
2022	Cinderford	October	18 (96)	18.7	14	23.3-32.3
2023	Cinderford	August	7 (54)	12.9	16	13.7-38.3

Table 5-1. qRT-PCR results across the years for the Cinderford roost.

After identification of RhGB01 at the roost in Cinderford, extended sampling was conducted across 2021, 2022 and 2023. The faecal samples from 2020 were also retrospectively screened for detection of positives with viral loads outside of mNGS sensitivity. qRT-PCR positivity suggest prevalence is higher in 2020 and

2021 compared to 2022 and 2023 and that prevalence is higher earlier in the year (June, July) compared to later months (August, October) (Table 5-1). The positive sample from which RhGB01 was characterised was found to be Ct 32.4, which is likely an underestimate resulting from repeated use and freeze-thawing prior to PCR screening.

The Cinderford roost was sampled twice in 2022. In September faecal samples were bulk collected (10 pellets per tube) compared to the October sampling which was individual dropping collection (1 pellet per tube). From the September collection, 169/176 (96%) faecal samples were positive with Ct values ranging from 21 – 37, which is likely an artificially inflated positivity rate.

Across the 4-year period 274 faecal samples were found to be *Sarbecovirus* positive by RdRp qRT-PCR WHO assay. All faecal samples were negative using a Wuhan Hu 1 SARS-CoV-2 specific probe. All negative controls were negative by both probes and SARS-CoV-2 positive controls positive by Wuhan Hu 1 SARS-CoV-2 specific probe.

5.1.2 Extended screening for sarbecoviruses outside of the Cinderford roost

In addition to the longitudinal screening of the Cinderford roost, 40 faecal samples were collected from a roost in Plymouth in 2022. 8/40 (20%) faecal samples were identified as positive by qRT-PCR with Ct values ranging from 26.8-32.9.

In 2021, 77 faecal samples were collected from a selection of locations in Somerset and Devon from bats of different genus (Table 5-2). No *Sarbecovirus* positive samples were reported. In 2023, 49 faecal samples were collected from

a *Rhinolophus ferrumequinum* roost in Dorset. No positives were identified by qRT-PCR.

Species	N
Myotis daubentonii	18
Pipistrellus pygmaeus	13
Myotis nattereri	12
Myotis mystacinus	6
Barbastella barbastellus	5
Myotis bechsteinii	5
Nyctalus noctula	5
Plecotus auritus	4
Pipistrellus pipistrellus	3
Nyctalus leisleri	2
Myotis alcathoe	2
Eptesicus serotinus	2
Rhinolophus ferrumequinum	1
Rhinolophus hipposideros	1

Table 5-2. Faecal samples collected from across Somerset and Devon

5.1.3 Faecal samples from Cinderford with <30 Ct value were selected for metagenomic NGS

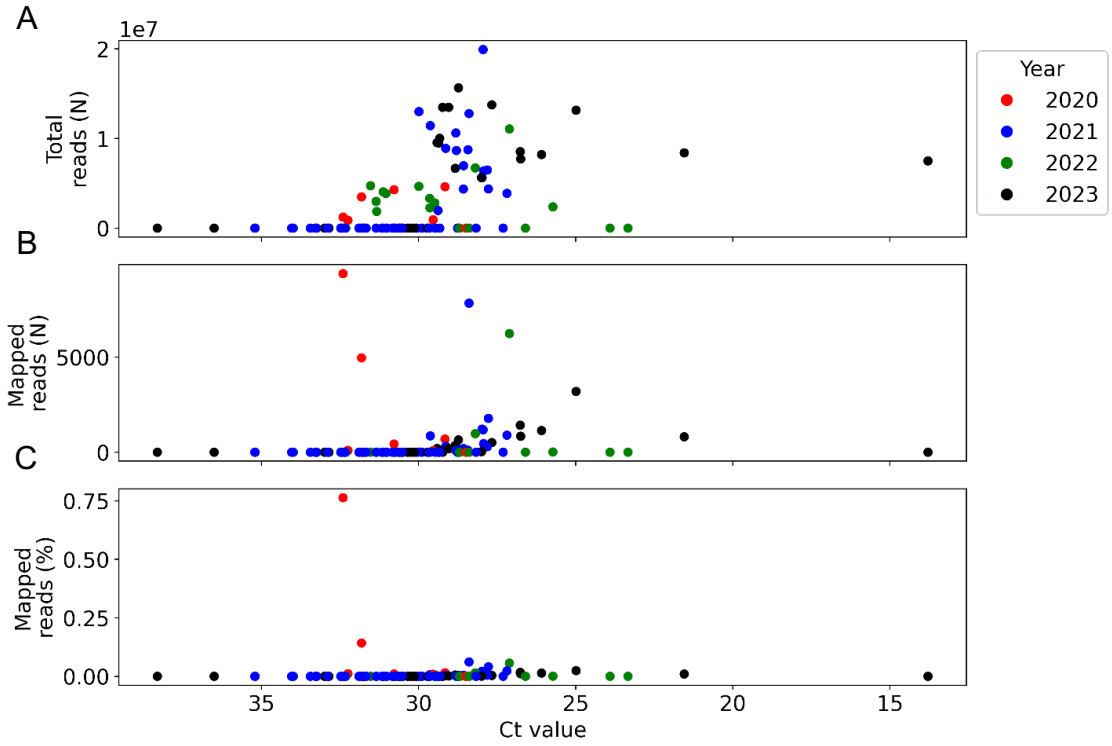


Figure 5-1. Sequencing statistics for mNGS.

Sequencing and mapping statistics for the faecal samples selected for mNGS. The total read number (A), mapped reads to RhGB01 (B) and percentage mapped reads to RhGB01 (C) according to Ct value recorded from the qRT-PCR screen. Points are coloured by year of collection. The X-axis scale is reversed.

All 53 samples from 2020 were metagenomically sequenced as an initial viral screening methodology prior to *Sarbecovirus* discovery. The characterisation of RhGB01 altered the criteria for selecting samples and the qRT-PCR RdRp pan-Sarbecovirus assay was selected for screening. In total 100 samples underwent mNGS for 2020-2023 from the Cinderford roost, comprising of 36% of all positives. Additionally, all 40 samples from the Plymouth roost were

metagenomically sequenced. An increase in total reads and mapped reads was observed as Ct value decreases (Figure 5-1).

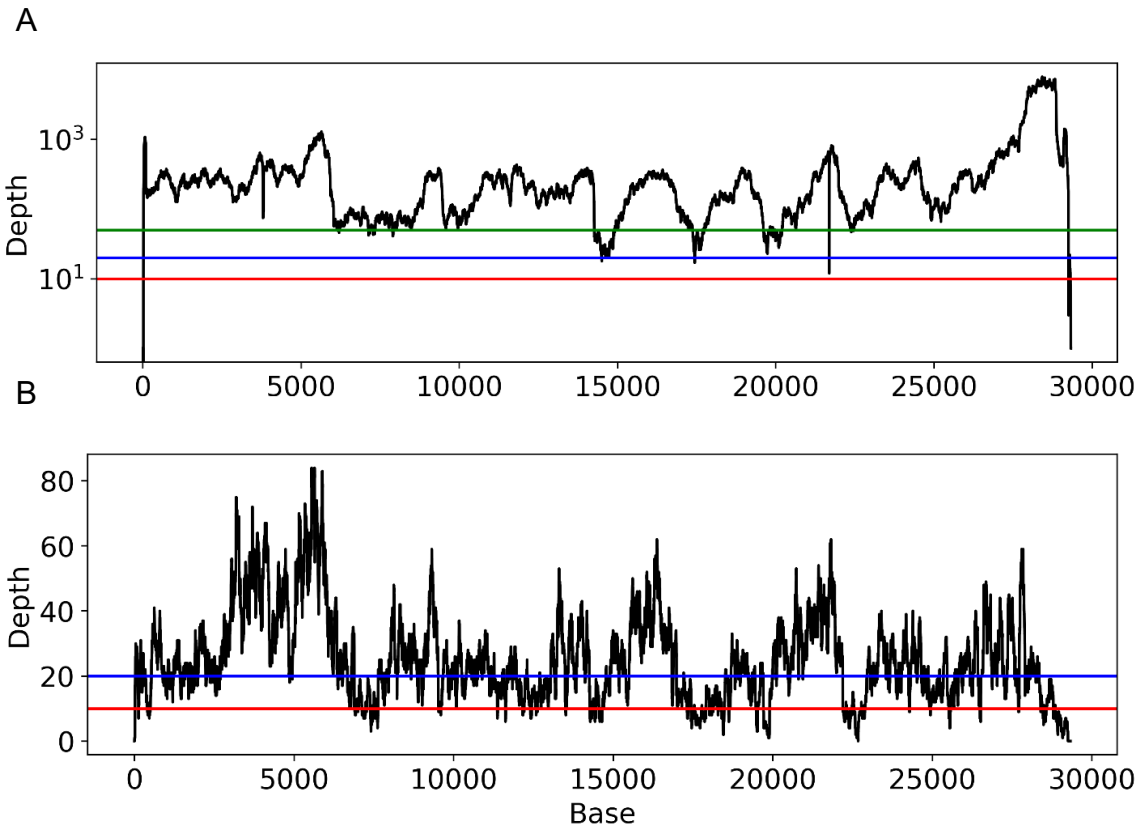


Figure 5-2. CI63 and IM328 read depths.

Raw reads from CI63_2022 (A) and IM328_2021 (B) when mapped to the RhGB01 genome. Different thresholds of depths of 10, 20 and 50 are demonstrated by red, blue, and green horizontal lines. The Y-axis in panel A is on a log10 scale.

5.1.4 A complete and near complete *Sarbecovirus* genome were recovered from CI63_2022 and IM328_2021

All samples metagenomically sequenced were mapped to RhGB01 to investigate the distribution of raw reads across the *Sarbecovirus* genome. Two samples showed significant genome sequence recovery, CI63_2022 collected in 2022 and IM328_2021, collected in 2021. 139,870 (1.45%) reads from CI63 mapped to the

RhGB01 genome compared to 7842 (0.06%) reads from IM328. Consensus sequences were assembled requiring ≥20 read depth and 80% conservation to call a consensus base. 89.16% of the IM328 coding sequence was recovered compared to 100% for CI63. Raw reads from CI63 demonstrate >50x depth across the RhGB01 genome, strongly supporting the presence of a similar *Sarbecovirus* within this sample (Figure 5-2).

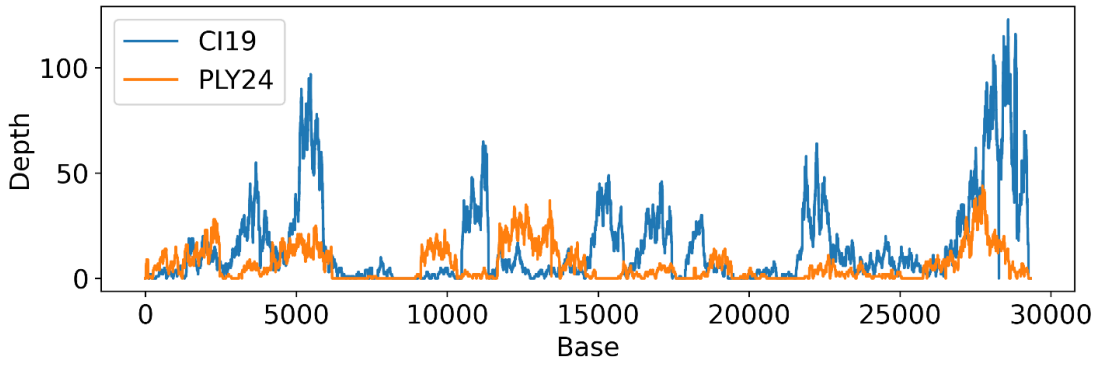


Figure 5-3. CI19 and PLY24 read depths.

Reads from positive faecal samples from Cinderford and Plymouth mapped to RhGB01.

In addition to CI63 and IM328, samples from Cinderford (CI19_2020) and Plymouth (PLY24, 2022) generated partially complete or complete nucleotide sequence for selected genes. CI19 generated full coding sequences for the ORF6, ORF7a and N, as well as partial sequences for M (77.6%) and the RBD (63.98%). PLY24 provides a full coding sequence for ORF6 and partial ORFs for ORF7a (96.7%) and N (68.7% N) (Figure 5-3).

A

Target	Accession	ID (%)	Alignment length	Mis matches	Start	End
BM48-31	NC_014470	80.4	23670	4326	1	23498
SARS-CoV Tor2	NC_004718	81.0	17253	3144	6279	23464
SARS-CoV-2	NC_045512	79.9	17368	3271	6281	23545

B

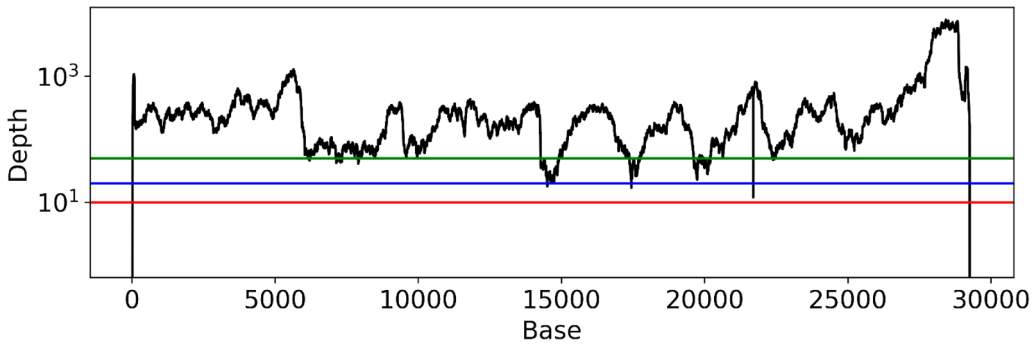


Figure 5-4. Blastn and raw reads mapped to the assembly.

Blastn hits for the CI63 de novo assembled contigs filtered for matches with alignment length >10 kb. Target refers to blastn match with accession number. Start and end refers to coordinates in the query sequence which are used to align to target (A). Reads for sample CI63_2020 mapped to the reference guided consensus assembly (B).

All samples sequenced by mNGS underwent de novo assembly as the gold standard for the detection of viral contigs without a reference genome bias. A de novo assembly for CI63 generated a 27.7 kb contig with an average depth of 96.7X and homology to BM48-31, SARS-CoV and SARS-CoV-2 as assessed by blastn (Figure 5-4). Genome annotation of the contig identified all ORFs, except the nucleocapsid. An additional contig 1131 bases long with >2000x depth was generated which translates to the first 372 amino acids of the CI63 nucleocapsid protein. 45 amino acids are missing from the end of the protein sequence.

To circumvent this a consensus sequence was generated using the RhGB01 genome as a reference. Reads mapped to the consensus sequence demonstrate at least 20X depth across the consensus sequence. 140,054 (1.45%) reads from CI63 map to the consensus genome (Figure 5-4). The consensus sequence was predicted to have nine ORFs in the correct orientation for a *Sarbecovirus* genome.

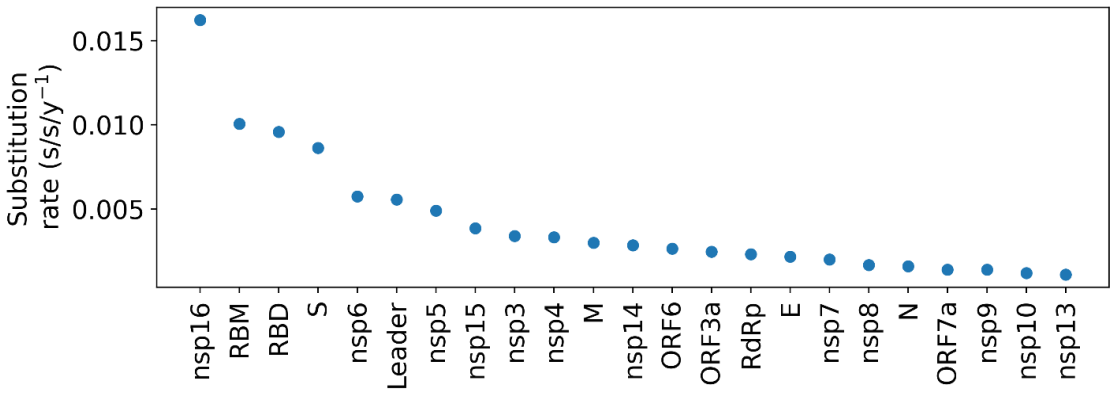


Figure 5-5. Substitution rates for the different open reading frames.

Substitution rates (s/s/y^{-1}) for the different open reading frames between RhGB01 and CI63 as an initial investigation to rate heterogeneity across the genome within the Cinderford roost.

5.1.5 RhGB01 differs from CI63 by 277 single nucleotide polymorphisms

No insertions or deletions are observed between RhGB01 and CI63 and the same number of ORFs are predicted. Across the coding sequence there are 277 nucleotide SNPs (0.96%) and 157 amino acids (1.63%) mutations. This suggests a nucleotide evolutionary rate of 0.00472 s/s/y^{-1} at the Cinderford roost between RhGB01 and CI63. nsp16 exhibits the highest nucleotide substitution rate between RhGB01 and CI63 (0.0162 s/s/y^{-1}). The spike, RBD and RBM also

exhibited elevated nucleotide substitution rates (0.0086, 0.009, 0.01 $s/s/y^{-1}$) (Figure 5-5).

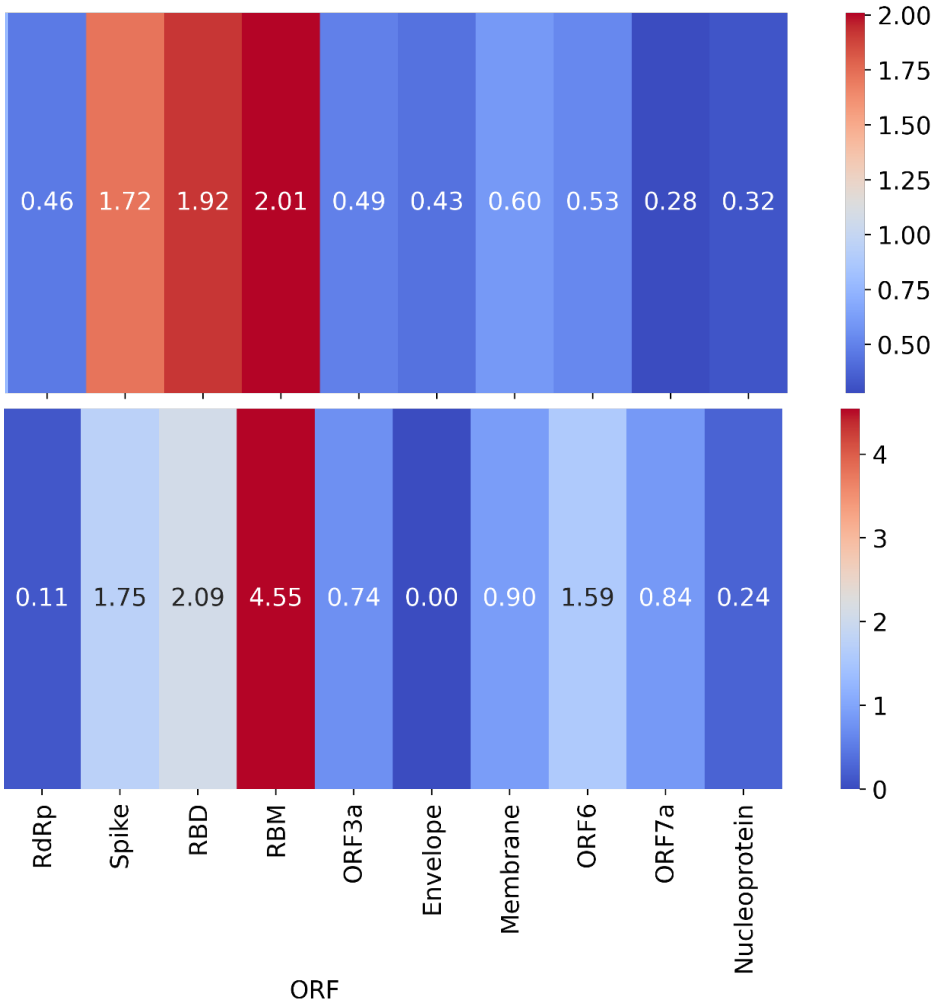


Figure 5-6. Nucleotide and amino acid differences.

The percentage differences between the nucleotide (A) and amino acid (B) sequences of RhGB01 and CI63. Red indicates a high percentage difference observed within that ORF. The receptor binding domain, receptor binding motif and RNA dependent RNA polymerase are included because of their functional importance to sarbecoviruses.

A disproportionate number of the amino acid differences are contained with the RBM with 3 amino acid alterations (4.55%) (Figure 5-6). Outside of the spike

glycoprotein, ORF6 demonstrates the highest percentage of amino acid mutations (1.59%).

5.1.6 Comparing the predicted structures of the structural proteins in RhGB01 and CI63

Structural proteins were predicted by homology modelling and compared using RMSD calculations (Appendix 18). Minimal structural disparities were observed in the spike and nucleocapsid proteins (RMSD=0.77, 0.003 Å) and the membrane and envelope proteins were found to be structurally identical.

22 amino acid differences are observed in the spike protein between RhGB01 and CI63 (Figure 5-7). 21 are within the S1 domain and one in the S2 domain (Appendix 19). Four amino acid differences are contained within the RBD. Using an alignment with Wuhan Hu 1 (NC_045512) the receptor binding loops within the ACE2 binding interface were identified. Binding loops 1 and 2 for RhGB01 and CI63 were identified at residues 462-476 and 483-494. The receptor binding loop 3 in SARS-CoV-2 possesses an insertion at sites 474-450, which contains the loop 3 residues 449-453. Therefore loop 3 in RhGB01/CI63 is hypothesised to be residues 436-441. Three amino acid differences are found within the RBM region (469, 471, and 491). The amino acid difference in the S2 domain is within the HR1 domain, Q901L (Figure 5-7).

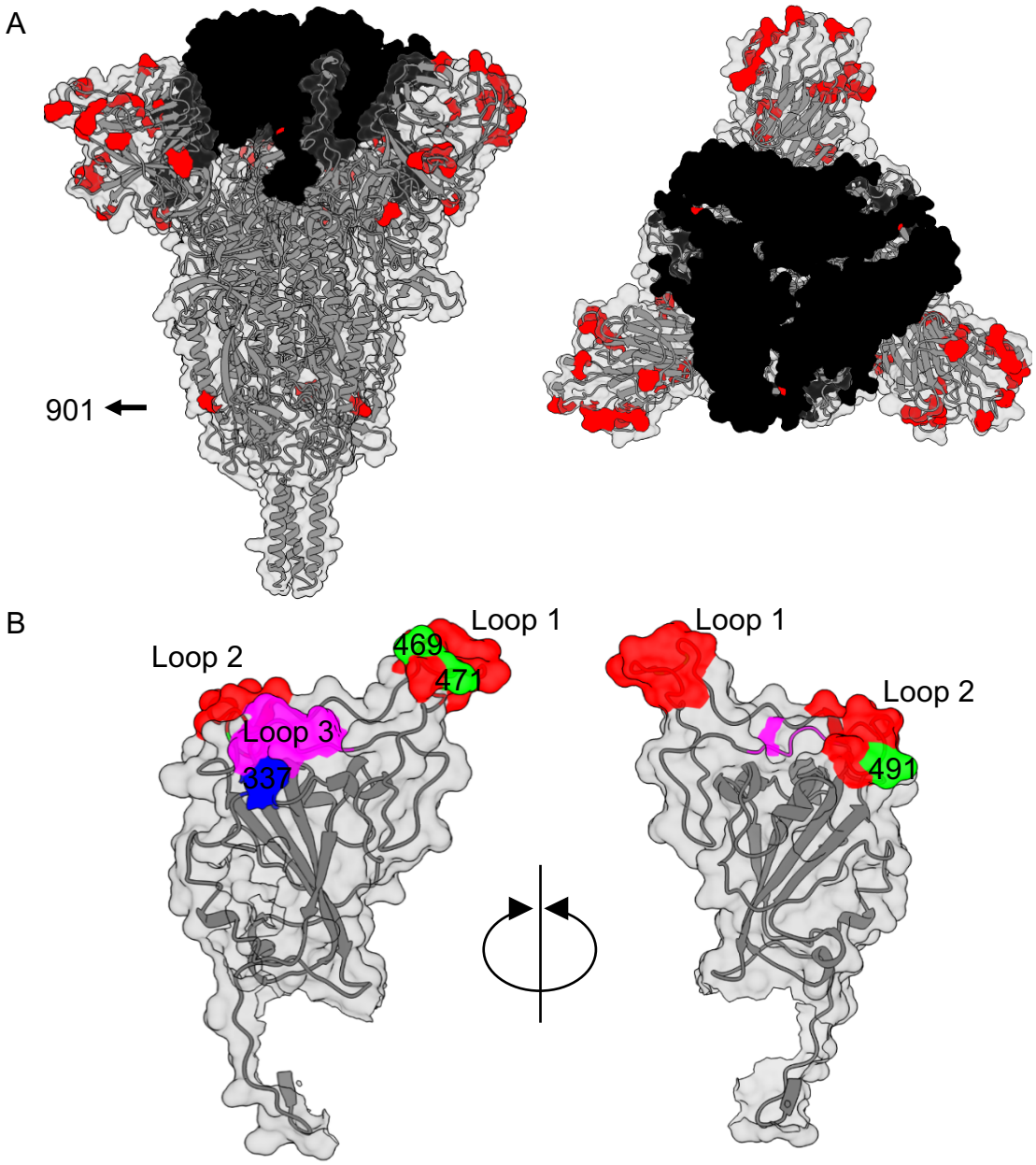


Figure 5-7. Spike glycoprotein comparison between RhGB01 and CI63.

Homology models for the RhGB01 spike protein trimer with amino acids differences outside the receptor binding domain highlighted in red. The receptor binding domain is shown in black (A). The receptor binding domain for RhGB01 with receptor binding loops 1 and 2 highlighted in red, and loop 3 in pink. Residues 469, 471 are highlighted in loop 1 (green), residue 491 is highlighted in loop 2 (green) and residue 337 is highlighted near loop 3 (blue) (B).

The RBD exhibits greater structural variation compared to the entire spike monomer (RMSD=1.49 Å vs 0.77 Å) (Figure 5-8). RBDs were modelled

separately from the spike protein to facilitate the selection of the most representative RBD as a template structure. In comparing RhGB01 and CI63, five local regions within the RBD demonstrate RMSD >2 Å. Among these regions, Regions 4 and 5, which lie outside the receptor binding motif, exhibit RMSD values of 4.87 Å and 3.09 Å, respectively. Conversely, regions 1, 2, and 3 (named as such because of proximity to binding loops 1, 2 and 3) situated within the RBM display RMSD distances of 2.3 Å, 3.9 Å, and 5.6 Å, respectively (Figure 5-8).

5.1.7 Investigation into conservation and homology modelling of non-structural proteins

As with the structural proteins, high levels of similarity are observed across the non-structural proteins between RhGB01 and CI63 (Appendix 21). Nsp16 demonstrates the highest nucleotide variation with 29 (3.24%) SNPs. However, only 3 amino acids are altered (1.01%). Nsp13 has the lowest number of SNPs with 4 differences (0.22%) and 1 (0.17%) amino acid change. The highest amino acid variation was observed in nsp2 with 15 (2.35%) differences. However, the nucleotide sequence for nsp2 contains 28 ambiguous bases: 16 Y (C/T), 7 R (A/G), 3 K (G/T), 1 S (C/G) and 1 W (A/T). All possible bases suggested by the IUPAC code are observed in corresponding nucleotide locations in RhGB01. For investigation purposes the RhGB01 nucleotide was inserted at ambiguous bases in the CI63 nsp2 nucleotide sequence, leaving only SNPs which are unambiguous. When ambiguous bases are removed nsp2 has 27 nucleotide SNPs (1.41%) and 9 amino acid differences (1.41%) (Figure 5-9).

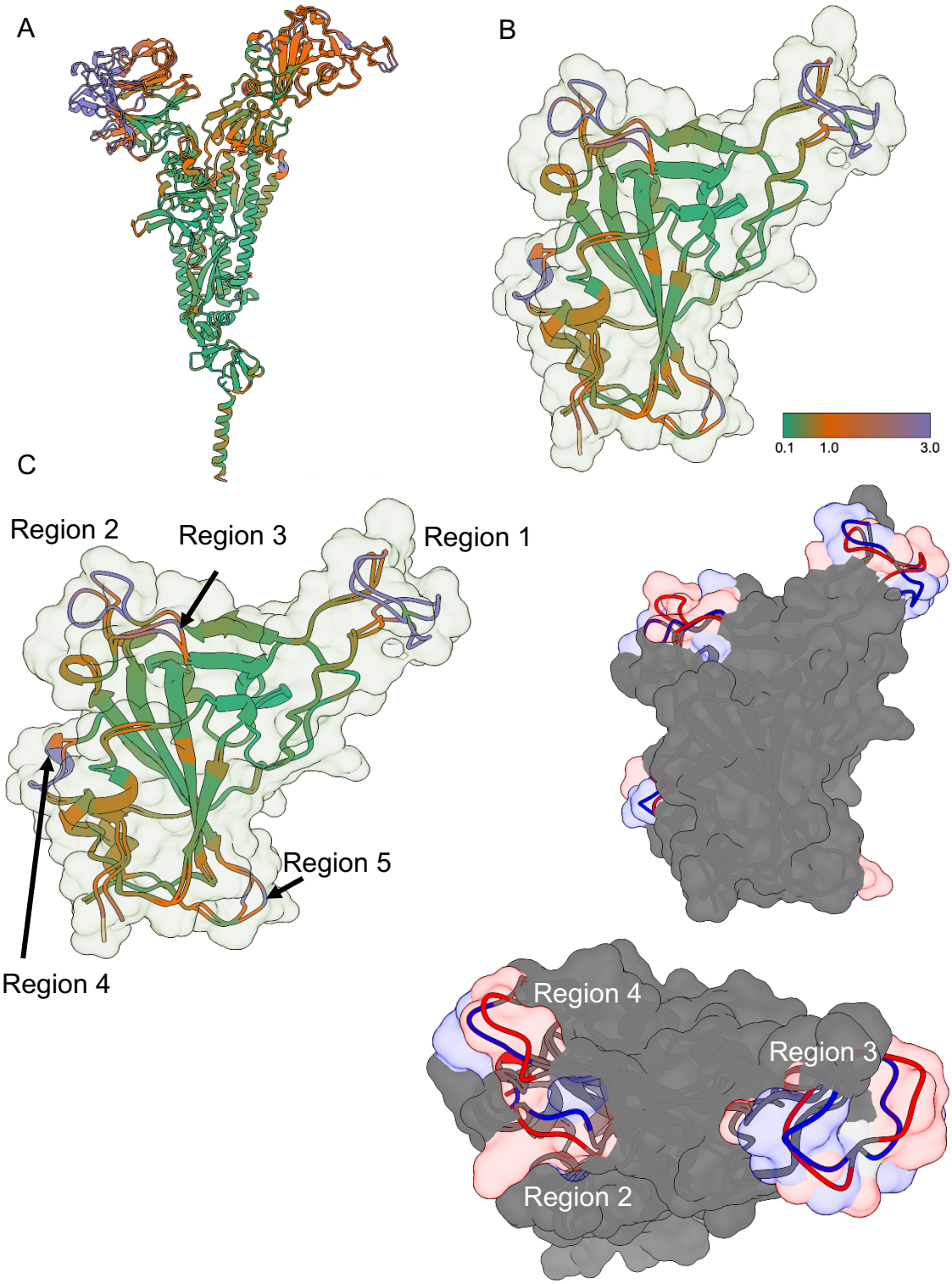


Figure 5-8. RBD differences in RhGB01 and CI63.

The predicted structures of the RhGB01 and CI63 spike glycoproteins structurally aligned and coloured by RMSD distance. The scale used ranged from 0.1-3 Å (A). Structurally aligned RBD when modelled separately (B). RhGB01 and CI63 RBD structurally aligned with regions of structural difference (>2 Å) highlighted. RhGB01 structure is coloured in red, and CI63 in blue. Structural difference is represented by RMSD score with purple representing more structurally divergent regions (C).

To investigate structural conservation between the nsps for RhGB01 and CI63, homology modelling was employed. For nsp2-6 no fully resolved structures suitable for homology modelling are available. The amino acid at site 46 in nsp15 is represented as a “X” because of an unresolved base at nucleotide 19658. This “X” was replaced with “K” as seen in RhGB01 for modelling purposes. The template for nsp1 was selected with GMQE 0.66 as this was the best scoring model.

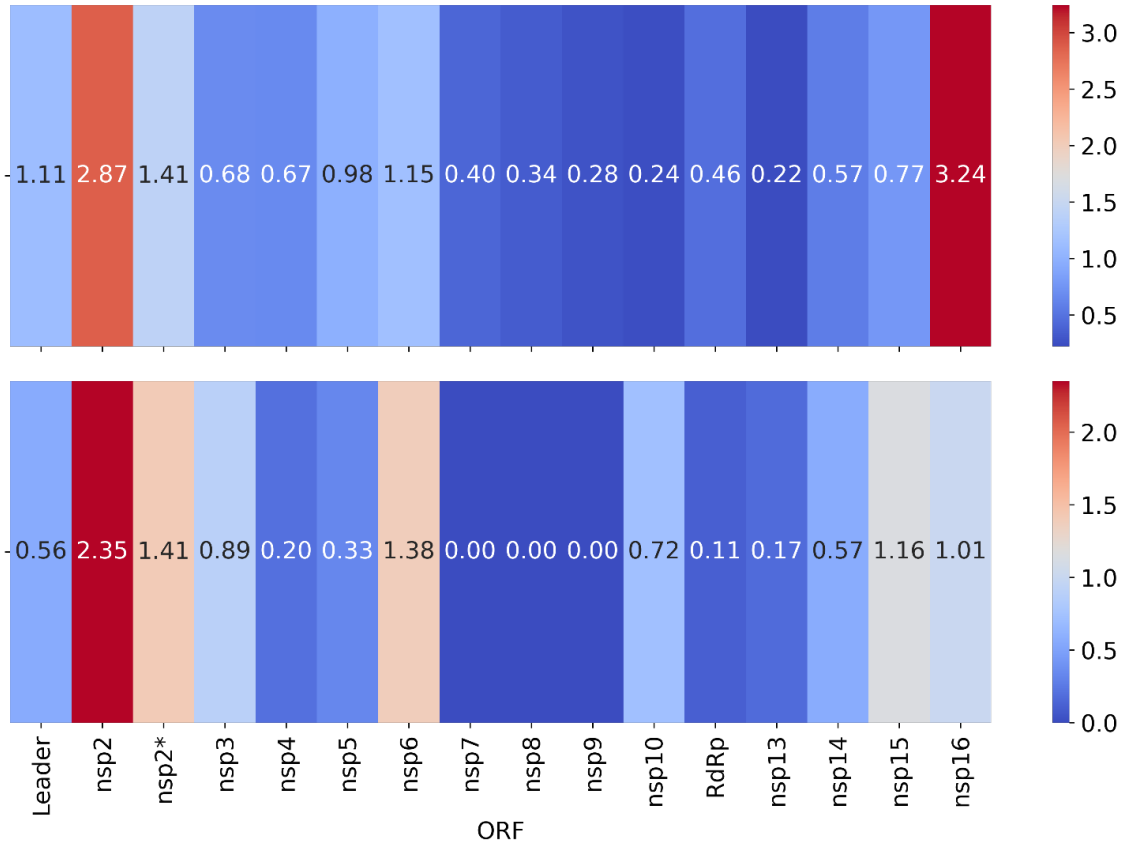


Figure 5-9. RhGB01 and CI63 nonstructural proteins.

Percentage differences between nucleotide (A) and amino acid (B) sequences between RhGB01 and CI63_2022 nonstructural proteins. Red indicates a high percentage difference observed within that ORF. The receptor binding domain, receptor binding motif and RNA dependent RNA polymerase are included because of functional importance to Sarbecoviruses. Nsp2* refers to the nsp2 sequence with ambiguous bases replaced with the RhGB01 nucleotide if the ambiguity code is inclusive of the nucleotide in RhGB01.

All other proteins were modelled using templates with GMQE >0.7. All models are of moderate quality, except nsp1 which has scores of 63.8 and 74.5% suggesting a poorer model quality (Appendix 20). Using the predicted structures, the conservation of structure between RhGB01 and CI63 can be evaluated. All non-structural proteins are structurally conserved when comparing RhGB01 and CI63 with RMSD <0.023 Å for all structures (Appendix 21) suggesting highly conserved structures.

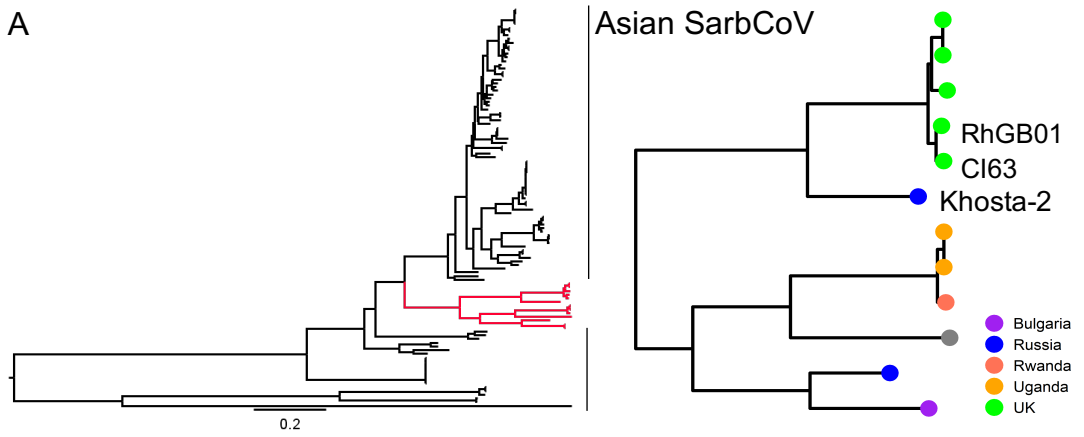


Figure 5-10. A phylogeny for *Rhinolophus* associated sarbecoviruses.

A maximum likelihood phylogeny for the coding sequence of all open reading frames of sarbecoviruses with full genomes available pre-2024. All sarbecoviruses from the public domain with full genomes were included in the phylogeny. The highlighted clade in the tree on the left is enlarged to highlight clustering of sarbecoviruses from different locations within the non-asian monophyletic clade.

5.1.8 RhGB01 and CI63 form a UK-specific monophyletic clade with other sarbecoviruses identified in lesser and greater horseshoes in the UK

RhGB01 and CI63 form a monophyletic group with the other UK sarbecoviruses and Khosta-2 from Russia for the phylogeny inferred using the coding sequence. The European clade is divided into two monophyletic clades, one containing

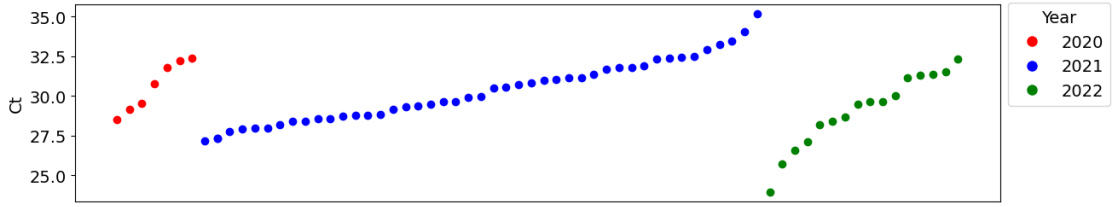
UK/Russian sarbecoviruses and the other containing sarbecoviruses from Bulgaria, Russia, and Africa. Outside of this clade, the remaining sequences in the *Sarbecovirus* phylogeny is a large range of sarbecoviruses from China, Japan, Vietnam, South Korea, and Malaysia (Figure 5-10).

5.1.9 86 samples were selected from 2020-22 for amplification by an amplicon scheme

A

Year	Month	Location	N
2020	June	Cinderford	6
2021	July	Cinderford	6
2021	June	Cinderford	40
2022	October	Cinderford	16
2022	October	Plymouth	8

B



C

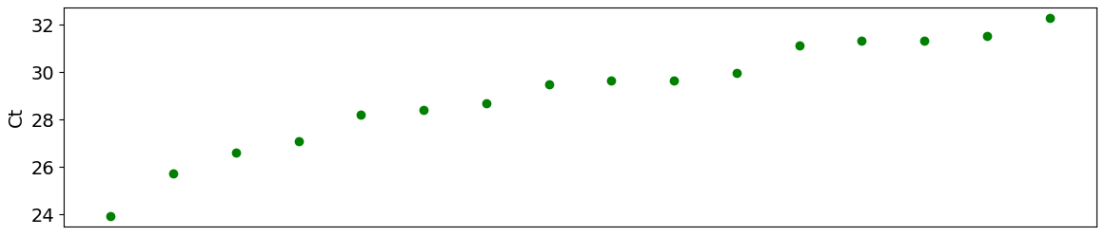


Figure 5-11. Samples for whole genome tiling approach.

A table showing the sample numbers used for amplicon scheme sequencing (A). The Ct values for samples selected for amplification by amplicon approach for the Cinderford roost (B) and the Plymouth roost (C). Points are coloured by year of sampling. On the x-axis is sample id, removed for visualisation purposes.

A whole genome tiling amplicon scheme was designed using the consensus sequence for CI63. Samples were selected based on raw material availability and Ct value.

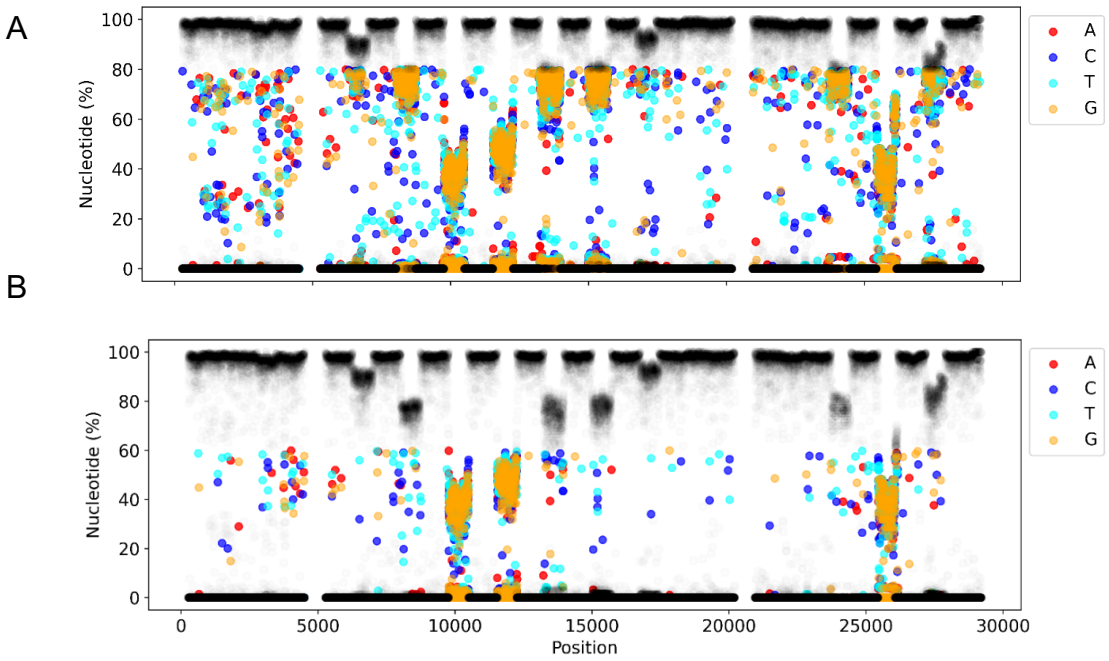


Figure 5-12. Minor variant populations from the tiling approach.

The minor variation observed at each nucleotide across the reference genome when raw reads for the amplicon scheme for sample IM303 are mapped. In the top panel, an 80% agreement is required to call a consensus base and bases with >80% agreement are coloured black. In the bottom panel, a 60% agreement is required to call a consensus base and bases with >60% observed are coloured black (B). Bases with missing points represent bases at which no raw reads are mapped to the reference genome. Sequencing was completed using ONT.

For most samples which had been metagenomically sequenced raw material was limited and amplicon scheme amplification not possible. The Ct value range used for the amplicon approach was 23.9-35.2 (Figure 5-11). A 0-9.7% increase in genome recovery per sample was observed when stringency is dropped from an 80% to 60% conservation for raw reads at each base to accommodate increased

variation using ONT sequencing. For sample IM303 collected in June 2021, an additional 2810 (9.7%) bases were recovered from reducing the stringency (Figure 5-12). This methodology was used going forward for amplicon scheme genomic data to maximise sequence recovery.

5.1.10 Differential amplicon success was observed for different open reading frames across the genome

Between 2020 and 2022, no complete genomes were obtained using an amplicon approach (Table 5-3). The nsp9 region was the best performing ORF being fully recovered in 60/86 faecal samples (69.7%). This is followed by the membrane, leader and RBM all recovered in ≥ 40 samples (48, 47, 40). The RBD was fully recovered in 14/86 samples (16.3%) and only one sample (1.2%) contained the full spike sequence (CI52_Oct_2022). The focus was placed on the 14 receptor binding domains for further analysis.

ORF	N	%
nsp9	60	69.8
Membrane	48	55.8
Leader	47	54.7
RBM	40	46.5
nsp7	35	40.7
ORF6	31	36.0
nsp2	23	26.7
nsp6	23	26.7
nsp8	21	24.4
nsp14	17	19.8
ORF7a	16	18.6
nsp4	14	16.3
RBD	14	16.3
nsp13	13	15.1
nsp10	11	12.8
Nucleoprotein	8	9.3
nsp5	6	7.0

RdRp	5	5.8
ORF3a	3	3.5
nsp15	2	2.3
Envelope	2	2.3
Spike	1	1.2
nsp3	0	0.0
nsp16	0	0.0

Table 5-3. ORFs with 100% recovery.

The percentage of samples with 100% ORF recovery using the amplicon approach.

5.1.11 14 receptor binding domains representing 6 different genotypes were retrieved from the three years

A

Site	RhGB01	CI63	Other	Sample(s) with mutation	Geno type	Binding loop
330	S	S	G	CI15_oct_2022	1	NA
337	M	N	N	RhGB01_2020	All	NA
364	I	I	T	80816_2020	6	NA
395	R	R	K	80816_2020	6	NA
432	H	H	L	80810_2020, CI15_oct_2022	1, 5	NA
437	Q	Q	R	80816_2020	6	3
458	R	R	M	80816_2020	6	NA
466	T	T	K	CI53_oct_2022	2	1
469	T	S	S	CI13_oct_2022, CI30_oct_2022, CI45_oct_2022,CI45_sept_2022, CI52_oct_2022, IM338_2021, IM341_2021	3	1
			T	80810_2020, 80816_2020, CI11_oct_2022, CI48_sept_2022, CI53_sept_2022	4, 5, 6	1
			I	CI53_oct_2022	2	1
			K	CI15_oct_2022	1	1
470	V	V	T	80810	5	1
			A	80816, CI15_oct_2022	1,6	1
			K	CI53_oct_2022	2	1
471	V	E	E	RhGB01	All	1
477	D	D	N	80816, CI53_oct_2022	2, 6	NA
491	I	T	T	RhGB01	All	2

B

Site	Genotype 1	Genotype 2	Genotype 3	Genotype 4	Genotype 5	Genotype 6
330	G	S	S	S	S	S
337	N	N	N	N	N	N
364	I	I	I	I	I	T
395	R	R	R	R	R	K
432	L	H	H	H	L	H
437	Q	Q	Q	Q	Q	R
458	R	R	R	R	R	M
466	T	K	T	T	T	T
469	K	I	S	T	T	T
470	A	K	V	V	T	A
471	E	E	E	E	E	E
477	D	N	D	D	D	N
491	T	T	T	T	T	T

C

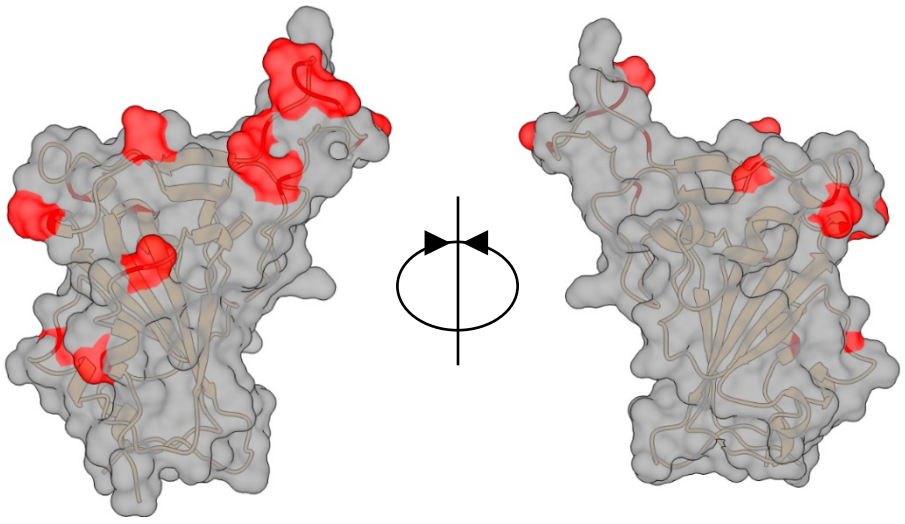


Figure 5-13. The six RBD genotypes from the Cinderford roost.

A table highlighting differences across the 13 amino acid sites among RhGB01, CI63, and six other RBD genotypes found at the Cinderford roost. It lists the amino acids present in RhGB01 and CI63, along with those from different mutations. Sample IDs with the mutated amino acid are included, along with the genotype bearing the mutation and whether the site falls within a binding loop (A). A second table with the amino acid present at each site for each genotype (B). The RhGB01 RBD predicted structure with the 13 sites with polymorphisms between the genotypes highlighted in red (C).

An in-frame nucleotide sequence for the RBD was recovered from 14 samples across the four years: 2 from 2020, 2 from 2021 and 10 from 2022. 13 amino acid sites demonstrate variation between the different RBDs: 5 in the RBD outside the RBM, and 8 in RBM. The 14 RBDs from the Cinderford roost can be represented by 6 genotypes. RhGB01 encodes distinct amino acids at sites 337, 471 and 491 compared with CI63 and the genotypes. The amino acids at sites 469 and 470 demonstrate multiple genotypes (Figure 5-13).

To investigate the structural implications of these amino acid mutations, homology modelling was used for structure prediction. For genotypes 1 and 6, a template RBD from a *Sarbecovirus* from *Rhinolophus sinicus* in China was selected as the template (SHC014). For genotypes 2-5 a SARS-CoV-2 and SARS-CoV RBD were selected, respectively. All predicted models have ERRAT2 scores >90 and all residues are in allowed regions (Appendix 22).

The genotype 4 RBD most closely resembles the RhGB01 RBD across the entire RBD (RMSD=0.008 Å), RBM (RMSD=0.010 Å), binding loop 1 (RMSD=0.010 Å), and binding loop 3 (RMSD=0.001 Å). The RhGB01 binding loop 2 exhibits most structural similarity to genotype 5 binding loop 2 (RMSD=0.002 Å). Genotype 3 diverges the most from RhGB01 across the entire RBD (RMSD=1.49 Å) and is identical to CI63. Genotype 2 showcases the most divergence from RhGB01 in the RBM and binding loops (RMSD=2.13 Å, 3.97 Å, 3.158 Å, 2.052 Å). Moreover, RBDs from genotypes 4 and 5 display considerable structural dissimilarity to CI63, with >3 Å for binding loops 1 and 2, and 1.45 Å for binding loops overall (Figure 5-14) (Appendix 23, Appendix 24).

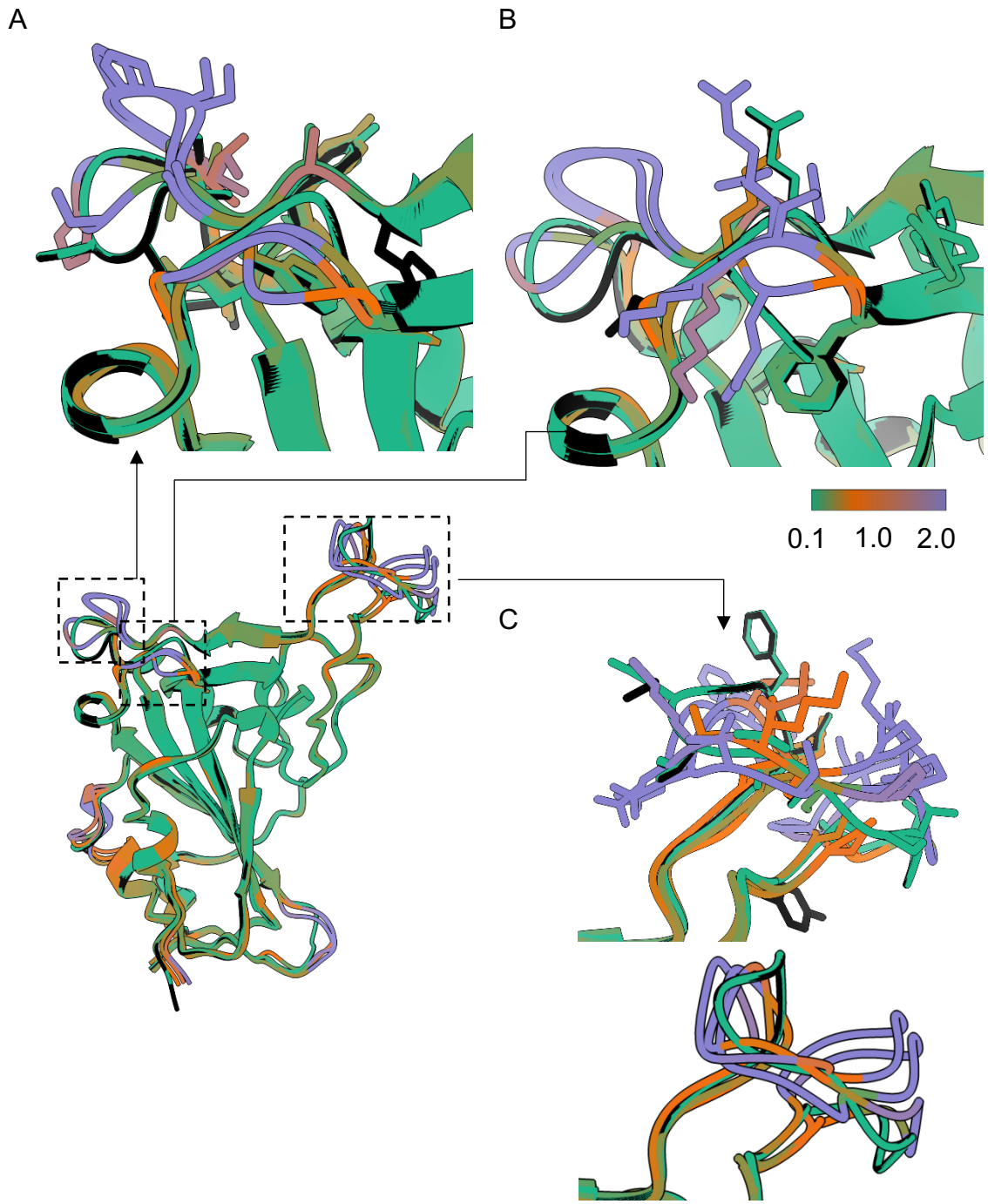


Figure 5-14. The six genotypes aligned to the RhGB01 RBD.

Predicted RBD structures for the 6 genotypes structurally aligned to RhGB01 (black). The structures are coloured by molecular distance from RhGB01. The three receptor binding loops within the receptor binding motif are highlighted in three snapshots: binding loop 1 (C) (138-151), binding loop 2 (B) (159-170) and binding loop 3 (A) (112-117). The atomic structure of the amino acids in each binding loop are shown to demonstrate the variation in structures as the amino acid changes.

5.1.12 RhGB01 and CI63 in context of other sarbecoviruses from UK lesser and greater horseshoe populations

Since the discovery and characterisation of RhGB01 in 2020, four additional sarbecoviruses have been discovered in UK lesser horseshoe (RhGB02, RhGB05-07) bats and two from greater horseshoe (RfGB01-02). Although RhGB01 and CI63 were found in the same roost, they have varying degrees of similarity to the sarbecoviruses identified elsewhere in the UK (Figure 5-15, Appendix 25-Appendix 27).

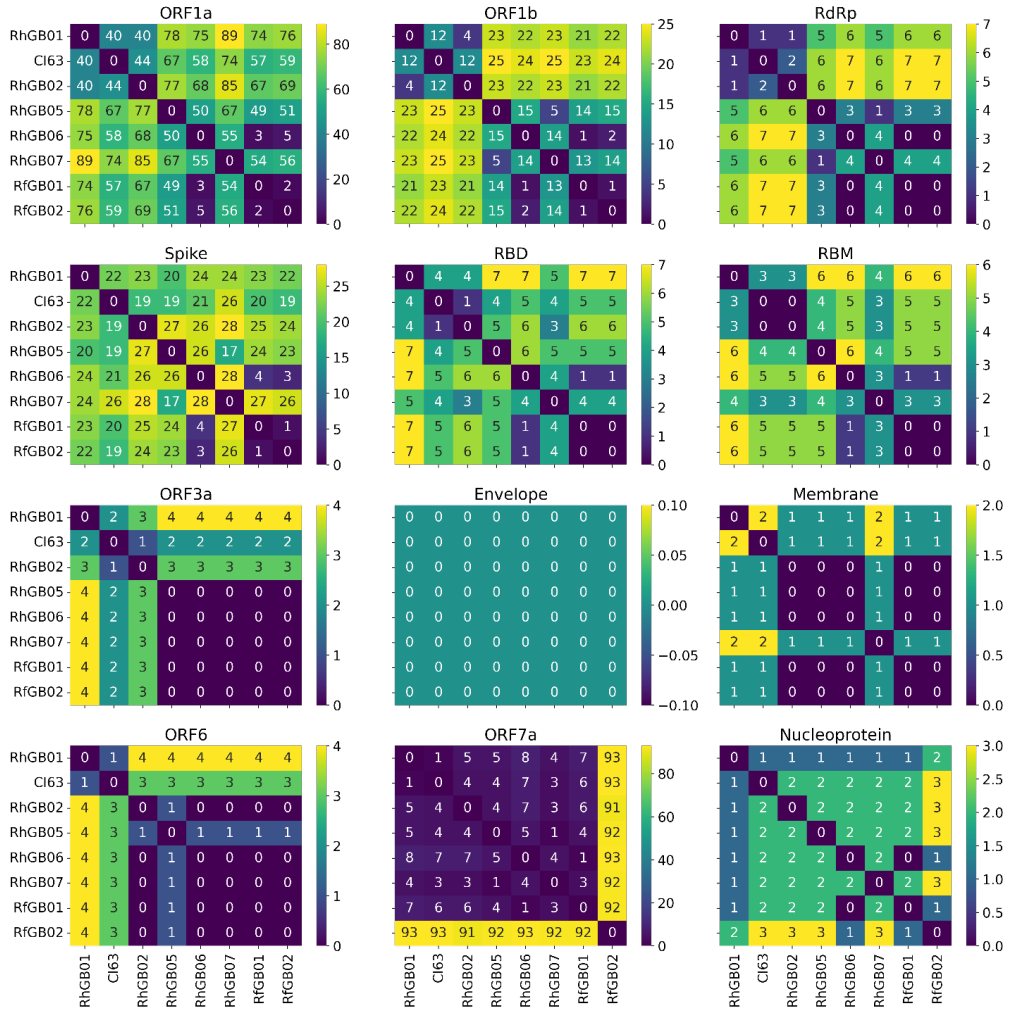


Figure 5-15. Differences between the sarbecoviruses in the UK.

This comparison includes sarbecoviruses identified in lesser and greater horseshoe bats.

5.1.13 The spike protein

Here, the focus will be on the spike protein, and specifically the receptor binding domain because of its importance for tropism and the observation of different ACE2 binding profiles for sarbecoviruses from the UK (C. C. S. Tan et al., 2023). In the complete spike protein, RhGB01 is most homologous to RhGB05, with 20 amino acid differences. Whereas CI63 is most homologous to RhGB02, RhGB05 and RfGB02 with 19 amino acid differences (Figure 5-15).

5.1.13.1 The receptor binding domain and receptor binding motif

Within the RBD, RhGB01 demonstrates four amino acid differences compared to CI63 and RhGB02. CI63 has one amino acid difference to RhGB02 (site 337). Three amino acid differences between RhGB01 and CI63/RhGB02 are within the RBM (Figure 5-16), two of which are within binding loop 1 (T469S, V471E), while the remaining two are positioned outside the RBM binding loops, specifically at site 149 (I491T) and site 337. At site 337, the methionine in RhGB01 is substituted with asparagine in CI63 and lysine in RhGB02. Notably, this mutation at site 337 also represents the sole amino acid difference between CI63 and RhGB02 (N337K) (Figure 5-16).

5.1.14 RhGB01 and CI63 in context of predicted hACE2 binding RhGB07

Another *Sarbecovirus* discovered in lesser horseshoe bats in the UK, RhGB07, has been demonstrated in vitro to bind and utilise hACE2 for cellular entry (C. C. S. Tan et al., 2023). Compared to RhGB01, RhGB07 differs by five amino acids in the RBD and four in the RBM. CI63 varies from RhGB07 by 4 amino acids in

the RBD and 3 in the RBM. Notably, site 337 in RhGB07 is identical to RhGB02 (337K). For sites 471 and 491 the amino acid in RhGB07 is identical to CI63 and RhGB02, and thus is different to RhGB01 (Figure 5-16).

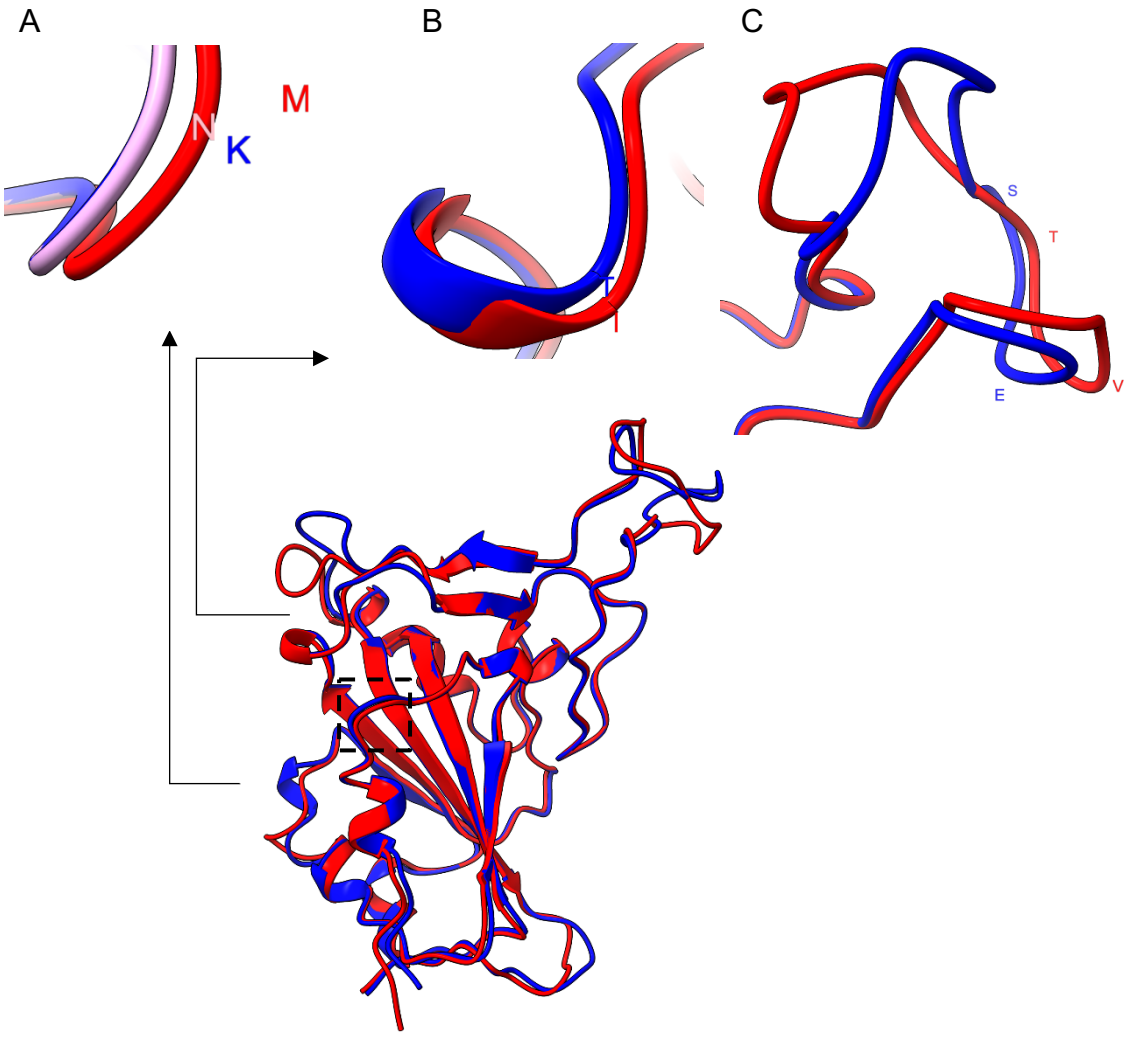


Figure 5-16. RBD differences between RhGB01 and CI63.

The receptor binding domain for RhGB01 (red), RhGB02 (blue) and CI63 (pink) for site 337 (A), site 491 (B) and sites 469 (T->S) and 471 (V->E) (C). The four amino acid mutations which separate RhGB01 from RhGB02 and CI63 are highlighted.

The additional varied sites in RhGB07 compared to RhGB01 are 435 (N435S) in binding loop 3 and 470 (V470T) in binding loop 1. The amino acid at site 469 in RhGB01 which is different in CI63 and RhGB02 is conserved in RhGB07. Despite the amino acid mutations observed between RhGB01 and RhGB07, the predicted structures exhibit more similar structures for the RDB (RMSD=0.231 Å) and in binding loops 1 and 2 (RMSD=0.052 Å, RMSD=0.065 Å) than CI63 and RhGB01. More structural difference between RhGB01 and RhGB07 was observed in binding loop 3 (RMSD=1.286 Å) (Figure 5-17).

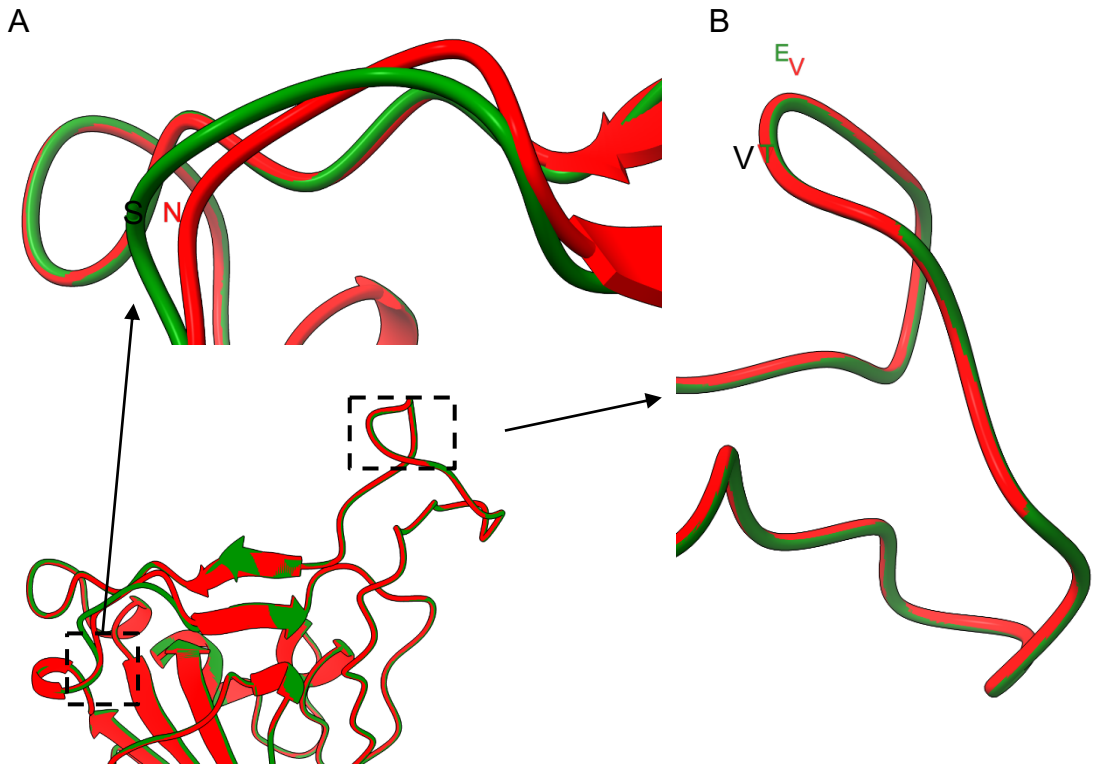
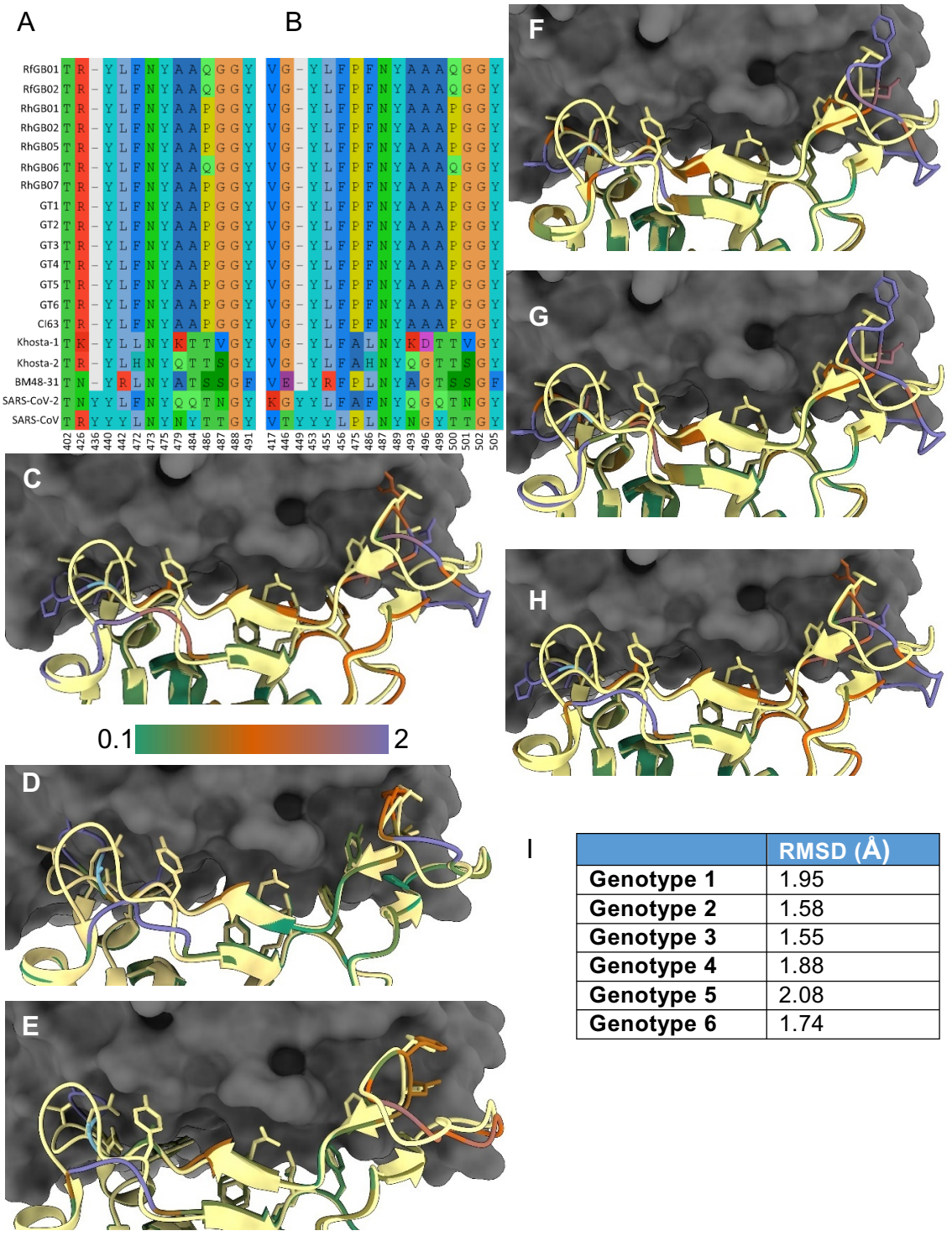


Figure 5-17. RBD differences between RhGB01 and RhGB07.

The receptor binding domain for RhGB01 (red), RhGB07 (green) for site 435 in binding loop 3 (A) and site 470 in binding loop 1 (B). These two amino acid mutations are in addition to sites 337, 471 and 491 in figure 5-16.

5.1.15 The six genotypes from the Cinderford roost in context of the SARS-CoV and SARS-CoV-2 receptor binding motifs

The amino acids observed at SARS-CoV and SARS-CoV-2 hACE2 contact residue in RhGB01 are fully conserved in CI63 and the six other genotypes from the Cinderford roost (Figure 5-18). However, there are other structural differences outside of the contact residues (Figure 5-14) which change the predicted conformation of the different RBM binding interfaces. From structural superimposition of the predicted genotype RBD structures onto SARS-CoV-2 in complex with ACE2 we can hypothesise as to the impact these mutations on the molecular space filled by the contact residues. The amino acid mutations in the receptor binding motif outside of the contact residues alter the spatial arrangement of the side chains at the contact residue sites. Genotype 3 has the most and genotype 5 the least structural similarity to the SARS-CoV-2 RBD in complex with hACE2 (RMSD=1.5-2.08 Å) (Figure 5-18).



	RMSD (Å)
Genotype 1	1.95
Genotype 2	1.58
Genotype 3	1.55
Genotype 4	1.88
Genotype 5	2.08
Genotype 6	1.74

Figure 5-18. UK sarbecoviruses compared to SARS-CoV-2 and its hACE2 binding.

The contact residues for SARS-CoV (A) and SARS-CoV-2 (B) for all sarbecoviruses from the UK, the six genotypes from the Cinderford roost and other European sarbecoviruses. Structural superimposition of the six genotypes (C-H) to SARS-CoV-2 (6M0J) in complex with hACE coloured by RMSD. Contact residues show the molecular structure of the amino acid at those sites. A table showing the RMSD for each genotype to SARS-CoV-2 (I).

5.2 Discussion

5.2.1 Longitudinal screening after the discovery of RhGB01 indicates continued circulation of sarbecoviruses in the Cinderford roost

After the characterisation of RhGB01 from the Cinderford roost sampling was continued across the following three years, to confirm the presence of this *Sarbecovirus* in the roost and longitudinally study the viral population circulating within the roost. Sarbecoviruses were detected upon each sampling of the roost by qRT-PCR. Rates of positivity were higher in the faecal samples collected in the summer months (June-July vs August-Sept-Oct), potentially due the presence of pups in the roost and an increase in excretion rate within the roost. However, the inverse could also explain the increased prevalence in the summer month. For example, in May-June the individual bats in the roost have had maximal exposure to the circulating *Sarbecovirus* in the roost compared to pups just after birth who may excrete *Sarbecovirus* negative droppings.

5.2.2 Additional sampling indicates *Sarbecovirus* circulation at other lesser horseshoe roosts

Additional sampling was conducted in non-horseshoe species in different locations in Somerset and Devon, in relatively proximity to the Cinderford roost (Southern England). No *Sarbecovirus* positivity was found in the other bat genus sampled. Additional sampling was also conducting at a lesser horseshoe roost in Plymouth, where sarbecoviruses were detected. Considering this in context of the small sample size and geographic area of the UK which has been sampled, this suggests lesser horseshoe bats are potentially disproportionately harbouring sarbecoviruses and increased surveillance is required. Despite *Sarbecovirus*

detection at the Plymouth roost and modera∆te success from mNGS for PLY24, no full genomic transcripts were returned by mNGS or amplicon amplification. The observed reduction of amplicon return from the scheme could suggest the presence of a divergent *Sarbecovirus* lineage within these lesser horseshoe bats which would prevent primer binding as in the Cinderford roost, amplicons were returned.

5.2.3 A second full genome was recovered from the Cinderford roost with 277 SNPs compared to RhGB01

The metagenomic sequencing of a low Ct sample (CI63, 23.3) collected in September 2022 generated a second full *Sarbecovirus* genome from the Cinderford roost. A de novo assembly generated the first 27.7 kb nucleotides as one contig but the nucleocapsid was absent. The nucleocapsid was 89% resolved in a second contig but the end of the transcript was truncated. The separation of the genome into two separate contigs suggests an area of low depth or sequence quality which prevents sufficient overlap to join to unitigs into a larger contig. This was circumvented by use of the RhGB01 genome for a reference guided assembly. Annotation of CI63 suggests an identical genome organisation to RhGB01 including the presence of the same accessory proteins as hypothesised for two sarbecoviruses from the same roost.

5.2.3.1 Evolutionary rate estimates indicate accelerated evolution in the receptor binding motif

A preliminary evolutionary rate of $1.4x10^{-3}$ was estimated for sarbecoviruses circulating at the Cinderford roost using the genetic distance between the complete CDS and different ORFs for RhGB01 and CI63. This evolutionary rate is within the standard range expected for an RNA virus but towards the quicker end of the scale. The evolutionary rate in nsp16, spike, RBD and RBD are all accelerated beyond $1x10^{-3}$ demonstrating more rapid evolution within the nucleotide sequence compared to other coding sequences. This observation is manifested in the disproportionate number of amino acid differences between RhGB01 and CI63 in the RBM. The observation of more amino acid differences suggests positive selection of non-synonymous SNPs in the RBM and the generation of novel genotypes with potentially different ACE2 tropisms. Future work regarding the Cinderford sarbecoviruses should focus on defining the ACE2 binding capacity of these different genotypes.

5.2.3.2 RhGB01 and CI63 demonstrate structural conservation across the majority of open reading frames

The envelope proteins of RhGB01, CI63 and other UK sarbecoviruses are identical at the amino acid level. The homology model of the envelope protein predicts the topology with the α-helical kink which has been reported in the literature. However, as stated previously, there is significant discussion in the literature regarding the correct envelope protein topology. All other structural ORFs are comparatively well conserved at the amino acid level because of functional importance for viral fitness.

5.2.3.2.1 The spike protein demonstrates variation in both the S1 NTD and CTD

The CI63 spike protein demonstrated 22 amino acid mutations compared to RhGB01 with 17 mutations in the S1-NTD and 4 within the RBD. The increased number of mutations within the S1-NTD could suggest reduced functional importance within a reservoir host with no purifying selection pressure forcing amino acid conservation. The sites are also distributed through the S1-NTD which further supports the stochasticity observed from antigenic drift. Since the RBD resides within the S1-CTD, the functional relevance of *Sarbecovirus* S1-NTD is understudied compared to the S1-CTD but since the onset of the COVID-19 pandemic the S1-NTD has been subject to investigations in context of SARS-CoV-2 infection. Early studies found little evidence for the glycan-protein binding in context of SARS-CoV-2 cellular entry. Since, studies have shown that the S1-NTD harbours epitopes for SARS-CoV-2 neutralising antibodies (Suryadevara *et al*., 2021) and further work identified α2-6-linked sialic acid containing compounds prevent SARS-CoV-2 infection by binding the S1 domain (Saso *et al*., 2022). This finding is corroborated by the observation that antigenic drift observed within the S1-NTD in small subgroups of variants of concern can result in the gain of function for sialic acid binding (Tomris *et al*., 2023) but that this is not a function under positive selection or observed in larger clades (α, β, δ, o). Thus, the amino acid disparities observed between RhGB01 and CI36 may represent antigenic drift in the S1-NTD and further investigation to elucidate the possible impact this could have on neutralisation or glycan binding phenotypes observed for the two viruses.

On the other hand, only four amino acid differences are observed within the entire RBD, with the majority (3/4) located within the RBM which is responsible for ACE2

binding. The other amino acid mutation is near binding loop 3. The amino acid mutations within the RBM are likely to impact the conformation of the receptor binding interface and should be focus of future investigation to characterise ACE2 binding alterations resulting from mutations in proximity to contact residues. As initial investigations using the RhGB01 spike suggested limited ACE2 binding capacity from the ACE2 proteins included in the screening (chapter 1), the additional of different RBD genotypes for the roost provided additional spike proteins for investigation.

5.2.3.2.2 Limitations of structural prediction using homology modelling

When considering the structural incongruencies between RhGB01 and CI63 it is important to consider that structures are predicted and may be an artefact of homology modelling. An inherent limitation of homology modelling is the availability of templates with comparable identity to the query protein. Within the public domain there is a shortage of resolved spike and RBD structures in complex with ACE2 from different species and this undoubted reduces the validity of structural comparisons using such models (Pan *et al*., 2022).

5.2.4 Increased surveillance in the UK and Europe suggests a European/African specific Sarbecovirus clade with differential ACE2 tropism profiles needs characterising

From increased surveillance in horseshoe bat populations in the UK and Europe, the non-Asian monophyletic clade has expanded to include sarbecoviruses from the UK, Russia, Bulgaria, and African countries. This expansion suggests it is

likely that there is a large, and genotypically diverse, population of sarbecoviruses circulating independently of Asian horseshoe bat populations in the Western world. As mentioned previously, this observation coupled with the inconsistent use of mammalian ACE2 receptors being elucidated through in vitro investigation highlights surveillance in horseshoe bat populations as an important public health measure going forwards.

5.2.5 Six RBD genotypes in circulation at the Cinderford roost across the four years suggest potential different ACE2 binding capabilities within one roost

In addition to metagenomics, an amplicon-based approach was employed with the aim of recovering additional transcripts from faecal samples with higher Ct values. The ORF with the greatest recovery rate was nsp9, an RNA binding protein involved in viral replication. This function could explain the increased prevalence of the transcript within the faecal samples. 14 RBD sequences were recovered from across the 4 years with 13 polymorphic amino acid sites between RhGB01 and CI63. The amino acid sites exhibiting the greatest variability are 469 and 470, with 4 and 3 distinct genotypes, respectively. These amino acids are downstream of the hACE2 contact residue for SARS-CoV-2 observed at site 475 at the end of binding loop 3.

One notable knowledge gap in the study of bat *Sarbecovirus* function pertains to two main factors: firstly, the uncertain ACE2 utilisation by most of these viruses, and secondly, the lack of molecular understanding regarding how the various RBDs interact with different ACE2 receptors. Without substantial in vitro

experimentation, the functional implications of observed genotypic variations in the Cinderford roost rely on comparisons with distantly related sarbecoviruses, such as SARS-CoV and SARS-CoV-2, and their attachment to hACE2. As mentioned previously, recent publications have started to elucidate patterns of ACE2 usage within bat associated sarbecoviruses.

5.3 Conclusions

In this chapter, we successfully characterised a second *Sarbecovirus* from the lesser horseshoe roost in the UK. This virus, CI63, differs from RhGB01 by 277 SNPs with a disproportionate representation of amino acid differences and an enhanced evolutionary rate observed in the receptor binding motif. Additionally, 14 RBD sequences were recovered from faecal droppings from across the sampling period. These sequences can be clustered into six genotypes with amino acid differences in the RBD. In silico modelling of the RBD genotypes observed at the Cinderford roost indicate each genotype presents different conformations at the ACE2 binding interface, respectively.

6 Summary

6.1 Molecular sequence analysis of Crimean-Congo haemorrhagic fever virus

This project aimed to recover novel CCHFv genomes from both archival and contemporary clinical samples from haemorrhagic fever patients. The majority of novel genomes were recovered from a single year of cases in Turkey in 2015. The genomes recovered from the subset of clinical samples from Turkey facilitated the identification of a reassorted strain, indicating the circulation of a Kosovo-like strain in Turkey. Additionally, one recovered genome clustered in the Russian subclade within the European clade without evidence of recombination or reassortment indicating the circulation of a Russia-like strain in Turkey. Together, all recovered genomes from the subset indicate a heterogeneous viral population in Turkey in 2015. Outside of the European clade, a whole genome was recovered from a clinical sample taken from a previously unrepresented outbreak in Pakistan in 1998/99. This genome provided evidence of the circulation of similar viruses in Pakistan 25 years ago to contemporary genomes characterised from Indian clinical samples.

There is no current widely employed vaccine or effective treatment for CCHFV. Thus, the identification of sites under positive selection could highlight novel therapeutic targets. In the nucleoprotein, 477 sites were identified as being under negative selection highlighting the functional importance. The M segment encoded the GPC and is an essential target for therapeutics because of the surface presentation and receptor binding ability. Two tools predict 31 sites in

total across the GPC as under positive selection, with some near mAb binding sites reported in the literature.

Additionally, this project aimed to identify geographic regions which are liable to circulation of strains with higher case fatality rates and the generation of novel reassorted/recombined strains. An important indicator of co-circulation is the generation of variants resulting from reassortment or recombination. For CCHFv, three clades within the M segment phylogeny are composed of heterogeneous geographic locations with extensive mixing of strains from Asia and Africa. This indicates that progenitor strains from both continents have co-circulated within geographic regions in the past.

The addition of additional genomes facilitated phylodynamic and phylogeographic analysis with the largest selection of CCHFv genomes to date. We identify that rate heterogeneity was observed between clades according to the small segment given the sequences which are available. Using country as a discrete trait, we predict Southern Africa as an ancestral location for CCHFv with low confidence.

6.1.1 Limitations and future work

Recovering CCHFv genomes is currently reliant on the use of mNGS because of the lack of an amplicon scheme capable of characterising all clades. This limited the recovery of genomes to low Ct clinical samples and recovery after propagation. It has become increasingly apparent that propagation in vitro can induce accelerated evolution and thus lead to uploaded sequences being poor representations of the circulating virus.

The sampling bias observed is a key limitation of both this study, and phylodynamic and phylogeographic studies in general. The overrepresentation of specific countries e.g., South Africa, Uganda, and Russia, can result in bias when estimating transitions between different traits and ancestral reconstruction across the phylogeny. However, the absence of sequence representation in regions interspacing those with representation prevents the prediction of dissemination into these regions using phylogeographic methods. In the future to improve phylogeographic estimates, the gold standard should be a coordinated effort to retrieve and systematically characterise archived clinical and wildlife samples and the formation of global consortiums to allow future outbreaks to be represented sufficiently, especially in endemic regions.

In absence of improved global coverage, there are computational methods to account for sampling bias. Regions without representation can be added into the dataset by “ghost” sequences, where the metadata is provided with an empty sequence. The number of sequences added per region could be normalised in line with reported CCHFv cases per region. This is a suitable alternative for countries with infrastructure for accurate reporting for infectious disease incidence. A second method can also be grouping countries into larger geographic regions as to reduce the bias introduced by country metadata.

6.2 Initial detection and longitudinal surveillance of lesser horseshoe bats in the UK

This project resulted from the onset of the COVID-19 pandemic. First, we used a mNGS screening on faecal droppings from a roost in Cinderford to identify any

RNA viruses. This resulted in the characterisation of the first *Sarbecovirus* from lesser horseshoe in Europe, and the first *Sarbecovirus* from the UK in any bat species. The characterisation of RhGB01 extends the western limit within which sarbecoviruses have been detected in horseshoe bat species.

The identification of RhGB01 resulted in additional sampling by other institutions in the UK, finding an additional six novel sarbecoviruses in UK horseshoe bats (C. C. S. Tan et al., 2023). This study predicted zoonotic risk by pseudoviral assays, highlighting different ACE2 tropisms within the sarbecoviruses. All sarbecoviruses from the UK also possess a bibasic FCS precursor, only one SNP from generating a polybasic FCS, as observed in SARS-CoV-2.

In the second part of this project, we longitudinally sampled the same lesser horseshoe roost to monitor the viral population in circulation. A second *Sarbecovirus* was characterised from the Cinderford roost in 2022, differing from RhGB01 by 277 SNPs across the coding sequence. We estimated a slightly elevated rate of evolution across the coding sequence, but within the 10^{-3}-10^{-4} range accepted as standard for RNA viruses. However, the RBM is identified as a mutational hotspot for the sarbecoviruses from the Cinderford roost with the most amino acid differences observed in this region.

The observation of sarbecoviruses within Europe that have different ACE2 binding profiles, the observation of different genotypes within the RBD for one roost and the bibasic motif highlights the need for continued and systematic surveillance in previously ignored geographic locations.

6.2.1 Limitations and future work

A limitation of this study is the absence of in vitro assessment of ACE2 binding profiles for all RBD genotypes, and the absence of isolation of either RhGB01 or CI63 from faecal droppings. This was mainly because of time constraints surrounding obtaining suitable cell lines. Thus, future work should be targeted at elucidating receptor binding and the production of a protocol for Sarbecovirus isolation from faecal material.

For the in-silico predictions in this project specifically, we focused on the RBD because of the inferred importance of the indels up and downstream of the receptor binding interface on ACE2 interface conformation. However, an additional 40 RBM sequences were recovered from the Cinderford roost across the longitudinal period. As all indels in the receptor binding domain are conserved within the UK *Rhinolophus* associated sarbecoviruses, these sequences provide an additional avenue of investigation when characterising the functional features of the viruses in the Cinderford roost.

Additionally, enhanced optimisation of the targeted scheme would likely recover genomes or more sequence from higher Ct droppings. However, obtaining the RBD from samples across the longitudinal period was sufficient in this study for an in silico analysis.

Bibliography

Abdiyeva, K., Turebekov, N., Dmitrovsky, A., Tukhanova, N., Shin, A., Yeraliyeva, L., Heinrich, N., Hoelscher, M., Yegemberdiyeva, R., Shapiyeva, Z., Kachiyeva, Z., Zhalmagambetova, A., Montag, J., Dobler, G., Zinner, J., Wagner, E., Frey, S., & Essbauer, S. (2019). Seroepidemiological and molecular investigations of infections with Crimean–Congo haemorrhagic fever virus in Kazakhstan. *International Journal of Infectious Diseases*, *78*, 121–127. https://doi.org/10.1016/j.ijid.2018.10.015

Adam, I. A., Mahmoud, M. A. M., & Aradaib, I. E. (2013). A seroepidemiological survey of Crimean Congo hemorrhagic fever among cattle in North Kordufan State, Sudan. *Virology Journal*, *10*, 178. https://doi.org/10.1186/1743-422X-10-178

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Agarwala, R., Bafna, V., Farach, M., Paterson, M., & Thorup, M. (1998). On the Approximability of Numerical Taxonomy (Fitting Distances by Tree Metrics). *SIAM Journal on Computing*, *28*(3), 1073–1085. https://doi.org/10.1137/S0097539795296334

Akuffo, R., Brandful, J. a. M., Zayed, A., Adjei, A., Watany, N., Fahmy, N. T., Hughes, R., Doman, B., Voegborlo, S. V., Aziati, D., Pratt, D., Awuni, J. A., Adams, N., & Dueger, E. (2016). Crimean-Congo hemorrhagic fever virus in livestock ticks and animal handler seroprevalence at an abattoir in Ghana. *BMC Infectious Diseases*, *16*, 324. https://doi.org/10.1186/s12879-016-1660-6

Al Hajjar, S., & McIntosh, K. (2010). The first influenza pandemic of the 21st century. *Annals of Saudi Medicine*, *30*(1), 1–10. https://doi.org/10.4103/0256-4947.59365

Alkhovsky, S., Lenshin, S., Romashin, A., Vishnevskaya, T., Vyshemirsky, O., Bulycheva, Y., Lvov, D., & Gitelman, A. (2022). SARS-like Coronaviruses in Horseshoe Bats (Rhinolophus spp.) in Russia, 2020. *Viruses*, *14*(1), 113. https://doi.org/10.3390/v14010113

Álvarez, E., DeDiego, M. L., Nieto-Torres, J. L., Jiménez-Guardeño, J. M., Marcos-Villar, L., & Enjuanes, L. (2010). The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated. *Virology*, *402*(2), 281–291. https://doi.org/10.1016/j.virol.2010.03.015

Angelini, M. M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. (2013). Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *MBio*, *4*(4), e00524-13. https://doi.org/10.1128/mBio.00524-13

Annan, A., Baldwin, H. J., Corman, V. M., Klose, S. M., Owusu, M., Nkrumah, E. E., Badu, E. K., Anti, P., Agbenyega, O., Meyer, B., Oppong, S., Sarkodie, Y. A., Kalko, E. K. V., Lina, P. H. C., Godlevska, E. V., Reusken, C., Seebens, A., Gloza-Rausch, F., Vallo, P., … Drexler, J. F. (2013). Human Betacoronavirus 2c EMC/2012–related Viruses in Bats, Ghana and Europe. *Emerging Infectious Diseases*, *19*(3), 3. https://doi.org/10.3201/eid1903.121503

Appannanavar, S. B., & Mishra, B. (2011). An Update on Crimean Congo Hemorrhagic Fever. *Journal of Global Infectious Diseases*, *3*(3), 285–292. https://doi.org/10.4103/0974-777X.83537

Ar Gouilh, M., Puechmaille, S. J., Diancourt, L., Vandenbogaert, M., Serra-Cobo, J., Lopez Roïg, M., Brown, P., Moutou, F., Caro, V., Vabret, A., Manuguerra, J.-C., & EPICOREM consortium. (2018). SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world. *Virology*, *517*, 88–97. https://doi.org/10.1016/j.virol.2018.01.014

Arbely, E., Khattari, Z., Brotons, G., Akkawi, M., Salditt, T., & Arkin, I. T. (2004). A Highly Unusual Palindromic Transmembrane Helical Hairpin Formed by SARS Coronavirus E Protein. *Journal of Molecular Biology*, *341*(3), 769–779. https://doi.org/10.1016/j.jmb.2004.06.044

Arenas, M. (2015). Trends in substitution models of molecular evolution. *Frontiers in Genetics*, *6*. https://doi.org/10.3389/fgene.2015.00319

Aris-Brosou, S., & Yang, Z. (2003). Bayesian Models of Episodic Evolution Support a Late Precambrian Explosive Diversification of the Metazoa. *Molecular Biology and Evolution*, *20*(12), 12. https://doi.org/10.1093/molbev/msg226

Atim, S. A., Niebel, M., Ashraf, S., Vudriko, P., Odongo, S., Balinandi, S., Aber, P., Bameka, R., Ademun, A. R., Masembe, C., Tweyongyere, R., & Thomson, E. C. (2023). Prevalence of Crimean-Congo haemorrhagic fever in livestock following a confirmed human case in Lyantonde district, Uganda. *Parasites & Vectors*, *16*, 7. https://doi.org/10.1186/s13071-022-05588-x

Attwood, S. W., Hill, S. C., Aanensen, D. M., Connor, T. R., & Pybus, O. G. (2022). Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature Reviews. Genetics*, *23*(9), 547–562. https://doi.org/10.1038/s41576-022-00483-8

August, T., Mathews, F., & Nunn, M. (2012). Alphacoronavirus Detected in Bats in the United Kingdom. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *12*, 530–533. https://doi.org/10.1089/vbz.2011.0829

Backovic, M., & Jardetzky, T. S. (2009). Class III viral membrane fusion proteins. *Current Opinion in Structural Biology*, *19*(2), 189–196. https://doi.org/10.1016/j.sbi.2009.02.012

Badawi, A., Ryoo, S. G., Vasileva, D., & Yaghoubi, S. (2018). Prevalence of chronic comorbidities in chikungunya: A systematic review and meta-analysis. *International Journal of Infectious Diseases*, *67*, 107–113. https://doi.org/10.1016/j.ijid.2017.12.018

Baker, M. L., Schountz, T., & Wang, L. -F. (2013). Antiviral Immune Responses of Bats: A Review. *Zoonoses and Public Health*, *60*(1), 104–116. https://doi.org/10.1111/j.1863-2378.2012.01528.x

Baker, S. C., & Lai, M. M. (1990). An in vitro system for the leader-primed transcription of coronavirus mRNAs. *The EMBO Journal*, *9*(12), 4173–4179.

Balboni, A., Gallina, L., Palladini, A., Prosperi, S., & Battilani, M. (2011). A Real-Time PCR Assay for Bat SARS-Like Coronavirus Detection and Its Application to Italian Greater Horseshoe Bat Faecal Sample Surveys. *The Scientific World Journal*, *2012*, 989514. https://doi.org/10.1100/2012/989514

Balboni, A., Palladini, A., Bogliani, G., & Battilani, M. (2011). Detection of a virus related to betacoronaviruses in Italian greater horseshoe bats. *Epidemiology and Infection*, *139*(2), 2. https://doi.org/10.1017/S0950268810001147

Balinandi, S., Patel, K., Ojwang, J., Kyondo, J., Mulei, S., Tumusiime, A., Lubwama, B., Nyakarahuka, L., Klena, J. D., Lutwama, J., Ströher, U., Nichol, S. T., & Shoemaker, T. R. (2018). Investigation of an isolated case of human Crimean-Congo hemorrhagic fever in Central Uganda, 2015. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *68*, 88–93. https://doi.org/10.1016/j.ijid.2018.01.013

Balinandi, S., von Brömssen, C., Tumusiime, A., Kyondo, J., Kwon, H., Monteil, V. M., Mirazimi, A., Lutwama, J., Mugisha, L., & Malmberg, M. (2021). Serological and molecular study of Crimean-Congo Hemorrhagic Fever Virus in cattle from selected districts in Uganda. *Journal of Virological Methods*, *290*, 114075. https://doi.org/10.1016/j.jviromet.2021.114075

Balinandi, S., Whitmer, S., Mulei, S., Nyakarahuka, L., Tumusiime, A., Kyondo, J., Baluku, J., Mutyaba, J., Mugisha, L., Malmberg, M., Lutwama, J., Shoemaker, T. R., & Klena, J. D. (2022). Clinical and Molecular Epidemiology of Crimean-Congo Hemorrhagic Fever in Humans in Uganda, 2013–2019. *The American Journal of Tropical Medicine and Hygiene*, *106*(1), 1. https://doi.org/10.4269/ajtmh.21-0685

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, *35*(3), 235–241.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 5. https://doi.org/10.1089/cmb.2012.0021

Baudoux, P., Carrat, C., Besnardeau, L., Charley, B., & Laude, H. (1998). Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. *Journal of Virology*, *72*(11), 11. https://doi.org/10.1128/JVI.72.11.8636-8643.1998

Beijerinck, M. W. (1898). *Ueber ein contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblatter*.

Belouzard, S., Chu, V. C., & Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences*, *106*(14), 5871–5876. https://doi.org/10.1073/pnas.0809524106

Belshaw, R., Pybus, O. G., & Rambaut, A. (2007). The evolution of genome compression and genomic novelty in RNA viruses. *Genome Research*, *17*(10), 1496–1504. https://doi.org/10.1101/gr.6305707

Berche, P. (2022). The enigma of the 1889 Russian flu pandemic: A coronavirus? *Presse Medicale (Paris, France : 1983)*, *51*(3), 104111. https://doi.org/10.1016/j.lpm.2022.104111

Bergeron, É., Albariño, C. G., Khristova, M. L., & Nichol, S. T. (2010). Crimean-Congo Hemorrhagic Fever Virus-Encoded Ovarian Tumor Protease Activity Is Dispensable for Virus RNA Polymerase Function. *Journal of Virology*, *84*(1), 1. https://doi.org/10.1128/JVI.01859-09

Bertolotti-Ciarlet, A., Smith, J., Strecker, K., Paragas, J., Altamura, L. A., McFalls, J. M., Frias-Stäheli, N., García-Sastre, A., Schmaljohn, C. S., & Doms, R. W. (2005). Cellular Localization and Antigenic Characterization of Crimean-Congo Hemorrhagic Fever Virus Glycoproteins. *Journal of Virology*, *79*(10), 10. https://doi.org/10.1128/JVI.79.10.6152-6161.2005

Bestle, D., Heindl, M. R., Limburg, H., Van Lam van, T., Pilgram, O., Moulton, H., Stein, D. A., Hardes, K., Eickmann, M., Dolnik, O., Rohde, C., Klenk, H.-D., Garten, W., Steinmetzer, T., & Böttcher-Friebertshäuser, E. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, *3*(9), e202000786. https://doi.org/10.26508/lsa.202000786

Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., Ciccozzi, M., & Pascarella, S. (2020). Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics? *BioMed Research International*, *2020*, 4389089. https://doi.org/10.1155/2020/4389089

Biek, R., Henderson, J. C., Waller, L. A., Rupprecht, C. E., & Real, L. A. (2007). A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7993–7998. https://doi.org/10.1073/pnas.0700741104

Biswal, M., Diggs, S., Xu, D., Khudaverdyan, N., Lu, J., Fang, J., Blaha, G., Hai, R., & Song, J. (2021). Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Research*, *49*(10), 5956–5966. https://doi.org/10.1093/nar/gkab370

Bob, N. S., Bâ, H., Fall, G., Ishagh, E., Diallo, M. Y., Sow, A., Sembene, P. M., Faye, O., El Kouri, B., Sidi, M. L., & Sall, A. A. (2017). Detection of the Northeastern African Rift Valley Fever Virus Lineage During the 2015 Outbreak in Mauritania. *Open Forum Infectious Diseases*, *4*(2), ofx087. https://doi.org/10.1093/ofid/ofx087

Bodur, H., Akinci, E., Ascioglu, S., Öngürü, P., & Uyar, Y. (2012). Subclinical Infections with Crimean-Congo Hemorrhagic Fever Virus, Turkey. *Emerging Infectious Diseases*, *18*(4), 4. https://doi.org/10.3201/eid1804.111374

Boere, G., & Stroud, D. (2006). The Flyway Concept: What it Is and What it Isn't. In *Waterbirds Around the World* (p. 47).

Bos, L. (1999). Beijerinck's work on tobacco mosaic virus: Historical context and legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *354*(1383), 675–685.

Bosch, B. J., van der Zee, R., de Haan, C. A. M., & Rottier, P. J. M. (2003). The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *Journal of Virology*, *77*(16), 8801–8811. https://doi.org/10.1128/JVI.77.16.8801-8811.2003

Bouaicha, F., Eisenbarth, A., Elati, K., Schulz, A., Ben Smida, B., Bouajila, M., Sassi, L., Rekik, M., Groschup, M. H., & Khamassi Khbou, M. (2021). Epidemiological investigation of Crimean-Congo haemorrhagic fever virus infection among the one-humped camels (Camelus dromedarius) in southern Tunisia. *Ticks and Tick-Borne Diseases*, *12*(1), 1. https://doi.org/10.1016/j.ttbdis.2020.101601

Boushab, B. M., Kelly, M., Kébé, H., Bollahi, M. A., & Basco, L. K. (2020). Crimean-Congo Hemorrhagic Fever, Mauritania. *Emerging Infectious Diseases*, *26*(4), 817–818. https://doi.org/10.3201/eid2604.191292

Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., & Decroly, E. (2012). RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9372–9377. https://doi.org/10.1073/pnas.1201130109

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, *368*(6470), 455–457. https://doi.org/10.1038/368455a0

Boyles, J. G., Cryan, P. M., McCracken, G. F., & Kunz, T. H. (2011). Economic Importance of Bats in Agriculture. *Science*, *332*(6025), 41–42. https://doi.org/10.1126/science.1201366

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*, *185*(20), 6220–6223. https://doi.org/10.1128/JB.185.20.6220-6223.2003

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14250–14255. https://doi.org/10.1073/pnas.202488399

Briese, T., Alkhovskiy, S., Beer, M., Calisher, C., Charrel, R., Ebihara, H., Jain, R., Kuhn, J., Lambert, A., Maes, P., Nunes, M., Plyusnin, A., Schmaljohn, C., Tesh, R., Yeh, S.-D., Elbeaino, T., Digiaro, M., Martelli, G., Muehlbach, H.-P., &

Junglen, S. (2016). *Create a new order, Bunyavirales, to accommodate nine families (eight new, one renamed) comprising thirteen genera*. https://doi.org/10.13140/RG.2.2.27230.23368

Briese, T., Calisher, C. H., & Higgs, S. (2013). Viruses of the family Bunyaviridae: Are all available isolates reassortants? *Virology*, *446*(1–2), 207–216. https://doi.org/10.1016/j.virol.2013.07.030

Brnić, D., Lojkić, I., Krešić, N., Zrnčić, V., Ružanović, L., Mikuletič, T., Bosilj, M., Steyer, A., Keros, T., Habrun, B., & Jemeršić, L. (2023). Circulation of SARS-CoV–Related Coronaviruses and Alphacoronaviruses in Bats from Croatia. *Microorganisms*, *11*(4), 959. https://doi.org/10.3390/microorganisms11040959

Bromham, L., Duchêne, S., Hua, X., Ritchie, A. M., Duchêne, D. A., & Ho, S. Y. W. (2018). Bayesian molecular dating: Opening up the black box. *Biological Reviews*, *93*(2), 1165–1191. https://doi.org/10.1111/brv.12390

Bromham, L., & Penny, D. (2003). The modern molecular clock. *Nature Reviews. Genetics*, *4*(3), 216–224. https://doi.org/10.1038/nrg1020

Brown, F. (2003). The history of research in foot-and-mouth disease. *Virus Research*, *91*(1), 3–7. https://doi.org/10.1016/S0168-1702(02)00268-X

Bryant-Genevier, J., Bumburidi, Y., Kazazian, L., Seffren, V., Head, J. R., Berezovskiy, D., Zhakipbayeva, B., Salyer, S. J., Knust, B., Klena, J. D., Chiang, C.-F., Mirzabekova, G., Rakhimov, K., Koekeev, J., Kartabayev, K., Mamadaliyev, S., Guerra, M., Blanton, C., Shoemaker, T., … Moffett, D. B. (2022). Prevalence of Crimean-Congo Hemorrhagic Fever Virus among Livestock and Ticks in Zhambyl Region, Kazakhstan, 2017. *The American Journal of Tropical Medicine and Hygiene*, *106*(5), 1478–1485. https://doi.org/10.4269/ajtmh.21-1092

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy*, *99*(1), 1–14. https://doi.org/10.1093/jmammal/gyx147

Burrell, C. J., Howard, C. R., & Murphy, F. A. (2017). Epidemiology of Viral Infections. *Fenner and White's Medical Virology*, 185–203. https://doi.org/10.1016/B978-0-12-375156-0.00013-8

Burt, F. J., Spencer, D. C., Leman, P. A., Patterson, B., & Swanepoel, R. (1996). Investigation of tick-borne viruses as pathogens of humans in South Africa and evidence of Dugbe virus infection in a patient with prolonged thrombocytopenia. *Epidemiology and Infection*, *116*(3), 353–361. https://doi.org/10.1017/s0950268800052687

Butenko, A. M., & Karganova, G. (2007). Crimean-Congo Hemorrhagic Fever in Russia and Other Countries of the Former Soviet Union. In *Crimean-Congo Hemorrhagic Fever: A Global Perspective* (pp. 99–114). https://doi.org/10.1007/978-1-4020-6106-6_9

Cajimat, M. N. B., Rodriguez, S. E., Schuster, I. U. E., Swetnam, D. M., Ksiazek, T. G., Habela, M. A., Negredo, A. I., Estrada-Peña, A., Barrett, A. D. T., & Bente, D. A. (2017). Genomic Characterization of Crimean-Congo Hemorrhagic Fever

Virus in Hyalomma Tick from Spain, 2014. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *17*(10), 10. https://doi.org/10.1089/vbz.2017.2190

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: Important Reservoir Hosts of Emerging Viruses. *Clinical Microbiology Reviews*, *19*(3), 3. https://doi.org/10.1128/CMR.00017-06

Caplat, P., Edelaar, P., Dudaniec, R. Y., Green, A. J., Okamura, B., Cote, J., Ekroos, J., Jonsson, P. R., Löndahl, J., Tesson, S. V., & Petit, E. J. (2016). Looking beyond the mountain: Dispersal barriers in a changing world. *Frontiers in Ecology and the Environment*, *14*(5), 261–268. https://doi.org/10.1002/fee.1280

Cappuccio, J. A., Pena, L., Dibárbora, M., Rimondi, A., Piñeyro, P., Insarralde, L., Quiroga, M. A., Machuca, M., Craig, M. I., Olivera, V., Chockalingam, A., Perfumo, C. J., Perez, D. R., & Pereda, A. (2011). Outbreak of swine influenza in Argentina reveals a non-contemporary human H3N2 virus highly transmissible among pigs. *The Journal of General Virology*, *92*(Pt 12), 2871–2878. https://doi.org/10.1099/vir.0.036590-0

Carpentier, M., & Chomilier, J. (2020). Analyses of displacements resulting from a point mutation in proteins. *Journal of Structural Biology*, *211*(2), 107543. https://doi.org/10.1016/j.jsb.2020.107543

Carroll, S. A., Bird, B. H., Rollin, P. E., & Nichol, S. T. (2010). Ancient common ancestry of Crimean-Congo hemorrhagic fever virus. *Molecular Phylogenetics and Evolution*, *55*(3), 3. https://doi.org/10.1016/j.ympev.2010.01.006

Carter, S. D., Surtees, R., Walter, C. T., Ariza, A., Bergeron, É., Nichol, S. T., Hiscox, J. A., Edwards, T. A., & Barr, J. N. (2012). Structure, Function, and Evolution of the Crimean-Congo Hemorrhagic Fever Virus Nucleocapsid Protein. *Journal of Virology*, *86*(20), 10914–10923. https://doi.org/10.1128/JVI.01555-12

Caruso, I. P., dos Santos Almeida, V., do Amaral, M. J., de Andrade, G. C., de Araújo, G. R., de Araújo, T. S., de Azevedo, J. M., Barbosa, G. M., Bartkevihi, L., Bezerra, P. R., dos Santos Cabral, K. M., de Lourenço, I. O., Malizia-Motta, C. L. F., de Luna Marques, A., Mebus-Antunes, N. C., Neves-Martins, T. C., de Sá, J. M., Sanches, K., Santana-Silva, M. C., … Almeida, F. C. L. (2022). Insights into the specificity for the interaction of the promiscuous SARS-CoV-2 nucleocapsid protein N-terminal domain with deoxyribonucleic acids. *International Journal of Biological Macromolecules*, *203*, 466–480. https://doi.org/10.1016/j.ijbiomac.2022.01.121

Casals, J. (1969). Antigenic Similarity between the Virus Causing Crimean Hemorrhagic Fever and Congo Virus. *Experimental Biology and Medicine*, *131*(1), 233–236. https://doi.org/10.3181/00379727-131-33847

Cassedy, A., Parle-McDermott, A., & O'Kennedy, R. (2021). Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.637559

Cator, L. J., Johnson, L. R., Mordecai, E. A., Moustaid, F. E., Smallwood, T. R. C., LaDeau, S. L., Johansson, M. A., Hudson, P. J., Boots, M., Thomas, M. B., Power, A. G., & Pawar, S. (2020). The Role of Vector Trait Variation in Vector-

Borne Disease Dynamics. *Frontiers in Ecology and Evolution*, *8*, 189. https://doi.org/10.3389/fevo.2020.00189

Causey, O. R., Kemp, G. E., Madbouly, M. H., & David-West, T. S. (1970). Congo virus from domestic livestock, African hedgehog, and arthropods in Nigeria. *The American Journal of Tropical Medicine and Hygiene*, *19*(5), 846–850. https://doi.org/10.4269/ajtmh.1970.19.846

Chang, C., Sue, S.-C., Yu, T., Hsieh, C.-M., Tsai, C.-K., Chiang, Y.-C., Lee, S., Hsiao, H., Wu, W.-J., Chang, C.-F., & Huang, T. (2005). The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *Febs Letters*, *579*(25), 5663–5668. https://doi.org/10.1016/j.febslet.2005.09.038

Chapman, L. E., Wilson, M. L., Hall, D. B., LeGuenno, B., Dykstra, E. A., Ba, K., & Fisher-Hoch, S. P. (1991). Risk factors for Crimean-Congo hemorrhagic fever in rural northern Senegal. *The Journal of Infectious Diseases*, *164*(4), 4. https://doi.org/10.1093/infdis/164.4.686

Chen, Y., Savinov, S. N., Mielech, A. M., Cao, T., Baker, S. C., & Mesecar, A. D. (2015). X-ray Structural and Functional Studies of the Three Tandemly Linked Domains of Non-structural Protein 3 (nsp3) from Murine Hepatitis Virus Reveal Conserved Functions. *The Journal of Biological Chemistry*, *290*(42), 25293–25306. https://doi.org/10.1074/jbc.M115.662130

Chiuya, T., Masiga, D. K., Falzon, L. C., Bastos, A. D. S., Fèvre, E. M., & Villinger, J. (2021). Tick-borne pathogens, including Crimean-Congo haemorrhagic fever virus, at livestock markets and slaughterhouses in western Kenya. *Transboundary and Emerging Diseases*, *68*(4), 2429–2445. https://doi.org/10.1111/tbed.13911

Choi, K. H. (2012). Viral Polymerases. *Advances in Experimental Medicine and Biology*, *726*, 267–304. https://doi.org/10.1007/978-1-4614-0980-9_12

Cholleti, H., de Jong, J., Blomström, A.-L., & Berg, M. (2022). Characterization of Pipistrellus pygmaeus Bat Virome from Sweden. *Viruses*, *14*(8), 1654. https://doi.org/10.3390/v14081654

Chornelia, A., & Hughes, A. C. (2022). The evolutionary history and ancestral biogeographic range estimation of old-world Rhinolophidae and Hipposideridae (Chiroptera). *BMC Ecology and Evolution*, *22*(1), 112. https://doi.org/10.1186/s12862-022-02066-x

Chumakov, M., Ismailova, S., Rubin, S., Smirnova, S., Zgurskaya, G., & Khankishiev, A. (1970). *Detection of crimean hemorrhagic fever foci in Azerbaijan SSR from results of serological investigations of domestic animals*.

Clementz, M. A., Chen, Z., Banach, B. S., Wang, Y., Sun, L., Ratia, K., Baez-Santos, Y. M., Wang, J., Takayama, J., Ghosh, A. K., Li, K., Mesecar, A. D., & Baker, S. C. (2010). Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *Journal of Virology*, *84*(9), 4619–4629. https://doi.org/10.1128/JVI.02406-09

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009).

Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Cockburn, W. C., Delon, P. J., & Ferreira, W. (1969). Origin and progress of the 1968-69 Hong Kong influenza epidemic. *Bulletin of the World Health Organization*, *41*(3-4–5), 343–348.

Cohen, L. E., Fagre, A. C., Chen, B., Carlson, C. J., & Becker, D. J. (2023). Coronavirus sampling and surveillance in bats from 1996–2019: A systematic review and meta-analysis. *Nature Microbiology*, *8*(6), 1176–1186. https://doi.org/10.1038/s41564-023-01375-1

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science : A Publication of the Protein Society*, *2*(9), 9.

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., … Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*, *25*(3), 2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

Corse, E., & Machamer, C. E. (2000). Infectious Bronchitis Virus E Protein Is Targeted to the Golgi Complex and Directs Release of Virus-Like Particles. *Journal of Virology*, *74*(9), 9.

Corse, E., & Machamer, C. E. (2003). The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology*, *312*(1), 25–34. https://doi.org/10.1016/S0042-6822(03)00175-2

Cottam, E. M., Whelband, M. C., & Wileman, T. (2014). Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*, *10*(8), 1426–1441. https://doi.org/10.4161/auto.29309

Couret, J., Schofield, S., & Narasimhan, S. (2022). The environment, the tick, and the pathogen – It is an ensemble. *Frontiers in Cellular and Infection Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fcimb.2022.1049646

Crook, J. M., Murphy, I., Carter, D. P., Pullan, S. T., Carroll, M., Vipond, R., Cunningham, A. A., & Bell, D. (2021). Metagenomic identification of a new sarbecovirus from horseshoe bats in Europe. *Scientific Reports*, *11*(1), 1. https://doi.org/10.1038/s41598-021-94011-z

Cross, R. K., & Fields, B. N. (1976). Use of an aberrant polypeptide as a marker in three-factor crosses: Further evidence for independent reassortment as the mechanism of recombination between temperature-sensitive mutants of reovirus type 3. *Virology*, *74*(2), 345–362. https://doi.org/10.1016/0042-6822(76)90341-X

Cunningham, C., Omland, K., & Oakley, T. (1998). Reconstructing ancestral states: A critical reappraisal. *Tree*, *13*.

Darwish, M. A., Imam, I. Z. E., Omar, F. M., & Hoogstraal, H. (1977). *A Sero-epidemiological survey for Crimean-Congo hemorrhagic fever virus in humans and domestic animals in Egypt. 52*, 156–163.

De Benedictis, P., Marciano, S., Scaravelli, D., Priori, P., Zecchin, B., Capua, I., Monne, I., & Cattoli, G. (2014). Alpha and lineage C betaCoV infections in Italian bats. *Virus Genes*, *48*(2), 2. https://doi.org/10.1007/s11262-013-1008-x

de Haan, C. A. M., Haijema, B. J., Schellen, P., Schreur, P. W., te Lintelo, E., Vennema, H., & Rottier, P. J. M. (2008). Cleavage of Group 1 Coronavirus Spike Proteins: How Furin Cleavage Is Traded Off against Heparan Sulfate Binding upon Cell Culture Adaptation. *Journal of Virology*, *82*(12), 6078–6083. https://doi.org/10.1128/JVI.00074-08

de Haan, C. A. M., Smeets, M., Vernooij, F., Vennema, H., & Rottier, P. J. M. (1999). Mapping of the Coronavirus Membrane Protein Domains Involved in Interaction with the Spike Protein. *Journal of Virology*, *73*(9), 7441–7452.

de Haan, C. A. M., Vennema, H., & Rottier, P. J. M. (2000). Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *Journal of Virology*, *74*(11), 11.

de O. Araújo, J., Pinheiro, S., Zamora, W. J., Alves, C. N., Lameira, J., & Lima, A. H. (2021). Structural, energetic and lipophilic analysis of SARS-CoV-2 non-structural protein 9 (NSP9). *Scientific Reports*, *11*(1), 1. https://doi.org/10.1038/s41598-021-02366-0

de Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. *The New Phytologist*, *198*(2), 347–385. https://doi.org/10.1111/nph.12150

Decaro, N., & Lorusso, A. (2020). Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses. *Veterinary Microbiology*, *244*, 108693. https://doi.org/10.1016/j.vetmic.2020.108693

DeDiego, M. L., Álvarez, E., Almazán, F., Rejas, M. T., Lamirande, E., Roberts, A., Shieh, W.-J., Zaki, S. R., Subbarao, K., & Enjuanes, L. (2007). A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. *Journal of Virology*, *81*(4), 1701–1713. https://doi.org/10.1128/JVI.01467-06

DeDiego, M. L., Nieto-Torres, J. L., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Álvarez, E., Oliveros, J. C., Zhao, J., Fett, C., Perlman, S., & Enjuanes, L. (2011). Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Regulates Cell Stress Response and Apoptosis. *PLoS Pathogens*, *7*(10), e1002315. https://doi.org/10.1371/journal.ppat.1002315

Del Amparo, R., & Arenas, M. (2022). Consequences of Substitution Model Selection on Protein Ancestral Sequence Reconstruction. *Molecular Biology and Evolution*, *39*(7), msac144. https://doi.org/10.1093/molbev/msac144

Dellicour, S., Gill, M. S., Faria, N. R., Rambaut, A., Pybus, O. G., Suchard, M. A., & Lemey, P. (2021). Relax, Keep Walking—A Practical Guide to Continuous

Phylogeographic Inference with BEAST. *Molecular Biology and Evolution*, *38*(8), 3486–3493. https://doi.org/10.1093/molbev/msab031

Deval, J., & Gurard-Levin, Z. A. (2022). Opportunities and Challenges in Targeting the Proofreading Activity of SARS-CoV-2 Polymerase Complex. *Molecules*, *27*(9), 2918. https://doi.org/10.3390/molecules27092918

Devaraj, S. G., Wang, N., Chen, Z., Chen, Z., Tseng, M., Barretto, N., Lin, R., Peters, C. J., Tseng, C.-T. K., Baker, S. C., & Li, K. (2007). Regulation of IRF-3-dependent Innate Immunity by the Papain-like Protease Domain of the Severe Acute Respiratory Syndrome Coronavirus. *The Journal of Biological Chemistry*, *282*(44), 32208–32221. https://doi.org/10.1074/jbc.M704870200

Deyde, V. M., Khristova, M. L., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2006). Crimean-Congo Hemorrhagic Fever Virus Genomics and Global Diversity. *Journal of Virology*, *80*(17), 17. https://doi.org/10.1128/JVI.00752-06

Dhar, A., & Minin, V. N. (2016). Maximum Likelihood Phylogenetic Inference. *Encyclopedia of Evolutionary Biology*. https://doi.org/10.1016/B978-0-12-800049-6.00207-9

Dickerson, R. E. (1971). The structures of cytochrome c and the rates of molecular evolution. *Journal of Molecular Evolution*, *1*(1), 1. https://doi.org/10.1007/BF01659392

Dieng, I., Barry, M. A., Diagne, M. M., Diop, B., Ndiaye, M., Faye, M., Ndione, M. H. D., Dieng, M. M., Bousso, A., Fall, G., Loucoubar, C., Sall, A. A., Faye, O., & Faye, O. (2020). Detection of Crimean Congo haemorrhagic fever virus in North-eastern Senegal, Bokidiawé 2019. *Emerging Microbes & Infections*, *9*(1), 2485–2487. https://doi.org/10.1080/22221751.2020.1847605

Dobzhansky, T., & Sturtevant, A. H. (1938). Inversions in the Chromosomes of Drosophila Pseudoobscura. *Genetics*, *23*(1), 28–64. https://doi.org/10.1093/genetics/23.1.28

Domingo, E., García-Crespo, C., Lobo-Vega, R., & Perales, C. (2021). Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses*, *13*(9), 9. https://doi.org/10.3390/v13091882

Drexler, J. F., Corman, V. M., Wegner, T., Tateno, A. F., Zerbinati, R. M., Gloza-Rausch, F., Seebens, A., Müller, M. A., & Drosten, C. (2011). Amplification of Emerging Viruses in a Bat Colony. *Emerging Infectious Diseases*, *17*(3), 3. https://doi.org/10.3201/eid1703.100526

Drexler, J. F., Gloza-Rausch, F., Glende, J., Corman, V. M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, A., Müller, M. A., Deng, H., Herrler, G., & Drosten, C. (2010). Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *Journal of Virology*, *84*(21), 21. https://doi.org/10.1128/JVI.00650-10

Drummon, A., Olivergpybus, & Rambaut, A. (2003). Inference of Viral Evolutionary Rates from Molecular Sequences. In *Advances in Parasitology* (Vol. 54, pp. 331–358). Elsevier. https://doi.org/10.1016/S0065-308X(03)54008-8

Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R., & Rodrigo, A. G. (2003). Measurably evolving populations. *Trends in Ecology & Evolution*, *18*(9), 481–488. https://doi.org/10.1016/S0169-5347(03)00216-7

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*, 214. https://doi.org/10.1186/1471-2148-7-214

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/molbev/msi103

du Plessis, L., McCrone, J. T., Zarebski, A. E., Hill, V., Ruis, C., Gutierrez, B., Raghwani, J., Ashworth, J., Colquhoun, R., Connor, T. R., Faria, N. R., Jackson, B., Loman, N. J., O'Toole, Á., Nicholls, S. M., Parag, K. V., Scher, E., Vasylyeva, T. I., Volz, E. M., … Pybus, O. G. (2021). Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*, *371*(6530), 708–712. https://doi.org/10.1126/science.abf2946

Duchêne, S., Lanfear, R., & Ho, S. Y. W. (2014). The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, *78*, 277–289. https://doi.org/10.1016/j.ympev.2014.05.032

Durie, I. A., Tehrani, Z. R., Karaaslan, E., Sorvillo, T. E., McGuire, J., Golden, J. W., Welch, S. R., Kainulainen, M. H., Harmon, J. R., Mousa, J. J., Gonzalez, D., Enos, S., Koksal, I., Yilmaz, G., Karakoc, H. N., Hamidi, S., Albay, C., Spengler, J. R., Spiropoulou, C. F., … Pegan, S. D. (2022). Structural characterization of protective non-neutralizing antibodies targeting Crimean-Congo hemorrhagic fever virus. *Nature Communications*, *13*(1), 7298. https://doi.org/10.1038/s41467-022-34923-0

Dutt, R. (1960). *The Economic History of India Under Early British Rule*. https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/dutt/EcHisIndia1.pdf

Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., & Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature*, *602*(7895), 142–147. https://doi.org/10.1038/s41586-021-04332-2

Edwards, A. W., & SForza, L. L. (1963). *The reconstruction of evolution*.

Eigen, M. (1993). The origin of genetic information: Viruses as models. *Gene*, *135*(1), 37–47. https://doi.org/10.1016/0378-1119(93)90047-7

el-Azazy, O. M., & Scrimgeour, E. M. (1997). Crimean-Congo haemorrhagic fever virus infection in the western province of Saudi Arabia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *91*(3), 3. https://doi.org/10.1016/s0035-9203(97)90072-9

Ellwanger, J. H., & Chies, J. A. B. (2021). Zoonotic spillover: Understanding basic aspects for better prevention. *Genetics and Molecular Biology*, *44*(1 Suppl 1), e20200355. https://doi.org/10.1590/1678-4685-GMB-2020-0355

England, M. E., Phipps, P., Medlock, J. M., Atkinson, P. M., Atkinson, B., Hewson, R., & Gale, P. (2016). Hyalomma ticks on northward migrating birds in southern Spain: Implications for the risk of entry of Crimean-Congo haemorrhagic fever virus to Great Britain. *Journal of Vector Ecology: Journal of the Society for Vector Ecology*, *41*(1), 128–134. https://doi.org/10.1111/jvec.12204

Epstein, J. H., Anthony, S. J., Islam, A., Kilpatrick, A. M., Ali Khan, S., Balkey, M. D., Ross, N., Smith, I., Zambrana-Torrelio, C., Tao, Y., Islam, A., Quan, P. L., Olival, K. J., Khan, M. S. U., Gurley, E. S., Hossein, M. J., Field, H. E., Fielder, M. D., Briese, T., … Daszak, P. (2020). Nipah virus dynamics in bats and implications for spillover to humans. *Proceedings of the National Academy of Sciences*, *117*(46), 29190–29201. https://doi.org/10.1073/pnas.2000429117

Ergünay, K. (2021). New viruses on the rise: A One Health and ecosystem-based perspective on emerging viruses. *Future Virology*, 10.2217/fvl-2021–0215. https://doi.org/10.2217/fvl-2021-0215

Erickson, B., Deyde, V., Sanchez, A., Vincent, M., & Nichol, S. (2007). N-linked glycosylation of Gn (but not Gc) is important for Crimean Congo hemorrhagic fever virus glycoprotein localization and transport. *Virology*, *361*, 348–355. https://doi.org/10.1016/j.virol.2006.11.023

Eriksson, K. K., Cervantes-Barragán, L., Ludewig, B., & Thiel, V. (2008). Mouse Hepatitis Virus Liver Pathology Is Dependent on ADP-Ribose-1″-Phosphatase, a Viral Function Conserved in the Alpha-Like Supergroup. *Journal of Virology*, *82*(24), 12325–12334. https://doi.org/10.1128/JVI.02082-08

Espunyes, J., Cabezón, O., Pailler-García, L., Dias-Alves, A., Lobato-Bailón, L., Marco, I., Ribas, M. P., Encinosa-Guzmán, P. E., Valldeperes, M., & Napp, S. (2021). Hotspot of Crimean-Congo Hemorrhagic Fever Virus Seropositivity in Wildlife, Northeastern Spain. *Emerging Infectious Diseases*, *27*(9), 2480–2484. https://doi.org/10.3201/eid2709.211105

Estrada-Peña, A., Ottowell, L. J., Medlock, J., Vatansever, Z., & Tishkova, F. (2012). Unraveling the Ecological Complexities of Tick-Associated Crimean-Congo Hemorrhagic Fever Virus Transmission: A Gap Analysis for the Western Palearctic. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *12*, 743–752. https://doi.org/10.1089/vbz.2011.0767

Estrada-Peña, A., Palomar, A. M., Santibáñez, P., Sánchez, N., Habela, M. A., Portillo, A., Romero, L., & Oteo, J. A. (2012). Crimean-Congo Hemorrhagic Fever Virus in Ticks, Southwestern Europe, 2010. *Emerging Infectious Diseases*, *18*(1), 1. https://doi.org/10.3201/eid1801.111040

Falcón, A., Vázquez-Morón, S., Casas, I., Aznar, C., Ruiz, G., Pozo, F., Perez-Breña, P., Juste, J., Ibáñez, C., Garin, I., Aihartza, J., & Echevarría, J. E. (2011). Detection of alpha and betacoronaviruses in multiple Iberian bat species. *Archives of Virology*, *156*(10), 10. https://doi.org/10.1007/s00705-011-1057-1

Faria, N. R., Suchard, M. A., Rambaut, A., & Lemey, P. (2011). Towards a quantitative understanding of viral phylogeography. *Current Opinion in Virology*, *1*(5), 5. https://doi.org/10.1016/j.coviro.2011.10.003

Featherstone, L. A., Zhang, J. M., Vaughan, T. G., & Duchene, S. (2022). Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications. *Virus Evolution*, *8*(1), veac045. https://doi.org/10.1093/ve/veac045

Fehr, A. R., Channappanavar, R., Jankevicius, G., Fett, C., Zhao, J., Athmer, J., Meyerholz, D. K., Ahel, I., & Perlman, S. (2016). The Conserved Coronavirus Macrodomain Promotes Virulence and Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection. *MBio*, *7*(6), e01721-16. https://doi.org/10.1128/mBio.01721-16

Fels, J. M., Maurer, D. P., Herbert, A. S., Wirchnianski, A. S., Vergnolle, O., Cross, R. W., Abelson, D. M., Moyer, C. L., Mishra, A. K., Aguilan, J. T., Kuehne, A. I., Pauli, N. T., Bakken, R. R., Nyakatura, E. K., Hellert, J., Quevedo, G., Lobel, L., Balinandi, S., Lutwama, J. J., … Chandran, K. (2021). Protective neutralizing antibodies from human survivors of Crimean-Congo hemorrhagic fever. *Cell*, *184*(13), 3486-3501.e21. https://doi.org/10.1016/j.cell.2021.05.001

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 6. https://doi.org/10.1007/BF01734359

Fereidouni, M., Apanaskevich, D. A., Pecor, D. B., Pshenichnaya, N. Y., Abuova, G. N., Tishkova, F. H., Bumburidi, Y., Zeng, X., Kuhn, J. H., & Keshtkar-Jahromi, M. (2023). Crimean-Congo hemorrhagic fever virus in Central, Eastern, and South-eastern Asia. *Virologica Sinica*, *38*(2), 171–183. https://doi.org/10.1016/j.virs.2023.01.001

Ferron, F., Weber, F., de la Torre, J. C., & Reguera, J. (2017). Transcription and replication mechanisms of Bunyaviridae and Arenaviridae L proteins. *Virus Research*, *234*, 118–134. https://doi.org/10.1016/j.virusres.2017.01.018

Filipe, A. R., Calisher, C. H., & Lazuick, J. (1985). Antibodies to Congo-Crimean haemorrhagic fever, Dhori, Thogoto and Bhanja viruses in southern Portugal. *Acta Virologica*, *29*(4), 4.

Fischer, K., Zeus, V., Kwasnitschka, L., Kerth, G., Haase, M., Groschup, M. H., & Balkema-Buschmann, A. (2016). Insectivorous bats carry host specific astroviruses and coronaviruses across different regions in Germany. *Infection, Genetics and Evolution*, *37*, 108–116. https://doi.org/10.1016/j.meegid.2015.11.010

Fisher-Hoch, S. P., McCormick, J. B., Swanepoel, R., Middelkoop, A. V., Harvey, S., & Kustner, H. G. V. (1992). Risk of Human Infections with Crimean-Congo Hemorrhagic Fever Virus in a South African Rural Community. *The American Journal of Tropical Medicine and Hygiene*, *47*(3), 337–345. https://doi.org/10.4269/ajtmh.1992.47.337

Freitas, N., Enguehard, M., Denolly, S., Levy, C., Neveu, G., Lerolle, S., Devignot, S., Weber, F., Bergeron, E., Legros, V., & Cosset, F.-L. (2020). The interplays between Crimean-Congo hemorrhagic fever virus (CCHFV) M segment-encoded accessory proteins and structural proteins promote virus assembly and infectivity. *PLoS Pathogens*, *16*(9), 9. https://doi.org/10.1371/journal.ppat.1008850

Frieman, M., Ratia, K., Johnston, R. E., Mesecar, A. D., & Baric, R. S. (2009). Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling. *Journal of Virology*, *83*(13), 6689–6705. https://doi.org/10.1128/JVI.02220-08

Frieman, M., Yount, B., Agnihothram, S., Page, C., Donaldson, E., Roberts, A., Vogel, L., Woodruff, B., Scorpio, D., Subbarao, K., & Baric, R. S. (2012). Molecular Determinants of Severe Acute Respiratory Syndrome Coronavirus Pathogenesis and Virulence in Young and Aged Mouse Models of Human Disease. *Journal of Virology*, *86*(2), 884–897. https://doi.org/10.1128/JVI.05957-11

Fujita, M. S. (1991). Flying Foxes (Chiroptera: Pteropodidae): Threatened Animals of Key Ecological and Economic Importance. *Conservation Biology*, *5*(4).

Galli, A., & Bukh, J. (2014). Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends in Microbiology*, *22*(6), 354–364. https://doi.org/10.1016/j.tim.2014.02.005

Gandhi, S., Dave, P., Patel, G., Khatri, H., Shah, N., & Mishra, U. (2011). An epidemiological investigation of a multisource outbreak of Crimean-Congo hemorrhagic fever in Gujarat. *The Journal of Communicable Diseases*, *43*, 161–167.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., Ge, J., Zheng, L., Zhang, Y., Wang, H., Zhu, Y., Zhu, C., Hu, T., Hua, T., Zhang, B., … Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (New York, N.y.)*, *368*(6492), 779–782. https://doi.org/10.1126/science.abb7498

Gargili, A., Estrada-Peña, A., Spengler, J. R., Lukashev, A., Nuttall, P. A., & Bente, D. A. (2017). The role of ticks in the maintenance and transmission of Crimean-Congo hemorrhagic fever virus: A review of published field and laboratory studies. *Antiviral Research*, *144*, 93–119. https://doi.org/10.1016/j.antiviral.2017.05.010

Garrison, A., Radoshitzky, S., Kota, K., Pegoraro, G., Ruthel, G., Kuhn, J., Altamura, L., Kwilas, S., Bavari, S., Haucke, V., & Schmaljohn, C. (2013). Crimean–Congo hemorrhagic fever virus utilizes a clathrin- and early endosome-dependent entry pathway. *Virology*, *444*, 45–54. https://doi.org/10.1016/j.virol.2013.05.030

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A. A., Zhu, G., Epstein, J. H., Mazet, J. K., Hu, B., Zhang, W., Peng, C., Zhang, Y.-J., Luo, C.-M., Tan, B., Wang, N., Zhu, Y., Crameri, G., Zhang, S.-Y., Wang, L.-F., Daszak, P., & Shi, Z.-L. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, *503*(7477), 7477. https://doi.org/10.1038/nature12711

Geilhausen, H. E., Ligon, F. B., & Lukert, P. D. (1973). The pathogenesis of virulent and avirulent avian infectious bronchitis virus. *Archiv Fur Die Gesamte Virusforschung*, *40*(3), 285–290. https://doi.org/10.1007/BF01242547

Gerber, M., Isel, C., Moules, V., & Marquet, R. (2014). Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends in Microbiology*, *22*(8), 446–455. https://doi.org/10.1016/j.tim.2014.04.001

Gerrish, P. J., & García-Lerma, J. G. (2003). Mutation rate and the efficacy of antimicrobial drug treatment. *The Lancet Infectious Diseases*, *3*(1), 28–32. https://doi.org/10.1016/S1473-3099(03)00485-7

Ghosh, A., Pithadia, A. S., Bhat, J., Bera, S., Midya, A., Fierke, C. A., Ramamoorthy, A., & Bhunia, A. (2015). Self-Assembly of a 9-Residue Amyloid-Forming Peptide Fragment of SARS Corona Virus E-protein: Mechanism of Self Aggregation and Amyloid-Inhibition of hIAPP. *Biochemistry*, *54*(13), 2249–2261. https://doi.org/10.1021/acs.biochem.5b00061

Gilardi, K. V. K., & Mazet, J. A. K. (2019). The United States Agency for International Development Emerging Pandemic Threats PREDICT Project—Global Detection of Emerging Wildlife Viral Zoonoses. *Fowler's Zoo and Wild Animal Medicine Current Therapy, Volume 9*, 110–116. https://doi.org/10.1016/B978-0-323-55228-8.00019-9

Gloza-Rausch, F., Ipsen, A., Seebens, A., Göttsche, M., Panning, M., Drexler, J. F., Petersen, N., Annan, A., Grywna, K., Müller, M., Pfefferle, S., & Drosten, C. (2008). Detection and Prevalence Patterns of Group I Coronaviruses in Bats, Northern Germany. *Emerging Infectious Diseases*, *14*(4), 4. https://doi.org/10.3201/eid1404.071439

Godet, M., L'Haridon, R., Vautherot, J.-F., & Laude, H. (1992). TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology*, *188*(2), 666–675. https://doi.org/10.1016/0042-6822(92)90521-P

Goffard, A., Demanche, C., Arthur, L., Pinçon, C., Michaux, J., & Dubuisson, J. (2015). Alphacoronaviruses Detected in French Bats Are Phylogeographically Linked to Coronaviruses of European Bats. *Viruses*, *7*(12), 12. https://doi.org/10.3390/v7122937

Golden, J. W., Shoemaker, C. J., Lindquist, M. E., Zeng, X., Daye, S. P., Williams, J. A., Liu, J., Coffin, K. M., Olschner, S., Flusin, O., Altamura, L. A., Kuehl, K. A., Fitzpatrick, C. J., Schmaljohn, C. S., & Garrison, A. R. (2019). GP38-targeting monoclonal antibodies protect adult mice against lethal Crimean-Congo hemorrhagic fever virus infection. *Science Advances*, *5*(7), 7. https://doi.org/10.1126/sciadv.aaw9535

González Gordon, L., Bessell, P. R., Nkongho, E. F., Ngwa, V. N., Tanya, V. N., Sander, M., Ndip, L., Morgan, K. L., Handel, I. G., Mazeri, S., Bronsvoort, B. M., & Kelly, R. F. (2022). Seroepidemiology of Crimean-Congo Haemorrhagic Fever among cattle in Cameroon: Implications from a One Health perspective. *PLoS Neglected Tropical Diseases*, *16*(3), e0010217. https://doi.org/10.1371/journal.pntd.0010217

Gonzalez, J. P., LeGuenno, B., Guillaud, M., & Wilson, M. L. (1990). A fatal case of Crimean-Congo haemorrhagic fever in Mauritania: Virological and serological evidence suggesting epidemic transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *84*(4), 4. https://doi.org/10.1016/0035-9203(90)90045-g

Gonzalez, J.-P., LeGuenno, B., Guillaud, M., & Wilson, M. L. (1990). A fatal case of Crimean-Congo haemorrhagic fever in Mauritania: Virological and serological evidence suggesting epidemic transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *84*(4), 4. https://doi.org/10.1016/0035-9203(90)90045-G

Gonzalez, M. E., & Carrasco, L. (2003). Viroporins. *FEBS Letters*, *552*(1), 28–34. https://doi.org/10.1016/s0014-5793(03)00780-4

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., … Krogan, N. J. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug-Repurposing. *Nature*, *583*(7816), 459–468. https://doi.org/10.1038/s41586-020-2286-9

Grard, G., Drexler, J. F., Fair, J., Muyembe, J.-J., Wolfe, N. D., Drosten, C., & Leroy, E. M. (2011). Re-Emergence of Crimean-Congo Hemorrhagic Fever Virus in Central Africa. *PLoS Neglected Tropical Diseases*, *5*(10), 10. https://doi.org/10.1371/journal.pntd.0001350

Gray, J. S., Dautel, H., Estrada-Peña, A., Kahl, O., & Lindgren, E. (2009). Effects of Climate Change on Ticks and Tick-Borne Diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases*, *2009*, 593232. https://doi.org/10.1155/2009/593232

Greiner, A. L., Mamuchishvili, N., Kakutia, N., Stauffer, K., Geleishvili, M., Chitadze, N., Chikviladze, T., Zakhashvili, K., Morgan, J., & Salyer, S. J. (2016). Crimean-Congo Hemorrhagic Fever Knowledge, Attitudes, Practices, Risk Factors, and Seroprevalence in Rural Georgian Villages with Known Transmission in 2014. *PLOS ONE*, *11*(6), 6. https://doi.org/10.1371/journal.pone.0158049

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*, *303*(5656), 327–332. https://doi.org/10.1126/science.1090727

Guan, Y. (2003). Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science*, *302*(5643), 5643. https://doi.org/10.1126/science.1087139

Guilherme, J. M., Gonella-Legall, C., Legall, F., Nakoume, E., & Vincent, J. (1996). Seroprevalence of five arboviruses in Zebu cattle in the Central African Republic. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *90*(1), 31–33. https://doi.org/10.1016/S0035-9203(96)90468-X

Guindon, S., Guo, H., & Welch, D. (2016). Demographic inference under the coalescent in a spatial continuum. *Theoretical Population Biology*, *111*, 43–50. https://doi.org/10.1016/j.tpb.2016.05.002

Guo, Y., Wang, W., Ji, W., Deng, M., Sun, Y., Zhou, H., Yang, C., Deng, F., Wang, H., Hu, Z., Lou, Z., & Rao, Z. (2012). Crimean–Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proceedings of the*

National Academy of Sciences of the United States of America, *109*(13), 13. https://doi.org/10.1073/pnas.1200808109

Gupta, M., Azumaya, C. M., Moritz, M., Pourmal, S., Diallo, A., Merz, G. E., Jang, G., Bouhaddou, M., Fossati, A., Brilot, A. F., Diwanji, D., Hernandez, E., Herrera, N., Kratochvil, H. T., Lam, V. L., Li, F., Li, Y., Nguyen, H. C., Nowotny, C., … Verba, K. A. (2021). CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *BioRxiv*, 2021.05.10.443524. https://doi.org/10.1101/2021.05.10.443524

Gupta, P., Singh, M. P., Goyal, K., Tripti, P., Ansari, M. I., Obli Rajendran, V., Dhama, K., & Malik, Y. S. (2021). Bats and viruses: A death-defying friendship. *VirusDisease*, *32*(3), 467–479. https://doi.org/10.1007/s13337-021-00716-0

Guy, C., Ratcliffe, J. M., & Mideo, N. (2020). The influence of bat ecology on viral diversity and reservoir status. *Ecology and Evolution*, *10*(12), 5748–5758. https://doi.org/10.1002/ece3.6315

Habjan, M., Andersson, I., Klingström, J., Schümann, M., Martin, A., Zimmermann, P., Wagner, V., Pichlmair, A., Schneider, U., Mühlberger, E., Mirazimi, A., & Weber, F. (2008). Processing of Genome 5′ Termini as a Strategy of Negative-Strand RNA Viruses to Avoid RIG-I-Dependent Interferon Induction. *PLoS ONE*, *3*(4), 4. https://doi.org/10.1371/journal.pone.0002032

Hagberg, A., Schult, D., & Swart, P. (2008). *Exploring network structure, dynamics, and function using NetworkX*. 11–15.

Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K. M., Smith, C. M., Rota, P. A., & Baker, S. C. (2004). Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. *Journal of Virology*, *78*(24), 13600–13612. https://doi.org/10.1128/JVI.78.24.13600-13612.2004

Hardmeier, I., Aeberhard, N., Qi, W., Schoenbaechler, K., Kraettli, H., Hatt, J.-M., Fraefel, C., & Kubacki, J. (2021). Metagenomic analysis of fecal and tissue samples from 18 endemic bat species in Switzerland revealed a diverse virus composition including potentially zoonotic viruses. *PLoS ONE*, *16*(6), e0252534. https://doi.org/10.1371/journal.pone.0252534

Hassanein, K. M., el-Azazy, O. M., & Yousef, H. M. (1997). Detection of Crimean-Congo haemorrhagic fever virus antibodies in humans and imported livestock in Saudi Arabia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *91*(5), 5. https://doi.org/10.1016/s0035-9203(97)90014-6

Hawman, D. W., & Feldmann, H. (2023). Crimean–Congo haemorrhagic fever virus. *Nature Reviews Microbiology*, *21*(7), 7. https://doi.org/10.1038/s41579-023-00871-9

Head, J. R., Bumburidi, Y., Mirzabekova, G., Rakhimov, K., Dzhumankulov, M., Salyer, S. J., Knust, B., Berezovskiy, D., Kulatayeva, M., Zhetibaev, S., Shoemaker, T., Nicholson, W. L., & Moffett, D. (2020). Risk Factors for and Seroprevalence of Tickborne Zoonotic Diseases among Livestock Owners, Kazakhstan. *Emerging Infectious Diseases*, *26*(1), 1. https://doi.org/10.3201/eid2601.190220

Head, J. R., Bumburidi, Y., Salyer, S. J., Knust, B., Kuralbekovna, M. G., & Moffett, D. B. (2018). Prevalence of CCHF Virus in Ticks and People and Public Awareness in Zhambyl Region, Kazakhstan. *Online Journal of Public Health Informatics*, *10*(1), e158. https://doi.org/10.5210/ojphi.v10i1.8951

Hemnani, M., da Silva, P. G., Thompson, G., Poeta, P., Rebelo, H., & Mesquita, J. R. (2023). First Report of Alphacoronavirus Circulating in Cavernicolous Bats from Portugal. *Viruses*, *15*(7), 1521. https://doi.org/10.3390/v15071521

Hewson, R., Gmyl, A., Gmyl, L., Smirnova, S. E., Karganova, G., Jamil, B., Hasan, R., Chamberlain, J., & Clegg, C. (2004). Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *The Journal of General Virology*, *85*(Pt 10), Pt 10. https://doi.org/10.1099/vir.0.80121-0

Ho, S. Y. W., Tong, K. J., Foster, C. S. P., Ritchie, A. M., Lo, N., & Crisp, M. D. (2015). Biogeographic calibrations for the molecular clock. *Biology Letters*, *11*(9), 20150194. https://doi.org/10.1098/rsbl.2015.0194

Hong, M., & DeGrado, W. F. (2008). Structural basis for proton conduction and inhibition by the influenza M2 protein. *Nature*, *451*(7178), 596–599. https://doi.org/10.1038/nature06528

Honigsbaum, M. (2020). Revisiting the 1957 and 1968 influenza pandemics. *The Lancet*, *395*(10240), 1824–1826. https://doi.org/10.1016/S0140-6736(20)31201-0

Hoogstraal, H. (1979a). Review Article: The Epidemiology of Tick-Borne Crimean-Congo Hemorrhagic Fever in Asia, Europe, and Africa. *Journal of Medical Entomology*, *15*(4), 4. https://doi.org/10.1093/jmedent/15.4.307

Hoogstraal, H. (1979b). Review Article: The Epidemiology of Tick-Borne Crimean-Congo Hemorrhagic Fever in Asia, Europe, and Africa. *Journal of Medical Entomology*, *15*(4), 4. https://doi.org/10.1093/jmedent/15.4.307

Horton, H., Vogel, T., O'Connor, D., Picker, L., & Watkins, D. I. (2002). Analysis of the immune response and viral evolution during the acute phase of SIV infection. *Vaccine*, *20*(15), 1927–1932. https://doi.org/10.1016/S0264-410X(02)00069-5

Horváth, L. B. (1976). Precipitating antibodies to Crimean haemorrhagic fever virus in human sera collected in Hungary. *Acta Microbiologica Academiae Scientiarum Hungaricae*, *23*(4), 4.

Hsieh, P.-K., Chang, S. C., Huang, C.-C., Lee, T.-T., Hsiao, C.-W., Kou, Y.-H., Chen, I.-Y., Chang, C.-K., Huang, T.-H., & Chang, M.-F. (2005). Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. *Journal of Virology*, *79*(22), 22. https://doi.org/10.1128/JVI.79.22.13848-13855.2005

Hu, B., Zeng, L.-P., Yang, X.-L., Ge, X.-Y., Zhang, W., Li, B., Xie, J.-Z., Shen, X.-R., Zhang, Y.-Z., Wang, N., Luo, D.-S., Zheng, X.-S., Wang, M.-N., Daszak, P., Wang, L.-F., Cui, J., & Shi, Z.-L. (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathogens*, *13*(11), 11. https://doi.org/10.1371/journal.ppat.1006698

Hu, D., Zhu, C., Ai, L., He, T., Wang, Y., Ye, F., Yang, L., Ding, C., Zhu, X., Lv, R., Zhu, J., Hassan, B., Feng, Y., Tan, W., & Wang, C. (2018). Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerging Microbes & Infections*, *7*, 154. https://doi.org/10.1038/s41426-018-0155-5

Hu, W., Yen, Y.-T., Singh, S., Kao, C.-L., & Wu-Hsieh, B. A. (2012). SARS-CoV Regulates Immune Function-Related Gene Expression in Human Monocytic Cells. *Viral Immunology*, *25*(4), 277–288. https://doi.org/10.1089/vim.2011.0099

Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W., & Olejniczak, E. T. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*, *43*(20), 6059–6063. https://doi.org/10.1021/bi036155b

Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 9. https://doi.org/10.1038/s41401-020-0485-4

Huelsenbeck, J. P., Larget, B., & Swofford, D. (2000). A Compound Poisson Process for Relaxing the Molecular Clock. *Genetics*, *154*(4), 4. https://doi.org/10.1093/genetics/154.4.1879

Hulswit, R. J. G., Paesen, G. C., Bowden, T. A., & Shi, X. (2021). Recent Advances in Bunyavirus Glycoprotein Research: Precursor Processing, Receptor Binding and Structure. *Viruses*, *13*(2), 353. https://doi.org/10.3390/v13020353

Hurst, K. R., Koetzner, C. A., & Masters, P. S. (2013). Characterization of a Critical Interaction between the Coronavirus Nucleocapsid Protein and Nonstructural Protein 3 of the Viral Replicase-Transcriptase Complex. *Journal of Virology*, *87*(16), 9159–9172. https://doi.org/10.1128/JVI.01275-13

Hurtado-Tamayo, J., Requena-Platek, R., Enjuanes, L., Bello-Perez, M., & Sola, I. (2023). Contribution to pathogenesis of accessory proteins of deadly human coronaviruses. *Frontiers in Cellular and Infection Microbiology*, *13*. https://doi.org/10.3389/fcimb.2023.1166839

Imbert, I., Snijder, E. J., Dimitrova, M., Guillemot, J.-C., Lécine, P., & Canard, B. (2008). The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. *Virus Research*, *133*(2), 136–148. https://doi.org/10.1016/j.virusres.2007.11.017

Irving, A. T., Ahn, M., Goh, G., Anderson, D. E., & Wang, L.-F. (2021). Lessons from the host defences of bats, a unique viral reservoir. *Nature*, *589*(7842), 7842. https://doi.org/10.1038/s41586-020-03128-0

Jahirul Islam, Md., Nawal Islam, N., Siddik Alom, Md., Kabir, M., & Halim, M. A. (2023). A review on structural, non-structural, and accessory proteins of SARS-CoV-2: Highlighting drug target sites. *Immunobiology*, *228*(1), 152302. https://doi.org/10.1016/j.imbio.2022.152302

Jankevicius, G., Hassler, M., Golia, B., Rybin, V., Zacharias, M., Timinszky, G., & Ladurner, A. G. (2013). A family of macrodomain proteins reverses cellular

mono-ADP-ribosylation. *Nature Structural & Molecular Biology*, *20*(4), 4. https://doi.org/10.1038/nsmb.2523

Jauréguiberry, S., Tattevin, P., Tarantola, A., Legay, F., Tall, A., Nabeth, P., Zeller, H., & Michelet, C. (2005). Imported Crimean-Congo hemorrhagic Fever. *Journal of Clinical Microbiology*, *43*(9), 9. https://doi.org/10.1128/JCM.43.9.4905-4907.2005

Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., & Rao, Z. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Research*, *47*(12), 6538–6550. https://doi.org/10.1093/nar/gkz409

Jin, Y., Ouyang, M., Yu, T., Zhuang, J., Wang, W., Liu, X., Duan, F., Guo, D., Peng, X., & Pan, J.-A. (2022). Genome-Wide Analysis of the Indispensable Role of Non-structural Proteins in the Replication of SARS-CoV-2. *Frontiers in Microbiology*, *13*. https://www.frontiersin.org/articles/10.3389/fmicb.2022.907422

Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A., Zou, J., Zhang, X., Juelich, T., Smith, J. K., Zhang, L., Bopp, N., Schindewolf, C., Vu, M., Vanderheiden, A., Winkler, E. S., Swetnam, D., Plante, J. A., Aguilar, P., … Menachery, V. D. (2021). Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*. https://doi.org/10.1038/s41586-021-03237-4

Johnson, B. A., Xie, X., Kalveram, B., Lokugamage, K. G., Muruato, A., Zou, J., Zhang, X., Juelich, T., Smith, J. K., Zhang, L., Bopp, N., Schindewolf, C., Vu, M., Vanderheiden, A., Swetnam, D., Plante, J. A., Aguilar, P., Plante, K. S., Lee, B., … Menachery, V. D. (2020). Furin Cleavage Site Is Key to SARS-CoV-2 Pathogenesis. *BioRxiv*, 2020.08.26.268854. https://doi.org/10.1101/2020.08.26.268854

Johnson, M. A., Chatterjee, A., Neuman, B. W., & Wüthrich, K. (2010). SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding. *Journal of Molecular Biology*, *400*(4), 724–742. https://doi.org/10.1016/j.jmb.2010.05.027

Jones, G., & Teeling, E. C. (2006). The evolution of echolocation in bats. *Trends in Ecology & Evolution*, *21*(3), 149–156. https://doi.org/10.1016/j.tree.2006.01.001

Jordahl, K. (2014). *GeoPandas: Python tools for geographic data*.

Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T., & Poon, A. F. Y. (2016). Ancestral Reconstruction. *PLOS Computational Biology*, *12*(7), e1004763. https://doi.org/10.1371/journal.pcbi.1004763

Jukes, T., & Cantor, C. (1968). *Evolution of protein molecules*. 21–132.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 7873. https://doi.org/10.1038/s41586-021-03819-2

Kajihara, M., Simuunza, M., Saasa, N., Dautu, G., Mori-Kajihara, A., Qiu, Y., Nakao, R., Eto, Y., Furumoto, H., Hang'ombe, B. M., Orba, Y., Sawa, H., Simulundu, E., Fukushi, S., Morikawa, S., Saijo, M., Arikawa, J., Kabilika, S., Monze, M., … Yoshimatsu, K. (2021). Serologic and molecular evidence for circulation of Crimean-Congo hemorrhagic fever virus in ticks and cattle in Zambia. *PLoS Neglected Tropical Diseases*, *15*(6), 6. https://doi.org/10.1371/journal.pntd.0009452

Kakavandi, S., Zare, I., VaezJalali, M., Dadashi, M., Azarian, M., Akbari, A., Ramezani Farani, M., Zalpoor, H., & Hajikhani, B. (2023). Structural and non-structural proteins in SARS-CoV-2: Potential aspects to COVID-19 treatment or prevention of progression of related diseases. *Cell Communication and Signaling*, *21*(1), 110. https://doi.org/10.1186/s12964-023-01104-5

Kalkauskas, A., Perron, U., Sun, Y., Goldman, N., Baele, G., Guindon, S., & De Maio, N. (2021). Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk. *PLoS Computational Biology*, *17*(1), e1008561. https://doi.org/10.1371/journal.pcbi.1008561

Kanjanahaluethai, A., Chen, Z., Jukneliene, D., & Baker, S. C. (2007). Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology*, *361*(2), 391–401. https://doi.org/10.1016/j.virol.2006.12.009

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 7. https://doi.org/10.1038/s41576-020-0233-0

Karasin, A. I., Schutten, M. M., Cooper, L. A., Smith, C. B., Subbarao, K., Anderson, G. A., Carman, S., & Olsen, C. W. (2000). Genetic characterization of H3N2 influenza viruses isolated from pigs in North America, 1977-1999: Evidence for wholly human and reassortant virus genotypes. *Virus Research*, *68*(1), 71–85. https://doi.org/10.1016/s0168-1702(00)00154-4

Karlberg, H., Tan, Y.-J., & Mirazimi, A. (2011). Induction of Caspase Activation and Cleavage of the Viral Nucleocapsid Protein in Different Cell Types during Crimean-Congo Hemorrhagic Fever Virus Infection. *The Journal of Biological Chemistry*, *286*(5), 5. https://doi.org/10.1074/jbc.M110.149369

Kasso, M., & Balakrishnan, M. (2013). Ecological and Economic Importance of Bats (Order Chiroptera). *International Scholarly Research Notices*, *2013*, e187415. https://doi.org/10.1155/2013/187415

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kaul, H. N., Shetty, P. S., Ghalsasi, G. R., & Dhanda, V. (1990). Survey of ticks (Acarina: Ixodidae) for Crimean haemorrhagic fever virus activity in Jammu & Kashmir state, India. *The Indian Journal of Medical Research*, *91*, 5–8.

Kaushal, N., & Baranwal, M. (2023). Mutational analysis of catalytic site domain of CCHFV L RNA segment. *Journal of Molecular Modeling*, *29*(4), 88. https://doi.org/10.1007/s00894-023-05487-7

Kawasaki, J., Kojima, S., Tomonaga, K., & Horie, M. (2021). Hidden Viral Sequences in Public Sequencing Data and Warning for Future Emerging Diseases. *MBio*, *12*(4), e0163821. https://doi.org/10.1128/mBio.01638-21

Keane, S. C., & Giedroc, D. P. (2013). Solution Structure of Mouse Hepatitis Virus (MHV) nsp3a and Determinants of the Interaction with MHV Nucleocapsid (N) Protein. *Journal of Virology*, *87*(6), 3502–3515. https://doi.org/10.1128/JVI.03112-12

Kemenesi, G., Dallos, B., Görföl, T., Boldogh, S., Estók, P., Kurucz, K., Kutas, A., Földes, F., Oldal, M., Németh, V., Martella, V., Banyai, K., & Jakab, F. (2014). Molecular Survey of RNA Viruses in Hungarian Bats: Discovering Novel Astroviruses, Coronaviruses, and Caliciviruses. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *14*, 846–855. https://doi.org/10.1089/vbz.2014.1637

Khamassi Khbou, M., Romdhane, R., Bouaicha Zaafouri, F., Bouajila, M., Sassi, L., Appelberg, S. K., Schulz, A., Mirazimi, A., Groschup, M. H., Rekik, M., Benzarti, M., & Gharbi, M. (2021). Presence of antibodies to Crimean Congo haemorrhagic fever virus in sheep in Tunisia, North Africa. *Veterinary Medicine and Science*, *7*(6), 2323–2329. https://doi.org/10.1002/vms3.597

Khan, A. S., Maupin, G. O., Rollin, P. E., Noor, A. M., Shurie, H. H., Shalabi, A. G., Wasef, S., Haddad, Y. M., Sadek, R., Ijaz, K., Peters, C. J., & Ksiazek, T. G. (1997). An outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates, 1994-1995. *The American Journal of Tropical Medicine and Hygiene*, *57*(5), 5. https://doi.org/10.4269/ajtmh.1997.57.519

Kilbourne, E. D. (2006). Influenza Pandemics of the 20th Century. *Emerging Infectious Diseases*, *12*(1), 9–14. https://doi.org/10.3201/eid1201.051254

Kimura, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences of the United States of America*, *54*(3), 731–736.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Kishino, H., Thorne, J. L., & Bruno, W. J. (2001). Performance of a Divergence Time Estimation Method under a Probabilistic Model of Rate Evolution. *Molecular Biology and Evolution*, *18*(3), 3. https://doi.org/10.1093/oxfordjournals.molbev.a003811

Kivistö, I., Tidenberg, E.-M., Lilley, T., Suominen, K., Forbes, K. M., Vapalahti, O., Huovilainen, A., & Sironen, T. (2020). First Report of Coronaviruses in Northern European Bats. *Vector-Borne and Zoonotic Diseases*, *20*(2), 155–158. https://doi.org/10.1089/vbz.2018.2367

Kizito, S., Okello, P. E., Kwesiga, B., Nyakarahuka, L., Balinandi, S., Mulei, S., Kyondo, J., Tumusiime, A., Lutwama, J., Ario, A. R., Ojwang, J., & Zhu, B.-P. (2018). Notes from the Field: Crimean-Congo Hemorrhagic Fever Outbreak — Central Uganda, August–September 2017. *Morbidity and Mortality Weekly Report*, *67*(22), 22. https://doi.org/10.15585/mmwr.mm6722a6

Kleib, A. S., Salihy, S. M., Ghaber, S. M., Sidiel, B. W., Sidiya, K. C., & Bettar, E. S. (2016). Crimean-Congo Hemorrhagic Fever with Acute Subdural Hematoma, Mauritania, 2012. *Emerging Infectious Diseases*, *22*(7), 1305–1306. https://doi.org/10.3201/eid2207.151782

Klempa, B. (2018). Reassortment events in the evolution of hantaviruses. *Virus Genes*, *54*(5), 638–646. https://doi.org/10.1007/s11262-018-1590-z

Koch, L. K., Cunze, S., Kochmann, J., & Klimpel, S. (2020). Bats as putative Zaire ebolavirus reservoir hosts and their habitat suitability in Africa. *Scientific Reports*, *10*(1), 1. https://doi.org/10.1038/s41598-020-71226-0

Koonin, E. V., Krupovic, M., & Agol, V. I. (2021). The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiology and Molecular Biology Reviews : MMBR*, *85*(3), e00053-21. https://doi.org/10.1128/MMBR.00053-21

Kubo, H., Yamada, Y. K., & Taguchi, F. (1994). Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *Journal of Virology*, *68*(9), 5403–5410. https://doi.org/10.1128/JVI.68.9.5403-5410.1994

Kubyshkin, V., & Budisa, N. (2019). The Alanine World Model for the Development of the Amino Acid Repertoire in Protein Biosynthesis. *International Journal of Molecular Sciences*, *20*(21), 5507. https://doi.org/10.3390/ijms20215507

Kumar, P., Kumar, A., Garg, N., & Giri, R. (2023). An insight into SARS-CoV-2 membrane protein interaction with spike, envelope, and nucleocapsid proteins. *Journal of Biomolecular Structure and Dynamics*. https://www.tandfonline.com/doi/full/10.1080/07391102.2021.2016490

Kuo, L., & Masters, P. S. (2003). The Small Envelope Protein E Is Not Essential for Murine Coronavirus Replication. *Journal of Virology*, *77*(8), 4597–4608. https://doi.org/10.1128/JVI.77.8.4597-4608.2003

Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., & Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication–transcription complex. *Virology*, *484*, 313–322. https://doi.org/10.1016/j.virol.2015.06.016

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 7807. https://doi.org/10.1038/s41586-020-2180-5

Lanave, C., Preparata, G., Saccone, C., & Serio, G. (1984). A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, *20*(1), 86–93. https://doi.org/10.1007/BF02101990

Laskowski, R., Macarthur, M. W., Moss, D. S., & Thornton, J. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*, 283–291. https://doi.org/10.1107/S0021889892009944

Lau, S. K. P., Li, K. S. M., Huang, Y., Shek, C.-T., Tse, H., Wang, M., Choi, G. K. Y., Xu, H., Lam, C. S. F., Guo, R., Chan, K.-H., Zheng, B.-J., Woo, P. C. Y., & Yuen, K.-Y. (2010). Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *Journal of Virology*, *84*(6), 6. https://doi.org/10.1128/JVI.02219-09

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Tsoi, H.-W., Wong, B. H. L., Wong, S. S. Y., Leung, S.-Y., Chan, K.-H., & Yuen, K.-Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 39. https://doi.org/10.1073/pnas.0506735102

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Wang, M., Lam, C. S. F., Xu, H., Guo, R., Chan, K.-H., Zheng, B.-J., & Yuen, K.-Y. (2007). Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology*, *367*(2), 2. https://doi.org/10.1016/j.virol.2007.06.009

Lauber, C., Goeman, J. J., Parquet, M. del C., Thi Nga, P., Snijder, E. J., Morita, K., & Gorbalenya, A. E. (2013). The Footprint of Genome Architecture in the Largest Genome Expansion in RNA Viruses. *PLoS Pathogens*, *9*(7), e1003500. https://doi.org/10.1371/journal.ppat.1003500

Lauring, A. S., & Andino, R. (2010). Quasispecies Theory and the Behavior of RNA Viruses. *PLOS Pathogens*, *6*(7), e1001005. https://doi.org/10.1371/journal.ppat.1001005

Lazov, C. M., Chriél, M., Baagøe, H. J., Fjederholt, E., Deng, Y., Kooi, E. A., Belsham, G. J., Bøtner, A., & Rasmussen, T. B. (2018). Detection and Characterization of Distinct Alphacoronaviruses in Five Different Bat Species in Denmark. *Viruses*, *10*(9), 9. https://doi.org/10.3390/v10090486

Leblebicioglu, H., Eroglu, C., Erciyas-Yavuz, K., Hokelek, M., Acici, M., & Yilmaz, H. (2014). Role of Migratory Birds in Spreading Crimean-Congo Hemorrhagic Fever, Turkey. *Emerging Infectious Diseases*, *20*(8), 1331–1334. https://doi.org/10.3201/eid2008.131547

Lechevalier, H. (1972). Dmitri Iosifovich Ivanovski (1864-1920). *Bacteriological Reviews*, *36*(2), 135–145.

Lecis, R., Mucedda, M., Pidinchedda, E., Pittau, M., & Alberti, A. (2019). Molecular identification of Betacoronavirus in bats from Sardinia (Italy): First detection and phylogeny. *Virus Genes*, *55*(1), 1. https://doi.org/10.1007/s11262-018-1614-8

Lecoq, H. (2001). [Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898?]. *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*, *324*(10), 929–933. https://doi.org/10.1016/s0764-4469(01)01368-3

Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research*, *149*, 58–74. https://doi.org/10.1016/j.antiviral.2017.11.001

Lelli, D., Papetti, A., Sabelli, C., Rosti, E., Moreno, A., & Boniotti, M. B. (2013). Detection of Coronaviruses in Bats of Various Species in Italy. *Viruses*, *5*(11), 11. https://doi.org/10.3390/v5112679

Letko, M. (2024). *Functional assessment of cell entry and receptor use for merbecoviruses*. https://doi.org/10.1101/2024.03.13.584892

Letko, M., Marzi, A., & Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, *5*(4), 562–569. https://doi.org/10.1038/s41564-020-0688-y

Li, F. (2016a). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, *3*(1), 1. https://doi.org/10.1146/annurev-virology-110615-042301

Li, F. (2016b). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, *3*(1), 237–261. https://doi.org/10.1146/annurev-virology-110615-042301

Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005a). Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science*, *309*(5742), 5742. https://doi.org/10.1126/science.1116480

Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005b). Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science*, *309*(5742), 1864–1868.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 18. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 16. https://doi.org/10.1093/bioinformatics/btp352

Li, M., Ayyanathan, K., Dittmar, M., Miller, J., Tapescu, I., Lee, J. S., McGrath, M. E., Xue, Y., Vashee, S., Schultz, D. C., Frieman, M. B., & Cherry, S. (2023). SARS-CoV-2 ORF6 protein does not antagonize interferon signaling in respiratory epithelial Calu-3 cells during infection. *MBio*, *14*(4). https://doi.org/10.1128/mbio.01194-23

Li, N., Rao, G., Li, Z., Yin, J., Chong, T., Tian, K., Fu, Y., & Cao, S. (2022). Cryo-EM structure of glycoprotein C from Crimean-Congo hemorrhagic fever virus. *Virologica Sinica*, *37*(1), 127–137. https://doi.org/10.1016/j.virs.2022.01.015

Li, W. (2005). Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science*, *310*(5748), 5748. https://doi.org/10.1126/science.1118391

Liao, Y., Lescar, J., Tam, J. P., & Liu, D. X. (2004). Expression of SARS-coronavirus envelope protein in Escherichia coli cells alters membrane permeability. *Biochemical and Biophysical Research Communications*, *325*(1), 374–380. https://doi.org/10.1016/j.bbrc.2004.10.050

Liao, Y., Yuan, Q., Torres, J., Tam, J. P., & Liu, D. X. (2006). Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology*, *349*(2), 264–275. https://doi.org/10.1016/j.virol.2006.01.028

Lin, X.-D., Wang, W., Hao, Z.-Y., Wang, Z.-X., Guo, W.-P., Guan, X.-Q., Wang, M.-R., Wang, H.-W., Zhou, R.-H., Li, M.-H., Tang, G.-P., Wu, J., Holmes, E. C., & Zhang, Y.-Z. (2017). Extensive diversity of coronaviruses in bats from China. *Virology*, *507*, 1–10. https://doi.org/10.1016/j.virol.2017.03.019

Lindeborg, M., Barboutis, C., Ehrenborg, C., Fransson, T., Jaenson, T. G. T., Lindgren, P.-E., Lundkvist, Å., Nyström, F., Salaneck, E., Waldenström, J., & Olsen, B. (2012). Migratory Birds, Ticks, and Crimean-Congo Hemorrhagic Fever Virus. *Emerging Infectious Diseases*, *18*(12), 2095–2097. https://doi.org/10.3201/eid1812.120718

Lindgren, E., Tälleklint, L., & Polfeldt, T. (2000). Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick Ixodes ricinus. *Environmental Health Perspectives*, *108*(2), 119–123.

Liu, D. X., Fung, T. S., Chong, K. K.-L., Shukla, A., & Hilgenfeld, R. (2014). Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Research*, *109*, 97–109. https://doi.org/10.1016/j.antiviral.2014.06.013

Liu, P., Song, Y., Colijn, C., & MacPherson, A. (2022). The impact of sampling bias on viral phylogeographic reconstruction. *PLOS Global Public Health*, *2*(9), e0000577. https://doi.org/10.1371/journal.pgph.0000577

Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A., & Rocklöv, J. (2014). Vectorial Capacity of Aedes aegypti: Effects of Temperature and Implications for Global Dengue Epidemic Potential. *PLOS ONE*, *9*(3), e89783. https://doi.org/10.1371/journal.pone.0089783

Loeffler, F., & Frosch, P. (1897). *Summarischer Bericht uber die Ergebnisse der Untersuchungen zur Erforschung der Maul- und Klauenseuche*.

Lorenzo Juanes, H. M., Carbonell, C., Sendra, B. F., López-Bernus, A., Bahamonde, A., Orfao, A., Lista, C. V., Ledesma, M. S., Negredo, A. I., Rodríguez-Alonso, B., Bua, B. R., Sánchez-Seco, M. P., Muñoz Bellido, J. L., Muro, A., & Belhassen-García, M. (2023). Crimean-Congo Hemorrhagic Fever, Spain, 2013–2021. *Emerging Infectious Diseases*, *29*(2), 252–259. https://doi.org/10.3201/eid2902.220677

Lotfollahzadeh, S., Boroujeni, G. R. N., Dezfouli, M. R. M., & Bokaei, S. (2011). A Serosurvey of Crimean-Congo Haemorrhagic Fever Virus in Dairy Cattle in Iran. *Zoonoses and Public Health*, *58*(1), 1. https://doi.org/10.1111/j.1863-2378.2009.01269.x

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., Zhang, Y., Zhang, W., Yuan, Y., Bao, J., Zhang, B., Shi, Y., Yan, J., & Gao, G. F. (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature*, *500*(7461), 227–231. https://doi.org/10.1038/nature12328

Luis, A. D., Hayman, D. T. S., O'Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R. C., Mills, J. N., Timonin, M. E., Willis, C. K. R., Cunningham, A. A., Fooks, A. R., Rupprecht, C. E., Wood, J. L. N., & Webb, C. T. (2013). A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proceedings of the Royal Society B: Biological Sciences*, *280*(1756), 20122753. https://doi.org/10.1098/rspb.2012.2753

Lule, S. A., Gibb, R., Kizito, D., Nakanjako, G., Mutyaba, J., Balinandi, S., Owen, L., Jones, K. E., Abubakar, I., Lutwama, J. J., & Field, N. (2022). Widespread exposure to Crimean-Congo haemorrhagic fever in Uganda might be driven by transmission from Rhipicephalus ticks: Evidence from cross-sectional and modelling studies. *Journal of Infection*, *85*(6), 683–692. https://doi.org/10.1016/j.jinf.2022.09.016

Lunn, T. J., Peel, A. J., McCallum, H., Eby, P., Kessler, M. K., Plowright, R. K., & Restif, O. (2021). Spatial dynamics of pathogen transmission in communally roosting species: Impacts of changing habitats on bat-virus dynamics. *Journal of Animal Ecology*, *90*(11), 2609–2622. https://doi.org/10.1111/1365-2656.13566

Lwande, O. W., Irura, Z., Tigoi, C., Chepkorir, E., Orindi, B., Musila, L., Venter, M., Fischer, A., & Sang, R. (2012). Seroprevalence of Crimean Congo Hemorrhagic Fever Virus in Ijara District, Kenya. *Vector Borne and Zoonotic Diseases*, *12*(9), 9. https://doi.org/10.1089/vbz.2011.0914

Lwande, O. W., Thalin, T., de Jong, J., Sjödin, A., Näslund, J., Evander, M., & Ecke, F. (2022). Alphacoronavirus in a Daubenton's Myotis Bat (Myotis daubentonii) in Sweden. *Viruses*, *14*(3), 556. https://doi.org/10.3390/v14030556

Lysholm, S., Lindahl, J. F., Dautu, G., Johansson, E., Bergkvist, P. K., Munyeme, M., & Wensman, J. J. (2022). Seroepidemiology of selected transboundary animal diseases in goats in Zambia. *Preventive Veterinary Medicine*, *206*, 105708. https://doi.org/10.1016/j.prevetmed.2022.105708

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., & Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), 9436–9441. https://doi.org/10.1073/pnas.1508686112

Mackenzie, J. S., & Jeggo, M. (2019). The One Health Approach—Why Is It So Important? *Tropical Medicine and Infectious Disease*, *4*(2), 88. https://doi.org/10.3390/tropicalmed4020088

Macleod, J. M. L., Marmor, H., García-Sastre, A., & Frias-Staheli, N. (2015). Mapping of the interaction domains of the Crimean–Congo hemorrhagic fever virus nucleocapsid protein. *The Journal of General Virology*, *96*(Pt 3), Pt 3. https://doi.org/10.1099/vir.0.071332-0

Madan, V., García, M. de J., Sanz, M. A., & Carrasco, L. (2005). Viroporin activity of murine hepatitis virus E protein. *FEBS Letters*, *579*(17), 3607–3612. https://doi.org/10.1016/j.febslet.2005.05.046

Maeda, J., Repass, J. F., Maeda, A., & Makino, S. (2001). Membrane Topology of Coronavirus E Protein. *Virology*, *281*(2), 163–169. https://doi.org/10.1006/viro.2001.0818

Makwana, D., Yadav, P. D., Kelaiya, A., & Mourya, D. T. (2015). First confirmed case of Crimean-Congo haemorrhagic fever from Sirohi district in Rajasthan State, India. *The Indian Journal of Medical Research*, *142*(4), 489–491. https://doi.org/10.4103/0971-5916.169221

Ma-Lauer, Y., Carbajo-Lozoya, J., Hein, M. Y., Müller, M. A., Deng, W., Lei, J., Meyer, B., Kusov, Y., von Brunn, B., Bairad, D. R., Hünten, S., Drosten, C., Hermeking, H., Leonhardt, H., Mann, M., Hilgenfeld, R., & von Brunn, A. (2016). P53 down-regulates SARS coronavirus replication and is targeted by the SARS-unique domain and PLpro via E3 ubiquitin ligase RCHY1. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(35), E5192–E5201. https://doi.org/10.1073/pnas.1603435113

Mancuso, E., Toma, L., Polci, A., d'Alessio, S. G., Di Luca, M., Orsini, M., Di Domenico, M., Marcacci, M., Mancini, G., Spina, F., Goffredo, M., & Monaco, F. (2019). Crimean-Congo Hemorrhagic Fever Virus Genome in Tick from Migratory Bird, Italy. *Emerging Infectious Diseases*, *25*(7), 7. https://doi.org/10.3201/eid2507.181345

Mangombi, J. B., Roqueplo, C., Sambou, M., Dahmani, M., Mediannikov, O., Comtet, L., & Davoust, B. (2020). Seroprevalence of Crimean-Congo Hemorrhagic Fever in Domesticated Animals in Northwestern Senegal. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *20*(10), 10. https://doi.org/10.1089/vbz.2019.2592

Margoliash, E. (1963). PRIMARY STRUCTURE AND EVOLUTION OF CYTOCHROME C. *Proceedings of the National Academy of Sciences of the United States of America*, *50*(4), 672–679.

Mari Saez, A., Cherif Haidara, M., Camara, A., Kourouma, F., Sage, M., Magassouba, N., & Fichet-Calvet, E. (2018). Rodent control to fight Lassa fever: Evaluation and lessons learned from a 4-year study in Upper Guinea. *PLoS Neglected Tropical Diseases*, *12*(11), e0006829. https://doi.org/10.1371/journal.pntd.0006829

Marimoutou, C., Poizot-Martin, I., Loundou, A.-D., Cassuto, J.-P., Obadia, Y., & Groupe d'étude MANIF 2000. (2003). [Causes of hospitalization and death in the MANIF 2000 cohort, composed of HIV-infected patients contaminated by intravenous drug use, 1995-1999]. *Presse Medicale (Paris, France: 1983)*, *32*(13 Pt 1), 587–594.

Mariner, J. C., Morrill, J., & Ksiazek, T. G. (1995). Antibodies to hemorrhagic fever viruses in domestic livestock in Niger: Rift Valley fever and Crimean-Congo hemorrhagic fever. *The American Journal of Tropical Medicine and Hygiene*, *53*(3), 3. https://doi.org/10.4269/ajtmh.1995.53.217

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), vev003. https://doi.org/10.1093/ve/vev003

Martini, M., GAZZANIGA, V., BRAGAZZI, N. L., & BARBERIS, I. (2019). The Spanish Influenza Pandemic: A lesson from history 100 years after 1918. *Journal of Preventive Medicine and Hygiene*, *60*(1), 1. https://doi.org/10.15167/2421-4248/jpmh2019.60.1.1205

McArthur, D. B. (2019). Emerging Infectious Diseases. *The Nursing Clinics of North America*, *54*(2), 297–311. https://doi.org/10.1016/j.cnur.2019.02.006

McBride, R., van Zyl, M., & Fielding, B. C. (2014). The Coronavirus Nucleocapsid Is a Multifunctional Protein. *Viruses*, *6*(8), 2991–3018. https://doi.org/10.3390/v6082991

Menachery, V. D., Yount, B. L., Debbink, K., Agnihothram, S., Gralinski, L. E., Plante, J. A., Graham, R. L., Scobey, T., Ge, X.-Y., Donaldson, E. F., Randell, S. H., Lanzavecchia, A., Marasco, W. A., Shi, Z.-L., & Baric, R. S. (2015). A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine*, *21*(12), 1508–1513. https://doi.org/10.1038/nm.3985

Mhamadi, M., Badji, A., Dieng, I., Gaye, A., Ndiaye, E. H., Ndiaye, M., Mhamadi, M., Touré, C. T., Mbaye, M. R., Barry, M. A., Ndiaye, O., Faye, B., Ba, F. A., Diop, B., Ndiaye, M., Fall, M., Sagne, S. N., Fall, G., Loucoubar, C., … Faye, O. (2022). Crimean–Congo Hemorrhagic Fever Virus Survey in Humans, Ticks, and Livestock in Agnam (Northeastern Senegal) from February 2021 to March 2022. *Tropical Medicine and Infectious Disease*, *7*(10), 324. https://doi.org/10.3390/tropicalmed7100324

Millet, J. K., & Whittaker, G. R. (2018). Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology*, *517*, 3–8. https://doi.org/10.1016/j.virol.2017.12.015

Mills, J. N., Gage, K. L., & Khan, A. S. (2010). Potential Influence of Climate Change on Vector-Borne and Zoonotic Diseases: A Review and Proposed Research Plan. *Environmental Health Perspectives*, *118*(11), 1507. https://doi.org/10.1289/ehp.0901389

Mirza, M. U., Vanmeert, M., Froeyen, M., Ali, A., Rafique, S., & Idrees, M. (2019). In silico structural elucidation of RNA-dependent RNA polymerase towards the identification of potential Crimean-Congo Hemorrhagic Fever Virus inhibitors. *Scientific Reports*, *9*(1), 1. https://doi.org/10.1038/s41598-019-43129-2

Mishra, A. C., Mehta, M., Mourya, D. T., & Gandhi, S. (2011). Crimean-Congo haemorrhagic fever in India. *The Lancet*, *378*(9788), 372. https://doi.org/10.1016/S0140-6736(11)60680-6

Mohamed, M., Said, A. R., Murad, A., & Graham, R. (2008). A serological survey of Crimean-Congo haemorrhagic fever in the Sharkia Governorate of Egypt. *Vet Ital*, *44*, 513–517.

Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, *2*(1), 63–77. https://doi.org/10.1016/j.coviro.2011.12.004

Monchatre-Leroy, E., Boué, F., Boucher, J.-M., Renault, C., Moutou, F., Ar Gouilh, M., & Umhang, G. (2017). Identification of Alpha and Beta Coronavirus in

Wildlife Species in France: Bats, Rodents, Rabbits, and Hedgehogs. *Viruses*, *9*(12), 12. https://doi.org/10.3390/v9120364

Monsalve Arteaga, L., Muñoz Bellido, J. L., Negredo, A. I., García Criado, J., Vieira Lista, M. C., Sánchez Serrano, J. Á., Vicente Santiago, M. B., López Bernús, A., de Ory Manchón, F., Sánchez Seco, M. P., Leralta, N., Alonso Sardón, M., Muro, A., & Belhassen-García, M. (2021). New circulation of genotype V of Crimean-Congo haemorrhagic fever virus in humans from Spain. *PLoS Neglected Tropical Diseases*, *15*(2), 2. https://doi.org/10.1371/journal.pntd.0009197

Moraga-Fernández, A., Ruiz-Fons, F., Habela, M. A., Royo-Hernández, L., Calero-Bernal, R., Gortazar, C., de la Fuente, J., & Fernández de Mera, I. G. (2021). Detection of new Crimean–Congo haemorrhagic fever virus genotypes in ticks feeding on deer and wild boar, Spain. *Transboundary and Emerging Diseases*, *68*(3), 993–1000. https://doi.org/10.1111/tbed.13756

Moraga-Fernández, A., Sánchez-Sánchez, M., Queirós, J., Lopes, A. M., Vicente, J., Pardavila, X., Sereno-Cadierno, J., Alves, P. C., de la Fuente, J., & Fernández de Mera, I. G. (n.d.). A study of viral pathogens in bat species in the Iberian Peninsula: Identification of new coronavirus genetic variants. *International Journal of Veterinary Science and Medicine*, *10*(1), 100–110. https://doi.org/10.1080/23144599.2022.2139985

Morrill, J. C., Soliman, A. K., Imam, I. Z., Botros, B. A., Moussa, M. I., & Watts, D. M. (1990). Serological evidence of Crimean-Congo haemorrhagic fever viral infection among camels imported into Egypt. *The Journal of Tropical Medicine and Hygiene*, *93*(3), 201–204.

Mourya, D. T., Viswanathan, R., Jadhav, S. K., Yadav, P. D., Basu, A., & Chadha, M. S. (2017). Retrospective analysis of clinical information in Crimean-Congo haemorrhagic fever patients: 2014-2015, India. *The Indian Journal of Medical Research*, *145*(5), 673–678. https://doi.org/10.4103/ijmr.IJMR_65_16

Mourya, D. T., Yadav, P. D., Shete, A. M., Gurav, Y. K., Raut, C. G., Jadi, R. S., Pawar, S. D., Nichol, S. T., & Mishra, A. C. (2012). Detection, Isolation and Confirmation of Crimean-Congo Hemorrhagic Fever Virus in Human, Ticks and Animals in Ahmadabad, India, 2010–2011. *PLOS Neglected Tropical Diseases*, *6*(5), e1653. https://doi.org/10.1371/journal.pntd.0001653

Mourya, D. T., Yadav, P. D., Shete, A. M., Sathe, P. S., Sarkale, P. C., Pattnaik, B., Sharma, G., Upadhyay, K. J., Gosavi, S., Patil, D. Y., Chaubal, G. Y., Majumdar, T. D., & Katoch, V. M. (2015). Cross-sectional Serosurvey of Crimean-Congo Hemorrhagic Fever Virus IgG in Livestock, India, 2013–2014. *Emerging Infectious Diseases*, *21*(10), 1837–1839. https://doi.org/10.3201/eid2110.141961

Mourya, D. T., Yadav, P. D., Shete, A., Majumdar, T. D., Kanani, A., Kapadia, D., Chandra, V., Kachhiapatel, A. J., Joshi, P. T., Upadhyay, K. J., Dave, P., & Raval, D. (2014). Serosurvey of Crimean-Congo hemorrhagic fever virus in domestic animals, Gujarat, India, 2013. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *14*(9), 9. https://doi.org/10.1089/vbz.2014.1586

Msimang, V., Weyer, J., le Roux, C., Kemp, A., Burt, F. J., Tempia, S., Grobbelaar, A., Moolla, N., Rostal, M. K., Bagge, W., Cordel, C., Karesh, W. B.,

Paweska, J. T., & Thompson, P. N. (2021). Risk factors associated with exposure to Crimean-Congo haemorrhagic fever virus in animal workers and cattle, and molecular detection in ticks, South Africa. *PLoS Neglected Tropical Diseases*, *15*(5), e0009384. https://doi.org/10.1371/journal.pntd.0009384

Muianga, A. F., Watson, R., Varghese, A., Chongo, I. S., Ali, S., Monteiro, V., Inalda, F., Chelene, I., António, V., Hewson, R., & Gudo, E. S. (2017). First serological evidence of Crimean-Congo haemorrhagic fever in febrile patients in Mozambique. *International Journal of Infectious Diseases*, *62*, 119–123. https://doi.org/10.1016/j.ijid.2017.07.024

Mukherjee, S., Bhattacharyya, D., & Bhunia, A. (2020). Host-membrane interacting interface of the SARS coronavirus envelope protein: Immense functional potential of C-terminal domain. *Biophysical Chemistry*, *266*, 106452. https://doi.org/10.1016/j.bpc.2020.106452

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. L. K. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genetics*, *8*(7), e1002764. https://doi.org/10.1371/journal.pgen.1002764

Nabeth, P., Cheikh, D. O., Lo, B., Faye, O., Vall, I. O. M., Niang, M., Wague, B., Diop, D., Diallo, M., Diallo, B., Diop, O. M., & Simon, F. (2004). Crimean-Congo Hemorrhagic Fever, Mauritania. *Emerging Infectious Diseases*, *10*(12), 12. https://doi.org/10.3201/eid1012.040535

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*, *156*(1), 297–304.

Nagarajan, N., & Kingsford, C. (2011). GiRaF: Robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Research*, *39*(6), e34. https://doi.org/10.1093/nar/gkq1232

Nahata, K. D., Bielejec, F., Monetta, J., Dellicour, S., Rambaut, A., Suchard, M. A., Baele, G., & Lemey, P. (2022). SPREAD 4: Online visualisation of pathogen phylogeographic reconstructions. *Virus Evolution*, *8*(2), veac088. https://doi.org/10.1093/ve/veac088

Nakagawa, S., Sakaguchi, S., Ogura, A., Mineta, K., Endo, T., Suzuki, Y., & Gojobori, T. (2023). Current trends in RNA virus detection through metatranscriptome sequencing data. *FEBS Open Bio*, *13*(6), 992–1000. https://doi.org/10.1002/2211-5463.13626

Narain, P., & Chakraborty, R. (1987). Genetic differentiation of quantitative characters between populations or species. II: Optimal selection in infinite populations. *Heredity*, *59 (Pt 2)*, 199–212. https://doi.org/10.1038/hdy.1987.114

Nascimento, F. F., Reis, M. D., & Yang, Z. (2017). A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution*, *1*(10), 10. https://doi.org/10.1038/s41559-017-0280-x

Negredo, A., de la Calle-Prieto, F., Palencia-Herrejón, E., Mora-Rillo, M., Astray-Mochales, J., Sánchez-Seco, M. P., Bermejo Lopez, E., Menárguez, J., Fernández-Cruz, A., Sánchez-Artola, B., Keough-Delgado, E., Ramírez de Arellano, E., Lasala, F., Milla, J., Fraile, J. L., Ordobás Gavín, M., Martinez de la Gándara, A., López Perez, L., Diaz-Diaz, D., … Crimean Congo Hemorrhagic Fever@Madrid Working Group. (2017). Autochthonous Crimean-Congo Hemorrhagic Fever in Spain. *The New England Journal of Medicine*, *377*(2), 154–161. https://doi.org/10.1056/NEJMoa1615162

Negredo, A., Sánchez-Ledesma, M., Llorente, F., Pérez-Olmeda, M., Belhassen-García, M., González-Calle, D., Sánchez-Seco, M. P., & Jiménez-Clavero, M. Á. (2021). Retrospective Identification of Early Autochthonous Case of Crimean-Congo Hemorrhagic Fever, Spain, 2013. *Emerging Infectious Diseases*, *27*(6), 1754–1756. https://doi.org/10.3201/eid2706.204643

Nelson, M. I., & Vincent, A. L. (2015). Reverse zoonosis of influenza to swine: New perspectives on the human-animal interface. *Trends in Microbiology*, *23*(3), 142–153. https://doi.org/10.1016/j.tim.2014.12.002

Neri, U., Wolf, Y. I., Roux, S., Camargo, A. P., Lee, B., Kazlauskas, D., Chen, I. M., Ivanova, N., Zeigler Allen, L., Paez-Espino, D., Bryant, D. A., Bhaya, D., Narrowe, A. B., Probst, A. J., Sczyrba, A., Kohler, A., Séguin, A., Shade, A., Campbell, B. J., … Gophna, U. (2022). Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell*, *185*(21), 4023-4037.e18. https://doi.org/10.1016/j.cell.2022.08.023

Neuman, B. W. (2016). Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antiviral Research*, *135*, 97–107. https://doi.org/10.1016/j.antiviral.2016.10.005

Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., Liao, L., Klaus, J. P., Yates, J. R., Wuethrich, K., Stevens, R. C., Buchmeier, M. J., & Kuhn, P. (2008). Proteornics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *Journal of Virology*, *82*(11), 11.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nieto-Torres, J. L., DeDiego, M. L., Álvarez, E., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Llorente, M., Kremer, L., Shuo, S., & Enjuanes, L. (2011). Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology*, *415*(2), 69–82. https://doi.org/10.1016/j.virol.2011.03.029

Nurmakhanov, T., Sansyzbaev, Y., Atshabar, B., Deryabin, P., Kazakov, S., Zholshorinov, A., Matzhanova, A., Sadvakassova, A., Saylaubekuly, R., Kyraubaev, K., Hay, J., Atkinson, B., & Hewson, R. (2015). Crimean-Congo haemorrhagic fever virus in Kazakhstan (1948-2013). *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *38*, 19–23. https://doi.org/10.1016/j.ijid.2015.07.007

Nyakarahuka, L., Kyondo, J., Telford, C., Whitesell, A., Tumusiime, A., Mulei, S., Baluku, J., Cossaboom, C. M., Cannon, D. L., Montgomery, J. M., Lutwama, J.

J., Nichol, S. T., Balinandi, S. K., Klena, J. D., & Shoemaker, T. R. (2023). Seroepidemiological investigation of Crimean Congo hemorrhagic fever virus in livestock in Uganda, 2017. *PLOS ONE*, *18*(11), e0288587. https://doi.org/10.1371/journal.pone.0288587

Nyakarahuka, L., Whitmer, S., Kyondo, J., Mulei, S., Cossaboom, C. M., Telford, C. T., Tumusiime, A., Akurut, G. G., Namanya, D., Kamugisha, K., Baluku, J., Lutwama, J., Balinandi, S., Shoemaker, T., & Klena, J. D. (2022). Crimean-Congo Hemorrhagic Fever Outbreak in Refugee Settlement during COVID-19 Pandemic, Uganda, April 2021. *Emerging Infectious Diseases*, *28*(11), 2326–2329. https://doi.org/10.3201/eid2811.220365

Obanda, V., Agwanda, B., Blanco-Penedo, I., Mwangi, I. A., King'ori, E., Omondi, G. P., Ahlm, C., Evander, M., & Lwande, O. W. (2021). Livestock Presence Influences the Seroprevalence of Crimean Congo Hemorrhagic Fever Virus on Sympatric Wildlife in Kenya. *Vector-Borne and Zoonotic Diseases*, *21*(10), 10. https://doi.org/10.1089/vbz.2021.0024

O'Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M. J., Perini, F. A., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., … Cirranello, A. L. (2013). The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science*, *339*(6120), 662–667. https://doi.org/10.1126/science.1229237

Örd, M., Faustova, I., & Loog, M. (2020). The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Scientific Reports*, *10*. https://doi.org/10.1038/s41598-020-74101-0

Orłowska, A., Smreczak, M., Thor, K., Niedbalska, M., Pawelec, D., Trębas, P., & Rola, J. (2022). The Genetic Characterization of the First Detected Bat Coronaviruses in Poland Revealed SARS-Related Types and Alphacoronaviruses. *Viruses*, *14*(9), 1914. https://doi.org/10.3390/v14091914

Ortego, J., Ceriani, J. E., Patiño, C., Plana, J., & Enjuanes, L. (2007). Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology*, *368*(2), 296–308. https://doi.org/10.1016/j.virol.2007.05.032

O'Shea, T. J., Cryan, P. M., Cunningham, A. A., Fooks, A. R., Hayman, D. T. S., Luis, A. D., Peel, A. J., Plowright, R. K., & Wood, J. L. N. (2014). Bat Flight and Zoonotic Viruses. *Emerging Infectious Diseases*, *20*(5), 741–745. https://doi.org/10.3201/eid2005.130539

Ottis, K., Sidoli, L., Bachmann, P. A., Webster, R. G., & Kaplan, M. M. (1982). Human influenza A viruses in pigs: Isolation of a H3N2 strain antigenically related to A/England/42/72 and evidence for continuous circulation of human viruses in the pig population. *Archives of Virology*, *73*(2), 103–108. https://doi.org/10.1007/BF01314719

Oude Munnink, B. B., Nieuwenhuijse, D. F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M., Kamga, S. K., Schapendonk, C., Pronk, M., Lexmond, P., van der Linden, A., Bestebroer, T., Chestakova, I., Overmars, R. J., van Nieuwkoop,

S., Molenkamp, R., van der Eijk, A. A., GeurtsvanKessel, C., Vennema, H., … Dutch-Covid-19 response team. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Medicine*, *26*(9), 1405–1410. https://doi.org/10.1038/s41591-020-0997-y

Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., Bouwmeester-Vincken, N., Harders, F., Hakze-van der Honing, R., Wegdam-Blans, M. C. A., Bouwstra, R. J., GeurtsvanKessel, C., van der Eijk, A. A., Velkers, F. C., Smit, L. A. M., … Koopmans, M. P. G. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, *371*(6525), 172–177. https://doi.org/10.1126/science.abe5901

Palomar, A. M., Portillo, A., Santibáñez, P., Mazuelas, D., Arizaga, J., Crespo, A., Gutiérrez, Ó., Cuadrado, J. F., & Oteo, J. A. (2013). *Crimean-Congo Hemorrhagic Fever Virus in Ticks from Migratory Birds, Morocco—Volume 19, Number 2—February 2013—Emerging Infectious Diseases journal—CDC*. https://doi.org/10.3201/eid1902.121193

Pan, Q., Nguyen, T. B., Ascher, D. B., & Pires, D. E. V. (2022). Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures. *Briefings in Bioinformatics*, *23*(2), bbac025. https://doi.org/10.1093/bib/bbac025

Papa (Άννα Παπά), A., Marklewitz, M., Paraskevopoulou (Σοφία Παρασκευοπούλου), S., Garrison, A. R., Alkhovsky (Альховский Сергей Владимирович), S. V., Avšič-Županc, T., Bente, D. A., Bergeron, É., Burt, F., Di Paola, N., Ergünay, K., Hewson, R., Mirazimi, A., Sall, A. A., Spengler, J. R., Postler, T. S., Palacios, G., & Kuhn, J. H. (2022). History and classification of Aigai virus (formerly Crimean–Congo haemorrhagic fever virus genotype VI). *The Journal of General Virology*, *103*(4), 001734. https://doi.org/10.1099/jgv.0.001734

Paredes, M. I., Perofsky, A. C., Frisbie, L., Moncla, L. H., Roychoudhury, P., Xie, H., Mohamed Bakhash, S. A., Kong, K., Arnould, I., Nguyen, T. V., Wendm, S. T., Hajian, P., Ellis, S., Mathias, P. C., Greninger, A. L., Starita, L. M., Frazar, C. D., Ryke, E., Zhong, W., … Bedford, T. (2022). Local-Scale phylodynamics reveal differential community impact of SARS-CoV-2 in metropolitan US county. *MedRxiv*, 2022.12.15.22283536. https://doi.org/10.1101/2022.12.15.22283536

Parsons, H. (1891). *Report on the influenza epidemic of 1889-90*. Local Government Board. https://iiif.wellcomecollection.org/pdf/b20413361

Pauly, M., Pir, J. B., Loesch, C., Sausy, A., Snoeck, C. J., Hübschen, J. M., & Muller, C. P. (2017). Novel Alphacoronaviruses and Paramyxoviruses Cocirculate with Type 1 and Severe Acute Respiratory System (SARS)-Related Betacoronaviruses in Synanthropic Bats of Luxembourg. *Applied and Environmental Microbiology*, *83*(18), 18. https://doi.org/10.1128/AEM.01326-17

Payne, S. (2017). Family Coronaviridae. *Viruses*, 149–158. https://doi.org/10.1016/B978-0-12-803109-4.00017-9

Peck, K. M., & Lauring, A. S. (2018). Complexities of Viral Mutation Rates. *Journal of Virology*, *92*(14), 14. https://doi.org/10.1128/JVI.01031-17

Peiris, J., Lai, S., Poon, L., Guan, Y., Yam, L., Lim, W., Nicholls, J., Yee, W., Yan, W., Cheung, M., Cheng, V., Chan, K., Tsang, D., Yung, R., Ng, T., & Yuen, K. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet (London, England)*, *361*(9366), 1319–1325. https://doi.org/10.1016/S0140-6736(03)13077-2

Penedos, A. R., Myers, R., Hadef, B., Aladin, F., & Brown, K. E. (2015). Assessment of the Utility of Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks. *PloS One*, *10*(11), e0143081. https://doi.org/10.1371/journal.pone.0143081

Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., & Li, F. (2011). Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proceedings of the National Academy of Sciences*, *108*(26), 10696–10701. https://doi.org/10.1073/pnas.1104306108

Peng, G., Xu, L., Lin, Y.-L., Chen, L., Pasquarella, J. R., Holmes, K. V., & Li, F. (2012). Crystal Structure of Bovine Coronavirus Spike Protein Lectin Domain. *The Journal of Biological Chemistry*, *287*(50), 50. https://doi.org/10.1074/jbc.M112.418210

Peng, T., Lee, K., & Tarn, W. (2008). Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *The Febs Journal*, *275*(16), 4152–4163. https://doi.org/10.1111/j.1742-4658.2008.06564.x

Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, *30*, 296–307. https://doi.org/10.1016/j.meegid.2014.12.022

Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F. L., Yu, D., Vararattanavech, A., Soong, T. W., Liu, D. X., & Torres, J. (2009). Structure and Inhibition of the SARS Coronavirus Envelope Protein Ion Channel. *PLoS Pathogens*, *5*(7), e1000511. https://doi.org/10.1371/journal.ppat.1000511

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Phelps, K. L., Hamel, L., Alhmoud, N., Ali, S., Bilgin, R., Sidamonidze, K., Urushadze, L., Karesh, W., & Olival, K. J. (2019). Bat Research Networks and Viral Surveillance: Gaps and Opportunities in Western Asia. *Viruses*, *11*(3), 240. https://doi.org/10.3390/v11030240

Philippe, H., Casane, D., Gribaldo, S., Lopez, P., & Meunier, J. (2003). Heterotachy and functional shift in protein evolution. *IUBMB Life*, *55*(4–5), 257–265. https://doi.org/10.1080/1521654031000123330

Phonera, M. C., Simuunza, M. C., Kainga, H., Ndebe, J., Chembensofu, M., Chatanga, E., Kanyanda, S., Changula, K., Muleya, W., Mubemba, B., Chitanga, S., Kajihara, M., Sawa, H., Njunga, G., Takada, A., & Simulundu, E. (2021). Seroprevalence and Risk Factors of Crimean-Congo Hemorrhagic Fever in Cattle of Smallholder Farmers in Central Malawi. *Pathogens*, *10*(12), 12. https://doi.org/10.3390/pathogens10121613

Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., Perera, L., Hayne, C. K., Gordon, J., Stewart, Z. D., Sobhany, M., Deterding, L. J., Hsu, A. L., Dandey, V. P., Borgnia, M. J., & Stanley, R. E. (2021). Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nature Communications*, *12*. https://doi.org/10.1038/s41467-020-20608-z

Piotrowski, Y., Hansen, G., Boomaars-van der Zanden, A. L., Snijder, E. J., Gorbalenya, A. E., & Hilgenfeld, R. (2009). Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. *Protein Science : A Publication of the Protein Society*, *18*(1), 6–16. https://doi.org/10.1002/pro.15

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews. Microbiology*, *15*(8), 502–510. https://doi.org/10.1038/nrmicro.2017.45

Poon, L. L. M., Chu, D. K. W., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H. C., Lau, S. K. P., Woo, P. C. Y., Suen, K. Y., Yuen, K. Y., Guan, Y., & Peiris, J. S. M. (2005). Identification of a Novel Coronavirus in Bats. *Journal of Virology*, *79*(4), 4. https://doi.org/10.1128/JVI.79.4.2001-2009.2005

Popov, I., Ohlopkova, O., Donnik, I., Zolotukhin, P., Umanets, A., Golovin, S., Malinovkin, A., Belanova, A., Lipilkin, P., Lipilkina, T., Popov, I., Logvinov, A., Dubovitsky, N., Stolbunova, K., Sobolev, I., Alekseev, A., Shestopalov, A., Burkova, V., Chikindas, M., & Ermakov, A. (2023). Detection of coronaviruses in insectivorous bats of Fore-Caucasus, 2021. *Scientific Reports*, *13*. https://doi.org/10.1038/s41598-023-29099-6

Pramod, R. K., Nair, A. V., Tambare, P. K., Chauhan, K., Kumar, T. V., Rajan, R. A., Mani, B. M., Asaf, M., & Pandey, A. K. (2021). Reverse zoonosis of coronavirus disease-19: Present status and the control by one health approach. *Veterinary World*, *14*(10), 2817–2826. https://doi.org/10.14202/vetworld.2021.2817-2826

Putics, Á., Gorbalenya, A. E., & Ziebuhr, J. (2006). Identification of protease and ADP-ribose 1''-monophosphatase activities associated with transmissible gastroenteritis virus non-structural protein 3. *The Journal of General Virology*, *87*(Pt 3), 651–656. https://doi.org/10.1099/vir.0.81596-0

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., … Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes

directly from clinical samples. *Nature Protocols*, *12*(6), 1261–1276. https://doi.org/10.1038/nprot.2017.066

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. A., Dijkman, R., Muth, D., Demmers, J. A. A., Zaki, A., Fouchier, R. A. M., Thiel, V., Drosten, C., Rottier, P. J. M., Osterhaus, A. D. M. E., Bosch, B. J., & Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*, *495*(7440), 7440. https://doi.org/10.1038/nature12005

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 5. https://doi.org/10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, *2*(1), vew007. https://doi.org/10.1093/ve/vew007

Randolph, S. (2001). The Shifting Landscape of Tick-borne Zoonoses: Tick-borne Encephalitis and Lyme borreliosis in Europe. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*, 1045–1056. https://doi.org/10.1098/rstb.2001.0893

Reed, W. (1901). *The Etiology of Yellow Fever*.

Ren, W., Qu, X., Li, W., Han, Z., Yu, M., Zhou, P., Zhang, S.-Y., Wang, L.-F., Deng, H., & Shi, Z. (2008). Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin. *Journal of Virology*, *82*(4), 4. https://doi.org/10.1128/JVI.01085-07

Reusken, C. B. E. M., Lina, P. H. C., Pielaat, A., de Vries, A., Dam-Deisz, C., Adema, J., Drexler, J. F., Drosten, C., & Kooi, E. A. (2010). Circulation of Group 2 Coronaviruses in a Bat Species Common to Urban Areas in Western Europe. *Vector-Borne and Zoonotic Diseases*, *10*(8), 8. https://doi.org/10.1089/vbz.2009.0173

Richman, D. D., Murphy, B. R., Belshe, R. B., Rusten, H. M., Chanock, R. M., Blacklow, N. R., Parrino, T. A., Rose, F. B., Levine, M. M., & Caplan, E. (1977). Temperature-sensitive mutants of influenza A virus. XIV. Production and evaluation of influenza A/Georgia/74-ts-1[E] recombinant viruses in human adults. *The Journal of Infectious Diseases*, *136*(2), 256–262. https://doi.org/10.1093/infdis/136.2.256

Rihtaric, D., Hostnik, P., Steyer, A., Grom, J., & Toplak, I. (2010). Identification of SARS-like coronaviruses in horseshoe bats (Rhinolophus hipposideros) in Slovenia. *Archives of Virology*, *155*(4), 4. https://doi.org/10.1007/s00705-010-0612-5

Rizzo, F., Edenborough, K. M., Toffoli, R., Culasso, P., Zoppi, S., Dondo, A., Robetto, S., Rosati, S., Lander, A., Kurth, A., Orusa, R., Bertolotti, L., & Mandola, M. L. (2017). Coronavirus and paramyxovirus in bats from Northwest Italy. *BMC Veterinary Research*, *13*, 396. https://doi.org/10.1186/s12917-017-1307-x

Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., & Ng, W.-L. (2020). Coronavirus RNA Proofreading: Molecular Basis and

Therapeutic Targeting. *Molecular Cell*, *79*(5), 710–727. https://doi.org/10.1016/j.molcel.2020.07.027

Rocklöv, J., & Dubrow, R. (2020). Climate change: An enduring challenge for vector-borne disease prevention and control. *Nature Immunology*, *21*(5), 479–483. https://doi.org/10.1038/s41590-020-0648-y

Rodriguez, L. L., Maupin, G. O., Ksiazek, T. G., Rollin, P. E., Khan, A. S., Schwarz, T. F., Lofts, R. S., Smith, J. F., Noor, A. M., Peters, C. J., & Nichol, S. T. (1997). Molecular Investigation of a Multisource Outbreak of Crimeancongo Hemorrhagic Fever in the United Arab Emirates. *The American Journal of Tropical Medicine and Hygiene*, *57*(5), 5. https://doi.org/10.4269/ajtmh.1997.57.512

Roelle, S. M., Shukla, N., Pham, A. T., Bruchez, A. M., & Matreyek, K. A. (2021). *Expanded ACE2 dependencies of diverse SARS-like coronavirus receptor binding domains* (p. 2021.12.25.474149). bioRxiv. https://doi.org/10.1101/2021.12.25.474149

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., Kim, Y., Jedrzejczak, R., Maltseva, N. I., Endres, M., Jaroszewski, L., Godzik, A., Joachimiak, A., & Satchell, K. J. F. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *BioRxiv*, 2020.04.17.047498. https://doi.org/10.1101/2020.04.17.047498

Ruch, T. R., & Machamer, C. E. (2012). The Coronavirus E Protein: Assembly and Beyond. *Viruses*, *4*(3), 3. https://doi.org/10.3390/v4030363

Rugarabamu, S., Mwanyika, G. O., Rumisha, S. F., Sindato, C., Lim, H.-Y., Misinzo, G., & Mboera, L. E. G. (2021). Seroprevalence and associated risk factors of selected zoonotic viral hemorrhagic fevers in Tanzania. *International Journal of Infectious Diseases*, *109*, 174–181. https://doi.org/10.1016/j.ijid.2021.07.006

Russell, C. D., Lone, N. I., & Baillie, J. K. (2023). Comorbidities, multimorbidity and COVID-19. *Nature Medicine*, *29*(2), 334–343. https://doi.org/10.1038/s41591-022-02156-9

Russell, C. J. (2014). Acid-induced membrane fusion by the hemagglutinin protein and its role in influenza virus biology. *Current Topics in Microbiology and Immunology*, *385*, 93–116. https://doi.org/10.1007/82_2014_393

Sahay, R. R., Dhandore, S., Yadav, P. D., Chauhan, A., Bhatt, L., Garg, V., Gupta, N., Nyayanit, D. A., Shete, A. M., Singh, R., Patil, S., Sharma, H., Meena, D., Gupta, J., Gawande, P., Kalele, K., Jain, R., Mali, D., Malhotra, B., … Gopalakrishnan, M. (2020). Detection of African genotype in Hyalomma tick pools

during Crimean Congo hemorrhagic fever outbreak, Rajasthan, India, 2019. *Virus Research*, *286*, 198046. https://doi.org/10.1016/j.virusres.2020.198046

Sahay, R. R., Shete, A. M., Yadav, P. D., Patil, S., Majumdar, T., Jain, R., Nyayanit, D. A., Kaushal, H., Panjwani, S. J., Upadhyay, K. J., Varevadiya, C. L., Vora, A., Kanani, A., & Gangakhedkar, R. R. (2021). Sequential determination of viral load, humoral responses and phylogenetic analysis in fatal and non-fatal cases of Crimean-Congo hemorrhagic fever patients from Gujarat, India, 2019. *PLoS Neglected Tropical Diseases*, *15*(8), e0009718. https://doi.org/10.1371/journal.pntd.0009718

Saidi, S., Casals, J., & Faghih, M. A. (1975). Crimean Hemorrhagic Fever-Congo (CHF-C) Virus Antibodies in Man, and in Domestic and Small Mammals, in Iran. *The American Journal of Tropical Medicine and Hygiene*, *24*(2), 2. https://doi.org/10.4269/ajtmh.1975.24.353

Saitou, N., & Nei, M. (1986). The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *Journal of Molecular Evolution*, *24*(1–2), 189–204. https://doi.org/10.1007/BF02099966

Saluzzo, J. F., Aubry, P., McCormick, J., & Digoutte, J. P. (1985). Haemorrhagic fever caused by Crimean Congo Haemorrhagic Fever virus in Mauritania. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *79*(2), 2. https://doi.org/10.1016/0035-9203(85)90356-6

Sanchez, A. J., Vincent, M. J., Erickson, B. R., & Nichol, S. T. (2006). Crimean-Congo Hemorrhagic Fever Virus Glycoprotein Precursor Is Cleaved by Furin-Like and SKI-1 Proteases To Generate a Novel 38-Kilodalton Glycoprotein. *Journal of Virology*, *80*(1), 1. https://doi.org/10.1128/JVI.80.1.514-525.2006

Sánchez-Seco, M. P., Sierra, M. J., Estrada-Peña, A., Valcárcel, F., Molina, R., de Arellano, E. R., Olmeda, A. S., San Miguel, L. G., Jiménez, M., Romero, L. J., & Negredo, A. (2022). Widespread Detection of Multiple Strains of Crimean-Congo Hemorrhagic Fever Virus in Ticks, Spain. *Emerging Infectious Diseases*, *28*(2), 394–402. https://doi.org/10.3201/eid2802.211308

Sanderson, M. J. (1997). A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy. *Molecular Biology and Evolution*, *14*(12), 12. https://doi.org/10.1093/oxfordjournals.molbev.a025731

Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, *19*(1), 1. https://doi.org/10.1093/oxfordjournals.molbev.a003974

Sang, R., Lutomiah, J., Koka, H., Makio, A., Chepkorir, E., Ochieng, C., Yalwala, S., Mutisya, J., Musila, L., Richardson, J. H., Miller, B. R., & Schnabel, D. (2011). Crimean-Congo hemorrhagic fever virus in Hyalommid ticks, northeastern Kenya. *Emerging Infectious Diseases*, *17*(8), 1502–1505. https://doi.org/10.3201/eid1708.102064

Sanjuán, R., & Domingo-Calap, P. (2021). Genetic Diversity and Evolution of Viral Populations. *Encyclopedia of Virology*, 53–61. https://doi.org/10.1016/B978-0-12-809633-8.20958-8

Santerre, M., Arjona, S. P., Allen, C. N., Shcherbik, N., & Sawaya, B. E. (2021). Why do SARS-CoV-2 NSPs rush to the ER? *Journal of Neurology*, *268*(6), 2013–2022. https://doi.org/10.1007/s00415-020-10197-8

Sas, M. A., Mertens, M., Isselmou, E., Reimer, N., El Mamy, B. O., Doumbia, B., & Groschup, M. H. (2017). Crimean-Congo Hemorrhagic Fever Virus-Specific Antibody Detection in Cattle in Mauritania. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *17*(8), 8. https://doi.org/10.1089/vbz.2016.2084

Sas, M. A., Mertens, M., Kadiat, J. G., Schuster, I., Pongombo, C. P. S., Maloba, A. G. K., & Groschup, M. H. (2017). Serosurvey for Crimean-Congo hemorrhagic fever virus infections in ruminants in Katanga province, Democratic Republic of the Congo. *Ticks and Tick-Borne Diseases*, *8*(6), 858–861. https://doi.org/10.1016/j.ttbdis.2017.07.001

Saso, W., Yamasaki, M., Nakakita, S., Fukushi, S., Tsuchimoto, K., Watanabe, N., Sriwilaijaroen, N., Kanie, O., Muramatsu, M., Takahashi, Y., Matano, T., Takeda, M., Suzuki, Y., & Watashi, K. (2022). Significant role of host sialylated glycans in the infection and spread of severe acute respiratory syndrome coronavirus 2. *PLoS Pathogens*, *18*(6), e1010590. https://doi.org/10.1371/journal.ppat.1010590

Schäffr, J. R., Kawaoka, Y., Bean, W. J., Süss, J., Senne, D., & Webster, R. G. (1993). Origin of the Pandemic 1957 H2 Influenza A Virus and the Persistence of Its Possible Progenitors in the Avian Reservoir. *Virology*, *194*(2), 781–788. https://doi.org/10.1006/viro.1993.1319

Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: Current knowledge. *Virology Journal*, *16*(1), 69. https://doi.org/10.1186/s12985-019-1182-0

Schubert, A. F., Nguyen, J. V., Franklin, T. G., Geurink, P. P., Roberts, C. G., Sanderson, D. J., Miller, L. N., Ovaa, H., Hofmann, K., Pruneda, J. N., & Komander, D. (2020). Identification and characterization of diverse OTU deubiquitinases in bacteria. *The EMBO Journal*, *39*(15), e105127. https://doi.org/10.15252/embj.2020105127

Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., Leibundgut, M., Thiel, V., Mühlemann, O., & Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural & Molecular Biology*, *27*(10), 959–966. https://doi.org/10.1038/s41594-020-0511-8

Schulz, A., Barry, Y., Stoek, F., Ba, A., Schulz, J., Haki, M. L., Sas, M. A., Doumbia, B. A., Kirkland, P., Bah, M. Y., Eiden, M., & Groschup, M. H. (2021). Crimean-Congo hemorrhagic fever virus antibody prevalence in Mauritanian livestock (cattle, goats, sheep and camels) is stratified by the animal's age. *PLoS Neglected Tropical Diseases*, *15*(4), 4. https://doi.org/10.1371/journal.pntd.0009228

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13), 3381–3385.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seifert, S. N., Bai, S., Fawcett, S., Norton, E. B., Zwezdaryk, K. J., Robinson, J., Gunn, B., & Letko, M. (2022). An ACE2-dependent Sarbecovirus in Russian bats is resistant to SARS-CoV-2 vaccines. *PLOS Pathogens*, *18*(9), e1010828. https://doi.org/10.1371/journal.ppat.1010828

Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., Neuman, B. W., Subramanian, V., Saikatendu, K. S., Buchmeier, M. J., Stevens, R. C., Kuhn, P., & Wüthrich, K. (2007). Nuclear Magnetic Resonance Structure of the N-Terminal Domain of Nonstructural Protein 3 from the Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology*, *81*(21), 12049–12060. https://doi.org/10.1128/JVI.00969-07

Shahhosseini, N., Wong, G., Babuadze, G., Camp, J. V., Ergonul, O., Kobinger, G. P., Chinikar, S., & Nowotny, N. (2021a). Crimean-Congo Hemorrhagic Fever Virus in Asia, Africa and Europe. *Microorganisms*, *9*(9), 1907. https://doi.org/10.3390/microorganisms9091907

Shahhosseini, N., Wong, G., Babuadze, G., Camp, J. V., Ergonul, O., Kobinger, G. P., Chinikar, S., & Nowotny, N. (2021b). Crimean-Congo Hemorrhagic Fever Virus in Asia, Africa and Europe. *Microorganisms*, *9*(9), 9. https://doi.org/10.3390/microorganisms9091907

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 7807. https://doi.org/10.1038/s41586-020-2179-y

Shanmugam, J., Smirnova, S. E., & Chumakov, M. P. (1976). Presence of antibody to arboviruses of the Crimean Haemorrhagic Fever-Congo (CHF-Congo) group in human beings and domestic animals in India. *The Indian Journal of Medical Research*, *64*(10), 1403–1413.

Sharp, P. M., & Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1552), 2487–2494. https://doi.org/10.1098/rstb.2010.0031

Shepherd, A. J., Swanepoel, R., Leman, P. A., & Shepherd, S. P. (1987). Field and laboratory investigation of Crimean-Congo haemorrhagic fever virus (Nairovirus, family Bunyaviridae) infection in birds. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *81*(6), 6. https://doi.org/10.1016/0035-9203(87)90379-8

Shi, M., Hou, X., He, Y., Fang, P., Mei, S.-Q., Xu, Z., Wu, W.-C., Tian, J., Zhang, S., Zeng, Z.-Y., Gou, Q.-Y., Xin, G.-Y., Le, S.-J., Xia, Y., Zhou, Y.-L., Hui, F.-M., Pan, Y.-F., Eden, J.-S., Yang, Z., … Li, Z. (2023). *Artificial intelligence redefines RNA virus discover*. https://doi.org/10.21203/rs.3.rs-2832379/v1

Sicheritz-Pontén, T., & Andersson, S. G. E. (2001). A phylogenomic approach to microbial evolution. *Nucleic Acids Research*, *29*(2), 545–552.

Singh, K., Mehta, D., Dumka, S., Chauhan, A. S., & Kumar, S. (2023). Quasispecies Nature of RNA Viruses: Lessons from the Past. *Vaccines*, *11*(2), 2. https://doi.org/10.3390/vaccines11020308

Singh, P., Chhabra, M., Sharma, P., Jaiswal, R., Singh, G., Mittal, V., Rai, A., & Venkatesh, S. (2016). Molecular epidemiology of Crimean-Congo haemorrhagic fever virus in India. *Epidemiology and Infection*, *144*(16), 3422–3425. https://doi.org/10.1017/S0950268816001886

Sit, T. H., Brackman, C. J., Ip, S. M., Tam, K. W., Law, P. Y., To, E. M., Yu, V. Y., Sims, L. D., Tsang, D. N., Chu, D. K., Perera, R. A., Poon, L. L., & Peiris, M. (2020). Canine SARS-CoV-2 infection. *Nature*, *586*(7831), 776–778. https://doi.org/10.1038/s41586-020-2334-5

Smith, G. (1882). *The Geography of British India, Political & Physical*. https://books.google.co.uk/books?id=C20DAAAAQAAJ&printsec=frontcover&source=gbs_atb&redir_esc=y#v=onepage&q&f=false

Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S. M., Guan, Y., & Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, *459*(7250), 7250. https://doi.org/10.1038/nature08182

Sow, A., Faye, O., Ba, Y., Ba, H., Diallo, D., Faye, O., Loucoubar, C., Boushab, M., Barry, Y., Diallo, M., & Sall, A. A. (2014). Rift Valley fever outbreak, southern Mauritania, 2012. *Emerging Infectious Diseases*, *20*(2), 296–299. https://doi.org/10.3201/eid2002.131000

Spencer, K.-A., & Hiscox, J. A. (2006). Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region. *Febs Letters*, *580*(25), 5993–5998. https://doi.org/10.1016/j.febslet.2006.09.052

Spengler, J. R., Patel, J. R., Chakrabarti, A. K., Zivcec, M., García-Sastre, A., Spiropoulou, C. F., & Bergeron, É. (2015). RIG-I Mediates an Antiviral Response to Crimean-Congo Hemorrhagic Fever Virus. *Journal of Virology*, *89*(20), 20. https://doi.org/10.1128/JVI.01643-15

Speranskaya, A. S., Artiushin, I. V., Samoilov, A. E., Korneenko, E. V., Khabudaev, K. V., Ilina, E. N., Yusefovich, A. P., Safonova, M. V., Dolgova, A. S., Gladkikh, A. S., Dedkov, V. G., & Daszak, P. (2023). Identification and Genetic Characterization of MERS-Related Coronavirus Isolated from Nathusius' Pipistrelle (Pipistrellus nathusii) near Zvenigorod (Moscow Region, Russia). *International Journal of Environmental Research and Public Health*, *20*(4), 4. https://doi.org/10.3390/ijerph20043702

Spielman, S. J., Weaver, S., Shank, S. D., Magalis, B. R., Li, M., & Kosakovsky Pond, S. L. (2019). Evolution of Viral Genomes: Interplay Between Selection, Recombination, and Other Forces. *Methods in Molecular Biology (Clifton, N.J.)*, *1910*, 427–468. https://doi.org/10.1007/978-1-4939-9074-0_14

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.-D., & Rappuoli, R. (2003). SARS — beginning to understand a new virus. *Nature Reviews. Microbiology*, *1*(3), 209–218. https://doi.org/10.1038/nrmicro775

Starr, T. N., Zepeda, S. K., Walls, A. C., Greaney, A. J., Alkhovsky, S., Veesler, D., & Bloom, J. D. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, *603*(7903), 7903. https://doi.org/10.1038/s41586-022-04464-z

Steel, M., & Penny, D. (2000). Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. *Molecular Biology and Evolution*, *17*(6), 839–850. https://doi.org/10.1093/oxfordjournals.molbev.a026364

Sultankulova, K. T., Shynybekova, G. O., Kozhabergenov, N. S., Mukhami, N. N., Chervyakova, O. V., Burashev, Y. D., Zakarya, K. D., Nakhanov, A. K., Barakbayev, K. B., & Orynbayev, M. B. (2022). The Prevalence and Genetic Variants of the CCHF Virus Circulating among Ticks in the Southern Regions of Kazakhstan. *Pathogens*, *11*(8), 841. https://doi.org/10.3390/pathogens11080841

Sumilo, D., Asokliene, L., Bormane, A., Vasilenko, V., Golovljova, I., & Randolph, S. E. (2007). Climate Change Cannot Explain the Upsurge of Tick-Borne Encephalitis in the Baltics. *PLoS ONE*, *2*(6), e500. https://doi.org/10.1371/journal.pone.0000500

Suryadevara, N., Shrihari, S., Gilchuk, P., VanBlargan, L. A., Binshtein, E., Zost, S. J., Nargi, R. S., Sutton, R. E., Winkler, E. S., Chen, E. C., Fouch, M. E., Davidson, E., Doranz, B. J., Chen, R. E., Shi, P.-Y., Carnahan, R. H., Thackray, L. B., Diamond, M. S., & Crowe, J. E. (2021). Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell*, *184*(9), 2316-2331.e15. https://doi.org/10.1016/j.cell.2021.03.029

Svinti, V., Cotton, J. A., & McInerney, J. O. (2013). New approaches for unravelling reassortment pathways. *BMC Evolutionary Biology*, *13*(1), 1. https://doi.org/10.1186/1471-2148-13-1

Swanepoel, R., Shepherd, A. J., Leman, P. A., & Shepherd, S. P. (1985). Investigations following initial recognition of Crimean-Congo haemorrhagic fever in South Africa and the diagnosis of 2 further cases. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, *68*(9), 638–641.

Swanepoel, R., Shepherd, A. J., Leman, P. A., Shepherd, S. P., McGillivray, G. M., Erasmus, M. J., Searle, L. A., & Gill, D. E. (1987). Epidemiologic and clinical features of Crimean-Congo hemorrhagic fever in southern Africa. *The American Journal of Tropical Medicine and Hygiene*, *36*(1), 120–132. https://doi.org/10.4269/ajtmh.1987.36.120

Swanepoel, R., Struthers, J. K., Shepherd, A. J., McGillivray, G. M., Nel, M. J., & Jupp, P. G. (1983). Crimean-congo hemorrhagic fever in South Africa. *The American Journal of Tropical Medicine and Hygiene*, *32*(6), 1407–1415. https://doi.org/10.4269/ajtmh.1983.32.1407

Sy, K. T. L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *PLOS ONE*, *16*(4), e0249271. https://doi.org/10.1371/journal.pone.0249271

Sylvain, D., Hugo, T., François, C., & François, J. (2021). Tick-borne disease risks and livestock management: Farmer's knowledge and practices in a Corsican valley (France). *NJAS: Impact in Agricultural and Life Sciences*, *93*(1), 74–97. https://doi.org/10.1080/27685241.2021.1975977

Tall, A., Sall, A. A., Faye, O., Diatta, B., Sylla, R., Faye, J., Faye, P. C., Faye, O., Ly, A. B., Sarr, F. D., Diab, H., & Diallo, M. (2009). [Two cases of Crimean-Congo haemorrhagic fever (CCHF) in two tourists in Senegal in 2004]. *Bulletin De La Societe De Pathologie Exotique (1990)*, *102*(3), 3.

Tan, C. C. S., Trew, J., Peacock, T. P., Mok, K. Y., Hart, C., Lau, K., Ni, D., Orme, C. D. L., Ransome, E., Pearse, W. D., Coleman, C. M., Bailey, D., Thakur, N., Quantrill, J. L., Sukhova, K., Richard, D., Kahane, L., Woodward, G., Bell, T., … Savolainen, V. (2023). Genomic screening of 16 UK native bat species through conservationist networks uncovers coronaviruses with zoonotic potential. *Nature Communications*, *14*(1), 1. https://doi.org/10.1038/s41467-023-38717-w

Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J. R., Schmidt, C. L., & Hilgenfeld, R. (2009). The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes. *PLoS Pathogens*, *5*(5), e1000428. https://doi.org/10.1371/journal.ppat.1000428

Tan, Y.-J., Lim, S. G., & Hong, W. (2005). Characterization of viral proteins encoded by the SARS-coronavirus genome. *Antiviral Research*, *65*(2), 69–78. https://doi.org/10.1016/j.antiviral.2004.10.001

Tang, X. C., Zhang, J. X., Zhang, S. Y., Wang, P., Fan, X. H., Li, L. F., Li, G., Dong, B. Q., Liu, W., Cheung, C. L., Xu, K. M., Song, W. J., Vijaykrishna, D., Poon, L. L. M., Peiris, J. S. M., Smith, G. J. D., Chen, H., & Guan, Y. (2006). Prevalence and Genetic Diversity of Coronaviruses in Bats from China. *Journal of Virology*, *80*(15), 15. https://doi.org/10.1128/JVI.00697-06

Tao, Y., & Tong, S. (2019). Complete Genome Sequence of a Severe Acute Respiratory Syndrome-Related Coronavirus from Kenyan Bats. *Microbiology Resource Announcements*, *8*(28), e00548-19. https://doi.org/10.1128/MRA.00548-19

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: The Mother of All Pandemics. *Emerging Infectious Diseases*, *12*(1), 15–22. https://doi.org/10.3201/eid1201.050979

Taylor, L. H., Latham, S. M., & Woolhouse, M. E. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1411), 1411. https://doi.org/10.1098/rstb.2001.0888

Tchesnokov, E. P., Bailey-Elkin, B. A., Mark, B. L., & Götte, M. (2020). Independent inhibition of the polymerase and deubiquitinase activities of the Crimean-Congo Hemorrhagic Fever Virus full-length L-protein. *PLOS Neglected*

Tropical Diseases, *14*(6), e0008283. https://doi.org/10.1371/journal.pntd.0008283

Templeton, A. R. (1998). Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology*, *7*(4), 381–397. https://doi.org/10.1046/j.1365-294x.1998.00308.x

Teoh, K.-T., Siu, Y.-L., Chan, W.-L., Schlüter, M. A., Liu, C.-J., Peiris, J. S. M., Bruzzone, R., Margolis, B., & Nal, B. (2010). The SARS Coronavirus E Protein Interacts with PALS1 and Alters Tight Junction Formation and Epithelial Morphogenesis. *Molecular Biology of the Cell*, *21*(22), 3838–3852. https://doi.org/10.1091/mbc.E10-04-0338

Thackray, L. B., Turner, B. C., & Holmes, K. V. (2005). Substitutions of conserved amino acids in the receptor-binding domain of the spike glycoprotein affect utilization of murine CEACAM1a by the murine coronavirus MHV-A59. *Virology*, *334*(1), 1. https://doi.org/10.1016/j.virol.2005.01.016

Thomas, S. (2020). The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. *Pathogens and Immunity*, *5*(1), 342–363. https://doi.org/10.20411/pai.v5i1.377

Tigoi, C., Lwande, O., Orindi, B., Irura, Z., Ongus, J., & Sang, R. (2015). Seroepidemiology of selected arboviruses in febrile patients visiting selected health facilities in the lake/river basin areas of Lake Baringo, Lake Naivasha, and Tana River, Kenya. *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, *15*(2), 2. https://doi.org/10.1089/vbz.2014.1686

Tomris, I., Unione, L., Nguyen, L., Zaree, P., Bouwman, K. M., Liu, L., Li, Z., Fok, J. A., Ríos Carrasco, M., van der Woude, R., Kimpel, A. L. M., Linthorst, M. W., Kilavuzoglu, S. E., Verpalen, E. C. J. M., Caniels, T. G., Sanders, R. W., Heesters, B. A., Pieters, R. J., Jiménez-Barbero, J., … de Vries, R. P. (2023). SARS-CoV-2 Spike N-Terminal Domain Engages 9-O-Acetylated α2–8-Linked Sialic Acids. *ACS Chemical Biology*, *18*(5), 1180–1191. https://doi.org/10.1021/acschembio.3c00066

Torres, J., Wang, J., Parthasarathy, K., & Liu, D. X. (2005). The Transmembrane Oligomers of Coronavirus Protein E. *Biophysical Journal*, *88*(2), 1283–1290. https://doi.org/10.1529/biophysj.104.051730

Tripathi, S., Bhati, R., Gopalakrishnan, M., Bohra, G. K., Tiwari, S., Panda, S., Sahay, R. R., Yadav, P. D., Nag, V. L., & Garg, M. K. (2020). Clinical profile and outcome of patients with Crimean Congo haemorrhagic fever: A hospital based observational study from Rajasthan, India. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *114*(9), 643–649. https://doi.org/10.1093/trstmh/traa014

Tsagkogeorga, G., Parker, J., Stupka, E., Cotton, J. A., & Rossiter, S. J. (2013). Phylogenomic Analyses Elucidate the Evolutionary Relationships of Bats. *Current Biology*, *23*(22), 2262–2267. https://doi.org/10.1016/j.cub.2013.09.014

Tsui, J. L.-H., McCrone, J. T., Lambert, B., Bajaj, S., Inward, R. P. D., Bosetti, P., Tegally, H., Hill, V., Pena, R. E., Zarebski, A. E., Peacock, T. P., Liu, L., Wu, N., Davis, M., Bogoch, I. I., Khan, K., Kall, M., Abdul Aziz, N. I. B., Colquhoun, R., …

Kraemer, M. U. G. (2023). Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. *Science (New York, N.Y.)*, *381*(6655), 336–343. https://doi.org/10.1126/science.adg6605

Urushadze, L., Babuadze, G., Shi, M., Escobar, L. E., Mauldin, M. R., Natradeze, I., Machablishvili, A., Kutateladze, T., Imnadze, P., Nakazawa, Y., & Velasco-Villa, A. (2021). A Cross Sectional Sampling Reveals Novel Coronaviruses in Bat Populations of Georgia. *Viruses*, *14*(1), 72. https://doi.org/10.3390/v14010072

Vaiente, M. A., & Scotch, M. (2020). Going back to the roots: Evaluating Bayesian phylogeographic models with discrete trait uncertainty. *Infection, Genetics and Evolution*, *85*, 104501. https://doi.org/10.1016/j.meegid.2020.104501

van Eeden, P. J., * van Eeden, S.F.,* Joubert, J.R.,* King, J.B.,* van de Wal, B.W.* & Michell W.L.**. (1985). A nosocomial outbreak of Crimean-Congo haemorrhagic fever at Tygerberg Hospital—Part II. Management of patients. *South African Medical Journal*, *68*(10), 718–721. https://doi.org/10.10520/AJA20785135_6302

van Seventer, J. M., & Hochberg, N. S. (2017). Principles of Infectious Diseases: Transmission, Diagnosis, Prevention, and Control. *International Encyclopedia of Public Health*, 22–39. https://doi.org/10.1016/B978-0-12-803678-5.00516-6

Vawda, S., Goedhals, D., Bester, P. A., & Burt, F. (2018). Seroepidemiologic Survey of Crimean-Congo Hemorrhagic Fever Virus in Selected Risk Groups, South Africa. *Emerging Infectious Diseases*, *24*(7), 1360–1363. https://doi.org/10.3201/eid2407.172096

Venkatagopalan, P., Daskalova, S. M., Lopez, L. A., Dolezal, K. A., & Hogue, B. G. (2015). Coronavirus envelope (E) protein remains at the site of assembly. *Virology*, *478*, 75–85. https://doi.org/10.1016/j.virol.2015.02.005

Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., & Rottier, P. J. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *The EMBO Journal*, *15*(8), 2020–2028. https://doi.org/10.1002/j.1460-2075.1996.tb00553.x

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, *439*(7074), 7074. https://doi.org/10.1038/nature04388

Vijaykrishna, D., Mukerji, R., & Smith, G. J. D. (2015). RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS Pathogens*, *11*(7), e1004902. https://doi.org/10.1371/journal.ppat.1004902

Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M., & Van Ranst, M. (2006). Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *Journal of Virology*, *80*(14), 7270–7274. https://doi.org/10.1128/JVI.02675-05

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2. *Nature Reviews Microbiology*, *19*(3), 3. https://doi.org/10.1038/s41579-020-00468-6

von Brunn, A., Teepe, C., Simpson, J. C., Pepperkok, R., Friedel, C. C., Zimmer, R., Roberts, R., Baric, R., & Haas, J. (2007). Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFeome. *PLoS ONE*, *2*(5), e459. https://doi.org/10.1371/journal.pone.0000459

Wang, L., Fu, S., Cao, Y., Zhang, H., Feng, Y., Yang, W., Nie, K., Ma, X., & Liang, G. (2017). Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China. *Emerging Microbes & Infections*, *6*(3), e14. https://doi.org/10.1038/emi.2016.140

Wang, M.-Y., Zhao, R., Gao, L.-J., Gao, X.-F., Wang, D.-P., & Cao, J.-M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fcimb.2020.587269

Wang, Y., Dutta, S., Karlberg, H., Devignot, S., Weber, F., Hao, Q., Tan, Y. J., Mirazimi, A., & Kotaka, M. (2012). Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: Superhelical homo-oligomers and the role of caspase-3 cleavage. *Journal of Virology*, *86*(22), 22. https://doi.org/10.1128/JVI.01627-12

Wang, Y., Kirkpatrick, J., Lage, S. zur, & Carlomagno, T. (2023). Structural insights into the activity regulation of full-length non-structural protein 1 from SARS-CoV-2. *Structure (London, England : 1993)*, *31*(2), 128-137.e5. https://doi.org/10.1016/j.str.2022.12.006

Wasfi, F., Dowall, S., Ghabbari, T., Bosworth, A., Chakroun, M., Varghese, A., Tiouiri, H., Ben Jemaa, M., Znazen, A., Hewson, R., Zhioua, E., & Letaief, A. (2016). Sero-epidemiological survey of Crimean-Congo hemorrhagic fever virus in Tunisia. *Parasite (Paris, France)*, *23*, 10. https://doi.org/10.1051/parasite/2016010

Welch, S. R., Scholte, F. E. M., Spengler, J. R., Ritter, J. M., Coleman-McCray, J. D., Harmon, J. R., Nichol, S. T., Zaki, S. R., Spiropoulou, C. F., & Bergeron, E. (2020). The Crimean-Congo Hemorrhagic Fever Virus NSm Protein Is Dispensable for Growth In Vitro and Disease in Ifnar-/- Mice. *Microorganisms*, *8*(5), 5. https://doi.org/10.3390/microorganisms8050775

Wells, H. L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J. A. K., Chandran, K., Munster, V. J., Gilardi, K., & Anthony, S. J. (2021). The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *Virus Evolution*, *7*(1), veab007. https://doi.org/10.1093/ve/veab007

White, J. M., Delos, S. E., Brecher, M., & Schornberg, K. (2008). Structures and Mechanisms of Viral Membrane Fusion Proteins. *Critical Reviews in Biochemistry and Molecular Biology*, *43*(3), 189–219. https://doi.org/10.1080/10409230802058320

Williams, R. K., Jiang, G. S., & Holmes, K. V. (1991). Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(13), 5533–5536.

Wilson, M. L., LeGuenno, B., Guillaud, M., Desoutter, D., Gonzalez, J. P., & Camicas, J. L. (1990). Distribution of Crimean-Congo hemorrhagic fever viral antibody in Senegal: Environmental and vectorial correlates. *The American Journal of Tropical Medicine and Hygiene*, *43*(5), 5. https://doi.org/10.4269/ajtmh.1990.43.557

Wittmann, W. (1999). The legacy of Friedrich Loeffler—The Institute on the Isle of Riems. *Archives of Virology. Supplementum*, *15*, 25–42. https://doi.org/10.1007/978-3-7091-6425-9_3

Wong, S. K., Li, W., Moore, M. J., Choe, H., & Farzan, M. (2004). A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *The Journal of Biological Chemistry*, *279*(5), 3197–3201. https://doi.org/10.1074/jbc.C300520200

Woo, P. C. Y., Lau, S. K. P., Chu, C., Chan, K., Tsoi, H., Huang, Y., Wong, B. H. L., Poon, R. W. S., Cai, J. J., Luk, W., Poon, L. L. M., Wong, S. S. Y., Guan, Y., Peiris, J. S. M., & Yuen, K. (2005). Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *Journal of Virology*, *79*(2), 884–895. https://doi.org/10.1128/JVI.79.2.884-895.2005

Woo, P. C. Y., Lau, S. K. P., Li, K. S. M., Poon, R. W. S., Wong, B. H. L., Tsoi, H., Yip, B. C. K., Huang, Y., Chan, K., & Yuen, K. (2006). Molecular diversity of coronaviruses in bats. *Virology*, *351*(1), 1. https://doi.org/10.1016/j.virol.2006.02.041

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 1. https://doi.org/10.1186/s13059-019-1891-0

Worobey, M., Telfer, P., Souquière, S., Hunter, M., Coleman, C. A., Metzger, M. J., Reed, P., Makuwa, M., Hearn, G., Honarvar, S., Roques, P., Apetrei, C., Kazanji, M., & Marx, P. A. (2010). Island Biogeography Reveals the Deep History of SIV. *Science*, *329*(5998), 5998. https://doi.org/10.1126/science.1193550

Wu, D., Wang, Z., Zhang, J., Robinson, A. G., Lyu, B., Chen, Z., Wang, C., Wei, B., Xia, X., Zhang, Q., & Zhou, X. (2022). Apoptotic caspase inhibits innate immune signaling by cleaving NF-κBs in both Mammals and Flies. *Cell Death & Disease*, *13*(8), 1–11. https://doi.org/10.1038/s41419-022-05156-2

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S., & Jin, Q. (2016). 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*, *10*(3), 3. https://doi.org/10.1038/ismej.2015.138

Xia, H., Beck, A. S., Gargili, A., Forrester, N., Barrett, A. D. T., & Bente, D. A. (2016). Transstadial Transmission and Long-term Association of Crimean-Congo

Hemorrhagic Fever Virus in Ticks Shapes Genome Plasticity. *Scientific Reports*, *6*, 35819. https://doi.org/10.1038/srep35819

Xia, S., Yan, L., Xu, W., Agrawal, A. S., Algaissi, A., Tseng, C.-T. K., Wang, Q., Du, L., Tan, W., Wilson, I. A., Jiang, S., Yang, B., & Lu, L. (2019). A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Science Advances*, *5*(4), eaav4580. https://doi.org/10.1126/sciadv.aav4580

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S., & Lu, L. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cellular and Molecular Immunology*, *17*(7), 765–767. https://doi.org/10.1038/s41423-020-0374-2

Xu, L., Zhang, F., Yang, W., Jiang, T., Lu, G., He, B., Li, X., Hu, T., Chen, G., Feng, Y., Zhang, Y., Fan, Q., Feng, J., Zhang, H., & Tu, C. (2016). Detection and characterization of diverse alpha- and betacoronaviruses from bats in China. *Virologica Sinica*, *31*(1), 1. https://doi.org/10.1007/s12250-016-3727-3

Xu, Z.-S., Du, W.-T., Wang, S.-Y., Wang, M.-Y., Yang, Y.-N., Li, Y.-H., Li, Z.-Q., Zhao, L.-X., Yang, Y., Luo, W.-W., & Wang, Y.-Y. (2024). LDLR is an entry receptor for Crimean-Congo hemorrhagic fever virus. *Cell Research*, *34*(2), 2. https://doi.org/10.1038/s41422-023-00917-w

Yadav, P. D., Cherian, S. S., Zawar, D., Kokate, P., Gunjikar, R., Jadhav, S., Mishra, A. C., & Mourya, D. T. (2013). Genetic characterization and molecular clock analyses of the Crimean-Congo hemorrhagic fever virus from human and ticks in India, 2010–2011. *Infection, Genetics and Evolution*, *14*, 223–231. https://doi.org/10.1016/j.meegid.2012.10.005

Yadav, P. D., Gurav, Y. K., Mistry, M., Shete, A. M., Sarkale, P., Deoshatwar, A. R., Unadkat, V. B., Kokate, P., Patil, D. Y., Raval, D. K., & Mourya, D. T. (2014). Emergence of Crimean-Congo hemorrhagic fever in Amreli District of Gujarat State, India, June to July 2013. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *18*, 97–100. https://doi.org/10.1016/j.ijid.2013.09.019

Yadav, P. D., Patil, D. Y., Shete, A. M., Kokate, P., Goyal, P., Jadhav, S., Sinha, S., Zawar, D., Sharma, S. K., Kapil, A., Sharma, D. K., Upadhyay, K. J., & Mourya, D. T. (2016). Nosocomial infection of CCHF among health care workers in Rajasthan, India. *BMC Infectious Diseases*, *16*, 624. https://doi.org/10.1186/s12879-016-1971-7

Yadav, P. D., Raut, C. G., & Mourya, D. T. (2013). Re-occurrence of Crimean-Congo haemorrhagic fever in Ahmedabad, Gujarat, India (2012): A fatal case report. *The Indian Journal of Medical Research*, *138*(6), 1027–1028.

Yadav, P. D., Raut, C. G., Patil, D. Y., D Majumdar, T., & Mourya, D. T. (2014). Crimean-Congo Hemorrhagic Fever: Current Scenario in India. *Proceedings of the National Academy of Sciences, India. Section B*, *84*(1), 9–18. https://doi.org/10.1007/s40011-013-0197-3

Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A., & Handu, S. (2021). Role of Structural and Non-Structural

Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*, *10*(4), 4. https://doi.org/10.3390/cells10040821

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, *367*(6485), 6485. https://doi.org/10.1126/science.abb2762

Yang, R., Zhao, Q., Rao, J., Zeng, F., Yuan, S., Ji, M., Sun, X., Li, J., Yang, J., Cui, J., Jin, Z., Liu, L., & Liu, Z. (2021). SARS-CoV-2 Accessory Protein ORF7b Mediates Tumor Necrosis Factor-α-Induced Apoptosis in Cells. *Frontiers in Microbiology*, *12*, 654709. https://doi.org/10.3389/fmicb.2021.654709

Yang, X.-L., Hu, B., Wang, B., Wang, M.-N., Zhang, Q., Zhang, W., Wu, L.-J., Ge, X.-Y., Zhang, Y.-Z., Daszak, P., Wang, L.-F., & Shi, Z.-L. (2016). Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology*, *90*(6), 6. https://doi.org/10.1128/JVI.02582-15

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution*, *10*(6), 1396–1401. https://doi.org/10.1093/oxfordjournals.molbev.a040082

Yang, Z., Goldman, N., & Friday, A. (1994). Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Molecular Biology and Evolution*, *11*(2), 316–324. https://doi.org/10.1093/oxfordjournals.molbev.a040112

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics*, *13*(5), 5. https://doi.org/10.1038/nrg3186

Yashina, L., Vyshemirskii, O., Seregin, S., Petrova, I., Samokhvalov, E., Lvov, D., Gutorov, V., Kuzina, I., Tyunnikov, G., Tang, Y.-W., Netesov, S., & Petrov, V. (2003). Genetic Analysis of Crimean-Congo Hemorrhagic Fever Virus in Russia. *Journal of Clinical Microbiology*, *41*(2), 2. https://doi.org/10.1128/JCM.41.2.860-862.2003

Yen, Y. C., Kong, L. X., Lee, L., Zhang, Y. Q., Li, F., Cai, B. J., & Gao, S. Y. (1985). Characteristics of Crimean-Congo hemorrhagic fever virus (Xinjiang strain) in China. *The American Journal of Tropical Medicine and Hygiene*, *34*(6), 6.

Yu, I.-M., Oldham, M. L., Zhang, J., & Chen, J. (2006). Crystal Structure of the Severe Acute Respiratory Syndrome (SARS) Coronavirus Nucleocapsid Protein Dimerization Domain Reveals Evolutionary Linkage between Corona- and Arteriviridae. *The Journal of Biological Chemistry*, *281*(25), 17134–17139. https://doi.org/10.1074/jbc.M602107200

Yuan, J., Hon, C.-C., Li, Y., Wang, D., Xu, G., Zhang, H., Zhou, P., Poon, L. L. M., Lam, T. T.-Y., Leung, F. C.-C., & Shi, Z. (2010). Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *Journal of General Virology*, *91*(4), 4. https://doi.org/10.1099/vir.0.016378-0

Yuan, Q., Liao, Y., Torres, J., Tam, J. P., & Liu, D. X. (2006). Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. *Febs Letters*, *580*(13), 3192–3200. https://doi.org/10.1016/j.febslet.2006.04.076

Zabiegala, A., Kim, Y., & Chang, K.-O. (2023). Roles of host proteases in the entry of SARS-CoV-2. *Animal Diseases*, *3*(1), 12. https://doi.org/10.1186/s44149-023-00075-x

Zakhashvili, K., Tsertsvadze, N., Chikviladze, T., Jghenti, E., Bekaia, M., Kuchuloria, T., Hepburn, M. J., Imnadze, P., & Nanuashvili, A. (2010). Crimean-Congo Hemorrhagic Fever in Man, Republic of Georgia, 2009. *Emerging Infectious Diseases*, *16*(8), 8. https://doi.org/10.3201/eid1608.100181

Zehender, G., Ebranati, E., Shkjezi, R., Papa, A., Luzzago, C., Gabanelli, E., Lo Presti, A., Lai, A., Rezza, G., Galli, M., Bino, S., & Ciccozzi, M. (2013). Bayesian Phylogeography of Crimean-Congo Hemorrhagic Fever Virus in Europe. *PLoS ONE*, *8*(11), 11. https://doi.org/10.1371/journal.pone.0079663

Zeller, H. G., Cornet, J. P., & Camicas, J. L. (1994). Crimean-Congo haemorrhagic fever virus infection in birds: Field investigations in Senegal. *Research in Virology*, *145*, 105–109. https://doi.org/10.1016/S0923-2516(07)80012-4

Zeller, H. G., Cornet, J. P., Diop, A., & Camicas, J. L. (1997). Crimean-Congo hemorrhagic fever in ticks (Acari: Ixodidae) and ruminants: field observations of an epizootic in Bandia, Senegal (1989-1992). *Journal of Medical Entomology*, *34*(5), 5. https://doi.org/10.1093/jmedent/34.5.511

Zhang, J., Ejikemeuwa, A., Gerzanich, V., Nasr, M., Tang, Q., Simard, J. M., & Zhao, R. Y. (2022). Understanding the Role of SARS-CoV-2 ORF3a in Viral Pathogenesis and COVID-19. *Frontiers in Microbiology*, *13*. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.854567

Zhang, Z., Nomura, N., Muramoto, Y., Ekimoto, T., Uemura, T., Liu, K., Yui, M., Kono, N., Aoki, J., Ikeguchi, M., Noda, T., Iwata, S., Ohto, U., & Shimizu, T. (2022). Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nature Communications*, *13*(1), 1. https://doi.org/10.1038/s41467-022-32019-3

Zhao, H., Lu, L., Peng, Z., Chen, L.-L., Meng, X., Zhang, C., Ip, J. D., Chan, W.-M., Chu, A. W.-H., Chan, K.-H., Jin, D.-Y., Chen, H., Yuen, K.-Y., & To, K. K.-W. (2022). SARS-CoV-2 Omicron variant shows less efficient replication and fusion activity when compared with Delta variant in TMPRSS2-expressed cells. *Emerging Microbes & Infections*, *11*(1), 277–283. https://doi.org/10.1080/22221751.2021.2023329

Zhou, P., Tachedjian, M., Wynne, J. W., Boyd, V., Cui, J., Smith, I., Cowled, C., Ng, J. H. J., Mok, L., Michalski, W. P., Mendenhall, I. H., Tachedjian, G., Wang, L.-F., & Baker, M. L. (2016). Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2696–2701. https://doi.org/10.1073/pnas.1518240113

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 7798. https://doi.org/10.1038/s41586-020-2012-7

Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Current Opinion in Microbiology*, *7*(4), 412–419. https://doi.org/10.1016/j.mib.2004.06.007

Zivcec, M., Scholte, F. E. M., Spiropoulou, C. F., Spengler, J. R., & Bergeron, É. (2016a). Molecular Insights into Crimean-Congo Hemorrhagic Fever Virus. *Viruses*, *8*(4), 4. https://doi.org/10.3390/v8040106

Zivcec, M., Scholte, F. E. M., Spiropoulou, C. F., Spengler, J. R., & Bergeron, É. (2016b). Molecular Insights into Crimean-Congo Hemorrhagic Fever Virus. *Viruses*, *8*(4), 4. https://doi.org/10.3390/v8040106

Zouaghi, K., Bouattour, A., Aounallah, H., Surtees, R., Krause, E., Michel, J., Mamlouk, A., Nitsche, A., & M'ghirbi, Y. (2021). First Serological Evidence of Crimean-Congo Hemorrhagic Fever Virus and Rift Valley Fever Virus in Ruminants in Tunisia. *Pathogens*, *10*(6), 6. https://doi.org/10.3390/pathogens10060769

Zuckerkandl, E., & Pauling, L. (1962). *Molecular Disease, Evolution, and Genic Heterogeneity*. Academic Press.

Zuckerkandl, E., & Pauling, L. (1965a). Evolutionary Divergence and Convergence in Proteins. In V. Bryson & H. J. Vogel (Eds.), *Evolving Genes and Proteins* (pp. 97–166). Academic Press. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6

Zuckerkandl, E., & Pauling, L. (1965b). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, *8*(2), 357–366. https://doi.org/10.1016/0022-5193(65)90083-4

Zumla, A., Hui, D. S., & Perlman, S. (2015). Middle East respiratory syndrome. *Lancet (London, England)*, *386*(9997), 995–1007. https://doi.org/10.1016/S0140-6736(15)60454-8

Zúñiga, S., Cruz, J. L. G., Sola, I., Mateos-Gómez, P. A., Palacio, L., & Enjuanes, L. (2010). Coronavirus Nucleocapsid Protein Facilitates Template Switching and Is Required for Efficient Transcription. *Journal of Virology*, *84*(4), 2169–2175. https://doi.org/10.1128/JVI.02011-09

Appendix

Sample	Host	Country	Ct	cDNA (ng/µL)
1001	Human	Turkey	23.98	1.02
1002	Human	Turkey	37.36	7.38
1003	Human	Turkey	30.6	1.96
1005	Human	Turkey	27.07	1.6
1009	Human	Turkey	29.11	0.532
1011	Human	Turkey	31.38	0.978
1012	Human	Turkey	25.42	1.51
1016	Human	Turkey	30.73	1.1
1018	Human	Turkey	32.9	0.382
1026	Human	Turkey	24.29	1.39
1027	Human	Turkey	21.43	18.5
2005	Human	Turkey	26.89	5.24
2009	Human	Turkey	31.27	12.7
2010	Human	Turkey	38.07	43
2011	Human	Turkey	30.04	76.6
2012	Human	Turkey	31.84	101
2013	Human	Turkey	33.21	54
2015	Human	Turkey	31.19	21.6
2018	Human	Turkey	37.07	106
2019	Human	Turkey	25.66	21.6
2020	Human	Turkey	31.61	106
2021	Human	Turkey	29.6	47
2022	Human	Turkey	31.81	69.4
2025	Human	Turkey	28.1	248
2028	Human	Turkey	25.15	143
2032	Human	Turkey	29.44	75
2033	Human	Turkey	31.4	19.3
2034	Human	Turkey	31.57	58.6
2037	Human	Turkey	24.89	73.4
2039	Human	Turkey	23.93	36.4
2043	Human	Turkey	29.05	62.6
2044	Human	Turkey	27.43	218

2047	Human	Turkey	26.75	14.8
2050	Human	Turkey	26.69	5.88
2052	Human	Turkey	32.54	62.4
2054	Human	Turkey	27.58	330
2056	Human	Turkey	32.26	19.4
2066	Human	Turkey	31.93	62.2
3007	Human	Turkey	21.67	100
3010	Human	Turkey	30.89	106
3014	Human	Turkey	29.94	77.2
3019	Human	Turkey	30.12	52.8
3020	Human	Turkey	24.93	212
3021	Human	Turkey	28.39	49.4
3022	Human	Turkey	21.44	159
3023	Human	Turkey	28.57	36.2
3024	Human	Turkey	37.73	50.4
3027	Human	Turkey	27.67	30.8
3033	Human	Turkey	30.29	29.4

Appendix 1. Sample ID, metadata, CT value and cDNA concentration from metagenomic amplification of CCHFv positive samples from Turkey.

Sample	Host	Country	Date	cDNA concentration
Kos2009	Human	Kosovo	2009	20.2
10248	NA	NA	NA	4.7
Dubai 616	Human	UAE	1979	15.6
UK1	Human	Afghanistan	2012	22.5
UK2	Human	Bulgaria	2014	42.4
3011	Human	Congo	1956	31.6
PYAT	Human	Russia	NA	20.1
JD206	Tick	Pakistan	1974	24.4
AR10200	Tick	Nigeria	NA	33.6
333.Rus	NA	Russia	NA	12.5
ARD15925	Tick	Senegal	1973	13.6
618	Tick	Russia	NA	11.5
SE7239	Human	Uganda	1971	13.5
AR7620	Goat	Nigeria	1974	21.6
Kos.1919	Human	Kosovo	2009	13.50
Kos.1921	Human	Kosovo	2009	15.30
Kos.1925	Human	Kosovo	2009	13.20

BH.2468	Human	Bosnia-Herzegovina	2009	6.50
Kos.2475	Human	Kosovo	2009	31.50
BH.2476	Human	Bosnia-Herzegovina	2009	5.20
K111.63	NA	Russia	NA	62.80
Khodzha	Human	Tajikistan	1975	89.20
SE7300	NA	NA	1967	16.30
Bagdhad12	Human	Iraq	1979	87.60
S3409	Human	Kosovo	2009	40.20
SPU4.81	Human	South Africa	1982	23.60
Semunya	NA	Uganda	NA	31.50
D1118	Human	Pakistan	1998	68.8
D1127	Human	Pakistan	1998	45.8
D1120	Human	Pakistan	1998	47
D1128	Human	Pakistan	1998	30.4
D1119	Human	Pakistan	1998	93
D1122	Human	Pakistan	1998	42.2
D1126	Human	Pakistan	1998	22.4
D1121	Human	Pakistan	1998	81.6

Appendix 2. Sample ID, metadata, and cDNA concentration from metagenomic amplification of archival CCHFV positive samples.

N	Sample id	cDNA (ng/µL)	Ct
1	80801	0.37	0
2	80802	0.46	0
3	80803	0.43	0
4	80804	<0.2	0
5	80805	1.49	0
6	80806	0.6	0
7	80807	97.6	0
8	80808	77.2	0
9	80809	18.8	0
10	80810	121	32.24643
11	80811	120	0
12	80812	79.2	0
13	170901	18.8	0
14	170902	121	0
15	80913	134	0
16	80914	126	0
17	80915	140	28.49175
18	80916	75.8	29.5338
19	80917	177	31.81352
20	80918	27.4	0
21	80919	102	0
22	80920	15.2	0
23	80901	81.2	0
24	80902	83.6	0

25	80903	120	0
26	80904	116	0
27	80905	91.12	0
28	80906	127	30.77917
29	80907	120	0
30	80908	108	0
31	80909	104	0
32	80910	151	0
33	80911	99.2	0
34	80912	0.362	0
35	80913	11.6	0
36	80914	92.8	0
37	80915	72.6	0
38	80916	91.2	0
39	80917	46.6	0
40	80918	176	0
41	80919	171	0
42	80920	77.4	0
43	80921	82.4	0
44	80922	102	0
45	80923	32.6	0
46	80924	83.6	32.40199
47	80925	68	0
48	80926	65.4	0
49	240801A	30.8	0
50	240802	1.47	0
51	240804	35.2	0
52	240805	114	29.15788
53	240803	40.6	0
NC	NA	2.7	0

Appendix 3. SISPA cDNA qubit concentrations and Ct value for 2020 faecal samples from Cinderford.

Sample	cDNA (ng/µL)	Ct
PLY1	4.8	0
PLY2	3.2	30.3551
PLY3	12.8	0
PLY4	8.64	0
PLY5	9.68	0
PLY6	9.68	0
PLY7	7.14	0
PLY8	8.82	0
PLY9	5.54	31.28791
PLY10	8.44	0
PLY11	20.2	0
PLY12	9.46	32.93995

PLY13	9.04	0
PLY14	29.06	0
PLY15	22.5	0
PLY16	19.6	0
PLY17	23.8	0
PLY18	9.44	0
PLY19	9.42	26.82015
PLY20	21.8	0
PLY21	20.8	0
PLY22	17.2	31.52971
PLY23	13.7	0
PLY24	20.2	28.33111
PLY25	9.14	0
PLY26	15.9	32.40744
PLY27	11.7	0
PLY28	18.9	0
PLY29	14.4	0
PLY30	36.6	29.70468
PLY31	24.4	0
PLY32	14.1	0
PLY33	7.52	0
PLY34	9.68	0
PLY35	29.4	0
PLY36	17.1	0
PLY37	23.4	0
PLY38	25.2	0
PLY39	17.8	0
PLY40	17.4	0
CI1_2023	60.6	29.03
CI2_2023	55.8	29.32
CI3_2023	68.6	26.74
CI9_2023	53.8	27.99
CI17_2023	50.2	26.08
CI18_2023	86.2	29.4
CI25_2023	69.6	21.5
CI26_2023	65.6	26.7
CI27_2023	92.8	28.7
CI31_2023	85	29.2
CI33_2023	76.8	27.6
CI34_2023	104	24.9
CI36_2023	68.8	0
CI37_2023	96.6	29.4
CI39_2023	77.2	28.8
CI44_2023	93.4	13.7
CI82_OCT_22	55	25.72
CI15_OCT_22	54	27.11

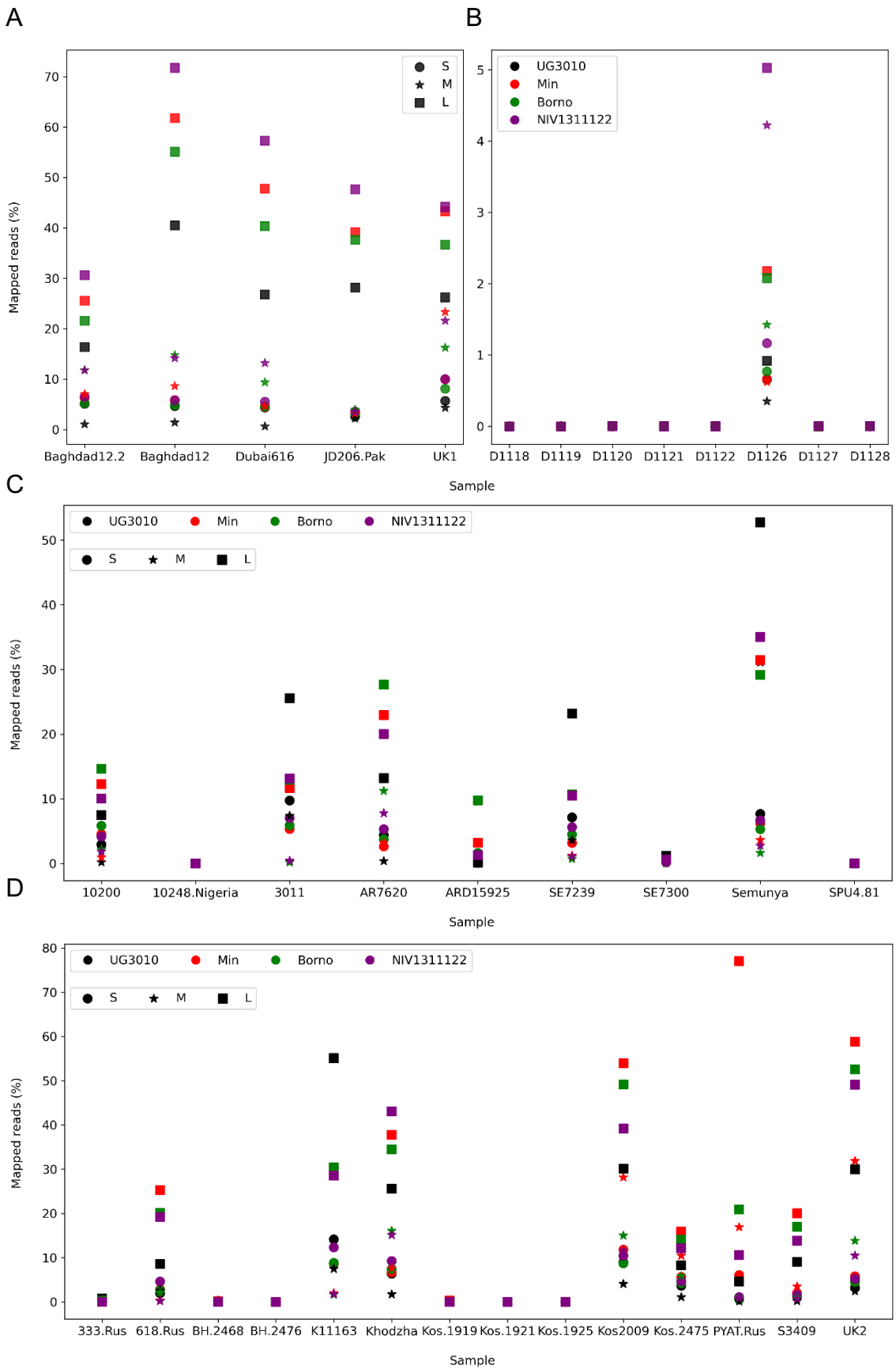
CI53_OCT_22	35	28.19
CI11_OCT_22	14	29.48
CI45_OCT_22	45	29.64
CI33_OCT_22	40	29.34
CI10_OCT_22	40	29.99
CI64_OCT_22	42	31.04
CI85_OCT_22	43	31.13
CI63_OCT_22	38	31.33
CI16_OCT_22	55	31.35
CI20_OCT_22	19	31.53
IM301	46	29.6
IM303	63	29.3
IM307	18.7	29.9
IM308	42	28.7
IM310	31	29.1
IM328	41	28.3
IM333	57.4	28.8
IM336	25.8	27.9
IM338	50	28.4
IM341	52	28.5
IM362	42	29.6

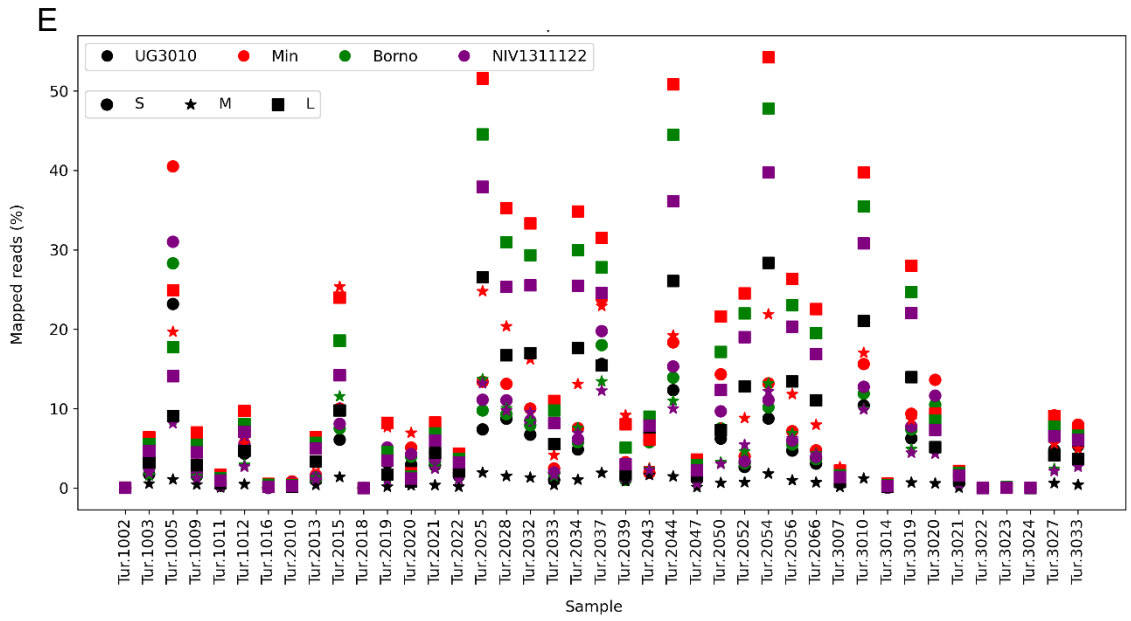
Appendix 4. Sample id, ct, cDNA concentration for sarbecovirus positive samples and faecal samples from Plymouth which underwent metagenomic NGS.

Reference	Start	End	Number	Pool
CI63	20	42	1	1
CI63	991	1013	1	1
CI63	877	899	2	2
CI63	1873	1893	2	2
CI63	1761	1783	3	1
CI63	2734	2757	3	1
CI63	2618	2640	4	2
CI63	3595	3617	4	2
CI63	3460	3482	5	1
CI63	4481	4503	5	1
CI63	4375	4397	6	2
CI63	5356	5378	6	2
CI63	5263	5285	7	1
CI63	6231	6253	7	1
CI63	6117	6141	8	2
CI63	7099	7121	8	2
CI63	7005	7027	9	1
CI63	7987	8011	9	1
CI63	7877	7899	10	2
CI63	8873	8894	10	2
CI63	8738	8759	11	1

CI63	9736	9758	11	1
CI63	9635	9657	12	2
CI63	10628	10650	12	2
CI63	10493	10515	13	1
CI63	11513	11536	13	1
CI63	11379	11401	14	2
CI63	12359	12383	14	2
CI63	12254	12280	15	1
CI63	13205	13227	15	1
CI63	13112	13134	16	2
CI63	14071	14093	16	2
CI63	13981	14003	17	1
CI63	14960	14985	17	1
CI63	14835	14859	18	2
CI63	15831	15853	18	2
CI63	15710	15734	19	1
CI63	16723	16746	19	1
CI63	16601	16623	20	2
CI63	17581	17603	20	2
CI63	17446	17469	21	1
CI63	18467	18489	21	1
CI63	18377	18399	22	2
CI63	19315	19337	22	2
CI63	19191	19213	23	1
CI63	20179	20201	23	1
CI63	20051	20074	24	2
CI63	21058	21080	24	2
CI63	20936	20959	25	1
CI63	21950	21972	25	1
CI63	21861	21883	26	2
CI63	22827	22849	26	2
CI63	22733	22755	27	1
CI63	23688	23711	27	1
CI63	23560	23582	28	2
CI63	24538	24560	28	2
CI63	24450	24472	29	1
CI63	25412	25434	29	1
CI63	25306	25328	30	2
CI63	26270	26294	30	2
CI63	26152	26174	31	1
CI63	27160	27183	31	1
CI63	27068	27090	32	2
CI63	28015	28037	32	2
CI63	27884	27906	33	1
CI63	28865	28888	33	1
CI63	28195	28217	34	2
CI63	29186	29208	34	2

Appendix 5. Primer pairs for the Sarbecovirus primer scheme.


A
B
C
D
Mapped reads (%)
Sample
S
M
L
UG3010
Min
Borno
NIV1311122
Baghdad12.2
Baghdad12
Dubai616
JD206.Pak
UK1
D1118
D1119
D1120
D1121
D1122
D1126
D1127
D1128
10200
10248.Nigeria
3011
AR7620
ARD15925
SE7239
SE7300
Semunya
SPU4.81
333.Rus
618.Rus
BH.2468
BH.2476
K11163
Khodzha
Kos.1919
Kos.1921
Kos.1925
Kos2009
Kos.2475
PYAT.Rus
S3409
UK2



Appendix 6. The percentage mapped reads for samples from Asia sequenced after tissue culture (A) or direct from sample (B), from Africa and sequenced after tissue culture (C), from Europe after tissue culture (D) or direct from sample (E). The colour depicts the reference sequence to which raw reads were mapped. The shape depicts the segment the raw reads were mapped to.

Country	Small	Medium	Large
Turkey	34	30	26
Kosovo	3	2	3
Uganda	2	0	2
Pakistan	1	1	1
Nigeria	1	0	0
Democratic Republic of the Congo	1	1	1
Tajikistan	1	0	0

Appendix 7. The number of full sequences generated per segment per country.

Country	Small	Medium	Large
India	68	59	58
Turkey	61	44	34
Russia	41	40	37
Uganda	21	15	19
China	20	17	10
Kosovo	18	16	15
South Africa	17	16	14
Iran	15	4	1
Oman	14	4	2
UAE	9	9	8
Pakistan	8	5	7
Tajikistan	6	2	2
Bulgaria	4	3	1
Senegal	4	4	4
Mali	1	0	0
Uzbekistan	3	2	1
Nigeria	3	2	2
Spain	3	3	3
Sudan	3	3	4
Afghanistan	2	2	2
Kazakhstan	2	2	1
Turkmenistan	2	2	2
Namibia	2	2	2
CAR	1	0	0
Iraq	1	0	1
Mauritania	1	1	1
Greece	1	0	0
Republic of the Congo	1	1	1
Democratic Republic of the Congo	1	1	1

Appendix 8. The total number of CCHFV sequences by country in the dataset.

Date	Small	Medium	Large
1956	3	3	3
1958	1	1	1
1960	1	1	1
1965	3	2	2
1966	2	2	1
1967	6	5	5
1968	5	5	4
1969	3	3	3
1970	1	1	0
1971	2	1	2
1972	1	1	1
1973	2	2	2
1975	4	2	1
1976	1	1	1

1978	2	1	0
1979	4	3	3
1981	5	5	4
1984	2	5	2
1985	4	3	3
1987	5	5	4
1988	4	3	3
1989	1	1	1
1990	4	3	3
1991	1	0	0
1995	1	1	1
1997	1	1	2
1998	3	4	2
2000	6	5	2
2001	3	2	1
2002	4	3	1
2003	1	1	2
2004	13	11	7
2005	4	1	1
2007	4	4	4
2008	6	5	6
2009	8	5	7
2010	3	2	1
2011	12	7	7
2012	13	8	7
2013	23	13	11
2014	12	8	8
2015	50	40	35
2016	12	11	10
2017	6	8	5
2018	13	9	12
2019	54	50	47
2021	11	1	1
2022	3	0	3

Appendix 9. Number of CCHFV sequences per year. Years are only shown if a sequence is present in a segment.

Host	Small	Medium	Large
Human	263	199	183
Tick	64	55	46
Goat	3	2	2
Sheep	1	1	1
Long Eared Jerboa	1	1	0

Camel	1	1	1

Appendix 10. The number of CCHFV sequences per host in the dataset.

M site	Prob [dS<dN]	Protein	dN/dS
341	0.989	GP38	4.80102
351	0.945	GP38	2.236111
432	0.971	GP38	2.488189
442	1	GP38	6.085366
1065	0.988	Gc	3.526316
1208	0.919	Gc	2.222222
1311	0.934	Gc	2.188482
1438	0.907	Gc	2.539568
1462	0.98	Gc	2.596639
1605	0.951	Gc	2.721154
1607	1	Gc	5.066667
533	0.996	Gn	4.121212
748	0.945	Gn	2.795699
42	0.913	MLD	2.43788
66	0.991	MLD	5.155009
72	0.935	MLD	3.492013
103	0.969	MLD	3.662568
129	0.98	MLD	2.877123
131	0.998	MLD	5.178891
185	0.997	MLD	4.802326
190	0.991	MLD	5.682081
214	1	MLD	9.975177
220	0.926	MLD	3.427711
828	0.949	NSm	2.659259
838	0.949	NSm	2.333333

Appendix 11. Codon sites under positive selection for the M segment predicted by FUBAR.

Site	p-value	Protein
1668	<0.0001	Gc
1316	<0.0001	Gc
1682	0.0001	Gc
1214	0.0003	Gc
1318	0.0003	Gc
1441	0.0004	Gc
1683	0.0006	Gc
1671	0.0007	Gc
1322	0.0008	Gc
1676	0.001	Gc
1684	0.0017	Gc
1662	0.0024	Gc
1664	0.0027	Gc
1237	0.003	Gc
1673	0.0035	Gc

1670	0.0062	Gc
1323	0.009	Gc
685	0.009	Gn
341	0.0053	GP38
111	0.0001	MLD
88	0.0005	MLD
214	0.0009	MLD
131	0.001	MLD
109	0.0014	MLD
113	0.0017	MLD
127	0.0025	MLD
95	0.003	MLD
89	0.0032	MLD
118	0.0052	MLD
962	0.0017	NSm

Appendix 12. Codon sites under episodic positive selection in the medium segment detected by MEME.

Contig ID	Length	Depth	Blastn match	% ID	Aln length
NODE.17	1610	193.9	BM48-31 (NC_014470.1)	88.66	1605
NODE.2	5319	11.4	BM48-31 (NC_014470.1)	72.755	2305
NODE.1	6039	12.2	BetaCoV Erinaceus (NC_039207.1)	70.52	3521
NODE.3	3978	12.8	SARS Tor2 (NC_00418.3)	75.946	2511
NODE.10	1839	21.9	SARS CoV 2 Wuhan (NC_045512.2)	77.926	1504
NODE.72	524	46.3	BM48-31 (NC_014470.1)	73.985	542

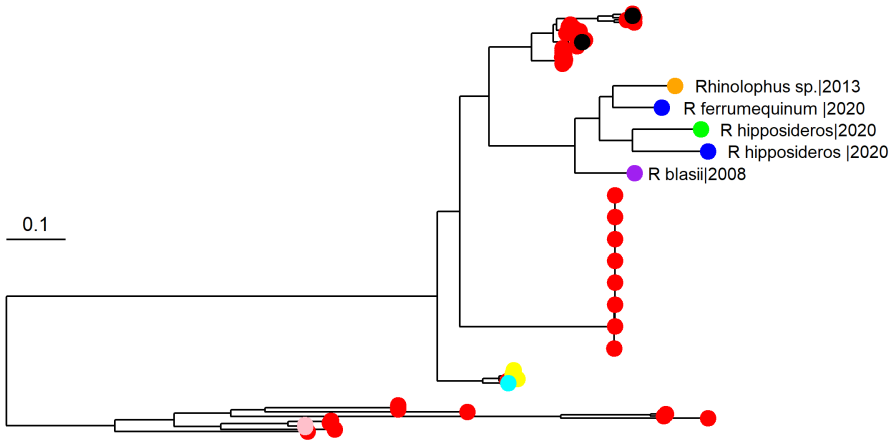
Appendix 13. Contigs generated from one round of illumina sequencing for Sample 46.

Contig ID	Length	Depth	Blastn match	% ID	Aln length
NODE.1	29290	25.6	BM48-31 (NC_014470.1)	80.748	25343
NODE.1	29290	25.6	SARS CoV 2 Wuhan (NC_045512.2)	79.747	17578
NODE.1	29290	25.6	SARS Tor2 (NC_00418.3)	80.86	17257
NODE.1	29290	25.6	Bat BetaCoV Zhejiang (NC_025217.1)	71.955	5099

Appendix 14. Blastn results for the one contig 29.2 kb in length generated from a hybrid de novo assembly.

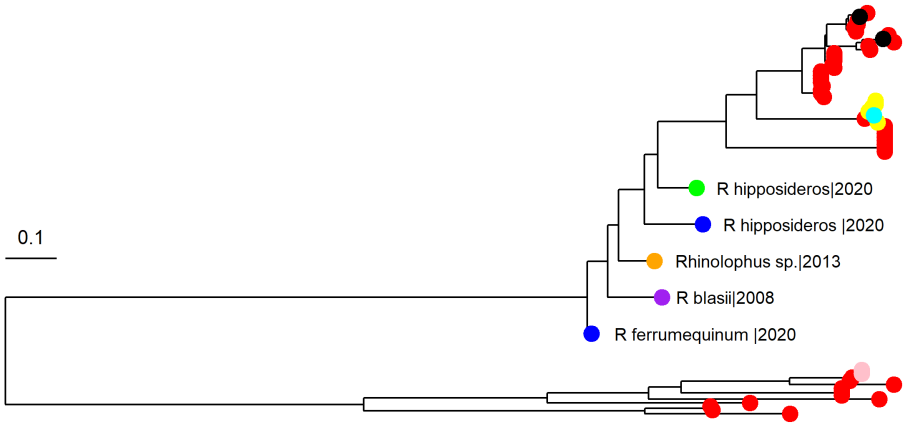
A

ORF3a



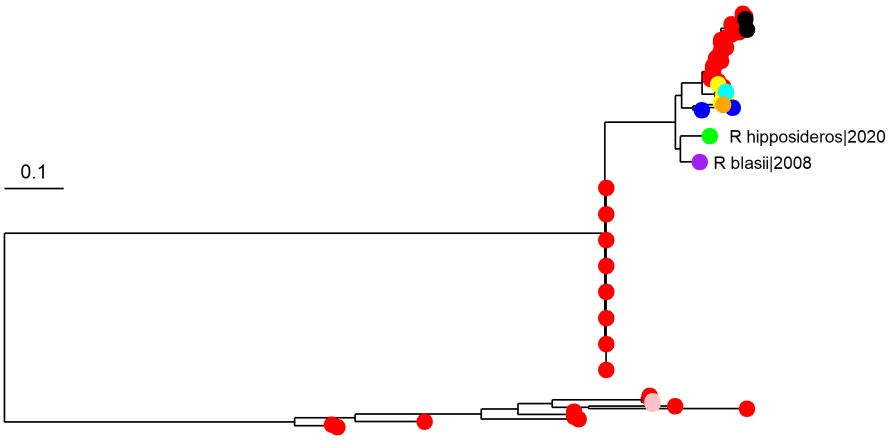
B

ORF3a

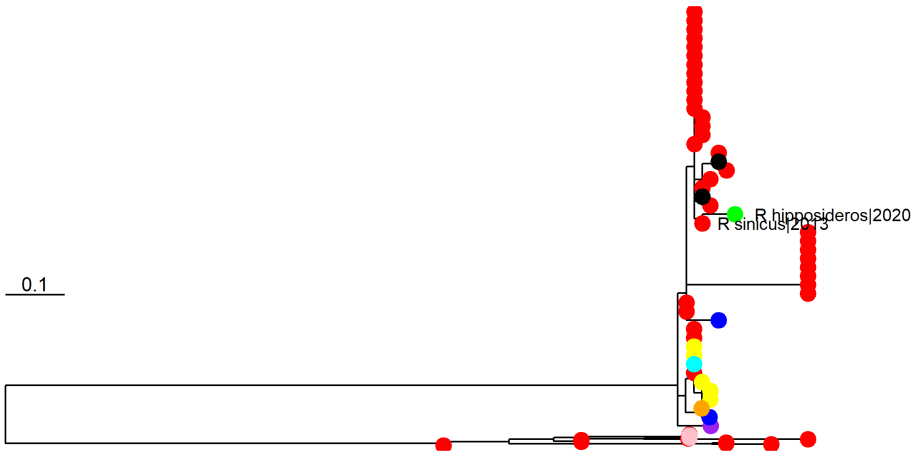


C

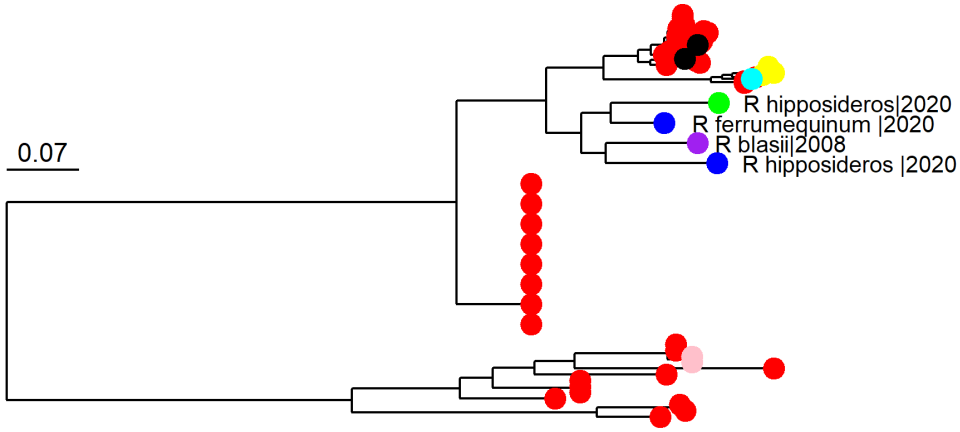
E



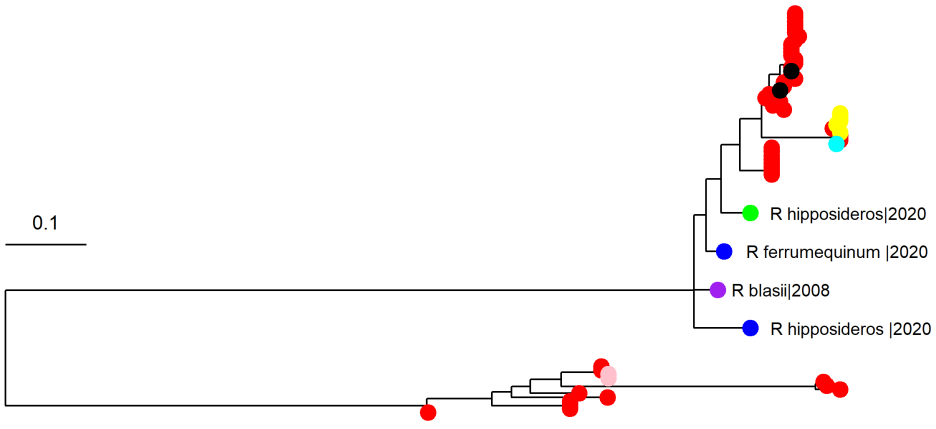
D


R hipposideros|2020
R sinicus|2013
0.1

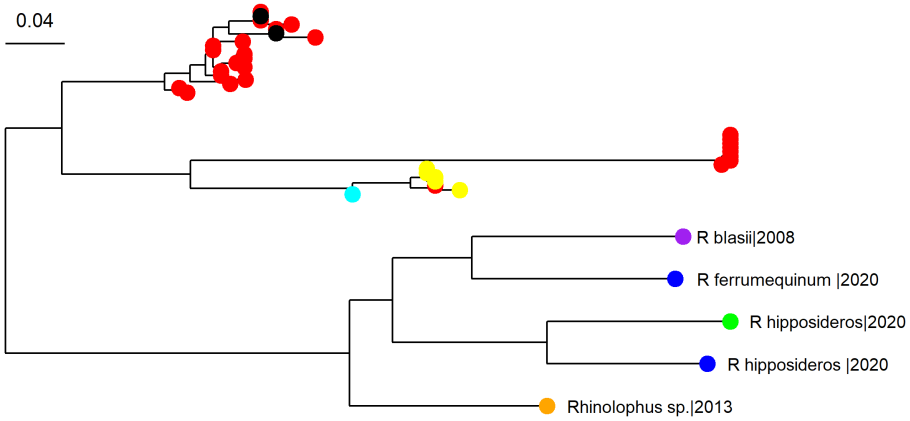
E


R hipposideros|2020
R ferrumequinum |2020
R blasii|2008
R hipposideros |2020
0.07

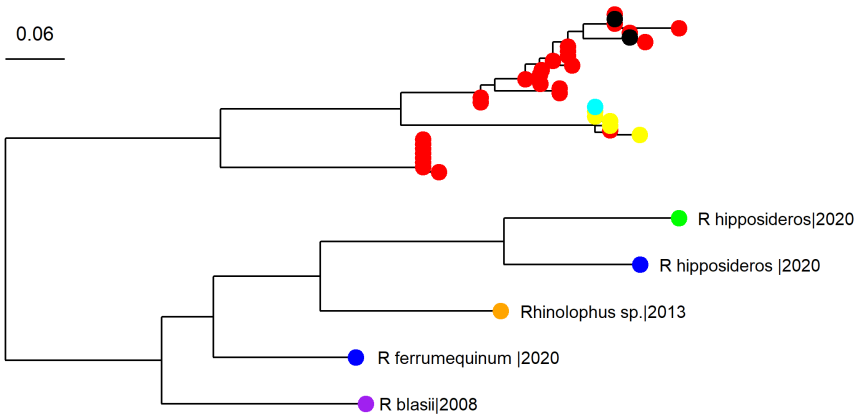
F


R hipposideros|2020
R ferrumequinum |2020
R blasii|2008
R hipposideros |2020
0.1

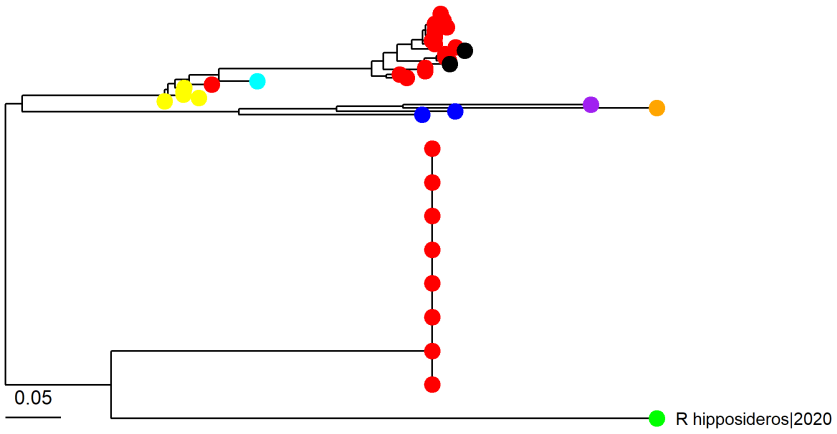
G

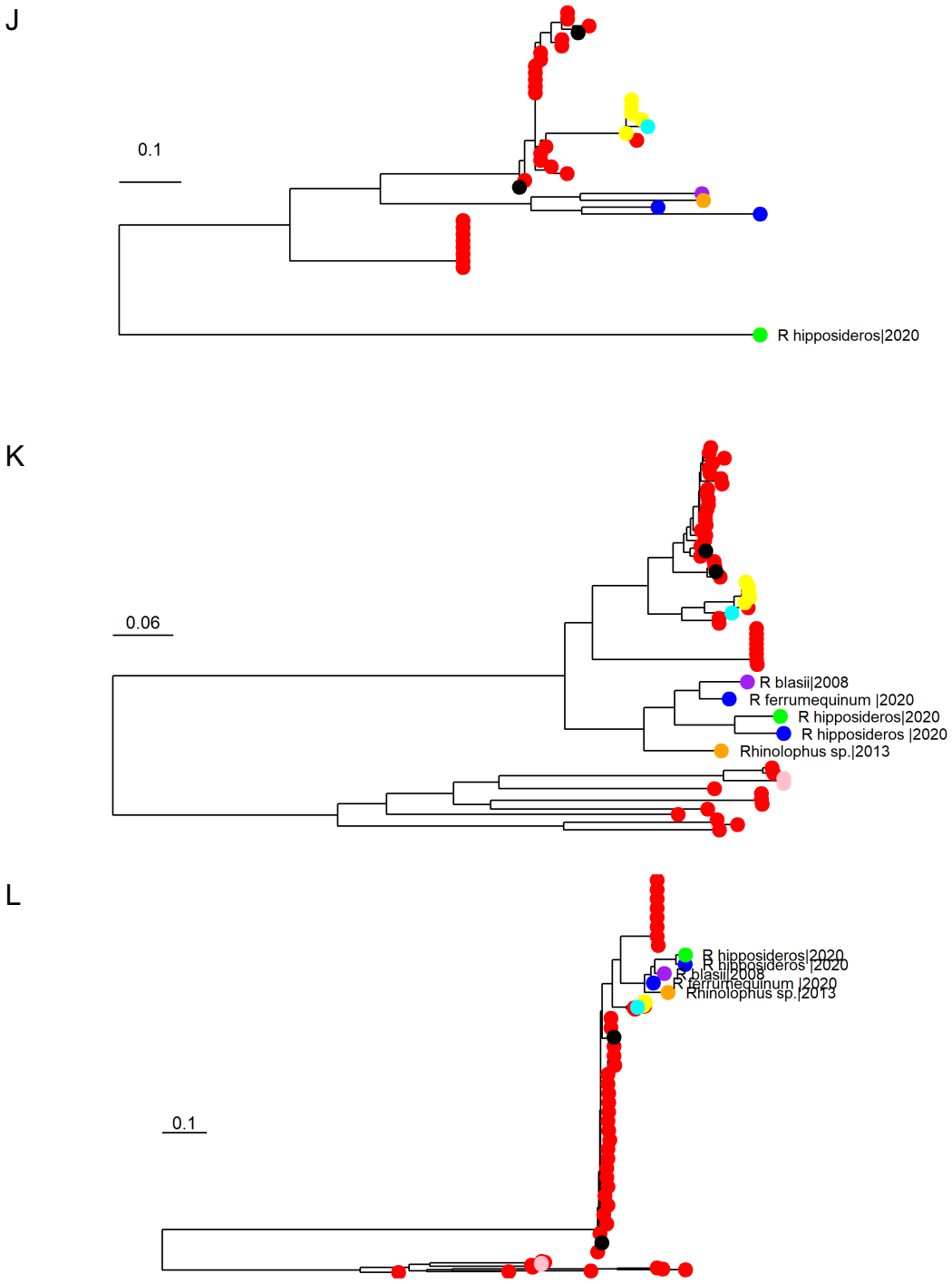

0.04
R blasii|2008
R ferrumequinum |2020
R hipposideros|2020
R hipposideros |2020
Rhinolophus sp.|2013

H


0.06
R hipposideros|2020
R hipposideros |2020
Rhinolophus sp.|2013
R ferrumequinum |2020
R blasii|2008

I


0.05
R hipposideros|2020



Appendix 15. Maximum likelihood phylogenies for the ORF3a nucleotide (A) and amino acid (B) sequence, Envelope nucleotide (C) and amino acid (D) sequence, Membrane nucleotide (E) and amino acid (F) sequence, ORF6 nucleotide (G) and amino acid (H), ORF7a nucleotide (I) and amino acid (J) sequence and nucleoprotein nucleotide (K) and amino acid (L) sequence. Nucleotide phylogenies are inferred under an HKY+G4+I nucleotide substitution model and amino acid phylogenies are inferred under a Dayhoff model.

CoV	ORF1ab	Spike	ORF3	E	M	ORF6	ORF7a	N
BtCoV.Khosta.2.Rh.Russia.2020	9.95479	16.39	17.24	8.66	16.07	14.23	11.46	11.46
BtCoV.Khosta.1.Rh.Russia.2020	18.73587	19.87	17.87	7.79	11.86	19.48	11.98	11.98
BtCoV.BM48.31.BGR.2008	19.14015	20.26	19.75	7.79	15.81	19.85	14.06	14.06
RaCH025	20.14432	23.77	26.33	9.09	15.81	28.46	14.06	14.06
Rs4255	20.03565	23.79	25.5	8.66	16.07	26.59	14.06	14.06
Rs4081	20.01826	23.88	25.5	9.09	16.07	26.59	13.8	13.8
BtRs.HuB2013	19.86176	24.07	25.29	9.09	15.42	27.34	14.32	14.32
Rm1	14.29751	24.09	25.39	9.96	15.02	27.34	14.32	14.32
Rs9401	20.0226	24.13	25.08	9.52	16.07	26.97	14.06	14.06
BtRs.GX2013	20.03999	24.16	26.02	9.96	15.28	27.34	14.06	14.06
RacCS203	21.41367	24.16	27.69	11.69	18.05	29.59	13.54	13.54
Rs4874	20.04434	24.2	24.97	8.23	16.21	26.97	14.32	14.32
BtRs.YN2013	20.00522	24.24	25.71	9.09	15.94	23.6	14.32	14.32
Rs4237	19.97044	24.26	25.29	9.52	15.68	26.22	13.8	13.8
Rs4231	19.89654	24.26	25.18	8.66	15.68	27.34	14.06	14.06
Rs4084	20.1052	24.29	24.97	8.66	15.94	27.34	14.06	14.06
BANAL.20.236.Laos.2020	21.30499	24.37	27.59	9.96	19.24	30.34	13.28	13.28
YNLF.31C	20.07912	24.46	26.44	9.52	16.6	26.97	14.32	14.32

Appendix 16. Nucleotide percentage differences per structural and accessory orfs.

Species 1	ORF1a	ORF1b	Spike	ORF3	E	M	ORF6	ORF7a	N
BtCoV.BM48.31.BGR.2008	18.61	6.08	14.58	16.36	9.09	7.14	42.62	19.67	19.67
BtCoV.Khosta.1.Rh.Russia.2020	17.64	6.19	14.02	13.94	7.79	5.95	36.07	19.67	19.67
BtCoV.Khosta.2.Rh.Russia.2020	6.94	3	12.7	13.94	7.79	9.52	26.23	18.85	18.85
BtRf.HeB2013	20.43	5.82	19.84	25.45	9.09	7.14	52.46	23.77	23.77
BtRf.JL2012	20.95	6.05	21.28	26.06	6.49	7.14	52.46	22.13	22.13
BtRf.SX2013	20.45	5.86	19.59	25.45	10.39	7.14	52.46	23.77	23.77
BtRs.GX2013	20.49	5.53	19.21	24.55	5.19	6.75	50.82	22.95	22.95
BtRs.HuB2013	20.02	5.75	19.21	23.33	6.49	7.14	49.18	22.13	22.13
RaCH025	20.7	5.71	19.46	24.24	5.19	7.54	50.82	22.13	22.13

Appendix 17. Amino acid percentage differences for structural and accessory orfs.

	PDB	GMQE	QMEAN DisCo Global	ID (%)	Virus	ERRAT 2	Favoured / Allowed (%)	Dis-allowed (%)
S	8D55	0.71	0.62 (±0.06)	73.09	RhGB01	84.9	87.9/10.7	0.7
RBD	7LM9	0.74	0.62 (±0.06)	70.53	RhGB01	92.5	89.1/10.9	0
E	2MM4	0.59	0.57 (±0.11)	91.38	RhGB01	72	92.6/7.4	0
M	7VGR	0.83	0.67 (±0.06)	88.74	RhGB01	86.5	81.5/16.9	0
N	8FD5	0.90	0.59 (±0.05)	86.54	RhGB01	78.1	85.9/12.4	0.3
S	7TNW	0.72	0.68 (±0.05)	73.31	CI63	89.1	87.7/11.3	0.2
RBD	8GZ5	0.75	0.64 (±0.06)	70.37	CI63	92.7	87.8/11.6	0
E	2MM4	0.59	0.57 (±0.11)	91.38	CI63	72	92.6/7.4	0
M	7VGR	0.86	0.65 (±0.06)	89.10	CI63	86.5	81.5/16.9	0
N	8FD5	0.9	0.60 (±0.05)	86.30	CI63	78.1	85.9/12.4	0.3

Appendix 18. Templates selected for homology modelling for the RhGB01 and CI63 structural proteins (PDB column). The Global model quality estimate (GMQE) is a quality estimate for the model considering the target-template amino acid alignment and the template structure. The QMEANDisCo global score is a average per residue score which correlates to IDDT per residue score. The estimated error for QMEANDisCo global score represents the root mean squared distance between QMEANDisCo score and IDDT. The ERRAT2 score is an overall quality factor metric of the non-bonded interactions in comparison with other experimentally resolved structures. Core refers to the percentage residues in the favored regions taken from PROCHECK Ramachandran plot.

Site	RhGB01	CI63	
11	N	T	S1
21	R	P	S1
24	K	Q	S1
75	P	S	S1
76	N	G	S1
77	V	K	S1
149	N	K	S1
158	N	S	S1
173	N	D	S1
175	Q	R	S1
178	G	D	S1
207	S	N	S1
235	R	K	S1
247	N	G	S1

260	Y	H	S1
277	I	V	S1
318	I	V	S1
337	M	N	RBD
469	T	S	RBM
471	V	E	RBM
491	I	T	RBM
905	Q	L	S2

Appendix 19. Amino acid differences between RhGB01 and CI63 in the spike protein.

	PDB	GMQE	QMEANDisCo Global	ID (%)	Virus	ERRAT 2	Favored/ Allowed (%)	Dis-Allowed (%)
nsp1	8AOU	0.87	0.66 (±0.06)	80	RhGB01	63.8	74.5/21.5	3.4
nsp7	7JLT	0.87	0.78 (±0.10)	93.98	RhGB01	100	98.7/1.3	0
nsp8	2AHM	0.92	0.75 (±0.06)	94.44	RhGB01	98.9	94.5/5.5	0
nsp9	6W9Q	0.86	0.73 (±0.05)	88.14	RhGB01	93.1	89.1/8.9	0
nsp10	5C8S	0.88	0.74 (±0.07)	94.24	RhGB01	83.9	94.4/5.6	0
nsp12	7KRN	0.86	0.86 (±0.05)	94.96	RhGB01	87.9	86.7/13.1	0.1
nsp13	5ROB	0.85	0.79 (±0.05)	94.01	RhGB01	91.3	88.8/10.8	0.2
nsp14	5C8S	0.86	0.82 (±0.05)	94.31	RhGB01	86.15	90.0/9.2	0
nsp15	7ME0	0.96	0.87 (±0.05)	89.88	RhGB01	97	92.6/7.3	0
nsp16	8BSD	0.96	0.86 (±0.05)	89.26	RhGB01	94.4	92.1/7.5	0.4
nsp1	8AOU	0.87	0.67 (±0.06)	80.56	CI63	63.8	74.5	3.2
nsp7	7JLT	0.87	0.78 (±0.10)	93.98	CI63	100	98.7	0
nsp8	2AHM	0.92	0.75 (±0.06)	94.44	CI63	98.9	94.5	0
nsp9	6W9Q	0.86	0.73 (±0.05)	88.14	CI63	93.1	89.1	0
nsp10	5C8S	0.88	0.74 (±0.07)	93.53	CI63	86.7	94.4	0
nsp12	7KRN	0.86	0.86 (±0.05)	94.85	CI63	87.3	86.8	0.12
nsp13	5ROB	0.85	0.78 (±0.05)	94.18	CI63	91.6	88.8	0.2
nsp14	5C8S	0.86	0.83 (±0.05)	94.50	CI63	85.9	90.1	0
nsp15	7ME0	0.96	0.86 (±0.05)	90.75	CI63	97.6	92.7	0
nsp16	8BSD	0.96	0.86 (±0.05)	89.26	CI63	94.4	92.1	0.

Appendix 20. Templates selected for homology modelling for the RhGB01 and CI63 nonstructural proteins.

	Amino acid (RhGB01:CI63)	RMSD (Å)	Domains	Domains
nsp1	N147D	0.003	N terminal domain tail (red) N terminal domain (green) Linker domain (purple) C terminal domain	1-10 10-127 127-148 149-180
nsp7	NA	0.0	No individual domains predicted	NA
nsp8	NA	0.0	No individual domains predicted	NA
nsp9	NA	0.0	No individual domains predicted	NA
nsp10	L92F	0.023	No individual domains predicted	NA
nsp12	A46S	0.004	NIRAN Interface RdRp RdRp catalytic	1-261 266-364 365-612,774-932 612-774
nsp13	A450D	0.004	Zinc-binding domain Alpha-helical stalk Beta-barrel 1B Helicase domain	1-95 99-146 150-228 251-590
nsp14	V87I, L327P, H455Y	0.009	N terminal 3'-5' exonuclease C terminal N7 Methyltransferase	72-287 296-527
nsp15	T218M, T328S R337H	0.014	N terminal domain Middle C terminal domain endoribonuclease	1-61 62-184 194-344
nsp16	V21M, R182K, I267V	0.0	No individual domains predicted	NA

Appendix 21. Metrics for comparing RhGB01 and CI63 for modelled non-structural proteins. Domains were predicted using InterPro server where structural information was not available in the literature.

	PDB	GMQE	QMEANDisCo Global	ID (%)	ERRAT2	Favored/ allowed (%)	Dis allowed (%)
1	7TTM	0.69	0.64 (±0.06)	69.84	95.8	88.3/11.7	0
2	7URQ	0.70	0.64 (±0.06)	69.84	92.2	86/14	0
3	8GZ5	0.70	0.64 (±0.06)	70.37	92.7	87.8/11.6	0
4	7LM9	0.68	0.62 (±0.06)	69.84	92.4	89/11	0
5	7LM9	0.69	0.63 (±0.06)	69.84	91.9	89/11	0
6	7TTM	0.69	0.63 (±0.06)	70.90	96.4	90.2/9.8	0

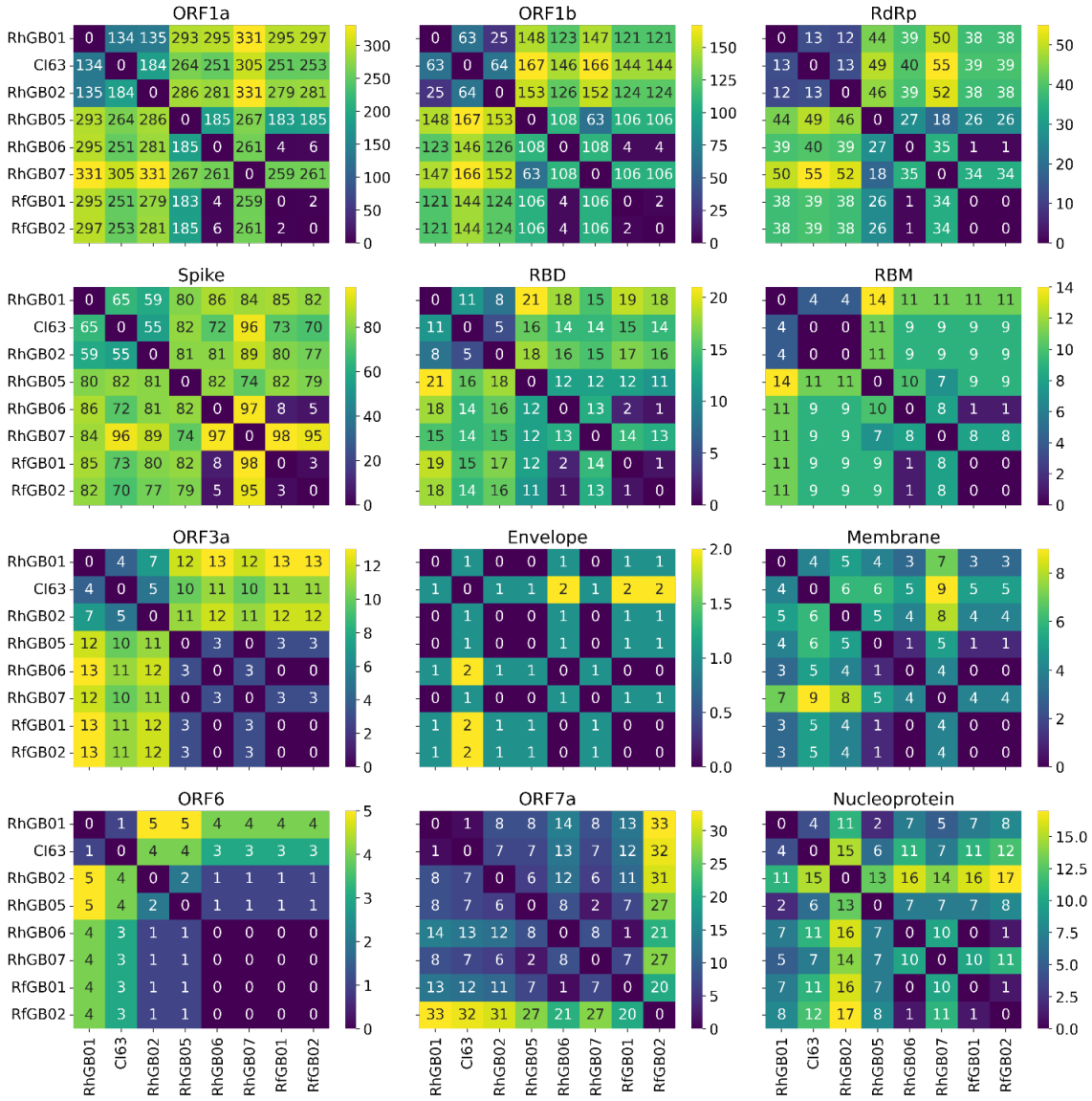
Appendix 22. Model statistics for RBD genotype models.

Virus	Genotype	Region	RMSD
RhGB01	1	RBD	0.836
RhGB01	1	RBM	1.044
RhGB01	1	Region 1	1.85
RhGB01	1	Region 2	1.187
RhGB01	1	Region 3	1.23
RhGB01	2	RBD	1.481
RhGB01	2	RBM	2.13
RhGB01	2	Region 1	3.79
RhGB01	2	Region 2	3.158
RhGB01	2	Region 3	2.052
RhGB01	3	RBD	1.490
RhGB01	3	RBM	1.96
RhGB01	3	Region 1	0.93
RhGB01	3	Region 2	3.356
RhGB01	3	Region 3	1.484
RhGB01	4	RBD	0.008
RhGB01	4	RBM	0.010
RhGB01	4	Region 1	0.011
RhGB01	4	Region 2	0.021
RhGB01	4	Region 3	0.001
RhGB01	5	RBD	0.01
RhGB01	5	RBM	0.13
RhGB01	5	Region 1	0.020
RhGB01	5	Region 2	0.022
RhGB01	5	Region 3	0.002
RhGB01	6	RBD	0.821
RhGB01	6	RBM	1.043
RhGB01	6	Region 1	1.850
RhGB01	6	Region 2	1.190
RhGB01	6	Region 3	1.214

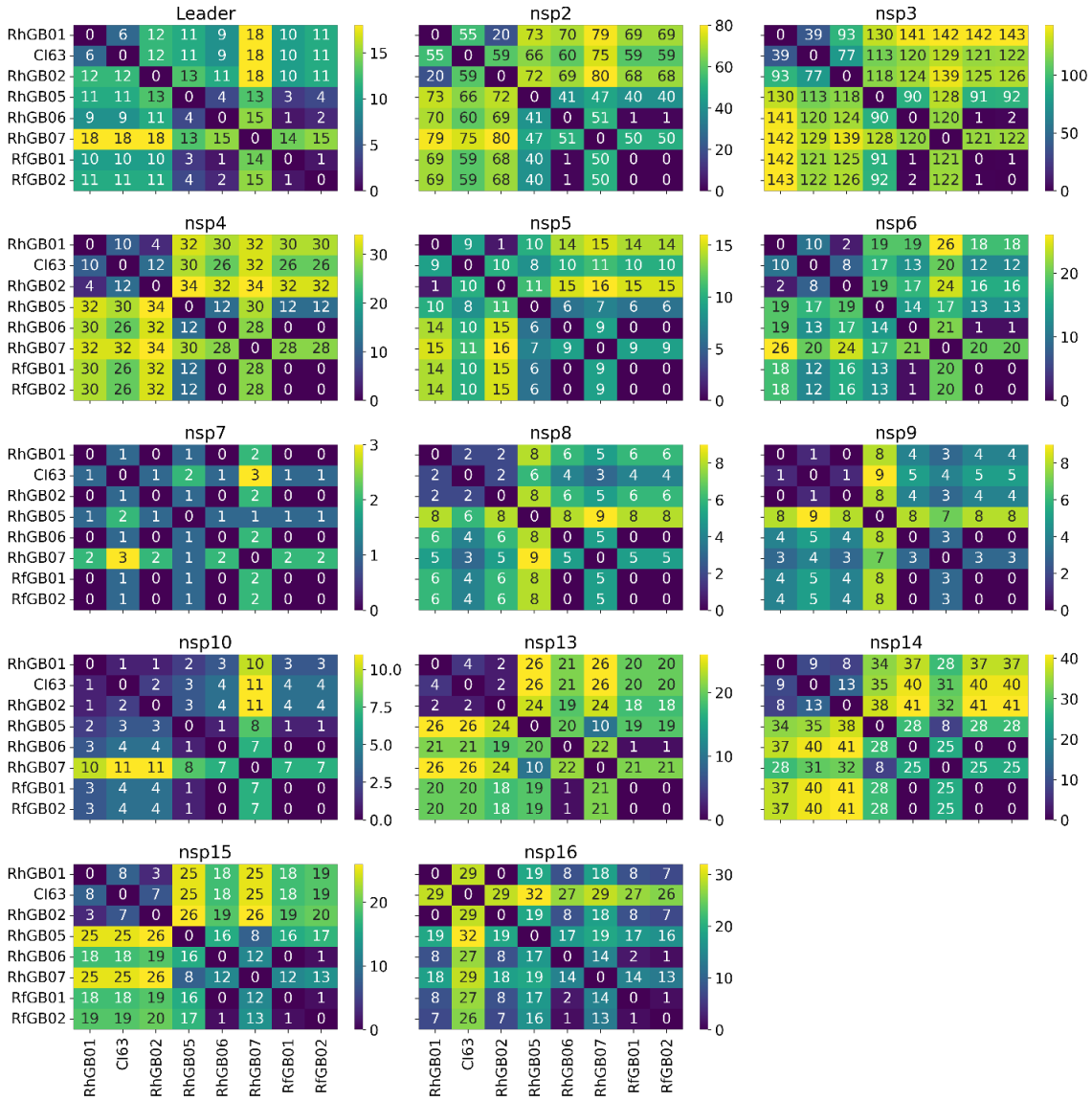
Appendix 23. Molecular distance between RBD genotypes versus RhGB01. With RMSD for RBD, RBM, and the three binding loops.

Virus	Genotype	Region	RMSD
CI63	1	RBD	1.33
CI63	1	RBM	1.74
CI63	1	Binding loop 1	2.66
CI63	1	Binding loop 2	3.19
CI63	1	Binding loop 3	0.47
CI63	2	RBD	0.93
CI63	2	RBM	0.78
CI63	2	Binding loop 1	1.18
CI63	2	Binding loop 2	0.54
CI63	2	Binding loop 3	1.55
CI63	3	RBD	0
CI63	3	RBM	0
CI63	3	Binding loop 1	0
CI63	3	Binding loop 2	0
CI63	3	Binding loop 3	0
CI63	4	RBD	1.49
CI63	4	RBM	1.97
CI63	4	Binding loop 1	3.24
CI63	4	Binding loop 2	3.4
CI63	4	Binding loop 3	1.45
CI63	5	RBD	1.48
CI63	5	RBM	1.97
CI63	5	Binding loop 1	3.23
CI63	5	Binding loop 2	3.58
CI63	5	Binding loop 3	1.45
CI63	6	RBD	0.63
CI63	6	RBM	1.739
CI63	6	Binding loop 1	2.67
CI63	6	Binding loop 2	3.16
CI63	6	Binding loop 3	0.49

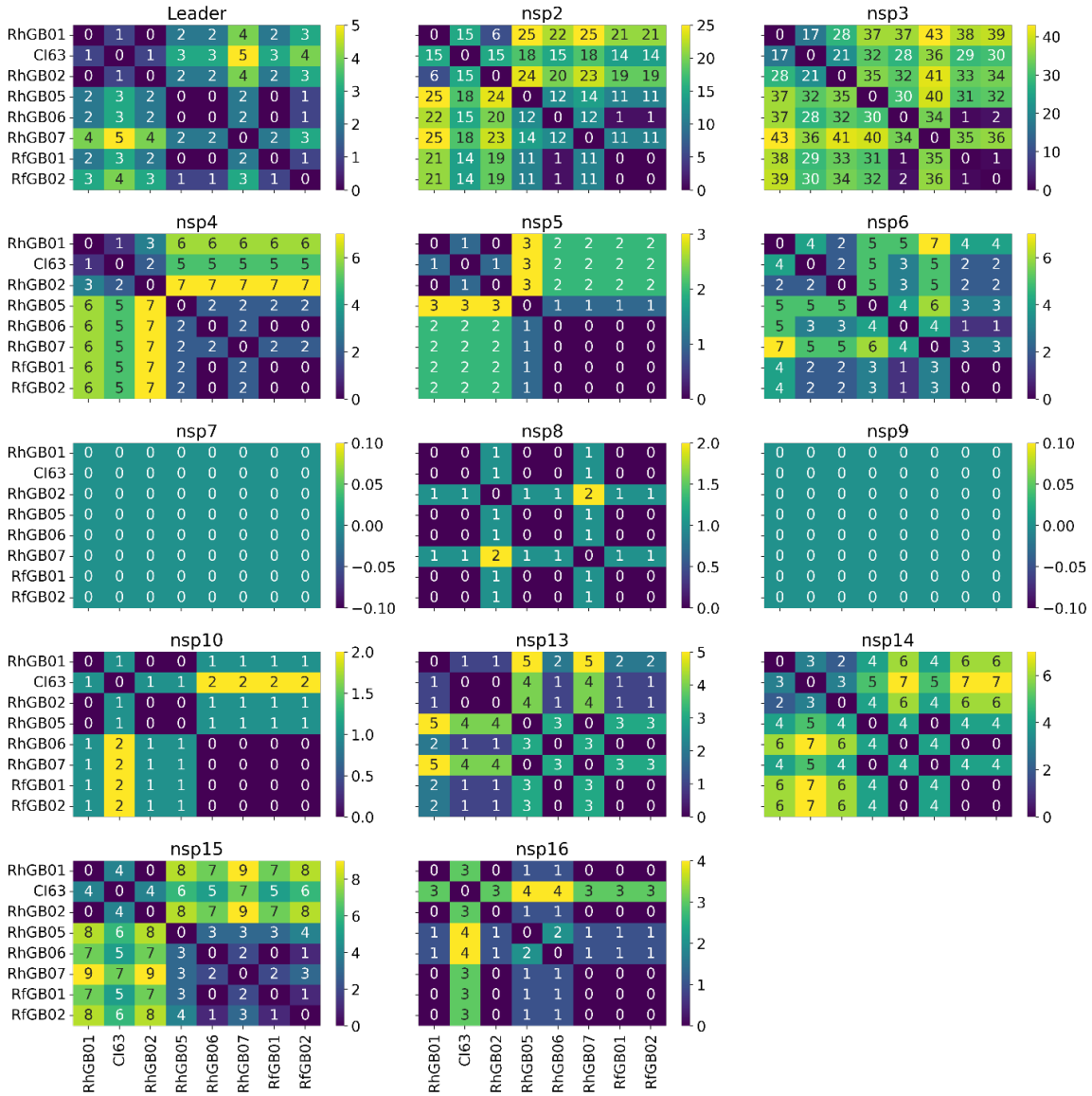
Appendix 24. The molecular distance between RBD genotypes versus CI63. With RMSD for RBD, RBM, and the three binding loops.



Appendix 25. Raw nucleotide differences between the sarbecoviruses identified in lesser and greater horseshoe bats in the UK.



Appendix 26. Raw nucleotide differences between nsps in the sarbecoviruses identified in lesser and greater horseshoe bats in the UK.



Appendix 27. The raw amino acid differences between nsps in the sarbecoviruses identified in lesser and greater horseshoe bats in the UK.