Genomic approaches to virus discovery and molecular epidemiology

Sarah Catherine Hill

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2017

Abstract

Genomic Approaches to Virus Discovery and Molecular Epidemiology

Sarah C Hill, *Doctor of Philosophy*, Trinity 2017

Viral sequence data has great potential for answering questions about the epidemiological dynamics and evolution of viruses. Classical approaches have sought amino acid changes that alter pathogenesis or transmissibility by influencing a virus's ability to enter or replicate within cells. However, this approach rarely recognises the fundamental impact of heterogeneous host contact structures and existing immunological responses on viral transmission. This thesis draws heavily on ecological and immunological concepts to explore the epidemiological dynamics, diversity and evolution of viruses using molecular sequence data.

A number of different research approaches and study systems are used in this thesis. I begin by describing a novel polyomavirus in a European badger, and apply phylogenetic techniques to analyze the evolutionary history of the *Polyomaviridae*. I subsequently describe a large metaviromic study in a population of wild mute swans, for which host demographic data are available. I describe nine new viral species and test whether age and season are associated with differences in abundance and prevalence of different viral taxonomic groups. The study highlights the potential of metaviromics for investigating viral epidemiological dynamics in natural populations.

Influenza A viruses of avian origin (AIV) threaten human and animal health. Using phylogeographic methods, I reconstruct the spatial spread of an H5N8 virus at a regional scale, and investigate how bird density and migration shaped this dispersal. Despite the importance of acquisition of humoral immunity to different strains throughout the lifespan of wild birds for epidemiological dynamics, this topic is poorly understood. I assess the accumulation of immune responses to AIV with age in mute swans. I consider how ecological factors, including age-structured immunity, might have affected the epidemiology of an H5N8 outbreak in the population.

Acknowledgements

Firstly, I would like to thank my supervisor, Oliver Pybus, for his advice and support over the past four years. I am extremely fortunate to have had such an excellent supervisor, who has guided me through my DPhil, and also encouraged me to engage in collaborative research that extends beyond the work presented in this thesis.

I am grateful to the Wellcome Trust, Natural Environment Research Council, and the John Fell Fund, for provision of financial support.

I would like to thank Ian Brown, who offered critical support and expertise throughout my research on avian influenza. Thanks to everyone at the Animal and Plant Health Agency who has assisted my work.

Particular thanks are due to Chris Perrins, for sharing the extensive demographic data on the Abbotsbury swans, without which much of this work would not have been possible. Thanks also to Ben Sheldon for providing guidance throughout the project.

Thanks to Lia van der Hoek and her team at the University of Amsterdam for their work on the swan virome project.

I am sincerely grateful to Adrian Smith and to Peter Simmonds for their generosity in allowing me to work in their laboratories throughout my DPhil.

Thanks to Nick Loman and Josh Quick for sharing their knowledge of the Nanopore MinION, and for kindly helping me to troubleshoot parts of the sequencing work described here.

Thanks to the Abbotsbury Swannery staff who assisted my work during my many visits, particularly Dave Wheeler, Charlie Wheeler and Steve Groves. In addition, I would like to thank Mrs Charlotte Townshend for allowing the study of the swans at Abbotsbury.

I have been very lucky to have had the support of some brilliant postdocs, from whom I have learnt a great deal. Thanks especially to Nuno and Jayna, for sharing their knowledge of phylodynamics, and to Julien, for his help with the swan virome project.

Many thanks are due to my friends and colleagues in the Department of Zoology and elsewhere, especially the current and past members of the Pybus team. I have thoroughly enjoyed the four years of my DPhil, and it is in no small part down to them.

Finally, I would like to thank Sean, my family, and friends, for their ceaseless love and support.

Statement of Contribution and Associated Publications

This work presented in this thesis includes research that has been published in peer-reviewed journals and papers that are currently in preparation. Below, I present a summary of my contributions to each chapter, and those of others. Asterisks in authorships of published papers denote equal contribution.

Chapter 2

Sarah C Hill*, *Aisling A Murphy**, *Matthew Cotten, Anne L Palser, Phillip Benson, Sandrine Lesellier, Eamonn Gormley, Céline Richomme, Sylvia Grierson, Deirdre Ni Bhuachalla, Mark A Chambers, Paul Kellam, María-Laura Boschiroli, Bernhard Ehlers, Michael A Jarvis, Oliver G Pybus. 2015. Discovery of a polyomavirus in European badgers (Meles meles) and the evolution of host range in the family Polyomaviridae. J Gen Virol 96(Pt 6): 14111422. doi: 10.1099/vir.0.000071*

I wrote the manuscript, performed genomic characterisation of the novel polyomavirus and all phylogenetic analyses of the evolutionary history of the family. OGP supervised the project and provided editorial assistance.

All laboratory work and the screening of captive badgers was conducted by other authors.

Chapter 3

Sarah C Hill*, *Youn-Jeong Lee**, *Byung-Min Song , Hyun-Mi Kang, Eun-Kyoung Lee, Amanda Hanna, Marius Gilbert, Ian H Brown, Oliver G Pybus. 2015. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. Infect Genet Evol. 34: 267277. doi: 10.1016/j.meegid.2015.06.014*

I wrote the manuscript and performed all analyses with the exception of generating the poultry-density maps, which were generated by MG. Other authors collected and sequenced all of the Korean isolates of the virus. OGP supervised the project and provided editorial assistance.

Chapter 4

I collected all of the samples used in the project. Viral nucleic acid was extracted from the fecal samples and dsDNA was synthesised by Lia van der Hoek, Maarten Jebbink and Martin Dejis at the University of Amsterdam. Library preparation and sequencing was conducted by the Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, Oxford.

The pipeline that I used for detecting viral reads through BLAST searching was developed by Julien Thézé. I performed all computational work, including assembly of viral genomes, phylogenetic analyses, host-determination and statistical analyses. I performed all laboratory work following primary sequencing, including confirmation of the novel genomes by MinION sequencing, qPCRs and specific PCRs.

Chapter 5

***Sarah C Hill**, Ruth J Manvell, Bodo Schulenburg, Wendy Shell, Paul S Wikramaratna, Christopher Perrins, Ben C Sheldon, Ian H Brown, Oliver G Pybus. 2016. Antibody responses to avian influenza viruses in wild birds broaden with age. Proc Biol Sci. 283(1845): 20162159. doi: 10.1098/rspb.2016.2159*

I wrote the manuscript and performed all analyses. BS performed the immunological assays. OGP supervised the project and provided editorial assistance.

Chapter 6

I collected samples from 70% of the dead birds (all birds sampled after January 1st 2017) and all live birds during the H5N8 outbreak, with the permission and logistical

support of the UK Animal and Plant Health Agency (APHA). The remaining dead birds were tested by APHA during initial confirmation of the outbreak. RNA extraction and qPCR screening for highly pathogenic H5N8 avian influenza was conducted by staff at the UK Animal and Plant Health Agency.

I conducted all other laboratory work, including cDNA synthesis, amplification and genomic sequencing.

I collected all of the fecal samples and cloacal swabs that were tested April 2015 to March 2016. I conducted RNA extraction and RT-qPCR screening for low pathogenic samples on these samples at the UK Animal and Plant Health Agency.

I conducted all of the statistical and phylogenetic analyses in this chapter.

Contents

1 Introduction **1**

1.1 Viral diversity and the effects of ecological processes on viral epidemiology and evolution . . . 1

1.1.1 Viral diversity and evolution . . . 1

1.1.2 Ecological processes shape viral epidemiology and evolution . . . 2

1.1.2.1 Contact structure . . . 2

1.1.2.2 Immunological landscape . . . 3

1.1.2.3 Seasonal drivers . . . 6

1.2 Molecular biological and phylogenetic tools for studying viral diversity and diversity . . . 7

1.2.1 Viral sequencing . . . 8

1.2.1.1 Classical approaches to virus discovery . . . 8

1.2.1.2 Virus discovery in the era of high-throughput sequencing 10

1.2.2 Phylogenetic and phylodynamic tools for studying viral epidemiology and evolution . . . 13

1.3 Introduction to the ecology and epidemiology of avian influenza virus 16

1.4 Structure of the thesis . . . 22

2 Discovery of a polyomavirus in European badgers (*Meles meles*) and the evolution of host range in the *Polyomaviridae* **24**

2.1 Introduction . . . 24

2.2 Materials and Methods . . . 28

2.2.1 Viral discovery and sequencing of UK isolate 28
2.2.2 Sequencing of French isolate 29
2.2.3 Phylogenetic analyses 31
2.3 Results 33
2.3.1 Virus discovery in cell culture 33
2.3.2 Virus discovery in wild badgers 34
2.3.3 Genome characterisation 37
2.3.4 Phylogenetic analysis of the *Polyomaviridae* 41
2.4 Discussion 44

3 Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. 48

3.1 Introduction 48
3.2 Materials and methods 51
3.2.1 Bird ecology data 51
3.2.1.1 Longitudinal data on bird counts and waterfowl numbers 51
3.2.1.2 Poultry density and temporal data 52
3.2.2 Sequence data 53
3.2.3 Bayesian molecular clock phylogeography 54
3.2.3.1 Model selection 54
3.2.3.2 Phylogeographic analyses 55
3.3 Results and Discussion 56
3.3.1 Density and temporal dynamics of poultry 56
3.3.2 Density and temporal dynamics of wild waterfowl 57
3.3.3 Phylogeographic analyses of H5N8 58
3.3.3.1 First wave of viral entry to ROK 58
3.3.3.2 H5N8 spread within ROK 59
3.3.3.3 Proposed second wave of virus transmission 61

3.4 Conclusions 69

4 Diversity of the mute swan (*Cygnus olor*) faecal virome 71

4.1 Introduction 71

4.2 Materials and Methods 77

4.2.1 Sample collection 77

4.2.2 Sample preparation and sequencing 78

4.2.3 Computation detection of viral reads 82

4.2.4 Assembly of novel viruses 82

4.2.5 Confirmation of *de novo* assembled viruses by specific PCR . . 84

4.2.6 Targeted sequencing 86

4.2.7 Phylogenetic analysis 86

4.2.8 Sensitivity test of Illumina sequencing data 87

4.2.9 Assignment of host group to faecal metagenomic reads 89

4.2.10 Statistical analysis 90

4.3 Results: Virus Discovery 92

4.3.1 Discovery of novel viral species 92

4.3.1.1 *Waterbird coronavirus 1* 92

4.3.1.2 Mute swan astroviruses 97

4.3.2 Additional novel viruses 102

4.3.2.1 *Mute swan picornavirus* 102

4.3.2.2 *Mute swan sapelovirus* 103

4.3.2.3 *Mute swan parvovirus* 105

4.3.2.4 *Mute swan megrivirus* 105

4.3.2.5 *Mute swan stool-associated circular virus* 105

4.3.2.6 *Mute swan stool-associated virus* 106

4.3.2.7 *Avian kobuvirus* 106

4.3.3 Genomic assembly of existing virus species 107

4.3.3.1 Existing species of *Caliciviridae* 107

4.3.3.2 Divergent strain of the *Avibirnavirus* IBDV 108
4.4 Results: Impact of Host Factors on Viral Carriage and Abundance . . 120
4.4.1 Effect of age on viral carriage and abundance 120
4.4.1.1 Comparison of Illumina read data and PCR data . . 120
4.4.1.2 Total viral read count 121
4.4.1.3 Age-specific variation in viral abundance and prevalence of different viral taxonomic groups 123
4.4.1.4 Seasonality of viral infection for *Coronaviridae* and *Astroviridae* 125
4.4.1.5 Stability of the viral repertoire in birds 126
4.5 Discussion 135

5 Antibody responses to avian influenza viruses in wild birds broaden with age 141

5.1 Introduction 141
5.2 Materials and Methods 145
5.2.1 Study population 145
5.2.2 Population sampling 145
5.2.3 Serological assays 146
5.2.4 Statistical analyses 147
5.3 Results 149
5.3.1 Seroprevalence varies by subtype 149
5.3.2 Subtype breadth increases with age 150
5.3.3 Effect of age, sex and sample year on each individual subtype 153
5.3.4 Longitudinal sampling in birds from both years 155
5.4 Discussion 157

6 Epidemiology of highly pathogenic H5N8 avian influenza virus in a population of long-lived birds 162

6.1 Introduction 162

6.2 Materials and Methods . . . 165
6.2.1 Site and sample collection . . . 165
6.2.2 Estimation of LPAI AIV prevalence . . . 166
6.2.3 Estimation of HPAI prevalence in healthy birds . . . 167
6.2.4 Screening for positive samples . . . 167
6.2.5 Amplification and sequencing of the H5N8 genome . . . 169
6.2.6 Phylogenetics . . . 170
6.2.7 Epidemiological analysis . . . 171
6.3 Results . . . 175
6.3.1 H5N8 outbreak . . . 175
6.3.1.1 H5N8 HPAI caused high mortality in the population 179
6.3.1.2 Mortality was skewed towards younger birds . . . 180
6.3.1.3 There is no effect of weight on avian deaths amongst cygnets . . . 184
6.3.2 LPAI prevalence in the population . . . 184
6.3.3 Molecular genetics and phylogenetics of HPAI H5N8 . . . 190
6.4 Discussion . . . 201

7 Discussion and Conclusions 206

7.1 Key Results, Implications and Future Direction of each Chapter . . . 206
7.1.1 Chapter 2: discovery of a novel polyomavirus . . . 206
7.1.2 Chapter 3: early spread of HPAI H5N8 in Asia . . . 207
7.1.3 Chapter 4: exploring the faecal viral diversity of mute swans . 209
7.1.4 Chapters 5 and 6: pre-existing immune responses to AIV in birds and mortality following HPAI viral infection . . . 213
7.2 General Implications and Future Directions for the Study of Viral Epidemiological Dynamics . . . 218

A Appendix to Chapter 2 222

A.1 Tables . . . 222

B Appendix to Chapter 3 **230**

B.1 Figures 230

B.2 Tables 236

C Appendix to Chapter 4 **243**

C.1 Figures 243

C.2 Tables 245

D Appendix to Chapter 5 **261**

D.1 Tables 264

E Appendix to Chapter 6 **270**

E.1 Figures 270

E.2 Tables 272

Bibliography **279**

List of Figures

1.1 Generation of influenza A viral diversity 18

2.1 Genome map of MmelPyV1 35
2.2 Gel electrophoresis of mink cell line NBL-7 36
2.3 Conserved ORFs of the two MmelPyV1 isolates and their nearest relative (CSLPyV1). 39
2.4 Maximum likelihood phylogeny of all known polyomavirus species or putative species, including MmelPyV1, estimated from the "genome-wide" alignment 40
2.5 Maximum likelihood phylogenies produced from different regions of the genome. 43

3.1 Maps showing domestic poultry density in ROK according to the Gridded Livestock of the World 2.0 64
3.2 Maps generated from ROK Ministry of Environment Wild Bird Census for winter 2014 data showing the number of overwintering waterfowl for the four most common species in ROK 65
3.3 Bird population dynamics in ROK 66
3.4 Map representing the estimated trajectory of the H5N8 spread in ROK. 67

4.1 Abbotsbury Swannery on the Fleet Lagoon 79
4.2 Uniquely identifiable leg rings on swans 80
4.3 Wild birds at Abbotsbury Swannery 81
4.4 Maximum likelihood phylogenetic tree of gammacoronaviruses 94

4.5 Waterbird coronavirus 1 genome 95
4.6 Gammacoronavirus maximum likelihood phylogenetic tree 96
4.7 Maximum likelihood tree of ORF1b of mute swan astrovirus 1 and mute swan astrovirus 2 99
4.8 Maximum likelihood tree of 3' end of ORF1b of mute swan astrovirus 1 and mute swan astrovirus 2 100
4.9 Maximum likelihood tree of 5' end of ORF2 of mute swan astrovirus 1 and mute swan astrovirus 2 101
4.10 Maximum likelihood tree of mute swan picornavirus and related viruses 103
4.11 Maximum likelihood tree of mute swan sapelovirus and related viruses 104
4.12 Maximum likelihood tree of mute swan parvovirus and related viruses 110
4.13 Maximum likelihood tree of replicase region of stool-associated circular viruses 111
4.14 Maximum likelihood tree of *Kobuvirus* genera and putative novel related genus 112
4.15 Midpoint rooted phylogenetic trees constructed from nucleotide-level alignments of IBDV 113
4.16 Midpoint rooted phylogenetic trees constructed from amino-acid-level alignments of IBDV 114
4.17 Midpoint rooted phylogenetic trees of avian circoviruses NS1 115
4.18 Midpoint rooted phylogenetic trees of avian orthoreovirus NS1 segment 116
4.19 Midpoint rooted phylogenetic trees of the goose parvovirus NS1 gene 117
4.20 Midpoint rooted phylogenetic trees of goose parvovirus VP1 gene . . 118
4.21 Midpoint rooted phylogenetic trees of aquatic bird bornavirus 1 . . . 119
4.22 qPCR Cts compared to Illumina read counts 122
4.23 Heatmap of read abundance by viral species for viruses believed to exclusively infect vertebrates 127
4.24 Log(10) proportion of viral reads for different age group birds 128

4.25 CSS normalised counts of viral families with differential read abundance in birds of different ages 129
4.26 CSS normalised counts of viral species with differential read abundance in birds of different ages 130
4.27 Seasonality of coronaviruses 131
4.28 Seasonality of astroviruses 132
4.29 Seasonality of viral reads from different families 133
4.30 Longitudinal metaviromic sampling of 5 birds 134

5.1 Immune responses of birds in the population 151
5.2 Cumulative fitted probabilities of effect of age on number of strains to which a swan exhibits a response (breadth of response) as determined by cumulative link models. 152
5.3 Fitted probability of effect of age and sex on response to H9N9 and H9N2 as determined by GLM with logit link 154
5.4 Proportion of birds observed to respond to H5N2 in 5 age groups for each sample year 156

6.1 Number of carcasses found at Abbotsbury Swannery over time 176
6.2 Comparison of mortalities during HPAI H5N8 outbreak period (2016-2017) and HPAI H5N1 outbreak period (2007-2008) 178
6.3 Estimated numbers of swans of each age in the population in December 2016 and proportion that died 183
6.4 Proportion of birds of each age that died during 23rd December to 23rd January in 2016-2017 and in previous years 185
6.5 No effect of weight or exact age on cygnet survival 186
6.6 Proportion of birds that died of HPAI AIV and their parental status . 188
6.7 Proportion of birds of different ages positive for LPAI influenza on 21st July 2015 189

6.8 Root-to-tip regression of genetic distance against sampling date for HA and NA 191
6.9 Maximum clade credibility tree estimated for HA segment 193
6.10 Maximum clade credibility tree estimated for NA segment 194
6.11 Maximum likelihood tree estimated for NS segment of H5N8 clade containing Abbotsbury sequences 195
6.12 Maximum likelihood tree of NP segment of H5N8 clade containing Abbotsbury sequences 196
6.13 Maximum likelihood tree of MP segment of H5N8 clade containing Abbotsbury sequences 197
6.14 Maximum likelihood tree of PA segment of H5N8 clade containing Abbotsbury sequences 198
6.15 Maximum likelihood tree of PB1 segment of H5N8 clade containing Abbotsbury sequences 199
6.16 Maximum likelihood tree of PB2 segment of H5N8 clade containing Abbotsbury sequences 200

B.1 Maximum clade credibility tree of avian influenza sequences for H5 clade 2.3.4.4. 231
B.2 Effect of different sampling and grouping of provinces with only one sequence on phylogeographic inference. 232
B.3 Maximum clade credibility tree for reconstruction without phylogeographic model. 233
B.4 Maximum clade credibility tree for reconstruction with phylogeographic model and without BSSVS. 234
B.5 Maximum clade credibility tree for reconstruction with phylogeographic model and with BSSVS. 235

C.1 Counts per sample of sequencing reads obtained, viral reads and vertebrate viral reads. 244

D.1 Cumulative proportion of birds sampled during each sampling year . 262
D.2 Change in HI titre for each antigen between 2007 and 2008 for eleven birds that were sampled in both years 263

E.1 Kernel density estimated distribution of posterior support for the time to most recent common ancestor of the Abbotsbury outbreak clade. 271

List of Tables

2.1 Results of the BaTS tests for association of virus phylogeny with host species taxonomy. 44

3.1 Phylogenetic statistics for various groups of isolates within clade C4 . 68

4.1 Previous metaviromic studies in wild bird species 76

5.1 Datasets used to calculate breadth of response 148

5.2 Population seroprevalence (number birds positive) for each test antigen. 150

6.1 Estimated age-adjusted and crude-mortalities during outbreak period per year 180

6.2 Estimated age-adjusted and crude-mortalities during outbreak period per year 181

6.3 Estimated proportion of birds that died by age 182

A.1 Tissue samples and PCR positivity for presence of polyomavirus using VP1 specific assay for 11 animals 223

A.2 Tissue samples and PCR positivity for presence of polyomavirus using NCCR specific assay for 11 animals 224

A.3 Host group categories for BaTS analysis 225

A.4 Primers used in sequencing of French badger polyomavirus isolate . . 228

B.1 Positive samples and selected isolates from poultry in ROK 237

B.2 Positive samples and selected isolates from wild birds in ROK 238

B.3 Sequences and accession numbers used in phylogenetic analyses . . . 239

C.1 Number of times that each uniquely identifiable bird was sampled at Abbotsbury per sample type 245
C.2 Number of samples collected from uniquely identifiable birds per visit 246
C.3 Whole genome PCR primers for novel viral genomes resequenced using Sanger sequencing 247
C.4 Primers for targeted sequencing PCRs 249
C.5 Primers for SYBR green qPCRs 250
C.6 Primers for whole genome multiplex PCRs 251
C.7 Viral read abundance between birds of different age per taxonomical unit 257
C.8 Viral read presence/absence differences birds of different age per taxonomical unit 258
C.9 Unique features that occur only in one age group 260

D.1 Generalised linear models for individual subtypes 265

E.1 Wild birds observed at site during Wetland Bird Surveys from November 2016 to January 2017 272
E.2 Primers used in N8 screening PCRs 274
E.3 Primers used in whole-genome multiplex PCRs 275
E.4 Percentage of each genomic segment sequenced at >20x coverage . . . 278

Chapter 1

Introduction

1.1 Viral diversity and the effects of ecological processes on viral epidemiology and evolution

1.1.1 Viral diversity and evolution

Viruses are probably the most diverse and most abundant replicative entity on the planet [426, 427]. We have no robust picture of how many different viruses are present across all global ecosystems, but we can be confident that the number is staggeringly large and vastly exceeds the several thousand viral species that have thus far been discovered and named. The host-species of currently identified viruses suggest that certain types of virus are more likely to occur in certain orders of life. Bacteria, for instance, appear to be far more commonly infected with dsDNA viruses than eukaryotes [304]. The distribution of viral diversity throughout different host ecosystems is far from clear.

The huge diversity of viruses and their evolutionary success results in part from their ability to rapidly replicate and undergo genomic change through mutation or recombination. Burst size (i.e., the average number of virions released during lysis of a cell following lytic infection) and replication time vary among virus species, but a virion that infects a single host cell can produce up to tens of thousands of progeny

within hours or days [517, 64]. Viruses, particularly RNA viruses that replicate using RNA-dependent RNA polymerases that lack proof-reading capabilities, have mutation rates that can be six orders of magnitude higher than the mutation rates of the hosts that they infect [111, 174]. Whilst viral mutation is the primary driver of the population level diversity that forms the substrate for natural selection, viral recombination that occurs through template jumping of polymerases between different viral genomes in a co-infected cell, or through reassortment that occurs when different segmented viruses infect a single cell, can also generate novel combinations of existing viral diversity. The continual generation of diversity and rapid rate of replication allow viruses to evolve rapidly in response to selective pressures.

1.1.2 Ecological processes shape viral epidemiology and evolution

The epidemiology and evolution of viruses are inextricably tied to the ecological landscape through which that virus is able to transmit. Evolutionary theory predicts that virus lineages should evolve traits that maximise the number of copies of that virus over long periods of time. Viruses are obligate intracellular parasites that lack the ability to replicate outside of a host cell. For a viral lineage to proliferate and survive, it may evolve traits that maximise its ability to encounter and enter a continual supply of new cells within an existing host organism and, ultimately, to allow effective transmission to a new host organism. Both host immunity and contact structure are thus critical in shaping the long-term evolution of epidemiologically important viral traits, including virulence and modes of transmission [144].

1.1.2.1 Contact structure

Contact structure between hosts includes when and how often contact occurs, what type of contact occurs and between which hosts. For many viruses, including all blood-borne and sexually transmitted pathogens, transmission is only possible through spe-

cific forms of direct contacts between infected and susceptible hosts. Network-based models of infectious disease transmission predict that network modularity, number of nodes (hosts), and degree of each node (number of contacts that each host has), can all affect how a virus is likely to transmit once it is introduced to the network [83]. Contacts between hosts in a network can be influenced by whether virulence of the pathogen (i.e., the ability of that pathogen to damage the host) significantly affects whether contact is more or less likely to occur. It has been theorised that viruses that rely primarily on direct transmission may evolve lower virulence than viruses that rely primarily on indirect transmission [123], because direct transmission relies heavily upon behaviours that are energetically demanding for the host, such as fighting or sex.

The existing contact structure of the host has huge capacity to shape transmission of a pathogen. For example, infection with simian immunodeficiency virus (SIV) has been repeatedly shown to be more common in adult African green monkeys than juveniles, and more common in females than males [207, 274]. Sex-based differences in prevalence are presumably an effect of higher variance in number of mating contacts in males than in female in the monkey's polygynous mating structure, in which non-dominant males are prevented from mating [440]. For indirectly transmitted disease, such as vector-borne viruses or respiratory viruses, pathogen transmission is no longer rigidly tied to specific direct contact patterns, yet host behaviour and ecology still critically influence observed patterns of virus evolution. For example, host population size critically determines whether measles is likely to persist in a population without reintroduction from outside [406, 19].

1.1.2.2 Immunological landscape

The availability of susceptible hosts within an existing contact network is a fundamental driver of how transmission occurs through that network. The extent to which a host is susceptible to infection upon exposure with a virus results from immunological protection as a result of previous exposure to an antigenically similar pathogen, and

the physiological ability of the host to mount an effective immune response. Susceptible hosts accumulate through the entrance of new naïve hosts into the population, for example, through birth or through antigenic evolution of the virus.

The effect of age and sex on the physiological aspects of the immune-function is extremely complex, and likely to vary by species according to the typical life history of that species. Infectious diseases progress slower and cause fewer fatalities in females across a wide range of mammalian species [135, 225, 448]. This paradigm is often attributed to the evolutionary benefit of males channeling energetic resources to mating efforts at the expense of immune defense [418], via the proximate mechanism of immunosuppressive effects of sex hormones such as testosterone [135, 418]. A general decline in immune function with age is documented for many animals, including humans and certain birds [172, 242]. Elderly humans show reduced ability to establish immunological memory to novel antigens and decline of the adaptive immune system, which increases the risks associated with infection in the elderly by three-fold compared to young adults [505]. Other factors such as immunosuppression as a result of infection with viruses which target the immune system, such as HIV which destroys lymphocytes, can also affect the host's ability to mount an effective immune response.

Immunological memory of specific encountered antigens is formed in vertebrate hosts as part of the adaptive immune response, such that particular antigens can be targeted more rapidly upon secondary than primary exposure. For viruses such as measles, against which permanent immunity typically follows primary infection [475], the gradual accumulation of naïve individuals through their birth leads to cyclic dynamics of large outbreaks that cause the development of immunity in most of the population, followed by troughs in prevalence during which the number of naïve individuals gradually accumulates again before the next outbreak can occur [74].

Acute viral infections such as measles are often associated with rapid viral replication, followed by elicitation of host immune responses that result in viral clearance and induction of long term immunity to the encountered viral antigens in the host.

Many viruses have evolved strategies that allow them to circumnavigate the potential problem of running out of susceptible hosts, including rapid antigenic evolution or establishment of long term, chronic infection of the host. For viruses such as avian influenza virus, the rapid rate of antigenic evolution means that hosts can be reinfected with the same strain within a few years of primary infection because antibodies developed against the primary infection are no longer capable of recognizing the virus [400, 478]. Cross-reactive responses can result in immune-mediated competition between different strains or species. Such competition has been proposed for subtypes and strains of avian influenza, in which subtype success is dependent on the degree of circulation of other subtypes in the population [238, 483].

Some viruses have evolved the ability to persist for long periods in a single host through effective evasion of the host immune system. This strategy can a small number of hosts to serve as reservoirs for a larger population, or can allow high viral prevalence in a population despite a relatively low rate of new infection. Persistently infecting viruses require a smaller critical community size (the minimum size of a population in which an infection can persist indefinitely [20]) than viruses that only cause short term, acute infection. There are two main non-exclusive strategies that can result in viral persistence, including continuous replication or establishment of latent viral infection. Continuously replicating viruses successfully evade host immune responses through continual production of extensive new antigenic diversity. Whilst the presence of the infection is detected by the host immune system and many virions are immediately neutralised, virions with antigens that are novel to the host immune system may evade immediate immunological detection, which leads to ongoing proliferation and evolution of the viral infection in the host in response to immune selection. In contrast, some viruses that establish persistent infection often endure by minimising detection by the host immune system. This can be achieved through a variety of strategies, including low levels of expression of viral antigens, infection of immunologically privileged sites, and/or by direct modulation of the immune response of the host. Viruses can lie dormant within cells in latent infection, reducing

or minimising the production of viral proteins that may be recognised by the immune system. Spread of the virus beyond the latently infected cell can be achieved by periodic reactivation of gene expression and the lytic cycle of viral replication. A classic example of a latently infecting virus is herpes simplex virus, which selectively expresses certain viral genes and can infect cells in the immunologically privileged nervous system [189]. For viruses that rely heavily on vertical transmission in order to be sustained in a host population, formation of persistent infection in an infected newborn allows the virus to be sustained until the host reaches reproductive age, at which point the virus can be transmitted onwards. Viruses that rely heavily on this mode of transmission including bovine diarrhea virus and hepatitis B virus (HBV) [264, 340]. Bovine diarrhea virus currently appears to be unique among viruses that cause persistent infection in that it establishes persistent infection following vertical transmission by infecting the foetus early enough that the foetus develops immunotolerance to the virus [340].

1.1.2.3 Seasonal drivers

A consequence of this reliance of transmission on host ecology, immunity and behaviour is that seasonal changes in any of these traits in the host can also drive seasonal waves of peaks in viral prevalence. Co-ordinated breeding of certain animals at a particular time of year leads to sudden leaps in the number of immunologically naïve hosts (for example, rabies in bats [419] and raccoons [112]). Alternatively, for vector-borne viruses such as Zika virus [126], seasonality determines the number of available vectors that can transmit disease between hosts. Seasonal changes in host sociality, such as gathering during breeding or because of changes in food or water availability, affect host-density and the number of contacts between different hosts. In humans, such seasonal changes include the oft-cited effect of school terms on increasing measles outbreaks in children [133]. Climatic conditions associated with particular calendar periods can help to drive seasonality of infection by causing differences in viral persistence in the environment or in the ability for viruses to transmit.

Avian influenza A viruses retain infectivity in the environment for longer periods in colder and more humid conditions [82, 271]. Flooding caused by monsoons or extreme weather has been associated with an increase in diarrhoea cases caused by the water-borne rotavirus [164, 280]. Host-susceptibility to infection is not well-explored, but modulation of the immune-system in response to energetic channeling towards reproduction or elsewhere might also cause seasonal differences in viral transmission [107].

1.2 Molecular biological and phylogenetic tools for studying viral diversity and diversity

The premise of this thesis is that the processes that determine viral epidemiological dynamics and the long-term evolution of viral diversity can, and should, be explored by studying viral distribution and transmission in natural ecological settings, in which heterogeneous contact patterns between hosts, competition between different viruses, and pre-existing immunological responses are all present and can potentially interact. There are several possible complementary approaches to studying genetic diversity that can be used to meet this goal, including classical epidemiological approaches, detection of immunological responses as a footprint of previous viral infection, and approaches that are reliant on molecular genetics. Although this thesis draws on traditional epidemiological and immunological approaches in Chapters 4 and 5, it is centered around the fact that a rapid revolution in sequencing technology has radically altered the ease with which viruses can be discovered and their evolutionary and epidemiological patterns explored.

1.2.1 Viral sequencing

1.2.1.1 Classical approaches to virus discovery

The development of sequencing technologies has resulted in a dramatic shift in the ease with which viruses can be discovered and their evolutionary and epidemiological patterns explored. Prior to the development of molecular technologies to detect viruses, viral identification was extremely laborious. The first discovery of a virus was achieved in the late 19th century with the identification of tobacco mosaic virus. Unlike bacteria, the causative agent of tobacco mosaic disease remained infectious upon passing through the extremely small pores in Chamberland filters, was capable of diffusing through agar, and was entirely unculturable - thus suggesting that the agent was significantly different from previously known bacteria [243, 510]. At the turn of the century, many important viruses were subsequently identified based on their ability to retain infectivity following passage through filters, including foot-and-mouth disease virus, yellow fever virus and influenza A virus [273, 370, 413]. Key discoveries in the mid 20th century, including the invention of the electron microscope and the development of tissue culture methods, allowed the subsequent characterisation of the physical structure and replicative cycle of different viruses (reviewed in [149]). Early virus classification and identification was often aided by structures identified via electron microscopy [5, 149, 300], to the extent that many viral taxonomic names are derived from the Latin etymological roots of their electromicroscopic structures; for example, astroviruses are named for their star-like shape (astro is derived from the Greek word for star). The discovery in 1949 that poliovirus could replicate in cell cultures [118] revolutionised virology by demonstrating that cultivation of viruses in the laboratory could be comparatively cheap and convenient, and did not have to rely on the availability of an appropriate, susceptible animal in order to grow the virus *in vivo*. Since the 1950s, cell culture has remained a fundamental cornerstone of virology, both for studying viral infectivity and pathobiology and for amplifying viruses prior to crystallisation for electron microscopy or for subsequent genomic sequencing.

Traditional approaches to viral discovery using cell culture methods are limited, in part because many viruses cannot yet be cultured [115]. The development of DNA sequencing technologies and associated molecular genetic techniques, such as PCR, have radically reshaped the way in which new, unculturable viruses can be identified, with genomic identification rapidly outstripping any other form of identification of new viruses. The history of the development of sequencing technologies is extensively documented elsewhere, for example, by Heather and Chain (2016) [165], but is reviewed very briefly here. The earliest widely adopted sequencing technology was pioneered by Sanger in 1977 [377], and relied upon the principle of synthesis of a DNA strand with dNTPs, followed by chain termination of a synthesising DNA strand with labelled ddNTPs that lack the 3' hydroxyl group required for further DNA extension. The second generation of sequencing technologies (sometimes termed "next generation sequencing") retained the principle of sequencing by synthesis, but introduced techniques in which huge numbers of sequencing reactions could be performed on different molecules in parallel. These include the bead-based emulsion PCRs required for 454 pyrosequencing, IonTorrent and SoLID, bridge-amplification processes utilised in Illumina sequencing, and electronic detection systems that would allow the high resolution, small scale detection of optical or pH changes resulting from base-addition during synthesis. All second generation techniques are limited by length of sequencing reads, which rarely exceed around 600bp. Between 2004 and 2010, the-per base cost of second generation sequencing halved every 5 months [414], democratising large-scale sequencing projects and resulting in a deluge of new genomic data. Third-generation sequencing technologies, including single-molecule sequencing and nanopore sequencing are still relatively young, but promise to further revolutionise the capabilities of genomic sequencing. The Oxford Nanopore MinION technology, and other nanopore platforms, represent a significant divergence from sequencing-by-synthesis, and instead sequence single strands of DNA by detecting electrical changes as the strands are channeled through pores in a charged plane. Although the error rate of the MinION platform is still very high at a per-base level and therefore currently unsuitable

for many applications, ultra-long read lengths of >800kb can be achieved.

1.2.1.2 Virus discovery in the era of high-throughput sequencing

Classical approaches to viral discovery using molecular genetic techniques have relied on sequence-dependent approaches, which use pre-existing information on the genomic sequences of other, related viruses to design degenerate or specific PCR primers that can amplify a range of sequences from a particular viral family or genus. Depending on the degree of viral diversity contained within different amplicons, this was followed by cloning of the amplicon into expression vectors in order to separate specific viral variants followed by colony PCR and Sanger sequencing of individual variants, or direct Sanger sequencing without cloning.

The rapid leap in sequencing capabilities over the past 15 years has heralded a new era of metagenomics, in which high-throughput sequencing can be used to investigate naturally occurring microbial diversity without prior need for culture in cells. Metagenomic studies can be classed into two groups; "shotgun metagenomics" and "marker genome metagenomics". The latter approach, sometimes also termed "metagenetics", is commonly used for identification of prokaryote or eukaryote species in a sample, as all prokaryotes and eukaryotes have conserved regions that can be amplified using PCR (16S or 18S ribosomal RNA regions, respectively, for prokaryotes and eukaryotes). Due to the absence of a universal genomic marker for all groups of viruses, a meta-genetic approach of this kind cannot be used to characterise all viruses in a sample. The earliest use of a metagenomics study specifically designed to characterise viral diversity, or "metaviromics", was a study that investigated seawater samples [42]. The seawater study demonstrated that viral diversity in seawater was extremely high and that most sequences (65%) had no significant sequence homology to any other previously characterised virus or other organism. The finding that most sequences were unlike any other known organism has been replicated in many subsequent studies across a wide range of sample types, including environmental samples [40, 187, 41, 162].

Although the relationships are not strictly linear [482], it has been estimated that in gastrointestinal and marine environments there are 10-100 times more viral particles present than cells [290, 482]. The longest viral genome belongs to viruses from the species *Pandoravirus salinus*, and, at 2.5Mb, is larger than certain bacterial or archaeal genomes [343, 361]. Most viral genomes are several orders of magnitude smaller than this [52]. Despite the ubiquity of viruses in the environment, the extremely small size of the typical viral genome means that, without enrichment for viral nucleic acid, any sequenced environmental sample of nucleic acid would be almost entirely dominated by nucleic acids co-extracted from prokaryotes and eukaryotes. Modern metaviromic approaches to detection and sequencing of viruses can be therefore be categorised as either sequence-dependent or sequence-independent, depending on whether prior information about the genomic sequence of that virus is relied upon during the enrichment process. Sequence-dependent approaches include probe-based capture followed by high-throughput sequencing. Sequence independent approaches almost always require enzymatic or mechanical enrichment of viral material by filtration through small pore filters, centrifugation, density gradient centrifugation to select for particular density particles, or nuclease treatments to remove unwanted host or environmental nucleic acid. These methods are sometimes followed by molecular amplification of viral material through amplification, such as in sequence-independent single-primer amplification (SISPA) [363] or multiple displacement amplification using phi29 DNA polymerase. Although sequence-independent approaches are often seen as less biased than sequence-dependent approaches, many stages in sequence-independent approaches have the potential to introduce unintended bias towards particular components of a sample (for example, biasing the true proportions of virions of different sizes in a sample towards smaller size virions through the application of high centrifugation speeds [73, 441]).

Currently, the detection of viruses and reconstruction of whole viral genomes from metaviromic data is a significant computational challenge for sequence independent or metagenomic approaches. Classification methods that determine which virus, if

any, is represented by each sequencing read (or contig) are a necessary step of many sequence-independent pipelines. Classification methods most commonly rely on detecting similarities between query sequences and sequences in an annotated reference dataset, such as the NCBI GenBank nucleotide reference collection, using alignment algorithms such as BLAST [6] or DIAMOND [49]. Alternative classification methods to similarity searching include classification techniques based on shared sequence composition (k-mer distributions), for example, PhyloPythiaS+ [151, 286], PhymmBL [38, 39] or Kraken [490]. These algorithms can be useful for classification of viral data that are significantly divergent to known viruses [38].

Construction of whole viral genomes from metagenomic data can be completed using reference based mapping if a template genome is already available, but *de novo* assembly using either *de Bruijn* graph [89] or overlap consensus assembly techniques must be performed if a closely related reference is unavailable. Both of these approaches are computationally intensive. Overlap consensus assemblers are impractical to apply to large datasets, because contig extension relies on exhaustive pairwise comparisons between all unassembled reads and existing contigs [369].

Beyond surveying existing diversity, the potential to use metaviromics to learn about viral ecology and epidemiology in the wild is almost entirely untapped. Given a population in which comparatively little is known about the types of viruses that cause natural infections but for which behavioural or demographic data are available for individuals, it should be possible to use metaviromic data to both discover new viruses and subsequently infer patterns in prevalence by age, sex, social group or season that might provide insights into the drivers of viral spread in the population. Only a handful of papers have ever explored the factors that affect viral presence in wildlife using metaviromics. Temporal variation in viral infection has been investigated for honey bees, although micro-arrays, rather than true metaviromics, were used to detect the presence of infection [372]. A further metaviromic study has demonstrated increased prevalence of simian immunodeficiency in adults and female monkeys [207], as had been observed in previous studies. Geographic differences between animals

at different sampling sites have been shown, with greater sharing of viruses between macaques at the same site than at different sites [9].

1.2.2 Phylogenetic and phylodynamic tools for studying viral epidemiology and evolution

As DNA sequencing and computational technologies have developed that can rapidly generate and process large volumes of sequencing data, the availability of viral sequence data has dramatically increased. Phylogenetic analyses, which attempt to infer genealogical relationships between species or individual viral strains, have proved invaluable for exploring the evolution of different viruses and inferring important epidemiological parameters related to viral transmission. The applications of phylogenetics for exploring viral epidemiological and ecological dynamics are extensive. At the simplest level, differences in tree topologies can be used to explore viral genealogies by providing evidence for host-viral codivergence or lack thereof [145], demonstrating recombination or reassortment events, or helping to ascribe a novel virus to a particular taxonomic group [217]. Phylodynamic approaches exploit the fact that infectious disease behaviour is jointly shaped by epidemiological, immunological and evolutionary processes, and that information about these processes can therefore be inferred from phylogenies estimated from viral genomes [460, 152, 353]. Complex phylodynamic models can be used to explore adaptation of viruses following host-switches [27, 106], outbreak origins and spread [110, 439], or intra-host structuring of viral populations [357].

Estimation of phylogenetic trees from an alignment of sufficiently closely related sequences can be performed using multiple methods, including genetic distance methods, maximum parsimony, maximum likelihood and Bayesian methods [255]. Genetic distance methods such as neighbor-joining require the creation of a pairwise genetic distance matrix that is subsequently transformed into phylogenetic branch lengths between pairs of sequences. Pairwise distance matrices are calculated using substitu-

tion models that account for the effect of incurring different nucleotide or amino acid substitutions at the same site. Using only the raw observed percentage of differences between sequences (p-distance) would underestimate the number of substitutions that has occured between two sequences, as it fails to account for the possibility of multiple substitutions occuring at the same site. However, the information on the exact characters at each aligned site is discarded prior to tree construction using the evolutionary distance matrix, and is not used to further assess support for ancestral relationships between different sequences. Estimated neighbour-joining tree topology is affected by the order of sequence addition [234], and a single phylogenetic tree is produced that is not formally compared to other possible trees. Genetic distance methods are therefore rapid, but are often considered inferior to maximum likelihood phylogenetic estimation techniques. Maximum likelihood trees use similar given nucleotide (or amino acid) substitution models to determine how likely a set of data are given a particular tree topology, with the aim of comparing multiple trees and selecting the tree that generates the highest likelihood. Searching through possible tree topology space is usually made faster through the use of topological rearrangement approaches such as subtree pruning and regrafting or nearest neighbour interchange [255]. For both maximum likelihood and genetic distance methods, confidence in the tree topology can be assessed through resampling approaches, such as bootstrapping. Bayesian methods estimate the posterior probability for a tree given a set of data and prior support for certain parameters. Several algorithms can be used to estimate the posterior distribution, including using Markov chain Monte Carlo methods to converge on the set of trees with the highest posterior probability. Methods for phylogenetic tree reconstruction are reviewed in detail elsewhere (for example, [255]).

Accurately characterising the viral transmission routes, date of introduction and mode of transmission is important for efficient targeting of control and surveillance methods during viral outbreaks. Determining routes and risk factors for virus transmission has traditionally been achieved using classical epidemiological approaches, including descriptive epidemiology, contact tracing, case-control methods and prospec-

tive cohort studies [37]. Reliable estimation of transmission routes is typically dependent on the availability of accurate data on host contact structures or using estimated time and location of viral acquisition to assess plausible routes, yet such data are often unavailable. This is particularly true for epizootics in wildlife, especially when the infected species is highly mobile and non-terrestrial (such as migratory birds or aquatic mammals), or when the infected species avoids humans, is highly camouflaged or nocturnal. Recent advances in genome sequencing and computational analysis present a viable alternative to using classical epidemiological approaches to infer transmission routes. The phylodynamic framework proposes that certain pathogens can mutate so fast that the accumulation of new genetic changes in the genome can occur on the same timescale as ecological changes in the environment of the pathogen. Hence, phylogenies constructed using viral genomes can contain distinct signatures that are jointly shaped by epidemiological, ecological and evolutionary processes.

Incorporating estimates of the evolutionary rate of a pathogen by making an assumption of a molecular clock helps to calibrate phylogenies in terms of real calendar time units [353], which is invaluable if inferences of the probable ecological drivers are to be made from the phylogenetic reconstruction of that outbreak. For example, molecular-clock reconstruction of avian influenza virus H5N8 [439] demonstrates that the timing of introduction of the virus to new geographic regions was compatible with viral carriage by migrating wild waterfowl.

Coalescent theory is an approach that directly links the rate at which branches in a phylogenetic tree coalesce with population-level processes, including changes in population size or structure, and evolutionary selection [255, 219]. Under certain conditions, changes in the size of outbreaks and growth parameters such as the basic reproduction number can be estimated from tree topology. Amongst other studies, this has been used to show that HIV-1 group M underwent an explosive growth in population size around 1960, correlating with a proposed increase in iatrogenic transmission resulting from the introduction of unsterilised injections [127]. Estimates of the basic reproductive number have been calculated from viral sequence data for

viruses such as hepatitis C virus and Ebola virus [351, 410].

Inferring the spatial dissemination and transmission route of a virus is key to designing effective intervention methods for ongoing or future outbreaks. Although maximum likelihood and maximum parsimony approaches are often used for estimation of the geographic route of viral spread [256, 396], Bayesian approaches are particularly amenable for inferring the spatial spread of a pathogen because geographic location at ancestral nodes can be jointly estimated along with the tree topology whilst taking into account phylogenetic uncertainty, rather than being inferred on a fixed tree topology after it has been constructed. Determining transmission routes can include inferring the "transmission tree", i.e., the history of direct transmission between specific individuals [77, 78, 291, 294, 506, 507], or inferring more general movement of a pathogen through geographic space or among host-species [114, 254, 253, 276, 408]. The effect of specific factors on geographic transmission between discrete locations can be explicitly calculated using generalised linear models that parameterise movement between locations as a function of specific factors such as geographic distance or travel-time between locations [252]. This approach has been used to show the importance of air traffic movement on the spread of avian influenza A virus [252]) and dengue virus [309].

1.3 Introduction to the ecology and epidemiology of avian influenza virus

One virus that I focus on particularly throughout this thesis is avian influenza A virus, and I therefore provide a short general introduction of the virus, its ecology and its epidemiology here. Influenza A viruses are one of seven genera of segmented, negative sense RNA viruses in the family *Orthomyxoviridae*. The family includes the genera *Influenzavirus A*, *B*, *C* and *D*, and the genera *Thogotovirus*, *Isavirus* and *Quaranjavirus*. The genome of influenza A virus is composed of 8 segments that

encode the glycoproteins haemagglutinin (encoded by HA) and neuraminidase (NA), the matrix proteins (a single segment encodes M1 and M2), non-structural proteins (a single segment encodes NS1 and NS2), nucleoprotein (NP) and the proteins that form the RNA-dependent RNA polymerase (including PB1, PB2, PA). Structurally, influenza A viruses have a lipid membrane that is taken directly from the host cell during viral budding, into which the HA, NA and M2 proteins are embedded. The core of the virion consists of viral ribonucleoprotein that is composed of the genomic RNA, the polymerase proteins and the nucleoprotein. Influenza viral diversity can be generated from three different processes, including recombination, mutation and reassortment (in which genomic segments originating from different virions in the same genus can be shuffled upon reinfection of the same cell to generate new genomic combinations of progeny, 1.1).

The replication strategy of influenza A virus has been reviewed extensively elsewhere (for example, [494]), but is briefly summarised here. The HA protein of influenza A virus binds to sialic acid that is present on the surface of host cells. Different HA proteins show preferences for binding to sialic acids that have different chemical linkages. The distribution of these sialic acids varies according to host or tissue type, so the specific HA protein can therefore strongly influence the host specificity or pathogenicity of the virus. Following binding and endocytosis into the cell, low pH in the endosome initiates a conformational change in the HA protein that allows the viral membrane to fuse with the endosomal membrane, and allows release of the viral ribonucleoproteins into the cell. The ribonucleoproteins are transported to the host nucleus. In the nucleus, the viral RNA-dependent RNA polymerase synthesises positive-sense mRNA templates for viral protein synthesis by host ribosomes within the cytoplasm and cRNA that is used as a template for transcription of new viral genomic RNA. The HA, NA and M2 proteins are transported to the host cell membrane, genomic RNA and assoicated proteins are packaged into the new virion, and the virus buds from the cell membrane. Initially, the new virion remains still bound to the host cell due to the binding of HA with sialic acid present on the cell surface.

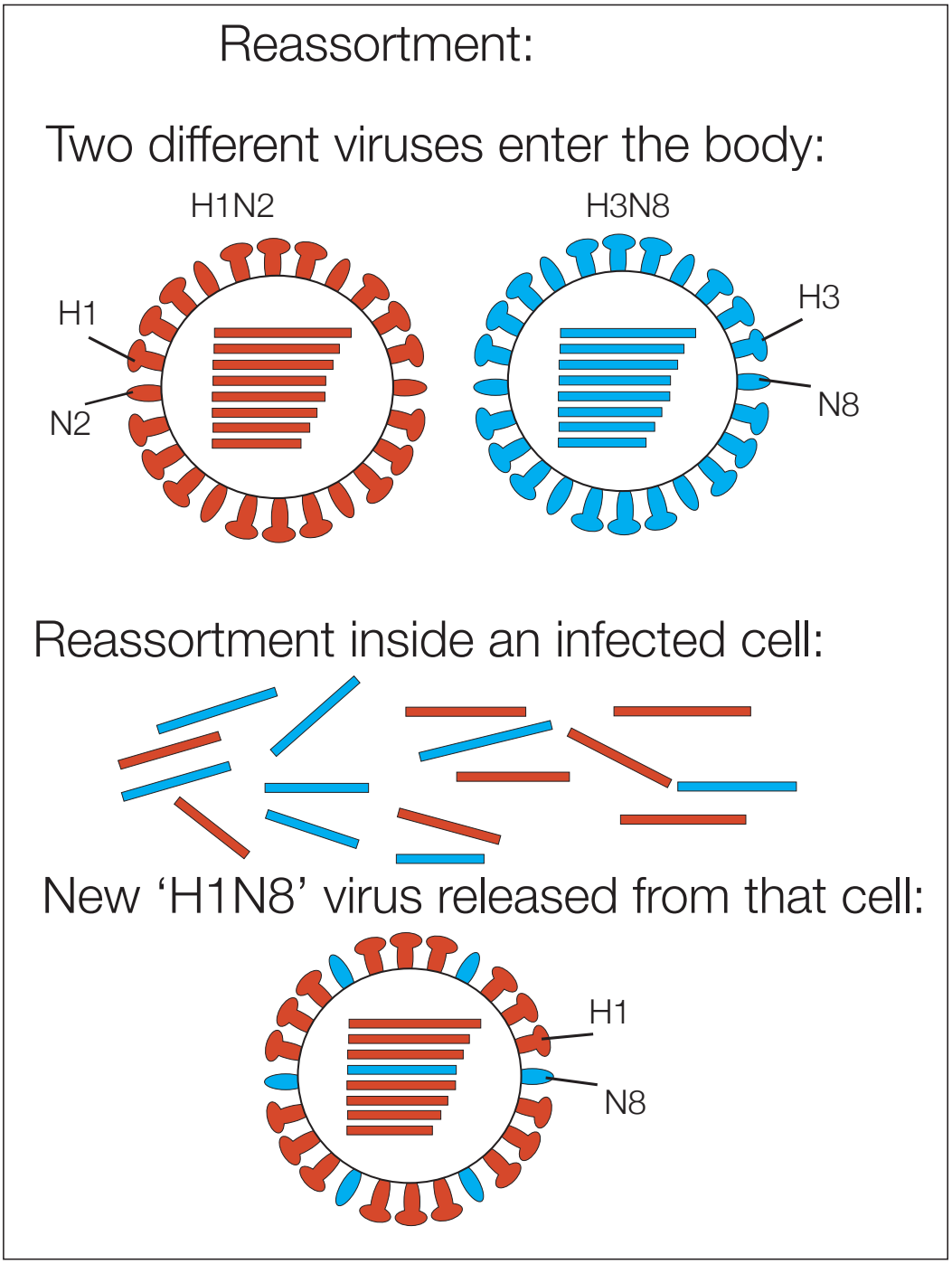

Figure 1.1: Generation of influenza A viral diversity via reassortment. Reassortment and selection acting on spontaneous mutations are the main drivers of new viral diversity for influenza viruses. Note that, whilst one reassortant virus is show here, theoretically reassortment between two different viruses could produce up to 256 different viral genotypes.

The NA protein possesses sialidase activity that removes sialic acid from the host cell, therefore allowing viral release from the cell surface.

As well as the main avian reservoir host, influenza A viruses can infect a wide variety of different animals, including pigs, horses, dogs, bats [445, 446] and humans. Influenza A viral dynamics in humans and birds are fundamentally different. Influenza A viruses can be subtyped on the basis of its two of its segments; the surface antigens, HA and neuraminidase NA, of which there are currently 18 and 11 different recognised forms, respectively. In human populations there is limited co-circulation of different subtypes, such that most human infections observed at any one time are all caused by the same subtype. Circulating subtypes gradually accrue genomic changes that may result in antigenic change as part of a process known as "antigenic drift". In this process, certain random mutations in the viral genome are selected for because they result in variants of the strain that has antigens that are not recognised by host antibodies. Continual antigenic change of the viral strain during its circulation in the population results in the ability of a human host to be infected by the same viral lineage multiple times, as immunity gained during an earlier infection of a virus from that lineage may not be specific enough to neutralise antigenically newly evolved strains. Every 10-50 years, the predominant circulating subtype is replaced by a different subtype ("antigenic shift"). In the past 100 years, there have been at least four major shifts of the dominant subtype, between H1N1 (1918-1957), H2N2 (1957-1968), H3N2 (1968-present) and a reintroduction of H1N1 in 1977 and 2009 [484] that is currently circulating with H3N2. These antigenic shifts are often associated with major pandemics due to immunological naivety of the majority of the global population to the new subtype.

The viral epidemiological dynamics in birds are very different to those in humans. Waterbirds, particularly Anseriformes (ducks, geese and swans) and Charadriiformes (waders, gulls and auks) represent the primary natural reservoir hosts of avian influenza A virus (AIV) [137]. The virus has been detected in at least 100 species from more than 25 families of bird [137, 411]. In birds, multiple different subtypes

co-circulate contemporaneously. The HA and NA segments can co-circulate in sixteen and nine forms respectively in aquatic birds [137], and therefore possible subtype diversity is significantly higher than that in humans, where only three HA types (H1, H2, H3) have ever been observed to cause very large scale human outbreaks. Some antigenic types are not detected evenly across all avian species that harbour AIV, especially H13 and H16 that are almost exclusively detected in gulls and other shorebirds [136, 208, 314]. Because the co-circulating diversity is much higher in birds than in humans, reassortment between subtypes occurs far more commonly in avian hosts than in humans.

AIV can be further described of as highly pathogenic (HPAI) or low pathogenic avian influenza virus (LPAI) based on the clinical phenotype in chickens. The HA protein is synthesised as a precursor HA0, that must be post-translationally cleaved into HA1 and HA2 subunits in order to mediate binding of the virus to host cells and fusion with endosomal membranes [415]. Low pathogenic AIVs can only be cleaved by trypsin-like proteases that are present in the respiratory or gastrointestinal tracts, thereby causing relatively localised infection. Although there is some evidence for differences in feeding and migration behaviour and reduced bodyweight upon infection with common influenza viruses [231, 239, 454], LPAI viruses do not appear to cause mortality or severe disease in birds. Most infections that occur in birds are caused by LPAI virus. However, LPAI H5 and H7 subtypes can sometimes gain multiple basic amino acids at the HA cleavage site through substitution or mutation. This allows the HA protein to be cleaved by proteases found in most cell types and hence the AIV can infect cells throughout the host, causing widespread infection that often results in high mortality rates [35, 417, 461]. Prior to 2005, most HPAI infections have been believed to cause little disease in aquatic birds with few exceptions [494]. The GsGd (A/goose/Guangdong/1/96) lineage of H5 viruses that I focus on in more detail in Chapters 3 and 6 appears to have become a clear exception to this rule. In 2005, an outbreak of H5N1 of this lineage killed more than 6000 wild birds at Qinghai Lake, China. Since then, this virus and reassortants of the subtype has spread widely and

been associated with thousands of deaths of wild birds globally [439].

Most infections in wild waterfowl are LPAI, and therefore significantly more is known about the epidemiology and ecology of LPAI viruses in wild birds than about HPAI viruses. In ducks, LPAI virus is replicated primarily in the gastrointestinal tract and shed in high viral loads via faeces [472], so the transmission route of LPAI viruses is considered to be primarily faecal-oral. The virus can remain infectious for weeks to months in water, lake sediment or faeces [45, 48, 305, 390, 412], with increasing salinity, pH and temperature being inversely proportional to virus persistence [45, 48, 412].

Substantial evidence suggests that prevalence of LPAI virus varies seasonally, which likely relates to the periodic introduction of susceptible, hatch-year individuals and, in migratory species, the effect of seasonal migration patterns. Most long-term studies show substantial inter-annual variation [240]. In ducks in the northern hemisphere, LPAI virus prevalence is believed to be highest immediately prior to and during autumn migration and lowest in spring [228, 240, 296, 325, 412, 462, 471]. Adult birds also typically have lower LPAI virus prevalence than juvenile birds (typically measured as hatch-year) [296, 325, 412, 462]. How birds gain and retain immunity to different subtypes throughout their lifespans has been theorised to cause sequential dominance of different subtypes in birds through cross-immunity induced competition [238] and contact between birds with different immune-profiles as a result of differences in lifespan might drive more rapid emergence of HPAI strains [483].

LPAI viruses sampled in the eastern and western hemispheres tend to form separate phylogenetic clades with rare trans-hemispheric transmission in accordance with structuring of the migratory flyways of the avian host [15, 183, 332]. Furthermore, at an intra-continental scale, viral movement within a flyway is significantly higher than movement between flyways [140, 142, 235, 381]. However, many researchers have claimed that higher virulence of HPAI on the host affects the host ability to migrate to the extent that wild bird migration should be less important for HPAI transmission than trade in live poultry [373].

1.4 Structure of the thesis

This thesis uses the newly developed technologies of sequencing and phylodynamics, described in detail in the previous sections, to improve our understanding of viral transmission and diversity. Drawing on concepts from immunology and ecology, I apply these methods to a number of different study systems, with a particular focus on viral infection in wildlife populations. I describe the novel diversity of viruses in wildlife and explore the processes that may have led to the epidemiological dynamics and macro-evolutionary patterns that we observe.

Chapter 2 describes the genome of a novel polyomavirus in the European badger (*Meles meles*), and uses phylogenetic methods to explore the evolution of host-range in the *Polyomaviridae*. Existing explanations of evolution of the *Polyomaviridae* proposed that the family's evolution was characterised by host-viral co-divergence [384]. The family *Polyomaviridae* has rapidly expanded within the last ten years, and the discovery of many new viral species that do not form clades with other viral species discovered in the same or similar hosts has helped to largely reject this theory. However, comprehensive exploration of alternative processes that could generate the observed evolutionary patterns has not been conducted. I explore the relative impact of cross-species transmission and recombination in shaping the diversity of the *Polyomaviridae*.

The fast rate of evolution of the RNA genome of avian influenza virus (AIV), particularly of the genes encoding the primary viral antigens (haemagglutinin and neuraminidase) means that phylodynamic methods are particularly appropriate for studing the transmission of this virus over short time-scales of months to years. Highly pathogenic avian influenza (HPAI) viruses threaten human and animal health yet their emergence is poorly understood, partly because sampling of the HPAI Asian-origin H5N1 lineage immediately after its identification in 1996 was comparatively sparse. The discovery of a novel H5N8 virus in 2013 (HA clade 2.3.4.4) provides a new opportunity to investigate HPAI emergence in greater detail. **Chapter 3**

investigates the emergence of the HA clade 2.3.4.4 AIV H5N8 in Asia, and how the ecological landscape of poultry and wild-bird density and migration patterns have affected the spread and persistence of the virus.

Chapters 4, 5 and 6 focus on using a single study population to explore the processes shaping viral distribution and diversity in wildlife. This study population consists of a well characterised group of wild, mute swans (*Cygnus olor*) that form a resident population at Abbotsbury, on the Fleet Lagoon in Dorset, UK. Chapter 4 describes the results of a large metaviromic study in this population. Beyond surveying existing diversity, the potential to use metaviromics to learn about viral ecology and epidemiology in the wild is almost entirely untapped. Chapter 4 describes the viromic diversity in faecal samples from 119 swans in this population of well-studied mute swans over a period of 6 months. I describe nine whole genomes of new viral species and use the metaviromic data to explore how age and sex are associated with differences in abundance and prevalence of different viral taxonomic groups.

Despite the importance for understanding epidemiological patterns, the acquisition of immune responses to AIV with age as a result of exposure to multiple strains has never been studied in detail for a long-lived bird. In **Chapter 5**, I perform detailed statistical analyses of the patterns of humoral immune response to different subtypes of AIV, as measured by haemagglutination inhibition assays conducted for five different HA antigens (belonging to three HA types). I study the gain in immunological responses to different avian influenza virus subtypes with age in the Abbotsbury mute swan population.

Chapter 6 presents data from an outbreak of HA clade 2.3.4.4 AIV H5N8 that struck the Abbotsbury mute swan population in winter 2016/2017, causing significant mortality. I describe the outbreak and present evidence for a significant age-bias in mortality that is consistent with the results of Chapter 5. I sequence 12 genomes and conduct phylogenetic analyses to explore the origins of the viral strain that entered the population, including the timing of entry of the virus into the population.

Chapter 2

Discovery of a polyomavirus in European badgers (*Meles meles*) and the evolution of host range in the *Polyomaviridae*

2.1 Introduction

Polyomaviruses are a family of small, non-enveloped icosahedral viruses, comprising approximately 70 putative species from three proposed genera [188, 195]. Polyomaviruses have been found in many different avian and mammalian hosts, including rodents, birds, bats, humans, non-human primates, carnivorans, elephants, dolphins, horses and Artiodactyla. In human populations, polyomavirus seroprevalence reaches up to 90% for BK polyomavirus and up to 80% for JC polyomavirus, and most infections are asymptomatic [453]. However, in immunocompromised individuals infection can cause an array of symptoms, including Merkel cell carcinoma, kidney disease, and progressive multifocal leukoencephalopathy [188]. Although renal and respiratory diseases, tumours and wasting have been reported in some other animals, the consequences of polyomavirus infection in many non-human species remains to be

determined.

All identified polyomaviruses share a similar genome organisation and virion structure. Polyomaviruses typically exhibit a 40-45nm diameter non-enveloped icosahedral capsid, comprised of 72 monomers [16, 188]. The capsid encloses a single double-stranded, circular DNA genome of approximately 5000bp [195]. All polyomavirus genomes are composed of early and late regions, and regulatory regions called the non-coding control region (NCCR) [188]. The NCCR contains the origin of replication, transcription factor binding sites, promoters and enhancers. Transcription from the NCCR is bidirectional, producing mRNA encoding either early proteins, or late proteins [195]. All polyomaviruses produce at least two early mRNA products encoding large (LT-Ag) and small tumour-antigens (St-Ag) [188]. These proteins are involved in viral genome replication and modulation of host cell conditions required for virus replication [416]. The mouse and hamster polyomaviruses also encode a middle tumour-antigen (MT-Ag) [188]. Recently, an alternate reading frame gene (called ALTO) overlapping the LT-Ag gene, has been identified in Merkel cell polyomavirus (MCPyV) and predicted to occur in polyomaviruses that are phylogenetically related to MCPyV [55]. Late polyomavirus mRNA transcripts encode capsid proteins (typically VP1, VP2 and VP3) [188, 195]. Several mammalian polyomaviruses also encode an agnoprotein or other short proteins between the NCCR and the VP2 open reading frame (ORF) [195]. Agnoproteins are involved in control of viral protein expression and have effects throughout the viral life cycle [146].

Most current knowledge about the replication strategy and epidemiology of polyomaviruses has been gained from experimentation with simian virus 40 (SV40), murine polyomavirus (MPyV) and the human polyomaviruses BK polyomavirus (BKPyV) and JC polyomavirus (JCPyV). Mechanisms of replication vary by different viral species, including differences in cell surface receptors and cell entry pathways. Many polyomaviruses recognise certain linkages of sialic acid on gangliosides and glycoproteins, and may also use a variety of co-receptors depending on the species. For example, it is believed that SV40 uses the sialic acid on the GM1 ganglioside

as a receptor, and MHC class I as a co-receptor, whereas MPyV1 uses α2,3-linked sialic acids on GD1a and GT1b gangliosides as a receptor, and α4β1 integrin as a co-receptor [447, 147, 306]. The major cell entry pathway for SV40, MPyV and BKPyV is via caveolin-dependent endocytosis, whereas JCPyV enters via clathrin-dependent endocytosis [306]. The viruses are trafficked from the endosome to the endoplasmic reticulum and on to the nucleus, where the capsid disassembles [188]. In the nucleus, the host RNA polymerase II produces a single mRNA transcript for each of the early and late regions for the genome, which are subsequently alternatively spliced to generate different transcripts. Following the production of the early transcripts and expression of the LT and St-antigens, the viral genomic DNA is produced. The LT-Ag protein fulfils important roles in allowing viral genomic replication, including inducing the host cell to enter the S-phase of its cell cycle that is necessary for viral genomic replication, and by acting as a helicase [188]. The production of LT-Ag controls the switch from expression of the early genes to the late genes [188], as it binds to motifs in the NCCR that prevent the host polymerase from completing further transcription of the LT-Ag and recruits transcription factors that promote late gene transcription. The capsid (late) proteins are synthesised in the cytoplasm and imported into the nucleus, where they assemble into viral particles into which genomic DNA is incorporated [188]. Virions are released via cell-lysis, and may also be released by lysis independent processes based on limited evidence [122].

Polyomavirus infection appears to be ubiquitous in human populations. Serological and genomic data suggests that infection can occur within the first few months of life, and that many infections persist asymptomatically throughout life. The probability of a serological response increases with age for many viruses [188, 459]. JCPyV sampled from the same individuals 4-7 years apart show the same genomic sequences [223]. In healthy humans, polyomaviruses appears to persist within the nucleii of certain host cells, with the kidney being commonly considered an important site of persistent infection for several polyomavirus species. The mechanisms that contribute to persistence are currently poorly understood, but may rely on epigenetic regulation,

immune regulation, or suppression of viral gene expression by viral microRNA [190].

The main mode of transmission of polyomaviruses has also not been conclusively resolved, although common detection of human polyomaviruses JCPyV and BKPyV in urine and sewerage have led to suggestion that urine may be important in the spread of these viruses [22, 29]. Some polyomaviruses, including BKPyV, have been associated with mother-to-child transmission, but this transmission strategy appears to be limited and restricted to only a small number of human polyomaviruses [30, 376]. It is probably that transmission mode and site of persistence varies between different polyomaviruses.

Virus cross-species zoonotic transmission is a frequent cause of emerging epidemics in humans (e.g., influenza A virus, Lassa fever virus, Ebola virus, severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus) [327]. Understanding the evolutionary history of viral families can help define the propensity of viruses to switch host species. However, little is known about the evolutionary history of polyomaviruses, or their capacity for host switching. In an early phylogenetic study [384], the similarity of mammalian and polyomavirus evolutionary trees led to the suggestion that polyomaviruses had co-evolved with their host species. This idea evolved to incorporate a combination of host-switching and virus-host co-divergence events (notably, the basal split between avian and mammalian polyomaviruses) [336]. However, recent statistical re-evaluation that incorporates novel polyomaviruses from a wider range of taxa has largely rejected the hypothesis of polyomavirus-host co-divergence [436, 469].

In the absence of evidence for co-divergence, alternative models have been proposed to explain the phylogenetic distribution of polyomavirus host species. In the first model, cross-species transmission of polyomaviruses is proposed to be a relatively common evolutionary event [436]. This hypothesis has consequences for human and animal health and raises the question of what factors determine the rate of virus transmission between species [419]. In the second model, the polyomavirus phylogeny is proposed to result from recombination among polyomaviruses, which results in the

appearance of host species switching. A recent study found support for recombination of several polyomaviruses, yet was unable to identify ancestors of the putative recombinant lineages with certainty [436]. In a third model, heterogeneity in the evolutionary rates of different genes amongst the polyomavirus lineages is used to explain their complex evolutionary history [436].

Here, I report MmelPyV1 as a novel polyomavirus present in wild European badger (*Meles meles*) populations, and attempt to clarify the evolutionary history of the mammalian polyomaviruses. To my knowledge, this represents the first description of a polyomavirus in *Mustelidae* family, expanding the host range of these viruses for which full length sequences exist to 18 distinct mammalian families. To explore the evolutionary history of this novel virus and the mammalian polyomaviruses as a whole, I performed comprehensive phylogenetic analyses and analyses of viral recombination using whole genome sequences from all available mammalian polyomavirus species.

2.2 Materials and Methods

2.2.1 Viral discovery and sequencing of UK isolate

MmelPyV1 was first discovered in a cell culture supernatant, derived from the apical lobe of a badger lung in Cornwall, England, in 1996 [18]. The culture was established using the mink cell line NBL-7 (ATC CCL 64) by Banks et al. (2002) [18]. The clarified supernatant of this cell culture had been stored at -80°C since 1997 (the "1997 supernatant").

The 1997 supernatant was enriched using a sucrose cushion according to published methods [466], and DNA was subsequently extracted. Randomly-primed amplification (REPLI-g UltraFast reagents) was used to amplify all sample DNA. This method provides robust amplification of all DNA in preparation for deep sequencing, but also works particularly well at amplifying small circular templates such as the polyomavirus through rolling circle replication (reviewed in [195]). This material was

sheared using a Covaris sonicator and standard Illumina paired-end libraries were prepared and sequenced on an Illumina MiSeq. The short read sequence data (1 x 10^7, 149nt reads) were filtered to remove low quality reads using QUASR [470] and reads mapping to mink repetitive sequences.

De novo assembly with Velvet (v1.2.7) and VelvetOptimiser-2.2.0 was used to assemble larger sequence contigs [512, 511]. The resulting contigs were processed using SLIM, an iterative BLAST algorithm [79], to identify viral sequences.

The 1997 supernatant from which the MmelPyV UK isolate was discovered was re-cultured to increase virus stock using mink cell line NBL-7 (ATC CCL 64). Following lysis by freeze-thawing to release any remaining cell associated virus, a new supernatant (henceforth called MusHV1) was produced by centrifugation. To establish whether the polyomavirus was of badger or mink cell line origin, cells from the mink cell line NBL-7 were cultured in Minimum Essential Medium (10% FCS/1% Pen/Strep/Glut), and were either infected with MusHV1 or mock infected. Infection was allowed to proceed until extensive cytopathic effect of the MusHV1 infected sample. Cells and supernatant were harvested, centrifuged (1200rpm x 5min) and the cell pellet was re-suspended. DNA was extracted using the Qiagen DNA Blood and Tissue kit, with lysates being passed through a QiaShredder column to reduce viscosity prior to loading onto the DNA binding columns. DNA was amplified using Invitrogen Accuprime Taq and primers listed in Table A.4. The following reaction conditions were used: 94°C for 5 min and 35 cycles of denaturation (94°C, 30s), annealing (58°C, 30s), and elongation (68°C, 5 min), and a final elongation at 68°C for 10 min. DNA products were resolved by agarose gel electrophoresis and visualised using ethidium bromide staining.

2.2.2 Sequencing of French isolate

Wild badger tissue samples were tested for the polyomavirus in an independent laboratory. Between March and May 2013, eleven badgers were captured in the Chavigny

commune, France, as part of a study into a vaccine for *Mycobacterium bovis*. Following capture, badgers were maintained together in open-air pens. Ethical approval was given by Anses/ENVA/UPEC, which is registered by the French national working group on animal ethics (CNREEA). Legal permissions were obtained from the French Ministry of Higher Education and Research (reference number 00611.02) and from regional governmental committees (Direction Départementale des Territoires de Meurthe-et-Moselle, Direction Départementale de la Protection des Populations de Meurthe-et-Moselle (reference C54-431-1) and Direction Départementale de la Protection des Populations d'Indre-et-Loire (reference C 37-175-3).

DNA was extracted from blood samples using QIAamp kits (Qiagen). Tissue samples were extracted according to published methods [80, 166]. Diagnostic PCRs targeting VP1 and the NCCR were performed in eight sample types from eleven different badgers (Tables A.1 and A.2). A template of 250ng of organ or whole blood DNA, or 5ul of DNA extracted from faeces, was added to the PCR master mix (AmpliTaq Gold reagents (Applied Biosystems)) (primers listed in Table A.4). PCRs were completed under the following conditions: 95°C for 12 min and 45 cycles of denaturation (95°C, 30s), annealing (60°C, 30s), and elongation (72°C, 2 min), and a final elongation at 72°C for 10 min. All PCR products of expected sizes were purified and sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystems) on a 377 automated DNA sequencer (Applied Biosystems). Bases were called using Sequencing Analysis v5.4 (Applied Biosystems) software and the resulting files analysed using Geneious 7.1.4 (Geneious).

One of the DNA samples positive for both the NCCR and VP1 (salivary gland of badger 6) was chosen for genome amplification. Long-distance nested PCR was performed using the TaKaRa-EX PCR system (Takara Bio Inc) and the primers listed in Table A.4. The amplicons were sequenced and analysed as above, producing a full-length genome. To validate the genome, independent overlapping PCRs were completed, and the new amplicons resequenced (Table A.4).

EMBOSS was used to search for tandem repeats and palindromes in the NCCR

[364]. Settings were; palindromes: minimum length 10 and 20, 1 and 3 mismatches respectively, and tandem repeats: maximum repeat size 600, threshold score 12.

2.2.3 Phylogenetic analyses

To construct an alignment of polyomaviruses, 63 whole genome polyomavirus sequences were obtained from GenBank (accessed August 2014), including reference sequences from all known mammalian polyomavirus species and putative novel mammalian polyomavirus species. Avian polyomaviruses were too divergent to be aligned at the whole genome level and were excluded. The two MmelPyV1 isolates were nearly identical, varying at 35 sites (0.67% of the genome), so only the UK isolate was included in the phylogenetic study.

Nucleotide sequences were aligned using Muscle 3.8.31 with default parameters [113], which generated a comparatively poor quality alignment that required manual editing using Se-Al (Rambaut, 1996). Regions where the alignment was highly uncertain (including non-coding regions) were removed, resulting in a "genome-wide" alignment of 3534bp that covered approximately 70% of the average polyomavirus genome length (regions retained are shown in Figure 2.1). Two further alignments were derived from the genome-wide alignment; (i) an "early-gene" alignment (2034nt) and (ii) a "late-gene" alignment (1500nt). Third codon positions were found to be significantly saturated using the test of Xia et al. (2003) as implemented in DAMBE [499, 500, 498], and were removed.

Mega 6.0 [434] was used to test nucleotide substitution models. For both early and late regions, the best-fit model was a general time reversible model of nucleotide substitution, with gamma-distributed among-site rate heterogeneity and a category of invariant sites (GTR+G+I). Phylogenetic trees were estimated using maximum likelihood as implemented in Garli-2.01 [519] from the genome-wide, early-gene and late-gene alignments outlined above. Ten heuristic searches for the maximum likelihood tree were repeated for each alignment and the tree with highest likelihood

retained. One thousand maximum likelihood bootstrap replicates were performed, and the support for each node annotated onto the maximum likelihood tree using the SumTrees functionality of DendroPy 3.12.0 [423].

MrBayes 3.2.2 [368] was used to generate a posterior sample of trees from the genome-wide alignment, using a Markov Chain Monte Carlo of 3 million steps. The first 25% of trees were discarded as "burn-in". Trait values that represent the taxonomic group of the host species of each viral sequence were assigned to the phylogenetic tree tips (see Table A.3). The program BaTS was used (with 200 randomisations of assigned trait values performed to define a null distribution) to statistically test whether viruses found in taxonomically-related hosts are clustered in the virus phylogeny [326].

To assess the presence of phylogenetic signal for recombination among polyomaviruses, the genome-wide alignment was divided into twelve overlapping subgenomic partitions. Starting from position 1, each partition was 600nt long with a neighbouring partition overlap of 200nt. Phylogenetic trees were estimated from each partition using the maximum likelihood approach detailed above (except that two heuristic search and 200 bootstrap replicates were performed for each partition). Visual inspection of these trees revealed phylogenetic inconsistencies between genome regions. Subsequently, one species was chosen randomly from every monophyletic clade that was present in all twelve partition phylogenies. Sixteen clades that were well supported in the whole-genome phylogeny were identified (fifteen clades supported by bootstrap support values of 100% and one by 91%). Ten species that did not belong to any clade were also included in the analysis, resulting in a reduced alignment of 26 polyomavirus species. These 26 taxa were analysed using Recco [282] to seek evidence of recombination. Recco attempts to find the minimum cost of reconstructing each sequence in an alignment from a combination of recombination and mutation of the other sequences, where recombination and mutation have a cost penalty [282]. Probable recombination breakpoints are generated based on the minimum cost solutions. To further investigate potential recombinants, putative

recombination breakpoints were used to define new genomic partitions and maximum likelihood phylogenies were again estimated from each partition.

To determine the divergence of the MmelPyV1 isolate genomes (UK and French), I created a whole genome alignment containing only the two isolates plus the closest related viral species (CSLPyV1) as an outgroup. All three genome positions and almost all regions except the NCCR could be retained (4812 sites). Pairwise genetic distances were calculated using Mega 6.0 [435]. The date of divergence of the UK and French MmelPyV1 isolates was estimated using previous estimates of polyomavirus substitution rates [134].

2.3 Results

2.3.1 Virus discovery in cell culture

The badger polyomavirus was first discovered in a cell culture supernatant, derived from tissues taken from a badger lung in Cornwall, England, in 1996 [18]. The original purpose of the culture was to maintain a badger herpesvirus, mustelid herpesvirus-1 (MusHV-1), present in the original tissue sample [18]. The clarified supernatant of this cell culture had been stored at -80 °C since 1997 (henceforth referred to as "1997 supernatant"). The 1997 supernatant was enriched using a sucrose cushion as part of an attempt to sequence the badger herpesvirus present. DNA was extracted from the sucrose cushion-purified supernatant, amplified using random priming, and sequenced using Illumina deep sequencing.

In addition to herpesvirus contigs, two polyomavirus contigs of 4495nt and 468nt were identified, which appeared to encode a full-length circular genome (Figure 2.1) (here denoted the MmelPyV1 UK isolate; GenBank accession number KP644239). Overlapping identical sequences of 110nt were found at both ends of the molecules. Furthermore, additional reads were present in the short read data that spanned both ends of the sequences. When the original MiSeq reads were mapped to the genome,

1278 reads (of 6.3 million) were identified with perfect homology to the genome. In alignable regions, nucleotide identity with any other polyomavirus genome was never more than 72%, and the closest known viral genome was the California sea lion polyomavirus 1 (CSLPyV1). Under the definition of less than 81-84% sequence identity suggested by Johne et al. (2011) [195], the polyomavirus described here represents a new species, and the species name *"Meles meles polyomavirus 1"* (MmelPyV1) is proposed.

To determine that the polyomavirus was not a contaminant in the mink cell line used to culture the badger herpesvirus isolate, the mink cell line was tested for presence of the virus. Primers designed from the MmelPyV1 genome generated strong PCR bands in a mink cell line sample that had been inoculated with material generated from the 1997 supernatant (Figure 2.2) (see Section 2.2.1 for details). Indistinct or no bands were present in the negative control mock-infected mink cell line sample. This finding is consistent with the polyomavirus being derived from the original badger lung tissue, and not the mink cell line or cell culture reagents.

2.3.2 Virus discovery in wild badgers

Fetal calf serum, which was used here throughout cell culturing, is the source of at least one species of polyomavirus, and other species have been isolated from bovine tissue [328, 334, 380, 513]. To establish conclusively that MmelPyV1 occurs in badgers, uncultured samples from wild badgers in France were tested in an independent laboratory for presence of the virus. Diagnostic PCRs targeting VP1 and the NCCR were performed on faecal samples and seven different tissues from eleven different badgers to determine prevalence. Eight badgers (73%) showed evidence of polyomavirus infection. PCR results were positive at least once in all sample types, except faecal samples which were all negative (Tables A.1 and A.2). Overall, 21 of 87 (24%) samples were positive for MmelPyV1. Samples from liver and lymph nodes exhibited the highest rate of positivity. Neither VP1 nor the NCCR showed any sequence variabil-

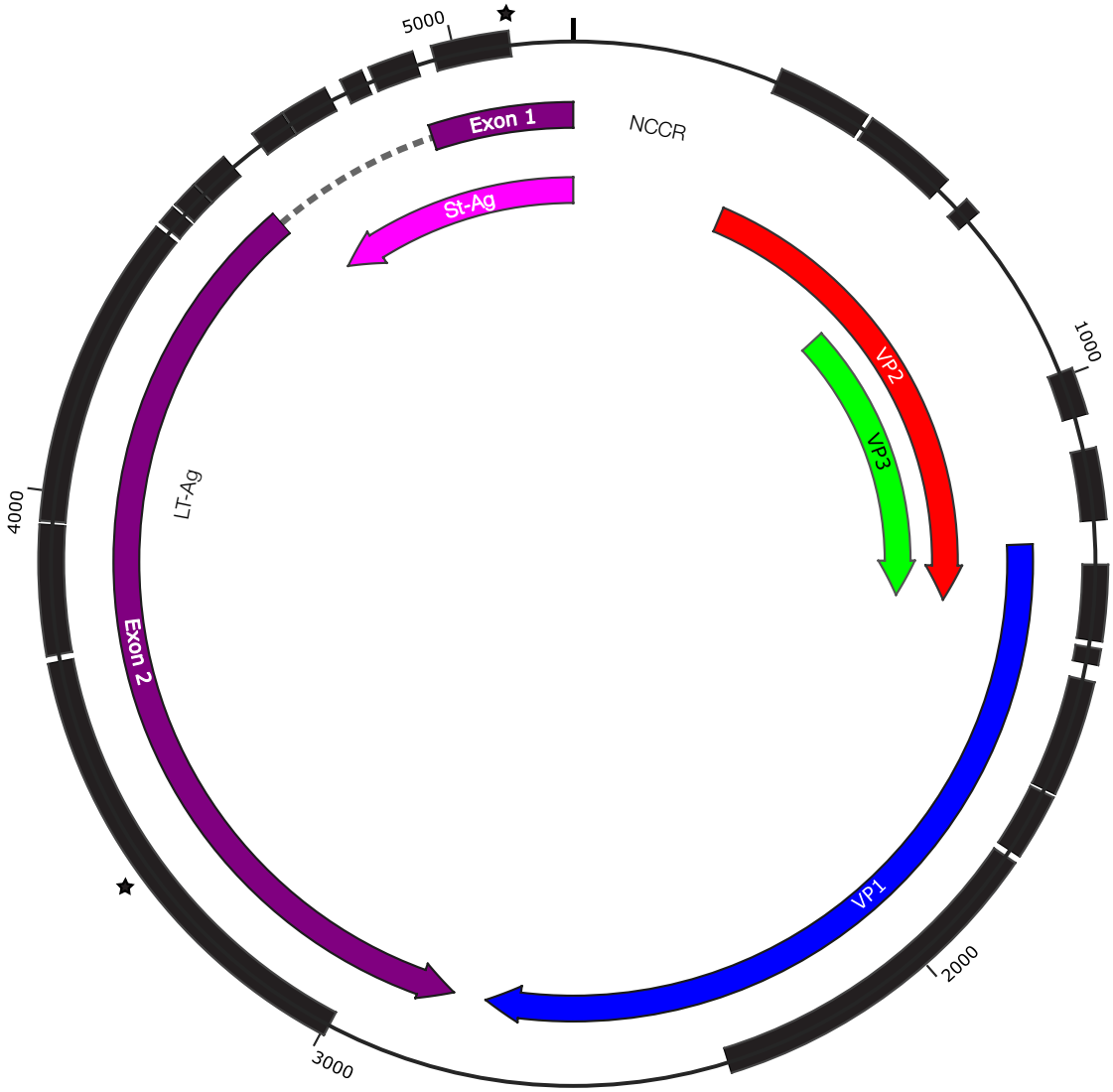

Figure 2.1: Genome map of MmelPyV1. Thick black blocks (outer circle) represent genomic regions that could be aligned across all mammalian polyomaviruses and were therefore included in the "genome-wide" alignment. Coloured arrows represent ORFs. Early proteins are in purple, late proteins in blue, red, and green. Breakpoints used in recombination analysis are marked with two stars.

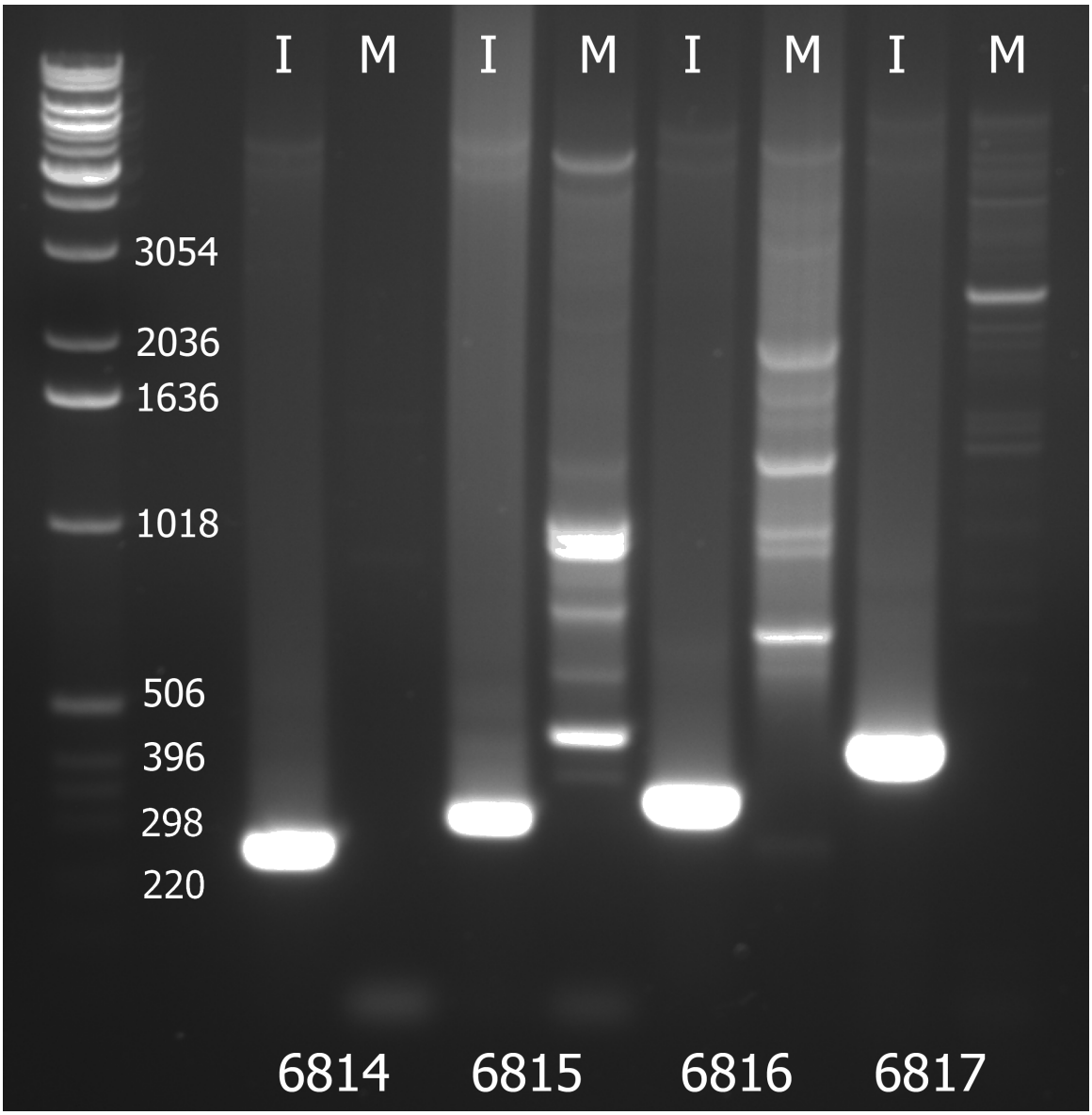

Figure 2.2: Gel electrophoresis of mink cell line NBL-7. Lanes marked "I" are infected with UK isolate cell line and lanes marked "M" are mock infected. Numbers below lanes represent the primers used. Numbers on left represent the ladder fragment size (bp). Controls not shown.

ity across the wild badger samples. Long-distance nested PCRs were used on one sample to generate an amplified genome, which was subsequently sequenced. This generated a further polyomavirus isolate, referred to here as the MmelPyV1 French isolate (GenBank accession number KP644238).

2.3.3 Genome characterisation

Full-length comparison of the French and UK isolates showed that they differ by 34 SNPs and one indel. Thirty of the 34 SNPs fell in coding regions, four of which were non-synonymous (three in VP1 and one in LT-Ag) (Ser1770Thr, Ile2520Val, Ala2557Gly, Ala3044Pro; nucleotide positions given relative to the MmelPyV1 UK isolate genome (comparison to the type species SV40 is inappropriate here due to sequence length variation between MmelPyV1 and SV40 in VP1). The single indel occurred in the NCCR.

The divergence time of the UK and French isolates was estimated based on previously published rates of polyomavirus nucleotide substitution. Assuming a substitution rate in coding regions of 4.34×10^{-5} substitutions/site/year (95% CIs 2.42-6.41 $\times 10^{-5}$ substitutions/site/year [134]), my observation of 30 SNPs in 4791bp of coding sequence suggests that the French and UK isolates diverged approximately 72 years ago (mean: 72 years ago, 95% CIs: 49-139 years ago). This estimate should be considered tentative as it is based on an evolutionary rate for the BK polyomavirus [134], and the rate for MmelPyV1 may be different.

The MmelPyV1 genome is a 5186-7bp circular molecule with a G+C content of 42.8%. These values fall within the range for known mammalian polyomaviruses (length 4697bp to 5722bp; G+C content 36.4 to 47.3%). The genome organisation of MmelPyV1 was typical of the *Polyomaviridae*, encoding ORFs from both strands of the genome separated by the NCCR (Figure 2.1). ORFs encoding known viral proteins were conserved between the MmelPyV1 isolates and their closest relative, CSLPyV1 (Figure 2.3). Proteins homologous to polyomavirus VP1, VP2, VP3, LT-

Ag and St-Ag are present.

The LT-Ag sequence shares many features with other polyomaviruses, including a DnaJ domain [7, 344] (including HPDKGG at positions 42-47; amino acid positions given relative to SV40), Rb binding motif [7, 344] (LRCDE at 103-107), ATPase motifs [344] (GPINSGKT at positions 426-433 and GCVKVNLE at 503-510), zinc finger motif [7, 344] (CMDCLEEQIITHYKYH at 302-317), a TPPK motif [92] (TPPK at 124-127), Bub-1 motif [7] (WERWW at 91-95) and a Cr1 domain [344] (LMQLL at 13-17). Other motifs identified in SV40 appear to be absent (Cul-7: FNXEX [116]). The St-Ag has motifs similar to other polyomaviruses, including two conserved PP2A binding motifs ([116] (CQRNVNPKCRCLMCRLKRKH at 103-122 and WGMCYCYSCYCQW at 135-148).

The NCCR contained three likely LT-Ag binding sites [7, 344]; one GAGGC at nucleotide positions 275-279 and two reverse complement GCCTC at positions 77-81, and 84-88 (positions relative to the UK isolate). Analyses using EMBOSS [364] showed no evidence of tandem repeats or palindromes in the NCCR.

Neither ALTO nor agnoprotein genes were identified in MmelPyV1. A 53 amino acid (aa) ORF was identified overprinting the LT-Ag exon 2 in a frameshift position, in a similar genome position to the previously proposed ALTO gene [55]. However, MmelPyV1 is phylogenetically distinct from the clade of ALTO-containing polyomaviruses (Figure 2.4) and the ORF is considerably shorter than identified ALTO proteins (53aa compared to 248-250aa [55]). Consequently, this ORF is unlikely to represent an ALTO gene. After searching for ORFs that (i) are conserved between the two MmelPyV1 isolate genomes, (ii) are >30aa long (corresponding to the shortest proposed current agnoprotein [288]) and (iii) are located in the forward direction upstream of the VP2 start codon, I found no evidence of an agnoprotein ORF. I discovered one ORF >100aa long in MmelPyV1 (location indicated by grey box in Figure 2.3). This ORF does not have sequence identity with any known polyomavirus proteins and is not conserved between MmelPyV1 and the closely related CSLPyV1, so is unlikely to represent a protein-encoding ORF.

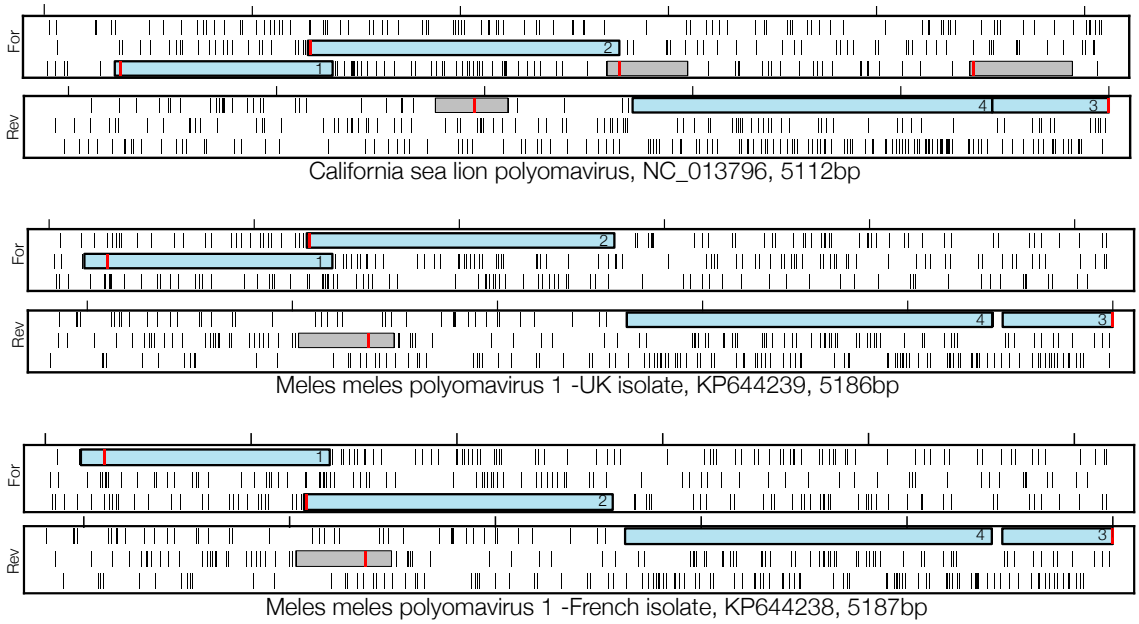

Figure 2.3: Conserved ORFs of the two MmelPyV1 isolates and their nearest relative (CSLPyV1). All 6 translation frames are shown; the three forward (For) and reverse (Rev) reading frames are shown in separate boxes. All ORFs >100 amino acids are shown as boxes, with the first methionine marked in red. Solid vertical lines indicate stop codons, and tick marks indicate 1000bp markers. Blue coloured ORFs represent probable protein coding ORFs identifiable by numbering as follows; 1: VP2 and VP3, 2: VP1, 3: St-Ag and Exon 1 of LT-Ag, 4: Exon 2 of LT-Ag.

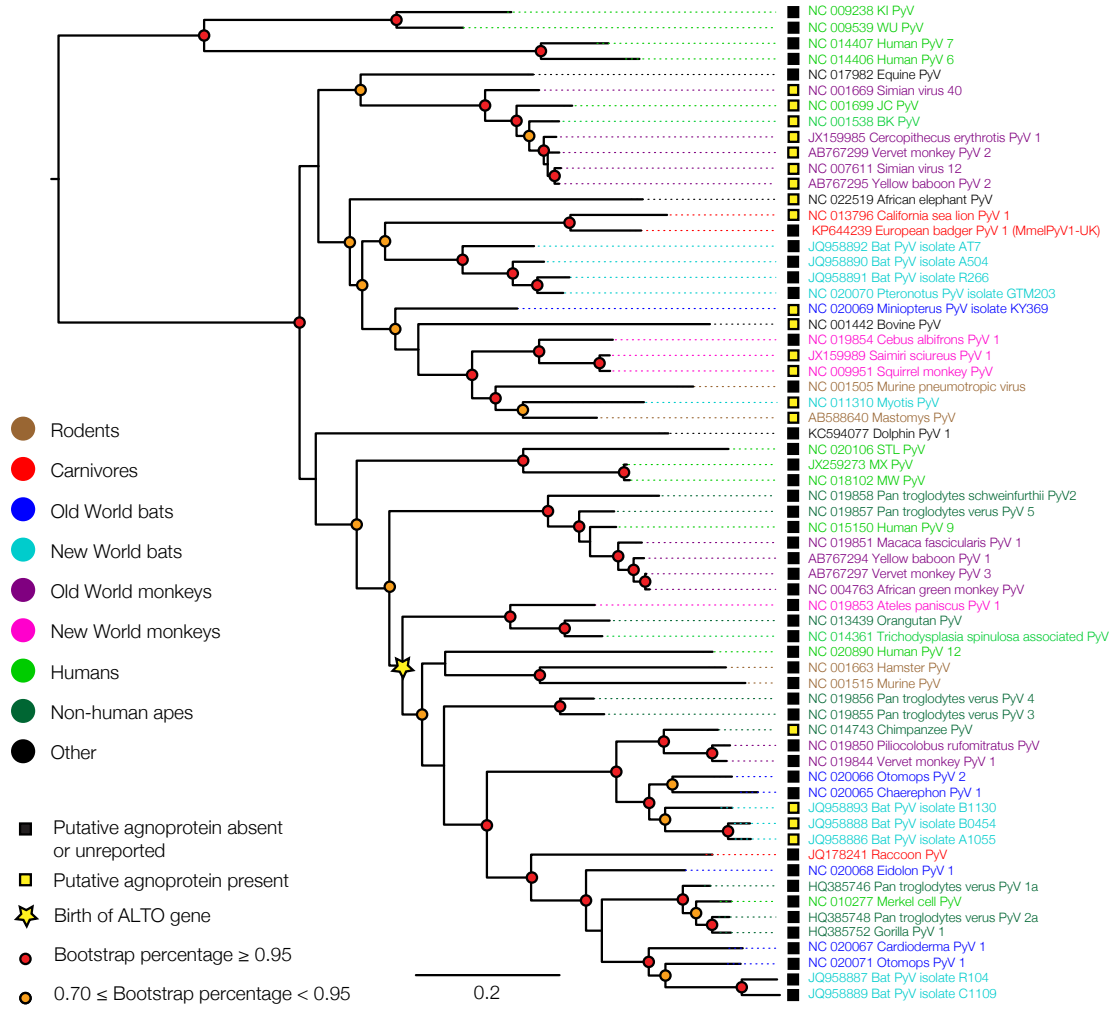

Figure 2.4: Maximum likelihood phylogeny of all known polyomavirus species or putative species, including MmelPyV1, estimated from the "genome-wide" alignment. Viral taxa are coloured according to host species and bootstrap support scores are indicated using coloured circles (details in legend). Scale bar represents expected number of substitutions per site (first and second codon positions only).

It is unknown whether MmelPyV1 causes disease, although none of the badgers identified to be carrying the disease in this studied showed serious clinical symptoms and it therefore appears likely that MmelPyV1 causes asymptomatic infections in animals that are otherwise healthy. The badger from which the UK isolate was derived was malnourished, harboured mustelid herpesvirus 1 (a virus found in most badgers [218]), and exhibited non-specific symptoms of inflammation and lesions in the kidneys, liver and lungs [18]. The badgers from which the French isolate was obtained showed no obvious disease symptoms when examined by a veterinarian.

2.3.4 Phylogenetic analysis of the *Polyomaviridae*

To investigate the evolutionary history of the mammalian polyomaviruses, I constructed maximum likelihood phylogenetic trees based on an alignment of 63 whole genome polyomavirus sequences (obtained from GenBank in August 2014). In a maximum likelihood phylogeny of polyomavirus genomes (Figure 2.4) MmelPyV1 clustered with CSLPyV1 [474]. This result is supported by 100% bootstrap support in all phylogenies I estimated. In the genome regions retained in the genome-wide alignment (but including all three codon positions) (Figure 2.1) MmelPyV1 and CSLPyV1 shared 72% nucleotide sequence identity. Based on a longer alignment of only the badger isolates and CSLPyV1 genomes across all coding regions, the estimated genetic p-distance between the French and UK badger isolates was 0.006 substitutions/site, and between each MmelPyV1 isolate and the CSLPyV1, was 0.334 and 0.335 substitutions/site, respectively.

The maximum likelihood phylogeny based on all known mammalian polyomavirus genomes is comprised of two well-supported clades; a smaller clade containing the KI, WU, HPyV6 and HPyV7 human polyomaviruses and a larger clade containing a broader range of species (Figure 2.4). MmelPyV1 and CSLPyV1 are clustered with monkey, bat, ape, rodent and dolphin viruses, among other hosts. This cluster is distinct from the Almipolyomavirus group that is defined by the presence of the

ALTO gene [55] (indicated by a yellow star in Figure 2.4).

The division into these two clades is not preserved when different parts of the genome are considered separately. Phylogenies estimated from overlapping 600bp partitions of the genome-wide alignment (not shown) show that the WU, KI, HPyV6 and HPyV7 polyomaviruses, which form a strongly supported monophyletic clade in Figure 2.4, are not consistently placed together when different sub-genomic regions are analysed. Such phylogenetic incongruence may indicate that these taxa have recombined in their evolutionary history. Analysis using Recco indicated statistical support for recombination of HPyV6, HPyV7, WU and KI, and also for the miniopterus polyomavirus (small black star in Figure 2.5) (alignment p-values <0.001). Recco also reported recombination in the equine polyomavirus lineage, although this was less well supported ($p=0.026$).

The putative breakpoints in the circular genomes of WU, KI, HPyV6 and HPyV7 were identified as occurring in the NCCR, and towards the 3' end of the LT-Ag exon 2 (breakpoints marked on Figure 2.1 with black stars). These breakpoint positions were used to define two partitions of the genome-wide alignment: (i) partition A, comprising the late region (approximately 1000 nucleotides) plus a short region of the LT-Ag exon 2 (approximately 360 nucleotides), and (ii) partition B, comprising the majority of the early region (approximately 1000 nucleotides of the genome-wide alignment). The phylogeny estimated from partition A (Figure 2.5) shows the WU, KI, HPyV6 and HPyV7 polyomaviruses to be a well-supported monophyletic cluster, placed as a sister group to the remainder of the mammalian polyomaviruses. The phylogeny estimated from partition B (Figure 2.5) does not support this cluster. Instead, WU and KI group together within one clade (red in Fig. 5b) and HPyV6 and HPyV7 group together in a separate clade (blue in Figure 2.5 panel B), in both instances with a bootstrap support of 93%.

Visual inspection of the mammalian polyomavirus phylogenies indicates that viruses isolated from the same host family or order are not always clustered together, yet closely related virus species are often from similar hosts (Figure 2.4). The

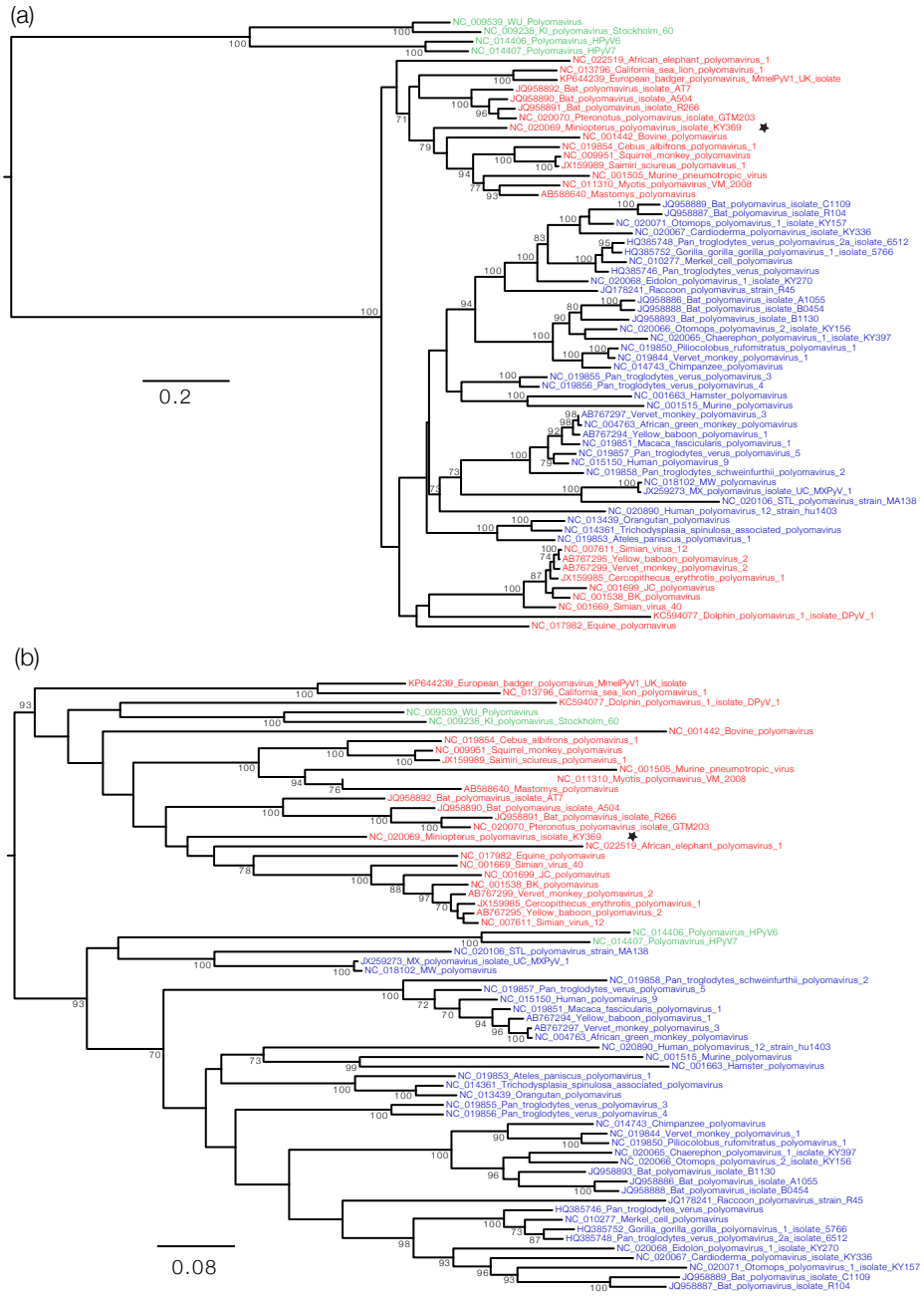

Figure 2.5: Maximum likelihood phylogenies produced from different regions of the genome. Fig. 5a (above): Partition A, comprising the late region, plus a short region of LT-Ag exon 2. Fig. 5b: Partition B, comprising the majority of the early region. Red and blue coloured branches indicate major the two clades as identified in the genome-wide alignment (see Fig. 1). Green branches indicate human polyomaviruses within inconsistent phylogenetic locations. The miniopterus polyomavirus is marked with a star to indicate identification in Recco as a further possible recombinant. Scale bar represents expected number of substitutions per site (first and second codon positions only).

program BaTS [326] was used to quantify the statistical support for an association between the virus phylogeny and host species taxonomy. Host species were assigned to categories in three different ways (Table A.3). In each case there was a significant correlation between host taxonomic category and viral phylogeny (Table 2.1). Thus polyomaviruses from related hosts are grouped together in the viral phylogenetic tree more often than expected by chance (Table 2.1).

Table 2.1: Results of the BaTS tests for association of virus phylogeny with host species taxonomy.

Host species grouping (number of categories)	Mean Association Index value	Mean Parsimony Score value	P-value
Mammalian Order (8)	0.47	13.79	<0.01
Mammalian Family (19)	1.69	26.00	<0.01
Modified Order grouping [1](10)	1.44	26.74	<0.01

2.4 Discussion

Under the species definition proposed by Johne et al. (2011) [195], the polyomavirus reported here represents a novel species and I suggest the name *"Meles meles polyomavirus 1"*. To my knowledge, MmelPyV1 represents the first complete polyomavirus species found in the family Mustelidae. A metagenomic analysis of ferret faeces [403] detected DNA fragments with sequence similarity to polyomaviruses in two ferrets, hinting that polyomaviruses may occur in mustelids [403], a hypothe-

[1]Host species are grouped according to mammalian order, except bats (split into Old/New World bats), primates (split into Old/New World monkeys, and apes), and ungulates (grouped into a single category). See Table A.3 for further details.

sis that has been confirmed here. Following the discoveries of the California sea lion and raccoon polyomaviruses [95, 474], MmelPyV1 represents the third fully described polyomavirus species in carnivorans and the first to be described in a European carnivoran.

The closest relative of MmelPyV1 is CSLPyV1 [474]. Of all known polyomavirus hosts, the sea lion is the most closely related to the European badger. Analyses using BaTS [326] showed significant support for preferential clustering of polyomaviruses from more similar hosts (Table 2.1). Although previous analyses have ruled out strict viral-host co-divergence during polyomavirus evolution [230, 436, 469], this is not inconsistent with the fact that more phylogenetically related polyomaviruses are isolated from more similar hosts than expected by chance. Instead, the evolutionary history of cross-species transmission in the polyomaviruses is likely to result from the interplay of multiple factors, such as preferential host switching among related hosts, viral-host co-divergence, and host species sympatry (e.g. [58, 419]). Understanding the evolutionary history is complicated by the fact that some polyomaviruses have been isolated from captive animals rather than from animals living within their natural host ranges (e.g., [382]). It is possible that animals living in artificially close proximity may share pathogens more frequently than in the wild. Further, such cross-species transmissions may be more likely to cause disease symptoms that are subsequently investigated and identified as caused by novel viruses.

Using a previously published nucleotide substitution rate for BK polyomaviruses [134], I estimate that the two MmelPyV1 isolates diverged around 72 years ago. The polyomavirus NCCR is the most variable genome region, both within and across virus species [90]. The recent divergence of the French and UK isolates is supported by a deficit of genetic differences in the NCCR. The substitution rate I employed is, to my knowledge, the only published rate derived from longitudinal sampling of polyomavirus genomes. There are caveats to the rate used here. First, Firth et al. [134] noted that the BK polyomavirus data set they analysed contained comparatively weak temporal structure, hence their estimated evolutionary rate may be poorly re-

solved. Second, BK polyomavirus is not closely related to MmelPyV1 (especially in the late region; Figure 2.5 panel A) and the extrapolation of evolutionary rates across different viral species may lead to errors in estimates of divergence times.

My phylogenetic analysis of all available genomes enabled us to explore the evolutionary relationships of the mammalian polyomaviruses. Four human polyomaviruses (HPyV6, HPyV7, WU and KI) were placed in different phylogenetic positions depending on the genome region analysed (Figure 2.5). Although this result may be compatible with a recombinant origin for this clade, it is known that distinguishing between genetic signatures of recombination and lineage-specific variation in evolutionary rate is difficult. Tao et al. [436] reported significant evolutionary rate heterogeneity between the early and the late regions in the HPyV6, HPyV7, WU and KI polyomaviruses, but observed no such heterogeneity in other polyomaviruses [436]. The ability of such rate heterogeneity to generate apparent signatures of recombination has been noted for other viruses, including influenza virus [492]. These observations, plus the observation that genetic distances and bootstrap support values are generally lower in the early region phylogeny (Figure 2.5 panel B), suggests that rate variation may be the more likely explanation for phylogenetic incongruence in the mammalian polyomaviruses.

Diagnostic PCRs to determine the presence of MmelPyV1 in different tissues suggested that the virus occurs at highest frequencies in the lymph nodes and the liver, and rarely in blood, faeces or kidney samples. In a study of primate polyomaviruses, viruses were also rarely present in these latter tissues [382]. A previous metagenomic analysis of badger faeces by van den Brand et al. [451] did not identify polyomavirus, supporting the suggestion that MmelPyV1 is rarely found in faeces. Future attempts to characterise MmelPyV1 or to discover new mammalian polyomaviruses should perhaps focus on solid tissue samples rather than blood or faeces.

The known diversity of the polyomavirus genus has increased substantially due to increased virus discovery efforts, yet unanswered questions remain. The true host range and genomic diversity of mammalian polyomaviruses is unknown. Evolution-

ary rates for the family are unclear and a molecular epidemiological investigation of polyomavirus transmission and diversity within a wild animal population has yet to be conducted. Further, it is unknown why only some polyomaviruses are associated with disease and information about the aetiology of polyomaviruses is mostly limited to the viruses infecting humans. This problem is unlikely to be solved without systematic virus sampling in natural host populations, thereby avoiding sampling biases arising from the convenient sampling of easy to catch, captive, high-profile or dead animals. [2]

[2]Sampling viruses exclusively from dead animals might bias our understanding of the true pathogenicity or evolutionary history of a viral family if the death of the animal was a result of the viral infection and under the circumstances that a) most species in the studied viral family are low pathogenic, such that the high pathogenicity represents the exception rather than the rule, b) high pathogenicity of the strain was caused by additional, unidentified host factors (for example, immunocompromisation), or c) the virus infection in that host species represents a recent spillover, such that high pathogenicity is symptomatic of the lack of adaptation of the virus to that new host species.

Chapter 3

Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea.

3.1 Introduction

Highly pathogenic avian influenza (HPAI) viruses are a threat to human and animal health and cause considerable economic damage. H5N1 viruses of the A/goose/Guangdong/1/96 (GsGd) lineage have become endemic in parts of Asia (including Bangladesh, China, India, Indonesia and Vietnam) and in Egypt, and have resulted in the culling of over 250 million birds worldwide [428, 502]. Previous research has tried to characterise where GsGd lineage H5N1 emerged and how it subsequently spread [63, 67, 155, 253, 262, 281, 391, 421, 502]. Unfortunately, comparatively poor sampling in the months immediately following the identification of GsGd lineage H5N1 in 1996 has prevented an accurate reconstruction of its

emergence.

Since 2009, there has been an unprecedented surge in the emergence of novel reassortant H5 viruses of the GsGd lineage [91, 268, 489, 514], most notably including H5N8 [515]. The emergence of these novel reassortant H5 viruses should be easier to investigate than that of the 1996 GsGd lineage H5N1 virus because more viral genetic data is available from the early phase of their emergence emergence. These novel viruses thus provide an fresh opportunity to investigate in detail the factors behind avian influenza virus emergence [456].

The first case of GsGd lineage H5N8 (henceforth referred to as H5N8) in Asia was reported in China in 2010 [515]. In November 2013, several viruses (clade 2.3.4.4; [105]) were isolated from domestic and wild ducks in China, which appeared to be novel reassortants of the 2010 H5N8 virus [124, 249, 495]. By January 2014, the first outbreaks outside of China were noted in the Republic of Korea (ROK; commonly known as South Korea) [249]. H5N8 viruses were found in domestic ducks and wild Anseriformes around the Donglim Reservoir (Jeonbuk province), an important habitat for wild migratory birds [193, 504]. During 2014, outbreaks of H5N8 were confirmed on at least 33 different commerical farms in ROK. The population sizes of birds in affected farms varied between 100 birds to many thousands of birds. A total of 296 H5N8 viruses were isolated in ROK, including 43 from wild birds and 253 from poultry farms. By the beginning of 2015 the H5N8 virus had been reported in wild and domestic bird populations throughout Asia, Russia, Europe, and, most notably, in North America.

The HPAI H5N1 and H5N2 viruses discovered in North America in December 2014 and January 2015 appear to be reassortants of the Eurasian H5N8 virus, carrying an HA segment of the GsGd lineage [330, 449]. Highly pathogenic Eurasian H5 viruses have never been identified before in North American birds [192, 194], so the appearance of Eurasian-like viruses in North America represents an unexpected shift in the global epidemiology of avian influenza.

Wild birds, particularly Anseriformes and Charadriiformes, are considered the

primary natural reservoir for low pathogenicity avian influenza viruses [314, 315]. In contrast, the relative contribution of wild birds versus domestic poultry to HPAI virus persistence, and to its introduction to new locations, remains unresolved [314]. It has been suggested that H5N8 may be maintained and spread by wild birds more readily than the earlier H5N1 viruses of the same HA lineage [121, 203], a difference that might in part explain the recent, notable changes in H5 epidemiology. For at least some species of wild birds, infection symptoms may be mild enough to allow migration whilst infected with H5N8 [193, 203, 213]. Recent studies have found that H5N8 viruses isolated in Europe, Russia and Asia are surprisingly genetically similar given the large intercontinental distances separating the locations from which they were sampled [36, 81, 161, 163, 316, 458]. These observations have prompted suggestions that migrating birds may have recently carried H5N8 from a common location [36, 81, 458]. Testing this hypothesis is of practical importance, as current surveillance measures were largely developed in the context of the earlier H5N1 subtype and may need adapting for H5N8 or other related reassortant viruses if transmission pathways are different.

Phylogeographic analyses can be used to reconstruct the geographic dispersal of a virus lineage from viral genome sequences [173]. Such analyses have been usefully applied to many rapidly-evolving viruses, including GsGd lineage H5N1 (e.g., [253]). Ecological data, for example on poultry, human or waterfowl population densities, can be used to identify risk factors for the emergence of avian influenza (e.g., H7N9 in China [148], H5N1 in China [277]), and can be combined with genetic sequence data from outbreaks [275, 507]. Although interpretation of phylogeographic analyses in the context of known ecology could help illuminate the causes of HPAI virus emergence and spread, no such studies have been attempted for H5N8, probably because insufficient sequences from any one affected region have been available for analysis.

Poultry production in ROK (including both duck and chicken farming for meat and eggs) has rapidly increased over the past 15 years. Census counts suggest over 150 million chickens and over 8 million ducks are farmed commercially in ROK (Re-

public of Korea Livestock Census 2014; http://kostat.go.kr) under high to moderate levels of biosecurity. Poultry are also produced for trading in live bird markets on small scale farms with low to moderate levels of biosecurity. Many species of wild birds regularly migrate to ROK to overwinter, creating distinct temporal dynamics in the available hosts for influenza viruses. Integrating virus genetic information with ecological information on host availability is critical for understanding the spread of avian influenza virus in the country.

Here I investigate the geographic spread of H5N8 with a particular focus on the Republic of Korea. ROK was one of the first countries to report outbreaks in both the 2003 GsGd lineage H5N1 outbreak and the ongoing H5N8 outbreak. I report 49 new sequences of the HA segment of H5N8 isolates from the country. Phylogeographic methods are used to reconstruct the spatial spread of the virus and to investigate how this dispersal was shaped by bird density and migration patterns. Understanding how H5N8 became established in ROK could help inform strategies to prevent future epidemics and to mitigate the risks arising from a greater diversity of HPAI viruses in poultry worldwide.

3.2 Materials and methods

3.2.1 Bird ecology data

3.2.1.1 Longitudinal data on bird counts and waterfowl numbers

Ecological information on seasonal and overwintering wild waterfowl counts was collated. Data was only obtained on wild birds from the family Anatidae, as the H5N8 virus has been mostly isolated from that family. A literature review was undertaken to obtain longitudinal bird count data, under the conditions that the data must include waterfowl counts taken for at least one year, with sampling at least every month (Web of Science search terms "Korea AND (longitudinal OR temporal OR seasonal OR changes OR dynamics) AND (bird OR waterfowl OR duck)", with papers in En-

glish and published after January 2000 considered). Longitudinally sampled waterfowl counts were obtained for four different sites in ROK: (i) Mokpo Namhang Urban Wetland, Jeonnam province (daily waterfowl counts April 2006-July 2010; Birds Korea), (ii) Sihwa Lake, Gyeongii province (monthly waterfowl counts January to December 2009; Park et al., 2011), (iii) Nakdonggang Estuary, Busan province (monthly Mergus albellus, Mergus serrator and Bucephala clangula count, 2002-2008; [176, 177], and (iiii) Junam Reservoir, near Changwon, Gyeongnam province. Province locations are shown in Figure 3.1. Waterfowl counts were scaled so maximum count at each site is equal to one. Data were plotted using R version 3.1.1 [356] and the package ggplot2 [481], as implemented in RStudio [371].

Maps of estimated numbers of the four most common waterfowl in ROK were obtained from the ROK Ministry of Environment Wild Bird Census for Winter 2014 report [211]. The Ministry simultaneously observed 195 national wildlife reserves for migratory birds from January 24th to 26th. At each reserve, bird numbers were counted by two specialists using line and point census methods. For the line census, researchers counted birds whilst walking along roads. For the point census, total bird numbers were counted using binoculars from a single observation point, usually on the water.

3.2.1.2 Poultry density and temporal data

Maps of livestock distributions for the Republic of Korea were produced for domestic ducks and chickens, using the Gridded Livestock of the World version 2 [488, 366]. Briefly, the GLW version 2 uses sub-national statistics and predictor variables of livestock density to model the density of livestock at a 1km by 1km scale. Data on seasonality of poultry counts throughout the year (not shown) were obtained from the Republic of Korea Livestock Census 2014 (published by Statistics Korea; http://kostat.go.kr). The census is based on poultry counts for a single day in each quarter, for all farms breeding >3000 chickens or >2000 ducks.

3.2.2 Sequence data

In total, 296 H5N8 viruses were isolated in ROK in 2014. The HA genes of 49 H5N8 isolates from ROK were sequenced to complement previously published sequences. Isolates were chosen for sequencing in order to generate a dataset that, when combined with previously published sequences, included at least 4 strains per month and 1 strain per province (Tables B.1 and B.2). Viral RNA was extracted from the allantoic fluid of embryonated eggs using the Viral Gene-spin viral DNA/RNA extraction kit (iNtRON) according to the manufacturer's instructions. The HA gene was amplified with gene-specific universal primers (Hoffmann et al., 2001), using the One Step RT-PCR Kit (Qiagen). PCR products were purified from agarose gels using the QIAquick gel extraction kit (Qiagen). Full genomic DNA was sequenced by Cosmo Genetech (Seoul, South Korea) with an ABI 3730 genetic analyzer (Applied Biosystems). Contig assembly was performed using CLC Main Workbench Ver. 6.8.2 (CLC bio). GISAID accession numbers for these new sequences are EPI573192, EPI573195-EPI573242 (Table B.3).

In addition to the 49 new sequences, 51 HA clade 2.3.4.4 sequences were downloaded from the Influenza Research Database (FluDB), and 22 HA sequences were downloaded from the Global Initiative on Sharing Avian Influenza Data (GISAID) EpiFlu. This resulted in a dataset of 122 HA sequences, of which 88 were from ROK (see Table B.3 for details). Acknowledgements of originating and submitting laboratories for GISAID sequences are provided in Table B.3. Geographic locations, collection dates and hosts are also provided. All sequences were combined and aligned using Muscle 3.8.31 [113] as implemented in Mega 6.0 [435]. The alignment was trimmed to coding regions only.

3.2.3 Bayesian molecular clock phylogeography

3.2.3.1 Model selection

Initial maximum likelihood trees were generated for the HA segment using Garli 2.01 [519]. Path-O-Gen [358] was used to confirm strong temporal signal and the appropriateness of a molecular clock approach. Coalescent and nucleotide substitution models were subsequently chosen following model comparison in BEAST [109, 108]. All combinations of the following models were compared: constant size versus Bayesian skyline (five groups) coalescent models and SRD06 versus General Time Reversible nucleotide substitution models, with gamma distributed rate heterogeneity. The SRD06 nucleotide substitution model and Bayesian skyline coalescent tree prior were chosen for all subsequent analyses based on model comparison using the harmonic mean estimator and Akaike information criteration, as implemented in Tracer 1.6 [360].

Early results obtained during model selection suggested that Maximum Clade Credibility (MCC) trees included a clade of European, Russian and Asian samples (clade C4), within which internal nodes were poorly supported. In an attempt to clarify this portion of the tree topology, available NA sequences for clade C4 strains were also included (see below; Table B.3). Maximum likelihood trees were generated using Garli 2.01 [519], and a concatenated alignment of HA and NA sequences was scanned for recombination using GARD [98, 347, 348] to check that no isolates within clade C4 had undergone reassortment in the NA segment.

A prior distribution for the rate of molecular evolution was chosen based on a preliminary analysis in BEAST. Specifically, 30 sequences from Europe and Asia spanning the period 2007-2014 were downloaded from the Influenza Research Database for both the HA (H5, clades 2.2 and 2.3) and NA segments (N8). For both segments, the 30 reference sequences and available H5N8 sequences were combined and aligned using Muscle, as implemented in Mega 6.0 [113, 435]. BEAST was used to generate estimates of the nucleotide substitution rate for the HA and NA segments from these temporally-structured alignments using the relaxed lognormal molecular clock, which

were then used as informative priors of the substitution rate parameters in subsequent analyses.

3.2.3.2 Phylogeographic analyses

Discrete phylogeographic analyses were performed to reconstruct the geographic spread of the virus, whilst simultaneously estimating the phylogenetic tree. These analyses were implemented in BEAST [109, 108, 253], using the BEAGLE library [12]. An asymmetric model was chosen to allow different rates of lineage movement in opposite directions between each pair of locations [114]. Analyses were performed under the SRD06 substitution model, using an uncorrelated lognormal relaxed clock and Bayesian skyline coalescent model. For most isolates, only the HA segment sequences were used, but for sequences in clade C4 NA sequences were also included, where available. Molecular clock and substitution models were unlinked for the two segments, whilst the phylogeny itself was linked. Sequences were coded by area (Figure 3.1). A fully annotated MCC phylogeny is provided in Figure B.1.

Discrete phylogeographic analyses were conducted both with and without Bayesian stochastic search variable selection (BSSVS). BSSVS identifies which geographic links are strongly supported by the data, by assuming that for many possible pairs of locations there will be limited or no observations of viral lineage movement [253]. Between two and four independent runs of 200 million steps of a Markov chain were performed for each analysis, with sampling every 20000 steps. Convergence of the runs and effective sample sizes were checked using Tracer (http://tree.bio.ed.ac.uk). Runs were combined after discarding 20 million steps as burnin. MCC trees were generated using TreeAnnotator and visualized using Figtree (http://tree.bio.ed.ac.uk).

A reduced dataset was generated to exclude sequences that were descended from specifically identified long branches in the MCC tree (see Results section 3.3 and Discussion), reflecting my view that these branches represent virus re-introduction from outside ROK, not cryptic persistence within ROK throughout 2014. Using a

subsampled empirical tree distribution of 1000 trees generated during discrete trait reconstruction with BSSVS on this reduced dataset, Markov jump analysis [287, 311, 433] was performed to estimate the number of jumps between pairs of locations, for each with a Bayes Factor >10. SPREAD was then used to calculate Bayes Factor support for the location rate indicators [28]. These values give an indication of the intensity of movement among locations.

Heterogeneous sampling may affect phylogenetic inference, with less well-sampled locations tending to be inferred as sink regions. Because H5N8 sequence data is relatively limited, it was not possible to use downsampling approaches (e.g., [127]) to create a dataset with equal sample sizes for each region. To assess the effects of including locations for which only one sequence is available, I re-ran the phylogeographic analysis with BSSVS whilst (i) removing locations with only one sequence from the dataset, or (ii) combining locations with only one sequence with the geographically closest province (as per the case map in [504]; Figure B.2).

To further investigate the phylogenetic structure of clade C4, monophyly statistics and TMRCAs for defined groups of viruses within this clade were estimated. Both HA and NA sequences were included where available for this clade (Table B.3). Statistics were estimated under three models (no phylogeographic reconstruction, phylogeographic reconstruction without BSSVS and phylogeographic reconstruction with BSSVS). The monophyly statistic represents the posterior probability that the defined groups of sequences are monophyletic. Each defined group (Table 3.1) includes all C4 isolates from the countries named.

3.3 Results and Discussion

3.3.1 Density and temporal dynamics of poultry

Relevant ecological data sets were collated for ROK. Estimates of livestock density were obtained for ROK [366] and are presented in Figure 3.1 panels A and B. Chicken

density is relatively homogeneous among provinces (Figure 3.1 panel A). In contrast, domestic duck densities (Figure 3.1 panel B) are noticeably higher in western mainland areas (Jeonbuk, Jeonnam, Chungnam, Chungbuk and Gyeonggi) than in eastern areas (Gyeongnam, Gangwon, Daegu, Gyeongbuk and Ulsan; see Figure 3.1 panel 1C for locations). Longitudinal data from commercial farms in ROK (Republic of Korea Livestock Census 2014) suggest that there are small but regular summer peaks in the number of domestic chickens in ROK. No seasonal changes in domestic duck intensity were observed in commercial farms. Although weak seasonality in domestic poultry numbers on commercial farms in ROK is unlikely to affect H5N8 dynamics, I cannot rule out seasonal variation in poultry numbers on smaller subsistence plots, as has been reported in China free-grazing duck poultry production systems [53].

3.3.2 Density and temporal dynamics of wild waterfowl

Maps generated from the Ministry of Environment Wild Bird Census for Winter 2014 indicate that overwintering waterfowl mostly inhabit the west of ROK (Figure 3.2). The most common waterfowl in winter in ROK, the Baikal teal (Anas formosa), occurs in high numbers only on the west coast (Figure 3.2 panel A). The other three most common waterfowl (mallard, *Anas platyrhynchos*; bean goose, *Anser fabalis*; spot billed duck, *Anas poecilorhyncha*) also mostly inhabit the west of the country (Figure 3.2 panels B-D). Counts of waterfowl show that the period of peak migration into ROK occurs between October and January each year, with the greatest number of waterfowl present in January (Figure 3.3 panel A). Due to a paucity of published raw longitudinal data, the wild bird counts are derived from only four sites in ROK. These data therefore will not reflect the full diversity of local flyways or of migration patterns for different species. Fluctuations that reflect local conditions, such as the effects of frozen wetlands on bird density, or the effect of averaging bird migration patterns across a variety of waterfowl species, may also be present. However, annual trends in waterfowl migration into ROK are consistent among the four sites. The observation

that high numbers of birds migrate to ROK to overwinter between October and March is further supported by other published longitudinal studies from which suitable raw data was not available [186, 244, 303, 389].

3.3.3 Phylogeographic analyses of H5N8

To reconstruct the phylogenetic history and geographic dissemination of H5N8 in ROK, I performed discrete phylogeographic molecular clock analyses using BEAST. The maximum clade credibility phylogeny (Figure 3.3 panel B) strongly supports two lineages, denoted Groups A and B following the nomenclature of Jeong et al., 2014 [193]. Group B includes viruses that have a different internal gene constellation to Group A isolates. Viruses from Group A dominated the outbreaks in ROK [193, 249].

3.3.3.1 First wave of viral entry to ROK

My data indicate that H5N8 first entered ROK via the province of Jeonbuk (Figure 3.4). The ancestral node of all Group A Korean isolates (black square in Figure 3.3 panel B) has a high probability of being located in Jeonbuk (location posterior probability = 0.99). This result appears to be robust to potential undersampling because the basal sequences of several well-supported clades were sampled in Jeonbuk and exhibit short terminal branch lengths. Furthermore, the only two Korean sequences in Group B are also both from Jeonbuk.

My analyses suggest that H5N8 first entered ROK at a time and in a place associated with the entry of wild waterfowl during winter migration. Specifically, the most recent common ancestor (MRCA) of the Group A sequences from ROK (black square in Figure 3.3 panel B) is estimated to have existed in Jeonbuk in mid-November 2013 (95% highest posterior density (HPD) interval = mid-October to mid-December 2013). This interval is coincident with the season during which overwintering waterfowl arrive in greatest numbers into ROK (Figure 3.3 panel A). Large numbers of wild waterfowl overwinter in Jeonbuk province, especially Baikal teal (Figure 3.2). Conse-

quently, the arrival of H5N8 to Jeonbuk is consistent with migrating wild waterfowl carrying the novel H5N8 subtype to ROK during their winter migration. That the first cases of H5N8 in both poultry and wild birds were identified near an important habitat of wild migratory birds (Donglim Reservoir [193]), supports the conclusion that wild waterfowl likely introduced H5N8 to ROK. The index cases of the November 2006 [248], April 2008 [214] and December 2010 [212] outbreaks of GsGd lineage H5N1 in ROK were also from Jeonbuk, suggesting that risk factors for the introduction of HPAI may be common to both strains.

I cannot formally exclude the possibility that this first wave of virus entry may have involved multiple introduction events to ROK, as opposed to a single introduction followed by local dissemination. However, isolates from the first wave were detected over a very short time window and are typically linked by short branch lengths in the molecular clock phylogeny (Figure 3.3 panel B), which supports a single epidemic origin. Furthermore, epidemiological reports of the spatio-temporal incidence of H5N8 in ROK support within-country transmission from a single epidemic origin in Jeonbuk to neighbouring provinces to the north and south [193].

3.3.3.2 H5N8 spread within ROK

Following its establishment in Jeonbuk, H5N8 spread widely across the west of ROK (Figure 3.4). There is strong support for lineage migration between Jeonbuk and Jeonnam in the southwest (Bayes factor = 141). Northwards dispersal of the virus from Jeonbuk to Chungnam and Chungbuk along the west of ROK is also well supported by the analysis (Figure 3.4; Bayes factors = 12208 and = 269, respectively). These western provinces, among which most transmissions have occurred (Figure 3.4), are characterised by high domestic duck densities and high numbers of overwintering waterfowl (Figures 3.1 and 3.2).

By contrast, all eastern provinces (including Daegu, Ulsan, Gyeongbuk, Gangwon and Gyeongnam) are characterised by a combination of (i) few H5N8 outbreaks, (ii) little or no onward transmission, (iii) a lack of phylogenetic clustering between

geographically nearby isolates, and (iv) low domestic and wild duck density. Despite being geographically proximate in the southeast of ROK, the three sequences from Daegu, Ulsan and Gyeongbuk are clearly separated from each other on the MCC tree and therefore each represents a separate virus introduction event. These results suggest that these eastern provinces may be "sink" regions, within and from which there is little or no onward transmission (Figures 3.1 and 3.4). Five of the six isolates from these eastern provinces are from domestic chickens, which is an unusually high proportion for my dataset and reflects the lower density of Anatidae hosts there. I suggest that the low density of wild waterfowl and domestic ducks prevents the establishment of persistent chains of H5N8 transmission in the eastern regions of ROK.

While H5N8 appears to have persisted more in waterfowl-rich areas in the west and persisted less in waterfowl-poor areas in the east, it is difficult to determine the relative contributions of domestic ducks and wild waterfowl to this observed pattern. This is because, at the provincial level, areas of high domestic duck density and areas harboring many overwintering wild waterfowl coincide. I note that the longest persisting clade in ROK, clade C1 in Figure 3.3 panel B, was isolated almost exclusively from domestic birds in a single province over a period of more than six months. This persistence in domestic birds in a single region contrasts with non-persisting Asian outbreaks in clades C2 and C4, which were sampled from four different East Asian locations and mostly from wild birds. Thus domestic poultry appear more important in the persistence of H5N8 within ROK, whilst wild birds appear key to viral introduction to new locations.

My observations that waterfowl densities are more important than domestic chicken density for regional emergence, persistence and dissemination of H5N8 are consistent with known pathological effects of the virus. H5N8 causes lower mortality in ducks than in chickens [202, 203, 216, 405, 495, 515]. Infections are harder to identify in ducks than chickens because symptoms are typically less severe, so duck cases may be detected later and culled less rapidly than those in chickens. Thus

waterfowl may be more effective vectors of H5N8 than chickens, congruent with my observations that waterfowl-rich areas are more important in the emergence of new outbreaks than areas rich in domestic chickens. As was noted previously in relation to GsGd lineage H5N1 prevalence [193], the observed higher prevalence of H5N8 in ROK compared to Japan may thus be explained in part by the larger size of the domestic duck populations in ROK.

My dataset included different numbers of sequences for each location, which has the potential to affect phylogeographic inference, particularly for undersampled locations which are more commonly inferred as sink regions (Figure B.2). Whilst the general trend of widespread transmission in the west of ROK and rare transmission to the east of ROK appears relatively robust and is supported by epidemiological information, the exact inferred trajectory of H5N8 spread to the east of ROK appears sensitive to sampling (Figure B.2). Larger and more comprehensive datasets would enable us to sub-sample isolates so that sample sizes are homogeneous among locations, which would allow us to definitely rule out any potential sample size effects.

3.3.3.3 Proposed second wave of virus transmission

Internal branches in the molecular clock phylogeny (Figure 3.3 panel B) that represent H5N8 transmission within ROK following its introduction in 2014, are generally short and almost all are of less than 8 weeks duration (Figure 3.3 panel B). However, three long branches are notable outliers from this pattern, each representing an internal branch of more than 17 weeks duration (dashed branches in Figure 3.3 panel B). The sequences from ROK (and Japan) that are descended from these three long branches were all sampled between 18th November and 19th December 2014.

There are two interpretations of these long branches, depending on whether the phylogeographic results are interpreted naïvely, or if they are interpreted in the context of known bird ecological data. Considered naïvely, the long branches would represent unsampled persistence of H5N8 within ROK during summer 2014. If so, then coincident sampling of these lineages in late 2014 requires a specific explanation

(for example, a sudden increase in sampling at that time, which identified multiple infections that had been overlooked during the summer). However, evidence for the persistence of clade C1 in domestic ducks in Jeonnam throughout the summer of 2014 indicates that sampling was undertaken, at least in parts of ROK. I suggest that this ecologically uninformed interpretation is implausible, and I favour an alternative that is informed by the ecological data. It is important to note that each long branch spans a period when few waterfowl are present in ROK (grey shading in Figure 3.3 panel B), yet the tips of these long branches were sampled during the period of peak inward migration of overwintering waterfowl. It is therefore much more likely that these clade C2 and C4 strains represent re-introductions of H5N8 into ROK via wild birds, and do not result from cryptic or unsampled persistence of H5N8 within ROK. Hence the long branches most likely represent transmission from an unsampled reservoir. Existing satellite tracking, GPS and modelling data suggest that migratory waterfowl typically migrate to or via ROK from sites in northern or northeastern China, Russia and Mongolia [432, 442, 458], so the unsampled reservoir may be located in one of these regions. This interpretation implies at least three separate introductions of H5N8 to Asia in late 2014 from an as-yet uncharacterized source as part of a "second wave" of virus entry.

The same model of transmission from an unsampled reservoir can explain the long-branches leading to North American (denoted clade C3 on Figure 3.3 panel B) and European and Russian (denoted clade C4 on Figure 3.3 panel B), North American and Russian samples from the same time period. The long internal branch that leads to clade C4 (Figure 3.3 panel B) contains viruses isolated in Russia, ROK, Japan and Europe. The presence of isolates from West Europe, Russian and East Asia in this clade is very striking, especially given the large geographic distances between these locations. Careful study of clade C4 shows that the phylogenetic topology within it is not certain, as evidenced by overlapping TMRCAs for several internal nodes in C4 (Table 3.1, Figure B.3, B.4 and B.5) and low posterior support for many nodes within the clade (Figure 3.3 panel B). There is a very low probability that all the

European isolates within C4 form a strictly monophyletic cluster (Table 3.1), whereas there is greater statistical support for the grouping of three Asian samples with three European samples, to the exclusion of other European isolates (Figure 3.3 panel B; Table 3.1).

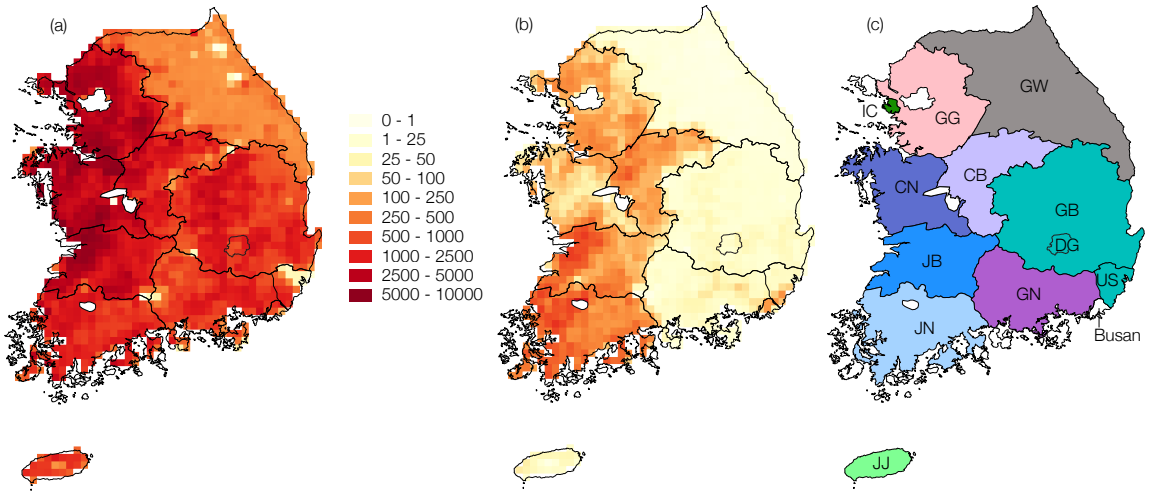

Figure 3.1: Maps showing domestic poultry density (number per kilometer, colours in key) in ROK according to the Gridded Livestock of the World 2.0 (Robinson et al., 2014). (A) Domestic chicken density. (B) Domestic duck density. (C) Map of provinces. Colours correspond to the branch colour scheme used in Figure 3.3. Province abbreviations are as follows; CB: Chungbuk, CN: Chungnam, DG: Daegu, GB: Gyeongbuk, GG: Gyeonggi, GN: Gyeongnam, GW: Gangwon, IC: Incheon, JB: Jeonbuk, JJ: Jeju, JN: Jeonnam, US: Ulsan.

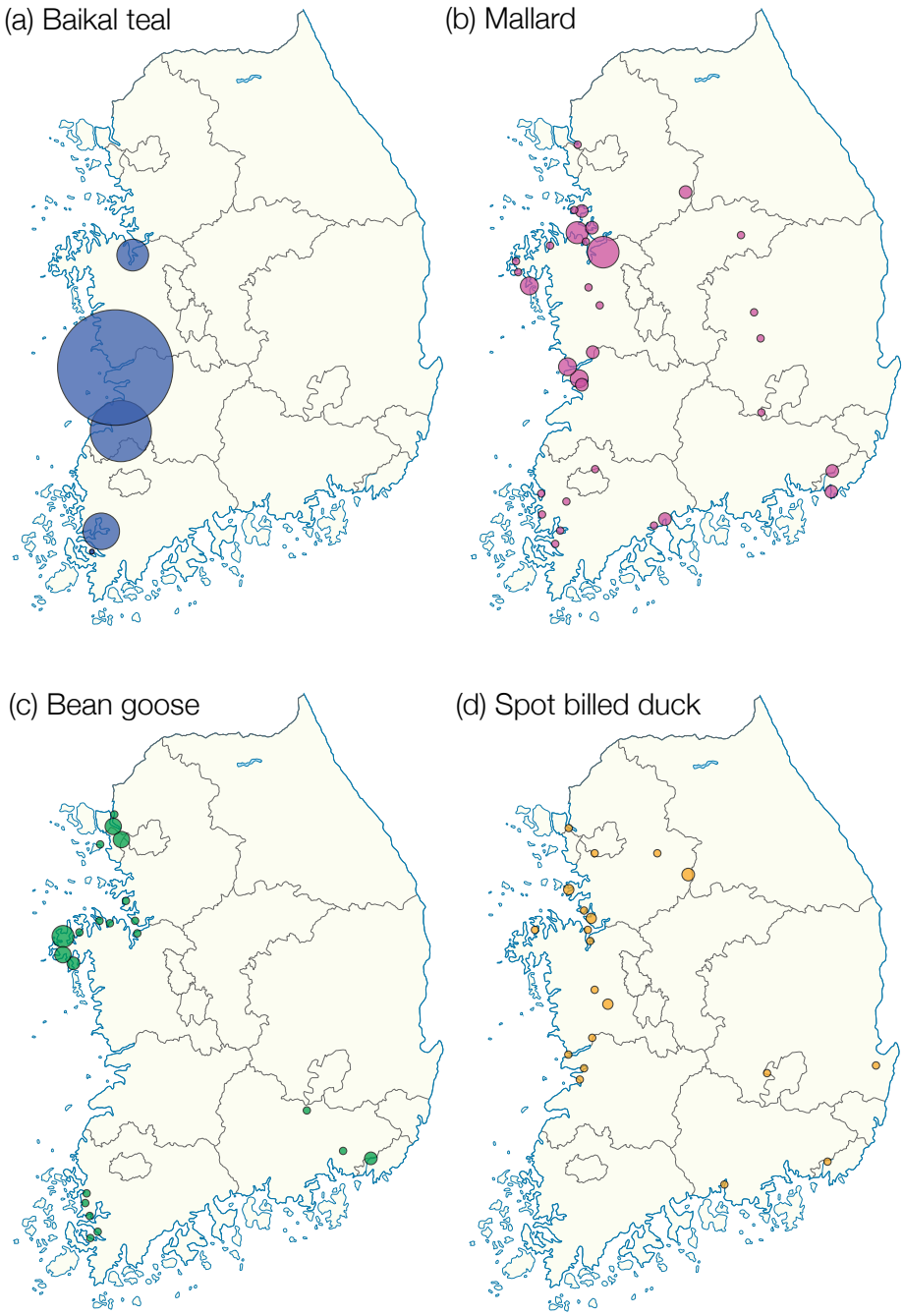

Figure 3.2: Maps generated from ROK Ministry of Environment Wild Bird Census for winter 2014 data showing the number of overwintering waterfowl for the four most common species in ROK. Circles are proportional to estimated bird numbers at sites. Geographic locations are approximate. Bird species and total observed numbers are as follows: (A) Baikal teal (*Anas formosa*) (365,641), (B) mallard (*Anas platyrhynchos*) (155,208), (C) bean goose (*Anser fabalis*) (72,225), (D) spot billed duck (*Anas poecilorhyncha*) (68,204).

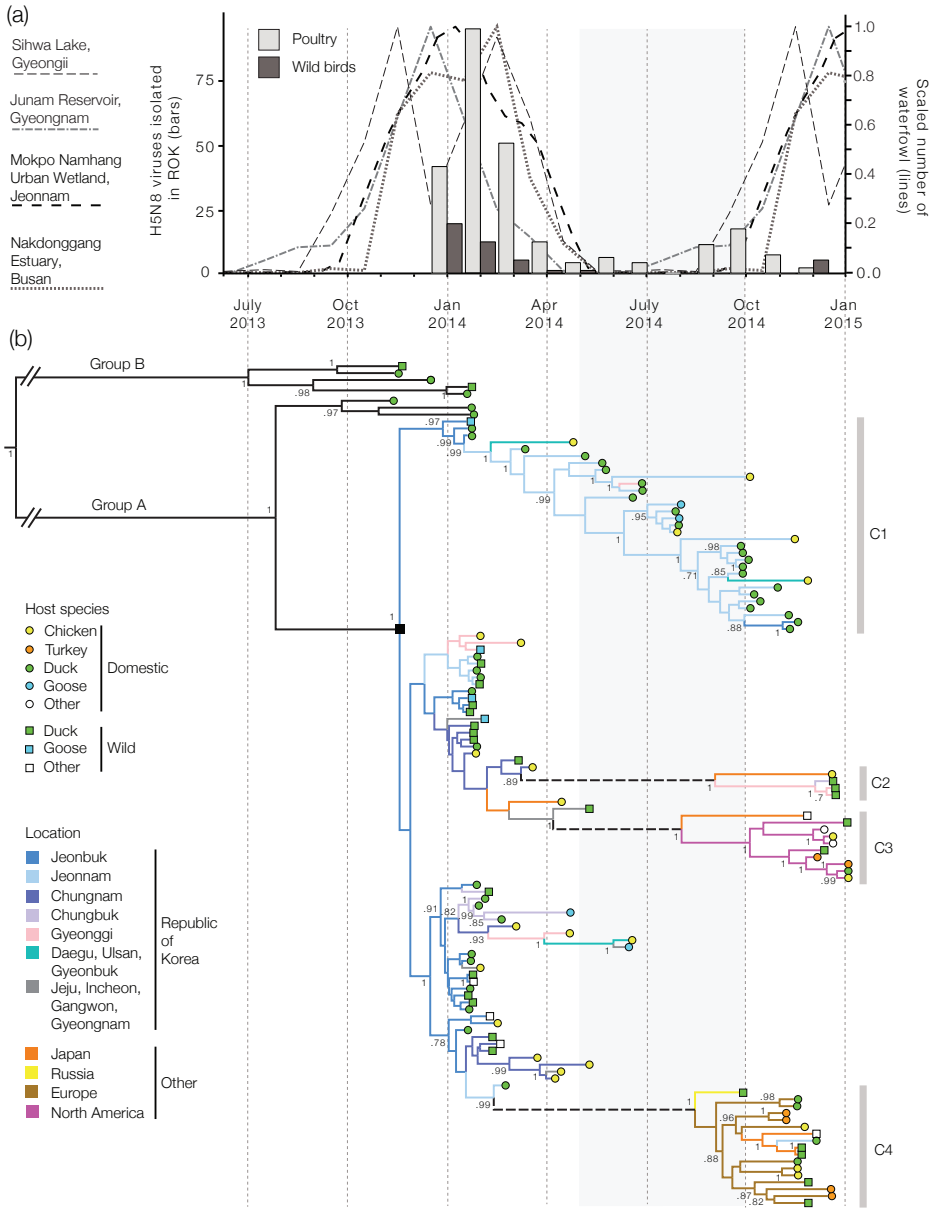

Figure 3.3: (A) Bird population dynamics in ROK. Lines represent the scaled average number of waterfowl observed at four sites across ROK (counts scaled so the maximum count at each site equals one). Bars represent the number of H5N8 viruses isolated in ROK in 2014 from wild birds (dark grey) and domestic birds (light grey). (B) The estimated MCC phylogeny of Eurasian H5N8 and American H5 strains (all H5 clade 2.3.4.4). Branch lengths represent time and the tree is placed on the same timescale as the plot in part (A) above. Branch colours represent locations inferred via discrete trait reconstruction using BSSVS (see key). "Long branches", as discussed in the main text, are dashed. Squares and circles at tips represent host species (see key). Provinces in ROK correspond to the locations and colours used in Figure 3.1. Numbers at nodes show posterior probabilities >0.7. Several sets of proximate provinces which are represented by few isolates have been grouped. A fully annotated tree is provided in Figure A.1.

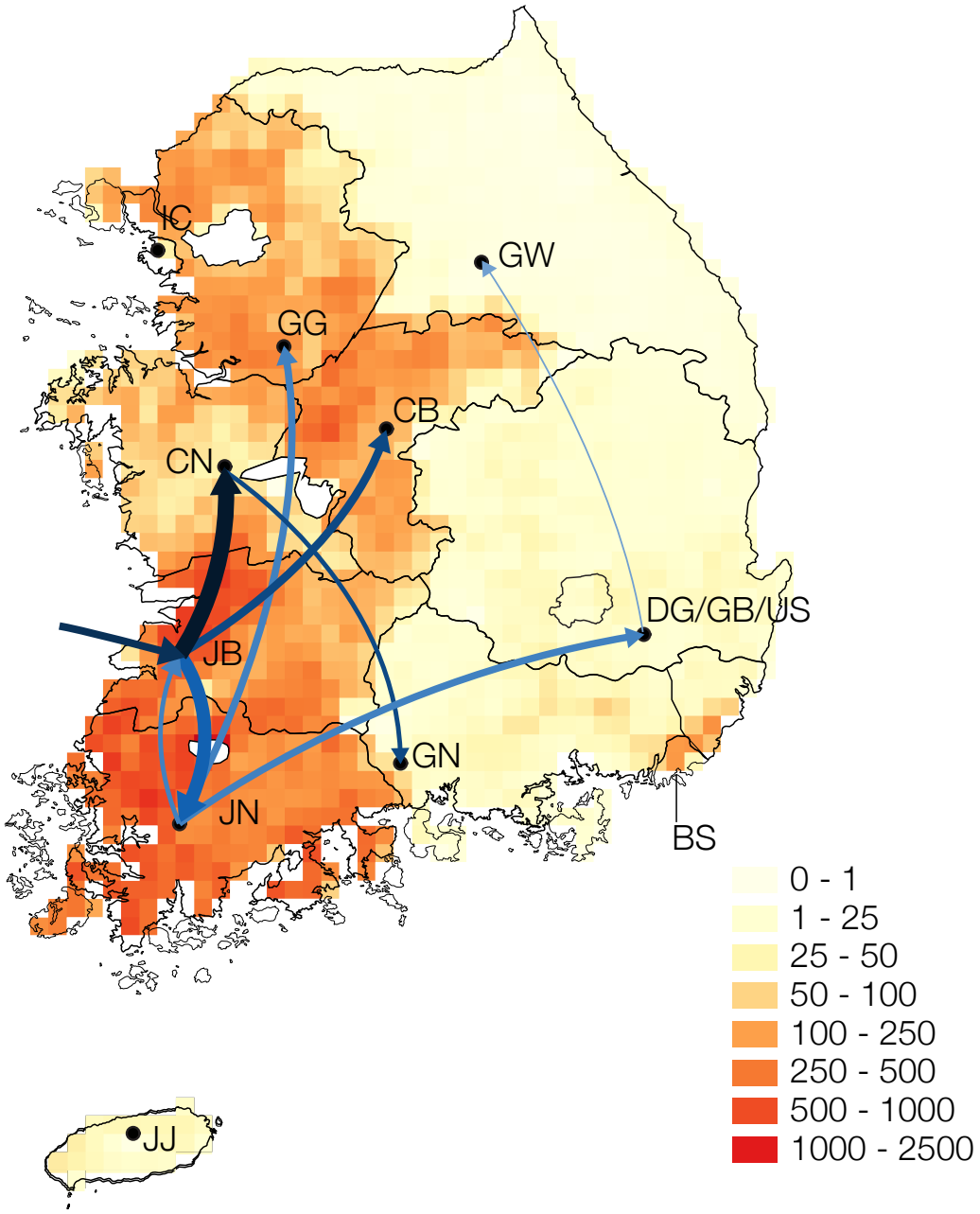

Figure 3.4: Map representing the estimated trajectory of the H5N8 spread in ROK. Arrows connecting locations represent directions of movement with Bayes Factor support >10. Arrow colours represent Bayes Factor support for rate indictors, with darker blue indicating better support. Arrow thicknesses are proportional to the inferred values of Markov jumps between locations, such that a wider arrow represents more migration between a pair of locations. Yellow and orange background background show the estimated density (numbers per kilometer) of domestic ducks (colours in key). Province abbreviations are the same as those used in Figure 3.1

Table 3.1: Phylogenetic statistics for various groups of isolates within clade C4

	Phylogeography with BSSVS		Phylogeography without BSSVS		No phylogeography	
Groups of isolates	Monophyly probability	TMRCA [1] (95% HPD interval)	Monophyly probability	TMRCA [1] (95% HPD interval)	Monophyly probability	TMRCA [1] (95% HPD interval)
Japan and Korea	1	0.21 (0.15, 0.27)	1	0.21 (0.15, 0.27)	1	0.22 (0.16, 0.29)
Europe	<0.001	0.34 (0.26, 0.44)	<0.001	0.35 (0.26, 0.44)	<0.001	0.37 (0.28, 0.457)
Russia and Europe	<0.01	0.38 (0.29, 0.47)	<0.01	0.38 (0.30, 0.46)	<0.01	0.38 (0.30, 0.464)
Europe, Japan and Korea	0.62	0.34 (0.26, 0.44)	0.55	0.35 (0.26, 0.44)	0.28	0.37 (0.28, 0.46)
Russia, Japan and Korea	0	0.38 (0.29, 0.47)	0	0.37 (0.23, 0.46)	0	0.37 (0.29, 0.46)

[1] Years before 30th December 2014

These results leave open two possible scenarios. In the first, there is a single introduction of C4 viruses into Europe (from an unsampled reservoir location) followed by rapid viral lineage movement from Europe to East Asia. In the second scenario, C4 viruses were independently dispersed from an unsampled reservoir to Western Europe and East Asia during July to September 2014. Because the European samples do not form a strictly monophyletic clade, this second scenario requires there to have been at least two, and possibly many more, separate introductions of C4 viruses into Europe from the reservoir population. Based on the patterns observed in ROK, I again posit that this may represent long distance transmission via migratory wild birds from an unsampled reservoir. This hypothesis follows recent suggestions that the almost simultaneous detection of H5N8 across the world could be a result of birds carrying the virus from unsampled breeding grounds in Russia and Beringia [247, 458, 456]. Ring recovery data from wild ducks suggests that intercontinental transmission by wild birds is plausible [458]. If this model is correct, then the phylogenetic data can only be reconciled if H5N8 entered Europe on at least two separate occasions, and entered Asia at least three times.

3.4 Conclusions

In this study I investigated the factors underlying the emergence and persistence of H5N8 HPAI in the Republic of Korea by analyzing new and previously available viral gene sequences using molecular clock and phylogeographic methods, and by interpreting the results in the context of data on avian ecology. I suggest two separate waves of migrating wild waterfowl contributed to the presence of H5N8 HPAI in ROK; the first during its initial emergence in Asia in 2013/2014, and the second during late 2014. I find that H5N8 initially emerged in ROK in an area of high wild bird and domestic duck density, at a time associated with the migration of overwintering wild waterfowl into the region. Although I cannot formally exclude the effect of climatic variables such as temperature, which are naturally co-linear

with bird migration patterns, these data suggest that migrating wild waterfowl were important in the establishment of H5N8. Despite several introductions, the virus did not become established in the east of ROK, which is an area characterised by low numbers of domestic ducks and wild waterfowl. Domestic duck distribution appears to have been important in regional persistence. I posit a model of a second wave of virus introductions by wild birds from an unsampled reservoir population, which may also explain the observed long distance transmissions to Europe, North America and Russia. The second wave of H5N8 introduction is supported by the presence of long-branches in the tree that coincide with the absence of wild bird migration to ROK.

My results support recent hypotheses that wild waterfowl and domestic ducks are important in the emergence and maintenance of HPAI H5N8 [458, 456]. I highlight the importance of interpreting phylogeographic analyses in the context of available ecological data, especially when sample sizes are small or when key locations are not sampled. Future studies formally integrating phylogenetic data with ecological data should be conducted in order to more clearly identify factors involved in H5 HPAI emergence and to help target surveillance resources to high-risk source areas. Publication of the exact geographic locations of viral isolates and collation of existing ornithological datasets would allow outbreaks to be considered in their proper ecological context. Determining the ecological drivers of HPAI transmission will become increasingly important for both human and animal health as domestic poultry production continues to intensify across the world [197].

Chapter 4

Diversity of the mute swan (*Cygnus olor*) faecal virome

4.1 Introduction

Viruses that affect wildlife are considerably under-studied compared to those in humans, yet improving our knowledge of animal viruses is important for reasons that extend far beyond their obvious veterinary significance. Viruses that originate in wildlife are the primary source of newly emerging pathogens in humans [198, 491]. Understanding the current host range, ecological niche and mode of transmission of a virus in animals can help predict when and why certain viruses might be more likely to transmit to humans [196]. Discovering animal homologs of key human viruses (for example, hepatitis C virus homologs in rodents [206]) can present new opportunities to develop effective animal-models that are useful for basic medical research, including vaccine and drug development [464]. Commercially, the spillover of wildlife viruses to farmed animals can cause huge direct and indirect losses through increases in mortality or reductions in trade; for example, recent H5 avian influenza outbreaks in the USA resulted in financial losses of $3.3 billion [373]. Infectious diseases can also pose a threat to conservation of existing biodiversity, as small populations of endangered species can be pushed closer to the brink of extinction by single outbreaks

[25, 241, 341]. The discovery and epidemiological characterisation of new viruses has great potential to improve our understanding of cross-species transmission and viral evolution, and contribute to efforts to protect human and animal health, food production systems, and global biodiversity.

Virus discovery during most of the 20th century involved virus isolation and amplification in cell culture [251]. As the cost of next generation sequencing has reduced, it has become more feasible to study viral genetic diversity using metagenomic techniques. Here, I use "metagenomics" (or "metaviromics") to refer to the direct genetic analysis of microbes (or viruses) in a sample without culturing or performing specific PCR amplification, sometimes termed shotgun metagenomics. Metaviromic studies in wildlife have been successfully employed to identify unknown disease-causing agents (for example, a retrovirus that caused unexplained unexplained mortality in chickens, or a picornavirus associated with hepatitis [17, 178]), or to catalogue viral diversity within a given taxonomic group (for example, large-scale ivestigations of viral diversity in invertebrates and bats [388, 497]).

Despite the existence of >10,000 species of birds, most metaviromic studies in birds have been conducted on only two species of domestic bird; turkeys and chickens [32, 33, 34, 85, 86, 87, 88, 99, 100, 215, 267, 385, 513, 518]. Single studies have investigated the viruses present in domestic ducks [62] and domestic guinea fowl [265]. To my knowledge, only six metaviromic studies have been conducted in wild species of bird; the European roller (*Coracias garrulus*) [320], mallard duck (*Anas playtrhynchos*) [129], northern fulmar (*Fulmarus glacialis*) [263], black-capped chickadee (*Poecile atricapillus*) [520], rock pigeon (*Columba livia*) [342] and captive birds of the endangered rowi kiwi (Okarito kiwi) species (*Apteryx rowi*) [476]. The contents of these studies is summarised in Table 4.1. Whilst all of these studies in wild species of bird are valuable in that they characterise at least one novel virus (sometimes associated with a particular disease) only two of the six studies publish any information on overall virus composition of tested samples. None of the six studies explore differences in the viromes of different birds that might result from differences

in ecology (such as seasonality or geography) or demography (such as age or sex, which is not published for birds in any of the studies). Both studies that publish viral composition data are based on sample pooling without barcoding prior to sequencing, so, despite reasonable numbers of birds sampled in each study (n=51 and n=23), it is not possible to assess within-population differences in viral carriage. Little is known about viral diversity for the majority of wild birds.

The potential to use metaviromics to learn about viral epidemiological dynamics in the wild is almost entirely untapped. Instead, current studies of viruses in wildlife are almost always specific to a single viral target, being reliant on molecular genetic approaches such as qPCR that can only be used to detect the presence of a single or small number of viruses. Such studies have been extremely important in increasing our understanding of temporal, geographic and age-related differences in the prevalence of many clinically important animal viruses, such as avian influenza virus (reviewed in [297]). Whilst PCR is extremely sensitive, simple and low cost per assay, the need for target-specific primers can bias detection, as related viruses that are divergent from the reference at the primer-binding site may not be amplified. Furthermore, targeting a single virus per PCR assay means that assaying more than a few different viruses becomes impractical because of the greater cost in time, consumables, and also due to the increased volume of sample required.

Metagenomic studies have been used to uncover seasonal, demographic and geographic factors that seem to affect viral prevalence in humans [261, 266, 295], farm animals [375, 516] and in the environment (e.g., water or air) [289, 317, 480]. However, to my knowledge, only a few studies have ever attempted to explore the factors that affect viral presence in wildlife using metagenomics. Temporal variation in viral infections have been detected for honey bees [372]. In monkeys, variation in the prevalence of simian immunodeficiency virus by age [207], and geographic differences between sampling sites have been shown [9]. Combining metaviromic techniques with detailed host information in order to study viral diversity in its natural ecological context is extremely rare, and, I believe, has never been conducted for birds. None of these

studies have been able to identify individual animals or follow them longitudinally through time.

The availability of susceptible hosts is fundamentally important to shaping how a virus transmits through a population, and, through necessitating viral evolutionary strategies to avoid immune recognition, is often a key driver of viral diversity. In wild birds, as in many other host groups, naïve hosts can be introduced into a population through hatching or through migration. 19% of all modern bird species are believed to be migratory [404, 220], following seasonal patterns of movement between breeding and wintering grounds. Breeding is also highly correlated with season in most birds [69], and is controlled by a physiological cascade triggered by environmental changes such as differences in the length of daylight or amount of rainfall. Avian influenza virus (AIV) is perhaps the best studied of all the viruses that infect wild birds. For low pathogenic AIV, we know that adult birds typically have lower AIV prevalence than juvenile birds (typically measured as hatch-year) [296, 325, 412, 462], and AIV prevalence is believed to be highest immediately prior to and during autumn migration and lowest in spring [228, 240, 296, 325, 412, 462, 471]. Viral prevalence in resident (non-migratory) species can be affected by the migration patterns of migratory bird species that share the same geographic environment if cross-species transmissions of that virus can occur and if migrant birds act as amplifiers of the virus. Beyond avian influenza, it is probable that age and seasonality are associated with viral prevalence of many viruses that infect wild birds, because, respectively, young birds are more likely to be immunologically naïve for any particular virus and because the timing of entry of these young birds to infection networks is determined through seasonal controls on hatching and migration. Despite their critical importance, the effects of these factors on the dynamics of most other viruses that infect wild birds, including many viruses that have been recently discovered through metagenomics, have not yet been addressed.

Here, I investigate the faecal metaviromes of 119 wild, mute swans (*Cygnus olor*) that form part of a well-studied population in Abbotsbury, Dorset, UK. The pop-

ulation has been studied intensively for over 60 years, and detailed demographic information (for example, age, sex and parentage) is available for most birds. Avian viruses are extremely undersampled compared to mammalian viruses, and in total only 9 viruses are known to infect swans [1]. I sample birds from this population over a period of 6 months, sequence the metavirome with help from colleagues at the University of Amsterdam and analyse the faecal metavirome of 119 wild mute swans.

The following results are presented in two distinct sections that reflect the two distinct aims of the study, discovery of novel viral diversity and analysis of the patterns that result in that diversity. Firstly, I describe the *de novo* assembly and genomic characterisation of 9 previously unknown viruses that probably represent natural infections in birds and investigate diversity within the population for two of these viruses using phylogenetic analyses. I also describe the presence of several additional viruses that have already been detected in birds. Secondly, I explore whether the metaviromic data is appropriate for analysing the factors that affect viral carriage in this population. Here, to provide a proof of principle that metaviromic data can be used effectively to explore the factors that affect viral carriage in wild birds, I focus only on the effect of bird age and season. As described above, these two factors are perhaps the most well established drivers of differences in viral carriage in wild waterfowl. Other possible correlates of infection that could be explored using the data available for this population, including family structure (defined by mating data and genealogies), last-known weight, data on observed illnesses or mortality, or fitness (as defined by number of cygnets that survive to breeding age), are not considered here but could perhaps be addressed with further analyses. In its entirety, this work begins to bridge the gap between genomic description and epidemiological analysis, and provides a possible blueprint for future studies.

[1] Low to moderate prevalences of aquatic bird bornavirus 1 [97], avian paramyxovirus serotype 1 [333], swan circovirus [159], duck viral enteritis (anatid herpesvirus) [493], coronavirus [185] and influenza A virus have been identified. Sporadic cases of goose parvovirus [387], West Nile virus [43] and avian poxvirus [157, 250, 293] are also reported.

Table 4.1: Previous metaviromic studies in wild bird species

Bird species	Number of individuals	Sample type	Health	Season	Location	Age [2]	Viruses present	Reference
Columba livia (rock dove)	51 samples, with no per bird information. Samples pooled in groups of 5	Faecal	Healthy	August (50) or April (1) 2011	Hong Kong (50) or Hungary (1)	Only 1 bird aged	Animal, insect and plant viruses described. Major animal viruses included *Circoviridae, Parvoviridae* and *Picornaviridae*.	[342]
Coracias garrulus (European roller)	2 birds from different locations.	Faecal	Healthy	July 2011	Hungary	Unknown	No species composition published. Single novel astrovirus described.	[320]
Anas playtrhynchos (mallard)	23 birds (pooled).	Cloacal swabs	Healthy	Unknown	India	Unknown	Primarily insect viruses (67%) with fewer plant (3%) or algae viruses (8%). Animal viruses are primarily *Parvoviridae* and *Nodaviridae*.	[129]
Fulmarus glacialis (northern fulmar)	1 bird.	Spleen and uropygial gland	Diseased	January 2014	California	Unknown	No species composition published. Single novel gyrovirus described.	[263]
Poecile atricapillus (black-capped chickadee)	8 birds.	Tissue (mandible)	Diseased	2001-2014	Alaska	Unknown	No species composition published. Single novel picornavirus described.	[520]
Apteryx rowi (Okarito kiwi)	8 birds (pooled).	Faecal	Broadly healthy	September 2013	New Zealand	Unknown	No species composition published. Single novel circovirus described.	[476]

[1](Note that "sex" is not also included in the table, as it is unknown across all six studies)

4.2 Materials and Methods

4.2.1 Sample collection

Non-invasive faecal samples were collected from swans at Abbotsbury Swannery on the Fleet Lagoon (Figure 4.1) (Dorset, United Kingdom; 50.6537°N, 2.6028°W)). Abbotsbury Swannery harbours a large population of wild mute swans (*Cygnus olor*) that have been subject to long-term study. The swans are not wing-clipped, are free to fly, and habitually mix with other species at the site (Figure 4.3). Population size varies seasonally due to immigration of swans to the site during the summer moult, but is typically between approximately 500 and 1000 birds. Birds hatched at the Swannery are sexed and marked with metal web-tags within 24 hours of hatching. Web-tags are replaced with adult rings at approximately 5 months of age (Figure 4.2) and birds are weighed. Every two years, all swans present on the Fleet Lagoon are caught, weighed, and year of hatching and/or sex is recorded where possible. Consequently, detailed data about date of hatching, sex and parentage are known for most birds.

Samples were collected on 19 occasions at approximately monthly or bimonthly intervals between April 2015 and June 2017, with each visit lasting 2-3 days. Samples collected between April 2015 and March 2016 were entirely swabs of feces (11 sampling visits), whereas samples collected after March 2016 were whole faecal samples (8 sampling visits). A summary of the number of samples collected per sample occasion and the number of times that each bird was sampled per sample type and in total are presented in Tables C.1 and C.2. Only whole faecal samples were used in this study as the higher volume of material in these samples was found to be more suitable for metagenomic sequencing, but earlier samples would be appropriate for investigating specific viruses in detail using PCR amplification methods in future. In total, 1440 samples from uniquely identifiable birds were collected, representing 427 different birds. Individual birds were sampled on up to 10 different sample occasions.

Birds were observed when on the land, and samples were taken from any ringed

bird seen to defecate. For each sample, approximately 0.5ml of faeces was collected with an individually wrapped, sterile spatula into a 1.5ml sample tube containing 1ml of Universal Transport Media (Sterilin). Each tube was shaken vigorously to promote mixing of the sample with the media. Samples were kept on ice in the field for up to an hour, before being placed at -80°C in a freezer at the field site. All subsequent transport was conducted on dry ice. Samples used in this study were collected on 5 sample occasions at approximately monthly intervals between May 2016 and October 2017.

None of the birds studied here are known to have died within 7 weeks following sampling, and no birds had obvious clinical symptoms of disease at the time of sampling.

4.2.2 Sample preparation and sequencing

Faecal samples were processed at the University of Amsterdam using previously published methods [31, 79]. Briefly, the method used attempts to enrich for viral nucleic acids without bias for the type of viruses that can be detected. Virions are enriched in the sample by centrifugation to remove most bacteria, cells and mitochondria. Residual DNA and RNA molecules that are not virion protected are degraded with a DNase treatment and by RNases naturally present in the sample. The Boom method is used for extraction of total nucleic acids [31]. To reduce reverse transcription of host ribosomal RNA and enrich for reverse transcription of viral RNA, the method uses non-ribosomal random hexamers during cDNA synthesis [119]. These primers are designed to bind to hexamer regions that occur more commonly in known viral genomes, and less commonly in ribosomal RNA and hence reduce reverse transcription of ribosomal RNA compared to what would be expected using standard random hexamers.

The quality of the dsDNA was checked on a Nanodrop, and concentrations measured using a Qubit high sensitivity dsDNA kit on a Qubit 3.0 fluorimeter. Samples

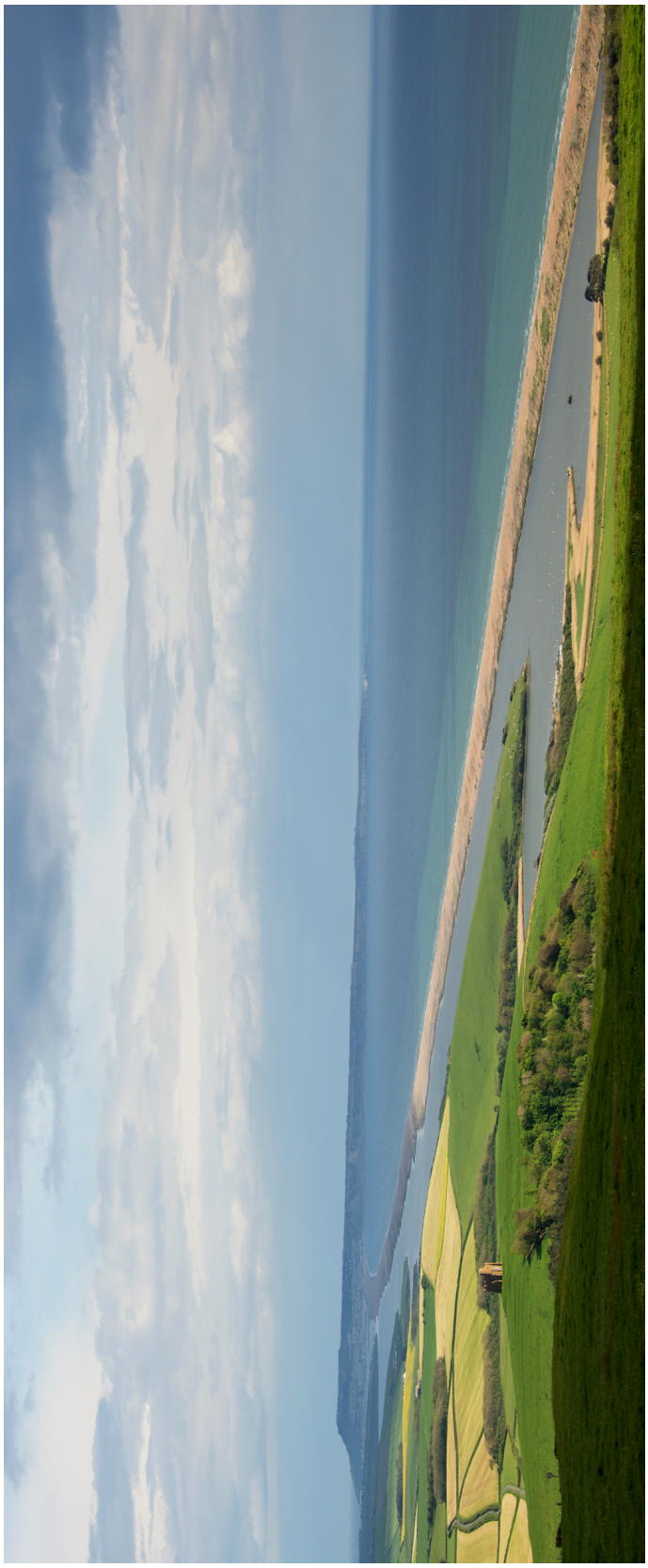

Figure 4.1: Position of Abbotsbury Swannery on the Fleet Lagoon (centre bottom: the swans are visible as white specks in the bay). The site is an important coastal stopover site for many migratory waterfowl. This photograph was taken by the Abbotsbury swanherd Charles Wheeler, and is included with his permission.

Figure 4.2: Uniquely identifiable leg rings on swans. Each swan is ringed with two legs rings: a metal ring with a code recognisable by the British Trust for Ornithology, and a coloured, plastic Darvic ring that is easy to read from a distance, allowing individual birds to be identified. Any unringed mute swans present on The Fleet Lagoon are able to be ringed during a large, biannual catching event held at the site ("roundup"). This photograph was taken by the Abbotsbury swanherd Charles Wheeler, and is included with his permission.

Figure 4.3: Wild birds at Abbotsbury Swannery. Note that the swans freely mix with other resident and migratory bird species, including many different species of waterbird. This photograph was taken by the Abbotsbury swanherd Charles Wheeler, and is included with his permission.

were barcoded and paired-end library preparation was performed for all samples using an Illumina Nextera XT kit by the sequencing team at the Wellcome Trust Centre for Human Genetics (Oxford, UK). The samples were sequenced in multiplexes on an Illumina HiSeq 4000 to generate 150bp paired end reads. The first batch was run using a multiplex of 16 samples (including 5 non-faecal samples that are not considered here) on a single lane to test the protocol (generating 115Gb), and 118 samples (including several unrelated samples that are not considered here) were sequenced across 5 lanes (generating 521Gb). In total, dsDNA generated from faecal samples from 119 wild mute swans was sequenced.

4.2.3 Computation detection of viral reads

Raw Illumina fastq files were subjected to a custom pipeline for classification of viral reads. All reads were aligned against a local viral protein database containing the GenBank complete viral protein database using DIAMOND BLASTX in fast mode with an E-value cutoff of 1 [49]. Reads that matched with an E-value of <1 were subjected to DIAMOND BLASTP searching against the entire *nr* protein reference [2] database to remove false positives. `Taxid` numbers of each *nr* best hit were used to extract the complete hierarchical taxonomy (that is to say, any available classification at all possible taxonomic levels) of the closest matching sequence for each read. All matching reads and their pairs were extracted using seqtk [258]. Low quality ends were trimmed to a quality score of 15 and adaptors were removed using cutadapt [278]. The resulting dataset was checked for quality with FastQC [8].

4.2.4 Assembly of novel viruses

The number of reads corresponding to each viral species was checked using the taxonomic metadata assigned to each putative viral read (as described in Section 4.2.3). Plausible viral candidates were chosen for assembly if they had high read numbers

[2]The nr database is a database of the unique sequences from the set of all translated GenBank sequences, and sequences from other databases such as RefSeq, SwissProt, PIR, PRF and PDB.

and were from a predicted vertebrate host species. Typically, this corresponded to viruses with approximately >150 putative reads (matching viruses with reads in the highest 10th percentile for the dataset). In addition, the ratio of number of uniquely occuring sequences corresponding to a species (i.e., grouping sequences that were unique at the amino acid level) to total viral reads corresponding to that species was considered, as considering only a low total read number matching to a particular species might obscure the occurence of high sequencing coverage but that reads could only be detected via BLAST aligning in a short (perhaps conserved) region of the genome. Choice of putative candidates for assembly was not blind to species names or entirely rigid to specific cut-offs of read-counts or unique/total read ratios, and it is probable that this manual component has biased the specific viruses that were assembled here.

Data were assembled in Geneious 8.1.7 using a combination of reference mapping and *de novo* assembly. For reference mapping, candidate reference genomes for mapping were chosen as the genome with the lowest BLAST E-value score (as described in Section 4.2.3). All viral reads were mapped against this reference. Where a novel virus was too divergent to the closest existing reference for reference mapping, *de novo* assembly using the Geneious assembler was performed. All viral reads with BLAST matches to the same target virus species or genus were extracted and *de novo* assembled, and the longest contigs were chosen. For many viruses, one or more gaps were present in the assembly (typically due to divergent regions or the presence of non-coding regions). If coverage at mapped or assembled sites was high, gap closure was attempted by iteratively mapping all viral reads to the end of the contig. Where gaps persisted, all reads (viral and non-viral) from the sample that contained the highest number of reads of the target virus were iteratively mapped to the ends of the existing contig until no further reads could be assembled. A BLASTX search was conducted on each assembled virus to check that all coding regions had been assembled, under the expectation that novel assembled species should have similar genome length and number of open reading frames to their closest matching reference. To check for the

absence of significant variation in coverage across the genome, which might indicate a poor assembly or artificial construction of chimaeras, all reads from the sample used to assemble the putative reference genome were re-mapped to assembled viruses using BWA [259]. The bam files were indexed and viewed in IGV 2.3 [365], and samtools [260] was used to extract measures of sequencing depth across the genome.

4.2.5 Confirmation of *de novo* assembled viruses by specific PCR

The assemblies of all putative new species of virus for which full genomes were assembled were confirmed by re-sequencing using virus specific primers. Two different approaches for re-sequencing were used. For shorter viruses (typically <6 kb), primers were designed using Primer-BLAST to amplify overlapping amplicons across the genome, and individual PCR reactions were set up for each target amplicon. 25ul reactions were performed using Q5 Hot Start High Fidelity Polymerase (NEB) according to the manufacturer's instructions and with 2.5ul cDNA per reaction and 45 seconds of extension per cycle. Amplification was conducted for each virus using the annealing temperature, sample and 0.5uM final concentration of each primer in the relevant primer pair in Table C.3. Forty cycles of denaturation, primer annealing and extension were performed for each assay. Following PCR, amplicons were purified for Sanger sequencing using Exonuclease I (NEB) and Shrimp Alkaline Phosphotase (NEB). 0.5U of Exonuclease I and 0.25U of Shrimp Alkaline Phosphotase were added to 20ul of each PCR product, and adjusted to 30ul using nuclease free water. The mixture was incubated at 37°C for 30 minutes and 85°C for 15 minutes to remove single-stranded DNA and free dNTPs. Amplicons were directly Sanger sequenced by a commercial company (Source Bioscience) using both the forward and reverse primers. Chromatograms were checked for disparities between the forward and reverse reads, and consensus sequences for each amplicon were aligned to the novel genome to check for the absence of any differences.

For longer viruses, multiplex PCRs were designed to generate overlapping amplicons across the whole of the novel genome using Primal Scheme, an online multiplex primer design tool. PCR amplification and all laboratory protocols were conducted according to previously published protocols [354]. All primers and assayed samples are given in Table C.6. Forty cycles of denaturation, annealing and extension were conducted. The negative controls used in cDNA synthesis and a PCR water control were included on each plate. PCR products were cleaned using 1x Ampure bead cleanups, and DNA concentration measured using a Qubit High Sensitivity dsDNA kit on a Qubit 3.0 fluorimeter. PCR products were standardised by concentration, and 200ng of total DNA per sequencing library was carried forwards into library preparation.

Amplified DNA and appropriate negative controls were sequenced in barcoded multiplexes of 6 - 8 samples per run on the MinION (Oxford Nanopore Technologies) using FLO-MIN106 flow cells. Library preparation was conducted using Ligation Sequencing 1D and Native Barcoding kits according to the manufacturer's instructions, but with the changes detailed in [354] (kits numbers; SQK-LSK108, EXP-NBD103). Libraries were loaded onto flow cells (FLO-MIN106) and sequencing was performed without basecalling for 48 hours using MinKnow 1.7.7.

Consensus sequences for each barcoded sample were generated following previously published methods [354]. Briefly, raw files were basecalled using Albacore 1.2.5 (Oxford Nanopore Technologies), demultiplexed and trimmed using Porechop, and then mapped with BWA to a reference genome (A/turkey/England/052131/2016; GISAID Isolate ID EPI_ISL_239801). Nanopolish variant calling was applied to the assembly to detect single nucleotide variants to the reference genome [269]. All sites where the coverage was <20 and all primer binding sites were masked with Ns during generation of the consensus sequences. Consensus sequences were aligned against the reference, and raw sequencing reads visually inspected in IGV 2.3 [365] to check consistency between overlapping regions of neighbouring amplicons.

4.2.6 Targeted sequencing

Following taxonomic assignment and mapping of viral reads, it was apparent that whilst certain samples clearly contained viral genomic material from a specific viral species, there were too few reads mapping to that virus to reliably assemble a long contig. In order to better explore the viral diversity in the population, PCRs generating amplicons of length approximately 1kb were conducted for a small selection of viruses in samples that had >20 reads of that virus. To reduce the probability of false-negatives occurring because of unaccounted divergence in the variable antigenic regions, a PCR targeting a conserved region of the virus species was performed prior to any PCRs in more divergent regions of the genome. Conserved regions were assessed based on alignment of any assembled viruses of that species detected at Abbotsbury and the most closely related species as estimated through BLAST searching. Positivity or negativity was defined by detection of a clear band of the correct size using gel electrophoresis on a 1.5% agarose gel containing 6ul SYBR Safe for every 100ml of TAE. If samples were PCR positive for the conserved region but were negative for the more divergent region, the primers for the divergent region were redesigned. PCRs were conducted using the conditions in Section 4.2.5 but using the annealing temperatures, extension times and primers given in Table C.4. 45 cycles of PCR were performed. Amplicons were purified and subject to direct Sanger sequencing using the methods detailed in Section 4.2.5.

4.2.7 Phylogenetic analysis

Phylogenetic trees were estimated using maximum likelihood methods for all putative novel species in order to place these species within the currently known viral diversity. For each new species, a BLASTX search against the NCBI viral reference database was conducted for each open reading frame of the putative new virus and the 7 reference genomes with the lowest E-values were extracted. A separate BLASTX search was conducted against the NCBI *nr* database and the sequence with the lowest

E-value was also extracted to ensure that a more closely related sequence would be included if appropriate, even if that sequence was not a species reference sequence on GenBank. All matches were combined with the query sequence and aligned using MUSCLE [113]. Due to high amino acid divergence, certain sequences sometimes could not be reliably aligned. These were removed and the dataset subsequently realigned. Alignments were manually trimmed to visually more conserved regions. Amino acid substitution model testing was performed using ProtTest3 [84] with the Subtree-Pruning-Regrafting (SPR) tree-searching method and the model with the best Bayesian Information Criterion (BIC) score chosen. Phylogenetic trees were constructed in PhyML [156] using the most appropriate model and 100 bootstrapped replicates.

Some mapped or assembled viral genomes had high homology to an existing virus species based on BLASTX or BLASTN searching, and hence it was considered likely that they belonged to this species. For these viruses, sequences from that species were downloaded from GenBank, and nucleotide alignments were constructed for each protein or partial protein. All full-length sequences were used where possible, but where there was a large number of sequences available data were manually sampled to try and achieve diverse sampling by species, location and time. Where possible, full length or near-full length protein-coding regions were used in alignments. Alignments were only trimmed to partial regions if excluding all partial sequences excluded a large amount of geographic or temporal sampling diversity. Model testing was performed using jModelTest [349] with SPR moves. The model with the best BIC chosen to use for phylogenetic tree construction in PhyML [156] with 100 bootstrapped replicates.

4.2.8 Sensitivity test of Illumina sequencing data

Later in the chapter I use raw viral read data to explore whether bird age is associated with higher viremia of certain taxa. This approach relies on an assumption that the number of viral reads sequenced per sample (or an appropriate normalisa-

tion thereof) reflects the amount of virus present in that sample. It was therefore necessary to first test that the sequencing protocol did not obscure genuine differences in viral load between different samples through standardisation of samples by nucleic acid concentration during sequencing library preparation. To explore whether Illumina read count is closely correlated with viral load for each virus, two qPCRs were designed against conserved regions of two commonly detected virus taxa in the population. Targeted viral taxa included the family *Astroviridae* (targeting conserved regions of 3 different species identified in this study) and the genus *Megrivirus* (targeting conserved regions of a single species identified in this study). New cDNA was made using the Protoscript II First Strand cDNA Synthesis Kit (New England Biolabs) and random hexamer primers (Bioline), rather than the non-random hexamers used in original cDNA synthesis [119]. cDNA was not subsequently standardised by concentration. qPCRs were set up using the PowerUp SYBR Green Master Mix kit Applied Biosystems according to the manufacturer's instructions, using 5ul of cDNA in a 20ul reaction and 0.5uM each primer. Technical replicates were run for all samples in triplicate. A 1:5 dilution series made from the sample with the highest read count for each virus was used in triplicate as a standard curve, with 6 points for the astrovirus assay and 7 points for megrivirus assay (in the expectation that the standard curve would then appropriately cover an appropriate range of sample viremia for the samples studied here, equivalent to 7500 to 2.4 reads for the astrovirus assay, and 20000 to 1.2 reads for the megrivirus assay). Due to the higher range of total read counts observed for each virus, 18 samples were tested for megrivirus, and only 7 samples were tested for astrovirus. Cts for each sample and standard were recorded. Primers and annealing temperatures are given in Table C.5.

To check that the Illumina read counts were representative of the amount of viral nucleic acid in a sample, Cts generated from qPCRs were correlated with the normalised read counts (normalized to reads per million) of the identified number of sequencing reads that had closest BLAST matches mapping to targeted viral taxa using simple linear models.

4.2.9 Assignment of host group to faecal metagenomic reads

Swan faecal samples contain large amounts of dietary material, bacteria and environmental material (for example, small insects or algae) consumed whilst feeding. Viruses that infect these components are therefore present in swan faecal samples along with any viruses that naturally replicate in the bird. I therefore attempted to classify reads based on which host they might naturally infect, with the intention of only analysing the carriage of viruses known to infect vertebrates. To make a reference dataset of associations between known viruses and their natural hosts, all viral reference sequences in the NCBI database (accessed 6th July 2017) were downloaded along with information about their known hosts (classified as one or multiple of; vertebrate, invertebrate, fungi, algae, plants, bacteria, human, protozoa, diatom, eukaryotes). This information was processed with a Python script that generated a "taxid-to-host" lookup file, comprising accession number, host classification and GenBank `taxids` for every viral reference sequence.

In parallel to this, all viral reads with BLASTX scores of <0.01 in the faecal dataset were extracted. Using the aforementioned taxid-to-host database and the `taxids` associated with each read (described in Section 4.2.3), each faecal viral read was associated with one or more possible natural hosts. Reads were compiled into several different datasets, including (a) reads with highest homology to viral taxonomic groups known to only infect vertebrates (including humans), and (b) reads with highest homology to viral taxonomic groups that are known to infect vertebrates but that may also infect other host groups. When the viral genus or family were unknown or the viral species was not included in the taxid-to-host file due to its absence from the viral reference sequence database, that read was excluded in genus, family or species-level analyses respectively. Viral reads counts were also summarised at the family, genus and species level without consideration of probable hosts.

4.2.10 Statistical analysis

Viral reads with BLAST E-values <0.01 to species that exclusively infected vertebrates were normalised to (a) the number of vertebrate viral reads per million total Illumina reads for that sample, and (b) to the number of vertebrate viral reads per million viral reads for that sample (i.e., including viral reads that mapped with E-values <0.01 to any virus regardless of probable host). Sampling of birds of different ages varied slightly throughout the year, with on average fewer hatch-year birds being sampled in the May and June when they were still hatching or were extremely young, due to my unwillingness to disturb nesting birds. Wilcoxon rank sum tests were used to test whether birds aged 2 or under were more likely than older birds to have significantly higher or lower normalised numbers of viral reads mapping to vertebrates. Birds born in 2015 or 2016 (in their first and second years) were chosen in order to compare the youngest reasonable sized group of birds for which samples from that group were available from all sample occasions. Specifically, birds < 1 year were not sampled in May or June 2016 (due to having not yet hatched, or being extremely young), so the next oldest age group was included to form a single group of "young" birds. This grouping also follows biologically justifiable definitions of juvenile vs adult mute swans, as mute swans typically begin to breed in their third year [60]. Furthermore, age-structured mortality rates are higher in the first two-years of life in this population than during most of the adult life [284], which may in part reflect an effect of a maturing immune system and differences in susceptibility to death because of infectious diseases. All birds were used in this analysis, as samples from all birds included reads mapping to vertebrate viruses.

To control for effects arising from the month of sampling, Wilcoxon rank sum tests were conducted for the above analysis on 100 subsampled datasets. Specifically, for each dataset, the number of older age-group birds sampled in each month was downsampled to match the number of younger age-group birds sampled in that month.

The R package metagenomeSeq was used to explore differences in viral read abun-

dance and the presence or absence of different taxa by age [331]. Although total read count normalisation to reads per million per taxa is commonly used in metagenomics, resulting normalised data can be heavily biased by a small number of high-count taxa [51, 102, 331]. To avoid this bias, metagenomeSeq implements normalisation of different OTUs by cumulative-sum-scaling, in which raw counts are divided by the cumulative sum of counts up to a certain percentile, chosen based on the data [331].

metagenomeSeq was used to explore the effect of bird age on virus detection at a number of different taxonomic levels (viral species, genus and family). Zero-inflated Gaussian mixture models were used to detect differences in viral read abundance between birds of different age, using month of sampling as a covariate (note that samples were collected on five separate trips, each lasting three days in different months (Section 4.2.1)).

The zero-inflated Gaussian mixture models are designed [331] to reduce error by distinguishing between an absence of reads matching to a particular microbe in a sample because of genuine absence of that microbe, and a false-negative absence of reads because of low coverage sequencing of that sample that is unable to detect the genuine presence of a microbe. Note that, becuase this approach models read count data for each tested viral taxonomic group as normally distributed log-abundances in each comparison group (e.g., "younger" or "older" birds), the models are only appropriate for testing differences in abundance where the tested viral taxonomic group occurs relatively commonly throughout the dataset, and rare taxonomic groups are removed as appropriate in the following analyses. Here, "rare" is considered to be anything that occurs less than 6 times in the dataset, a definition that is relatively liberal.

Presence-absence testing was completed using Fisher's exact tests in metagenomeSeq, but note that these tests do not explicitly control for sample occasion.

4.3 Results: Virus Discovery

In total, 108 million reads were produced from 119 samples, of which 7.5 million reads (7%) had BLAST E-values < 1 to known viruses (including phages and satellites [3]). Approximately 100,000 reads had closest BLAST hits to viral species that are believed to infect vertebrates at an E-value of less than 0.01 (approximately 0.1% of all reads). Counts per sample of sequencing reads obtained, reads that were similar to known viruses based on BLAST matching, and reads matching to viruses that exclusively occur in vertebrates are presented in Figure C.1. Whole genomes of 9 putative new species of virus that likely infect swans were generated during assembly. The assembled new species include four new picornaviruses that each appear most closely related to one of the three species *Duck picornavirus GL/12*, *Avian sapelovirus*, *Goose megrivirus* and to the genus *Kobuvirus*. I also assemble a novel parvovirus that is most closely related to viruses from the avian parvovirus species *Chicken parvovirus* and *Pigeon parvovirus A*, and a novel viral species that appears to be most closely related to the species *Turkey stool-associated circular virus*. First, I decribe in detail several viruses for which I can perform phylogenetic analysis using multiple samples in the population, including a novel gammacoronavirus and two new species of astrovirus. Secondly, I describe in limited detail all other putative new species for which I only have a single reference genome (Section 4.3.2).

4.3.1 Discovery of novel viral species

4.3.1.1 *Waterbird coronavirus 1*

A 28,875 bp genome (likely missing the terminal ends of both UTRs) was assembled that appeared to encode all the proteins of a gammacoronavirus. Coverage across the genome was extremely high (95% of the genome was covered by > 1011 reads, with a median depth of 2214 reads). The lengths of open-reading-frames in the assem-

[3]Satellites are subviral agents that do not encode their own polymerase and therefore require replication by a different virus that co-infects the host cell.

bled genome were conserved compared to the nearest available reference genome (duck coronavirus KM454473) (Figure 4.5), so only the 3' part of the genome, which contains accessory proteins, was re-sequenced to confirm the assembly. This novel coronavirus, which I call waterbird coronavirus 1 (WatCV1), clusters within the diversity of the genus *Gammacoronavirus* in the ORF1b region of the replicase (Figure 4.4). Coronaviruses that share more than 90% amino acid identity in seven conserved regions of the replicase are considered to belong to the same species (ADRP in nsp3, nsp5, nsp12, nsp13, nsp14, nsp15 and nsp16) [217]. The novel coronavirus shared <90% identity in all but one domain to existing species in the genus *Gammacoronavirus* and to the coronavirus with the lowest BLAST E-value (NC_0010646, NC_001451, NC_010800, KM454473 (from the putative species *Duck coronavirus*). Specifically, highest identity between WatCV1 and each of these four reference genomes across each definitive domain was; ADRP; 44%, nsp5; 58%, nsp12; 84%, nsp13; 90%, nsp14; 78%, nsp15; 59%, nsp16; 77%. The primary structural difference between WatCV1 and duck coronavirus is in the accessory protein region, in which two accessory proteins that appear similar at the amino acid level to ORFX are observed in the new coronavirus rather than just one in viruses from the species *Duck coronavirus* (Figure 4.5). This section of the genome was resequenced to confirm the assembly, as detailed in Section 4.2.5.

PCR and direct Sanger sequencing was used to amplify a short amplicon of the polymerase gene of WatCV1 across eight samples (primers in Table C.4). The polymerase gene was chosen because this region has been the target of previous studies exploring gammacoronavirus diversity, and hence the diversity of this region is better understood than the diversity of many other genomic regions. Phylogenetic analysis suggests that the partial polymerase of WatCV1 forms a clade with other gammacoronaviruses found in geese (Figure 4.6). The WatCV1 sequences from birds at Abbotsbury do not form a monophyletic clade, despite being sampled on 3 consecutive days in October 2016.

The existing gammacoronavirus sequences that form a major clade with the virus

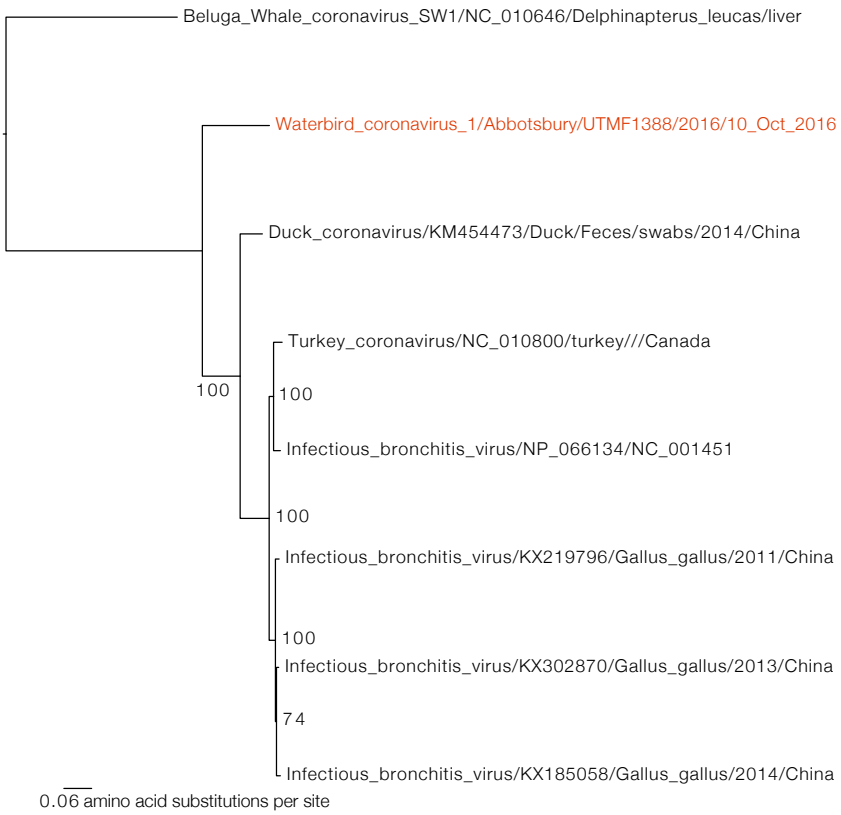

Figure 4.4: Midpoint rooted maximum likelihood phylogenetic tree estimated from an amino acid alignment of the ORF1b replicase region of several gammacoronaviruses. The tree was constructed using substitution model LG+G. Labels at nodes represent bootstrap scores >50. The sequence from Abbotsbury is in red.

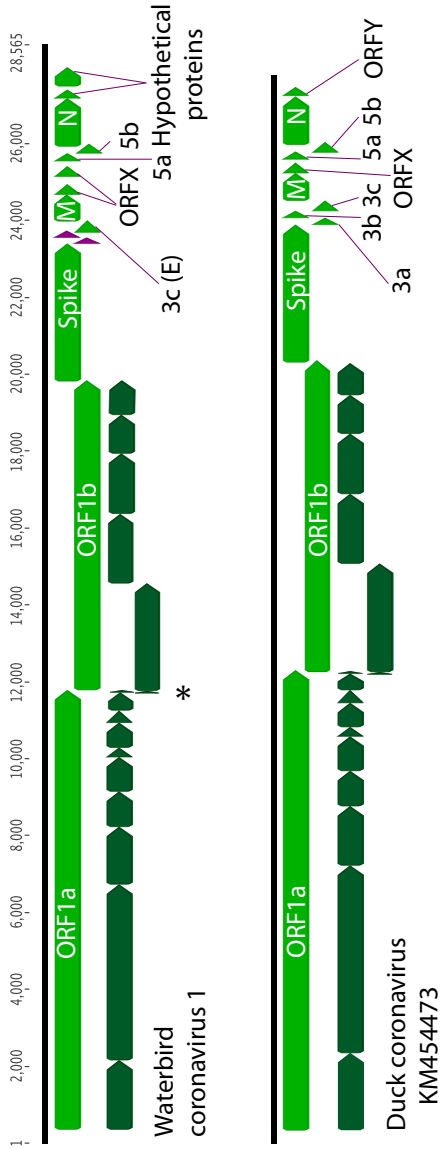

Figure 4.5: Comparison of waterbird coronavirus 1 (WatCV1) to a virus from the closely related putative species *Duck coronavirus*. Lighter green blocks in the duck coronavirus genome represent open reading frames (ORFs) and darker green blocks represent cleaved peptides. Lighter green blocks in the WatCV1 genome represent ORFs with BLASTX or BLASTN homology to existing gammacoronavirus ORFs and purple blocks represent other possible ORFs >100aa that do not show significant similarity to existing gammacoronavirus genes, yet are possible candidates for genes 3a and 3b based on their position in the genome. Darker green blocks in WatCV represent possible cleaved peptides based on alignment to the duck coronavirus. An asterix indicates the location of a single cleavage site that was annotated in the duck coronavirus, but that is not well conserved in WatCV1 and so the exact location of this cleavage site (if any) is unclear.

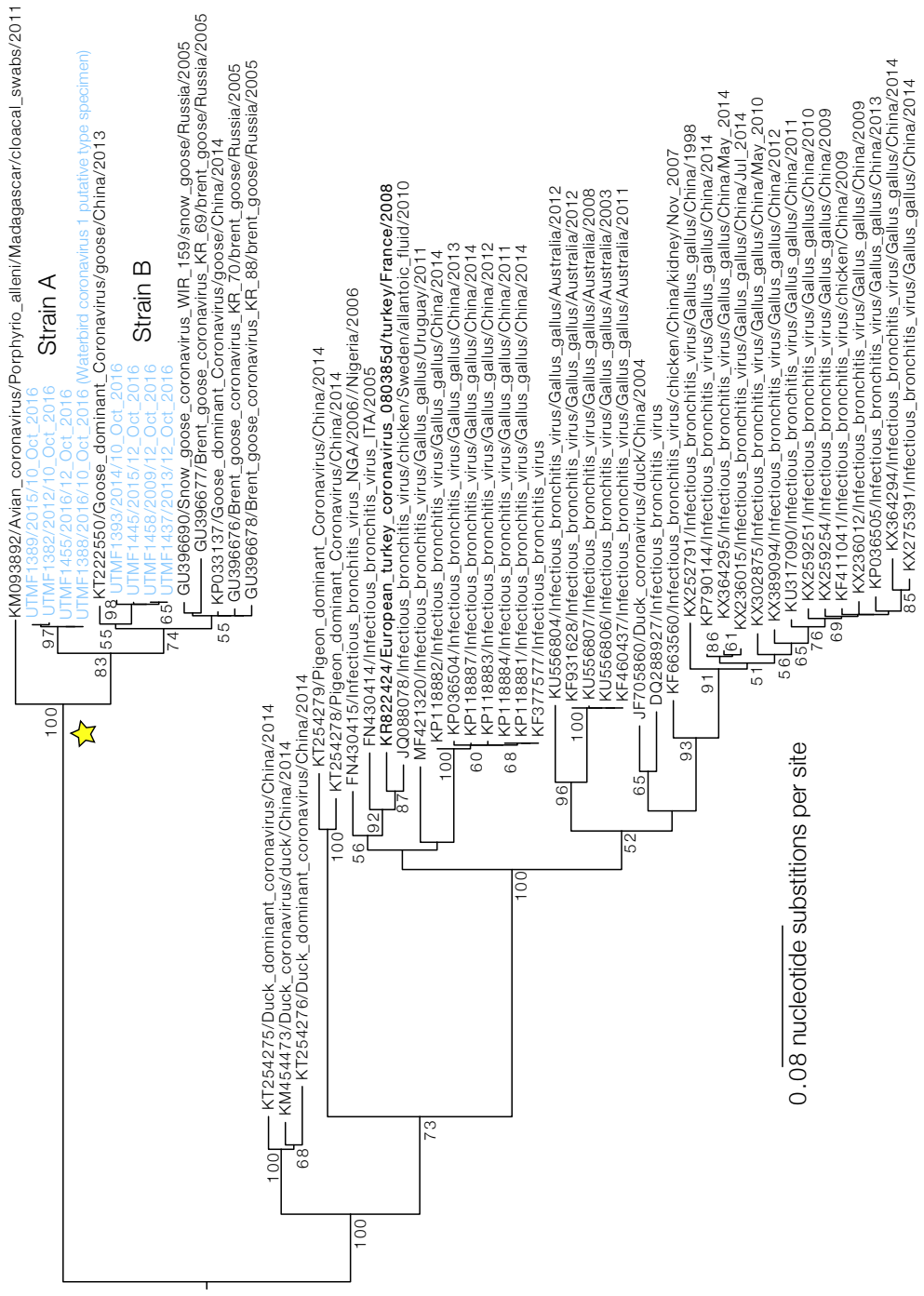

Figure 4.6: Midpoint rooted maximum likelihood phylogenetic tree estimated from a 996bp nucleotide region of the gammacoronavirus polymerase, estimated with substitution model TIM2+I+G. Labels at nodes represent bootstrap scores >50. Sequences from swans at Abbotsbury are in blue. Strains A and B are marked, as referred to in the text. A yellow star marks the major clade for which WATCV1 forms the first whole-genome identified in that clade.

identified here are all only partial genomes (starred clade on Figure 4.6). I therefore believe that the virus identified here is the first whole genome to be sequenced from this clade, and suggest that it be considered the type specimen of a new species of virus infecting waterbirds. I propose that this new species of coronavirus be named *Waterbird coronavirus 1*.

Targeted assembly of the second strain detected in the phylogenetic analyses (marked in Figure 4.6 as strain B) resulted in a partial assembly of a second genome at lower coverage. Amino acid level identity with WatCV1 is high in species-definitive regions, with >96% amino acid level identity in nsp12-16, 93% in the ADRP region and 86% in nsp5. I therefore suggest that this partial genome represents a related strain to the new species characterised here, rather than representing a second new species. The fact that strain B swan coronaviruses are more closely related to goose coronaviruses that they are to strain A swan coronaviruses could be explained by the occurence of at least one cross-species transfer in the history of these strains (Figure 4.6) . Whether cross-species transfer of this viral species occurs commonly amongst geese and swans (or, indeed, other untested waterbirds), should be investigated with further testing for this coronavirus species at Abbotsbury and in other populations.

4.3.1.2 Mute swan astroviruses

Several astroviruses were assembled that were highly similar to each other in the ORF1a and ORF1b region but had divergent capsid genes. One of these viruses was chosen for re-sequencing to confirm the assembly (mute swan astrovirus 1). This virus was originally assembled with a median sequencing depth of 533 reads, and 95% of the genome covered by at least 258 reads. Phylogenetically, the ORF1b of this virus groups with other bird viruses in the genus *Avastrovirus* (Figure 4.7). Currently, astrovirus species are defined based on the hosts that they infect, and as such, the finding of an astrovirus in a swan is sufficient for its inclusion as a new species. A new classification system proposed by ICTV in 2010 recommended that a new species should only be defined based on >75% amino acid identity in the capsid

region (ORF2) [104, 217]. The novel astrovirus has an amino acid identity of 46% to the most similar astrovirus based on pairwise genetic distance in the capsid region (chicken avastrovirus), and hence represents a new species of astrovirus; *Mute swan astrovirus 1*).

A second whole astrovirus genome was assembled at lower depth (median 139 bases per read, 95% of the genome with a coverage of 41 bases or more). The second astrovirus genome has 96% amino acid identity to mute swan astrovirus 1 in ORF1a and 98% identity in ORF1b, but is distinct in the ORF2 region with an amino acid identity of 46%. If the new proposed species definitions are accepted [104, 217], this astrovirus would therefore represent a second species infecting swans, *Mute swan astrovirus 2*.

Multiple PCR primers were designed against assembled sequences, and Sanger sequencing was used to generate an amplicon of approximately 1096 bp bridging the junction between ORF1b and ORF2. Sanger sequencing confirmed the presence of at least 3 distinct capsid proteins (including those matching the two new species, above), that were all associated with a more conserved ORF1b. Based on a phylogenetic tree estimated from amino acid level alignments of the last 411 nucleotides of ORF1b, all three distinct types of swan astrovirus formed a well supported single clade (bootstrap score of 81%) that groups with a monophyletic clade of astroviruses found in ducks, albeit with a low bootstrap support (52%) (Figure 4.8). A tree estimated from an alignment of the first 150 amino acids of ORF2 shows that, at least in this short region, the 3 distinct types of astrovirus found here do not form a monophyletic clade to the exclusion of all other astroviruses (Figure 4.9). Maximum amino acid identities to other astroviruses in this short region for the three distinct genotypes are 73% (Sample UTMF1420 (mute swan astrovirus 2) with JN582323), 58% (Sample UTMF1260 (mute swan astrovirus 1) with KY038163) and 82% (UTMF1123 with JQ307838). I suspect that the third distinct genotype, represented by UTMF1123, may be a third new species of astrovirus, but I cannot confirm this until the entire ORF2 region is assembled or sequenced in full.

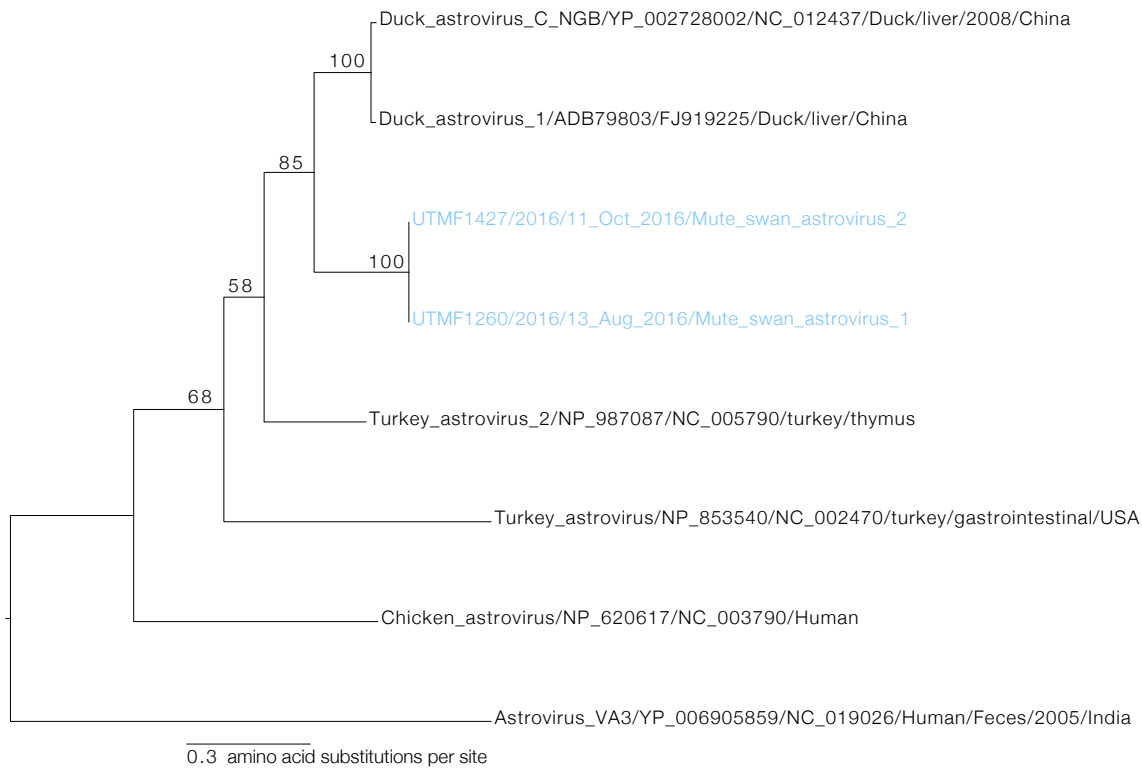

Figure 4.7: Midpoint rooted maximum likelihood phylogenetic tree estimated from an amino-acid alignment of the conserved region of the astrovirus ORF1b gene, using substitution mode LG+G. Labels at nodes represent bootstrap scores >50.

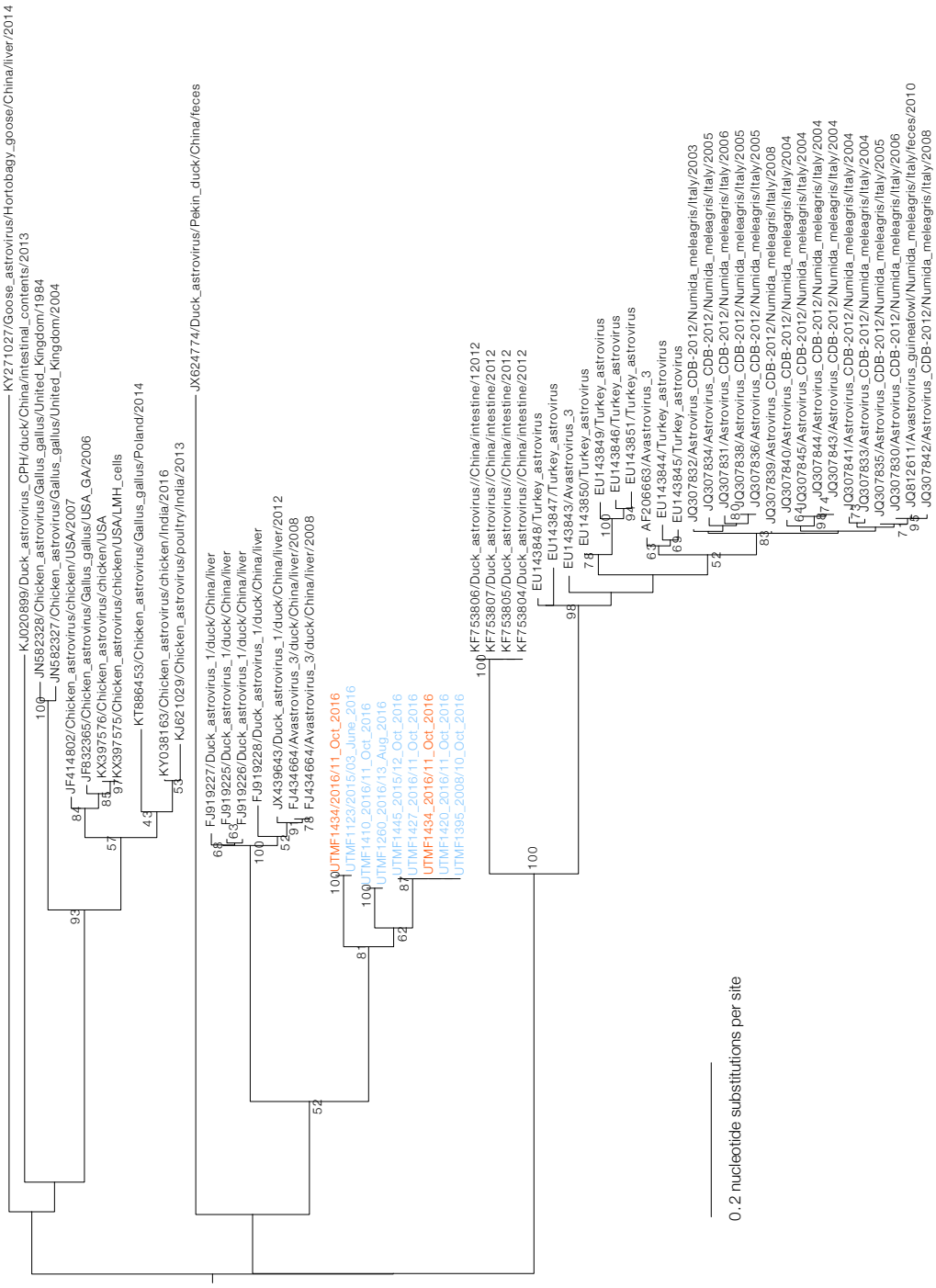

Figure 4.8: Midpoint rooted maximum likelihood phylogenetic tree estimated from an alignment of 411 nucleotides at the 3' end of the astrovirus ORF1b gene, estimated with substitution model TrN+I+G. Labels at nodes represent bootstrap scores >50. Samples from swans at Abbotsbury are in blue, except for one co-infected individual that is in red.

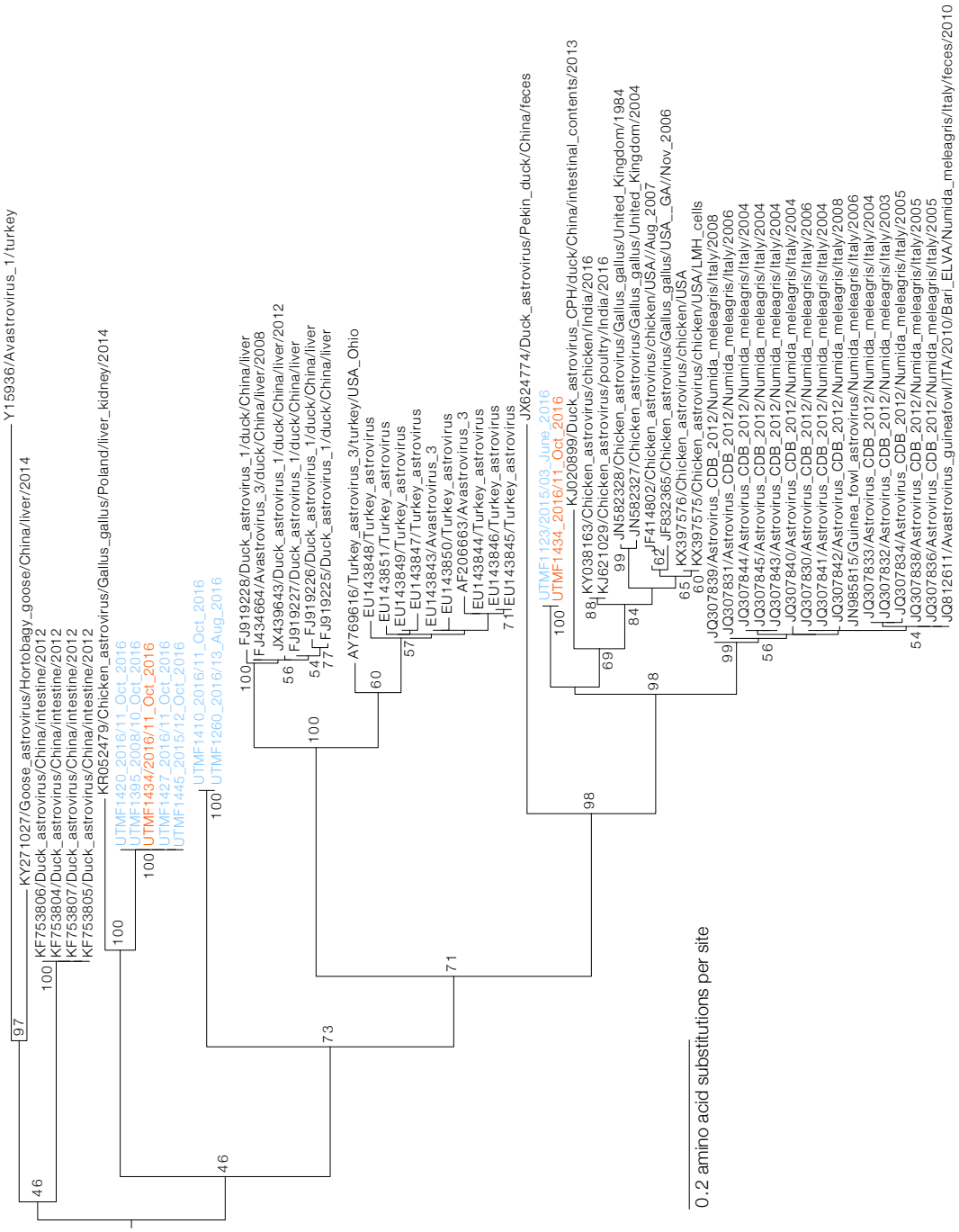

Figure 4.9: Midpoint rooted maximum likelihood phylogenetic tree estimated from an amino-acid alignment of the conserved region of the 5' end of the astrovirus ORF2 gene, using substitution model LG+G+F. Labels at nodes represent bootstrap scores >50. Samples from swans at Abbotsbury are highlighted in blue, except for one co-infected individual that is in red.

Astroviruses are present in the population on all five tested sample occasions. There is substantial diversity of astroviruses in a single time point, as all astrovirus types sequenced here are present during October 2016 (Figure 4.9). Although I did not systematically test for co-infections, two of the three different types were confirmed by amplicon sequencing as occurring within a single faecal sample from an individual that hatched in 2016 (highlighted in red on Figures 4.8 and 4.9).

4.3.2 Additional novel viruses

The following section presents other assembled genomes that appear to represent novel viral species. This section is descriptive and somewhat repetitive, detailing the basic properties of the viral genome assembly, nearest relatives as defined by phylogenetic analyses, and establising that each of these viruses fulfil the criteria for inclusion as a new species. No significant work has yet been undertaken to explore the within-population dynamics of these viruses.

4.3.2.1 *Mute swan picornavirus*

A near-complete genome (likely missing terminal ends of both UTRs) of a novel picornavirus was assembled, with a median depth of 31 bases (95% of the partial genome has a depth of by 13 or more bases). The polyprotein is 7248 bases long and is 42.9% GC, and is therefore within the typical range for a picornavirus (7-8.8kb and 35-60% GC, [217]). BLAST searching showed that is has highest identity to the virus duck picornavirus GL/12 (accession number NC_023985). An amino acid phylogenetic tree built from an alignment of conserved regions of the polyprotein groups the novel picornavirus in the same clade as duck picornavirus GL/12 with a strong bootstrap support of 100 (Figure 4.10). Species demarcation criteria are poorly defined for many recently discovered picornavirus genera, yet where criteria are defined members of a species are typically expected to have greater than 70% amino acid identity across the whole polyprotein [217]. In the more conserved (3') region of the polyprotein,

the novel picornavirus shares 67% amino-acid level identity with duck picornavirus GL/12. I hence tentatively suggest that this novel picornavirus be considered a new species in the same genus as the species *Duck picornavirus GL/12* [468], and propose the name *Mute swan picornavirus*.

Figure 4.10: Maximum likelihood tree of mute swan picornavirus and related viruses. The tree was estimated from an amino acid level alignment of the conserved region of the polyprotein using substitution model LG+G. Bootstrap scores >50 are shown, and the putative new species is in blue.

4.3.2.2 *Mute swan sapelovirus*

A novel 7996bp sapelovirus was assembled. The polyprotein is 7284bp long, and has a GC content of 42.6%. The assembly had a median depth of 152 bases, and 95% of the genome has a depth of 67 bases or more. An amino acid phylogenetic tree built from an alignment of conserved regions of the polyprotein clusters this new sapelovirus closely with other avian picornaviruses (NC_016403 and NC_006553)

(100% bootstrap support, Figure 4.11). Whilst the assembled sapelovirus appears closely related to the existing species, *Avian sapelovirus*, I suggest that under ICTV definitions it should be considered a new species of the genus *Sapelovirus*, because of the relatively high divergence across the polyprotein as a whole. Amino acid identity with avian sapelovirus is high in the 2C and 3CD region of the polyprotein (74% and 72%, respectively), but much lower in the VP1 region (41%) and across the whole polyprotein (57% across the easily alignable 2-3 region, but lower across the leader peptide and viral proteins). I suggest the species name *Mute swan sapelovirus*.

Figure 4.11: Maximum likelihood tree of mute swan sapelovirus and related viruses. The tree was constructed from an amino acid level alignment of the conserved region of the polyprotein using substitution model LG+G. Bootstrap scores >50 are shown, and the putative new species is highlighted in blue.

4.3.2.3 *Mute swan parvovirus*

A novel parvovirus was assembled that forms a strongly supported monophyletic clade with other viruses detected in avian faecal samples in amino acid trees of alignable regions of both the non-structural and capsid proteins (Figure 4.12). The depth of coverage was extremely high, with median depth >7810 (maximum depth limited to 8000 during reference mapping with BWA) and 95% of bases covered by 1817 reads or more. Amino acid identity in the non-structural protein with other avian parvoviruses is <50%. I propose the name *Mute swan parvovirus* for this new species.

4.3.2.4 *Mute swan megrivirus*

A whole genome of a megrivirus was assembled (median depth; 1737, 95% of bases covered by a depth of 416 bases or more). The polyprotein is 72% identical at the amino acid level to both the goose and duck megriviruses (NC_033793 and NC_024120). Based on the predicted cleavage sites reported in reference [33], the megrivirus shares a maximum of 44% amino acid identity in the VP1 region with other putative species of the genus, but is much more similar to existing species, with maximum 98% and 97% amino acid identity in the non-structural proteins 2C and 3CD, respectively.

Reads with significant BLAST hits to megrivirus were present in most samples (Figure 4.23).

4.3.2.5 *Mute swan stool-associated circular virus*

A partial genome was assembled that comprised two opposing direction open-reading frames. Based on amino acid identities of both open-reading frames, the most closely related species is *Turkey-stool associated circular virus* (TuSCV) (accession number KF880727) that forms part of a clade of currently unclassified, single-stranded circular DNA viruses. Phylogenetic analysis of the replicase region suggests that there is weak bootstrap support for the mute swan stool-associated virus being most closely related to TuSCV (Figure 4.13. The median depth of coverage 31, and 95% of the genome

had an assembly depth of 13 or more reads. Perhaps a result of this low depth, the long intergenic genomic region could not be circularized by iterative assembly. Similarly to the TuSCV, the longer ORF (1011bp) encodes a capsid protein, and the shorter ORF (724bp) encodes a replicase. The intergenic region between the genic 3' ends is 228bp. Limited effort was put into circularising the sequence using specific PCRs. These PCRs generated amplicons that would be consistent with an intergenic region of 450-500bp between genic 5' ends, but amplicon sequencing quality was poor and should be repeated.

4.3.2.6 *Mute swan stool-associated virus*

A 9209 bp contig was assembled that appeared to encode a single 8883 bp ORF with highest similarity to a posavirus. Most of the genome could not be aligned confidently with other posaviruses because of high divergence. Even in the most conserved regions of the encoded polyprotein amino acid identity to the next closest sequence (YP_009270628) was never more than 77% (77% in reverse-transcriptase-like region, 63-64% in two RHV-like domains, 54% in the RNA helicase region, based on alignment with previously curated posaviruses [386]). Detailed previous study on posaviruses in pigs has suggested that posaviruses represent infections of nematodes present in the pig's gastrointestinal tract, and not genuine infection of the pig [386]. I suggest that this novel species *Mute swan-stool associated virus* represents a new picornavirus in the same (currently unclassified) genus as the posaviruses, but I caution that current published evidence on other related species points towards it being a nematode-infecting virus.

4.3.2.7 *Avian kobuvirus*

A 7929 bp contig was assembled that encoded a 7482 bp ORF that was most similar to picornaviruses from the *Kobuvirus* genus. The species forms an outgroup to the current kobuvirus genera (Figure 4.14). New species in the genus *Kobuvirus* should share <70% amino acid identity in the polyprotein, including <70% identity in P1

and <80% identity in 2C and 2D. Based on cleavage sites predicted by alignment with representatives of *Kobuvirus*, maximum amino acid level identity to any other known *Kobuvirus* in the VP1, 2C and 3C regions is extremely low, at 32%, 45% and 29%, respectively. Amino acid identity is highest in the VP3 region (56% identity to mouse kobuvirus M-5, NC_015936) and 3D region (53% identity to bovine kobuvirus and to caprine kobuvirus). The kobuvirus-like swan virus is the first example of a kobu-like virus to be found in birds. Whilst the new virus appears to be related to the kobuviruses, I tentatively suggest that it is sufficiently divergent to merit inclusion in a new genus under the current classification scheme for genera in the *Picornaviridae*.

4.3.3 Genomic assembly of existing virus species

Several of the viruses that I was able to assemble have been previously characterised in waterfowl, including viruses from the species *Aquatic bird bornavirus 1* (ABBV1), *Goose calicivirus, Influenza A virus, Infectious bursal disease virus* (IBDV), *Avian orthoreovirus, Goose parvovirus, Swan circovirus*, and several short contigs of an adenovirus. Nucleotide alignments were constructed and maximum likelihood phylogenetic trees built for each virus to confirm that the nucleotide sequences assembled fell within the existing species diversity.

4.3.3.1 Existing species of *Caliciviridae*

A calicivirus of length 9599 bp was assembled that appeared to encode two open reading frames of 7905 bp and 852 bp, respectively. The genome was covered at a median depth of 42x, with 95% of the genome covered by 14 or more bases, and < 80% of the genome covered by 27 or more bases. The calicivirus forms a clade with a recently described goose calicivirus (H146) that has been proposed by the authors to be the founding member of a new genus within the *Caliciviridae* [465]. Species demarcation criteria within the *Caliciviridae* are poorly defined, relying on phylogenetic divergence, differences in structure and host-range [217]. Based on cleavage

site predictions, the new calicivirus shares amino acid identities with goose calicivirus H146 of 56% in VP1 and 85% in VP2 (ORF2). In non-structural regions, the caliciviruses share > 81% amino acid identity in the NTPase, VpG, P29 and Pro-Pol regions (respectively 92%, 81%, 86% and 85%). The majority of the Nterm protein is conserved between the two viruses (318 bp at the 5' end share 88% amino acid identity), yet the calicivirus assembled in this study appears to be significantly longer in this region (641 aa) than H146 (422 aa). On the basis of similar host and close identity in non-conserved regions, I conservatively suggest that this calicivirus should currently be grouped into a species together with the existing calicivirus found in geese, goose calicivirus H146.

4.3.3.2 Divergent strain of the *Avibirnavirus* IBDV

One virus, an avibirnavirus with high homology to IBDV, appears to represent a divergent strain of IBDV, sharing at most 84% nucleotide level identity (93% amino acid identity) with other IBDVs in the polyprotein region, and at most 88% (97%) in the VP1 region. In phylogenetic trees estimated from nucleotide alignments of a wide variety of other IBDV viral sequences, the Abbotsbury avibirnavirus falls outside of well-supported clades that encompass the majority of IBDV sequences (Figure 4.15). In phylogenetic trees estimated from amino acid alignments of coding regions of both gene segments, the Abbotsbury avibirnavirus falls with two other IBDV viruses within a well-supported, monophyletic clade. Both segments of the avibirnavirus with were assembled to low coverage [4], but the sequences of both open reading frames were confirmed by Sanger sequencing of overlapping PCR amplicons. Interestingly, the Abbotsbury avibirnavirus sequence was sampled from a cygnet that was only a few months old at the time of sampling. All other viruses included in Figures 4.15 and 4.16

[4]The ORF encoding the polyprotein had a median depth of 63 (95% of the genome coverage by 23 or more bases). Although the placement of the Abbotsbury avibirnavirus in these trees suggests that the virus probably is relatively unlike many other sampled IBDV viruses, the phylogenetic placement and support values shown in FIgure 4.15 should be cautiously interpreted, as sequences with long terminal branches have a tendency to group together regardless of their true genealogical relationships, often with high support ("long-branch attraction" [154, 130]). The VP1 segment was initially assembled to a median depth of 41 (95% of bases covered by 16 reads or more).

do not have host annotations on GenBank, or were sampled exlusively in chickens or turkeys. The diversity of IBDV in wild birds has not been well-explored, but in chickens the disease affects mostly young birds, < 6 weeks old [401]. It is easily possible that IBDV diversity in wild species might have been missed if the disease was similarly limited to extremely young birds in the wild, due to preferential sampling of birds that are not still on the nest.

Other genomes or partial genomes that I assembled appeared to fall more clearly within existing diversity. A swan circovirus sequence was assembled that appears closely related to other swan and goose circoviruses based on the conserved NS1 gene (Figure 4.17). For ABBV1, goose parvovirus and the sigma NS segment of avian orthoreovirus, phylogenetic trees demonstrate phylogenetic clustering of Abbotsbury sequences with other European isolates (Figures 4.18, 4.19, 4.20 and 4.21). Although all segments could not be assembled in their entirety due to low coverage, I believe that this represents the first identification of an orthoreovirus in a swan.

Several contigs mapping to the genera *Aviadenovirus* and *Siadenovirus* were assembled with extremely low coverage (only 936 reads with closest BLAST matches to adenoviruses were identified among all the raw reads). A 2,159 bp region of the polymerase mapped to goose adenovirus 4 (genus *Aviadenovirus*) with 96% nucleotide identity. The sample with the highest number of aviadenovirus reads does not also contain a correspondingly high number of siadenovirus reads, and vice versa. It is therefore possible that at least two distinct adenoviruses may be present in the population, although this should be explored further.

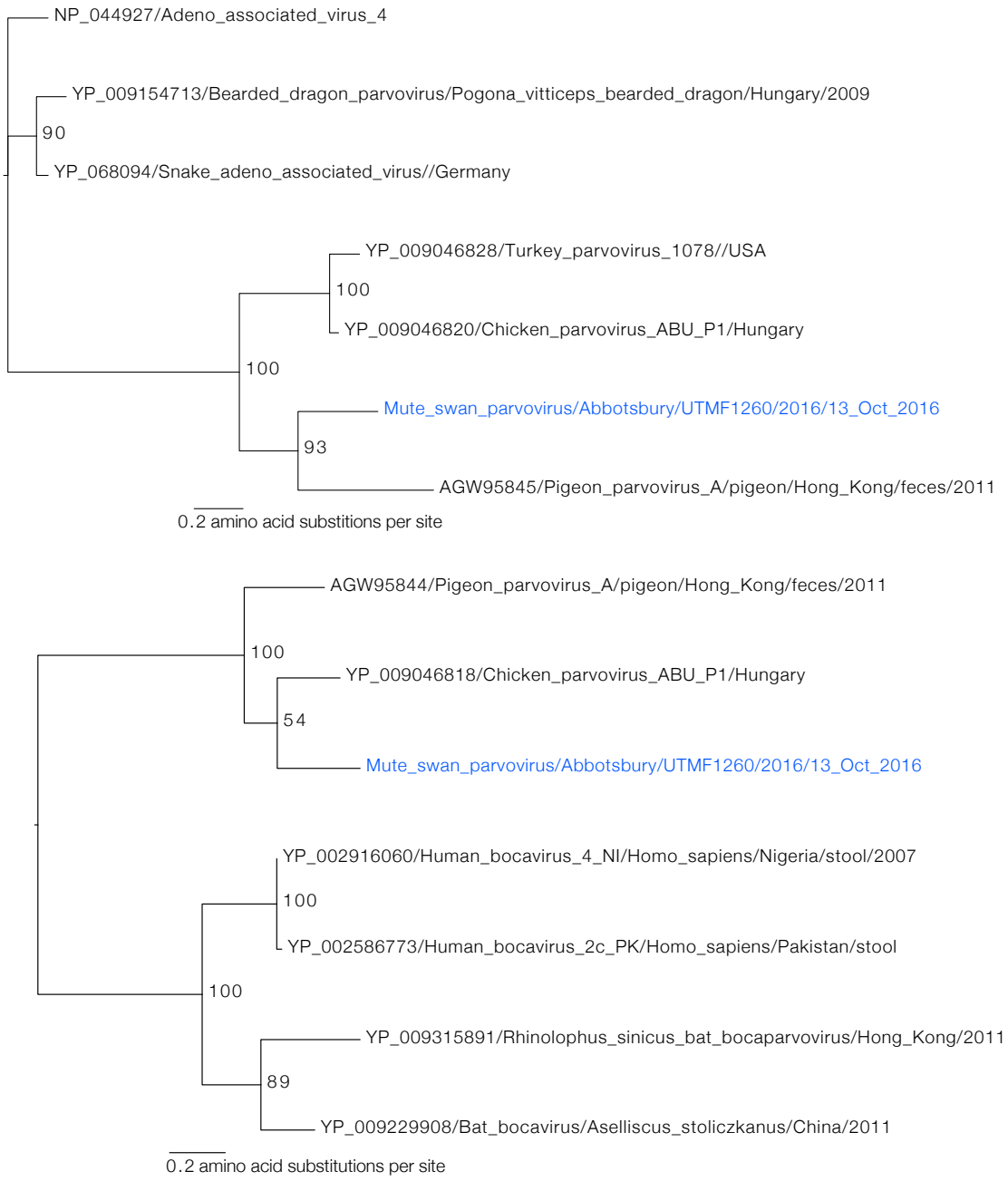

Figure 4.12: Maximum likelihood tree of mute swan parvovirus and related viruses. Bootstrap scores >50 are shown, and the putative new species are highlighted in blue. Upper: Capsid region. Lower: NS1 region. The trees were constructed from an amino acid level alignment of conserved regions using substitution models RtREV+G (capsid) and LG+G (NS1).

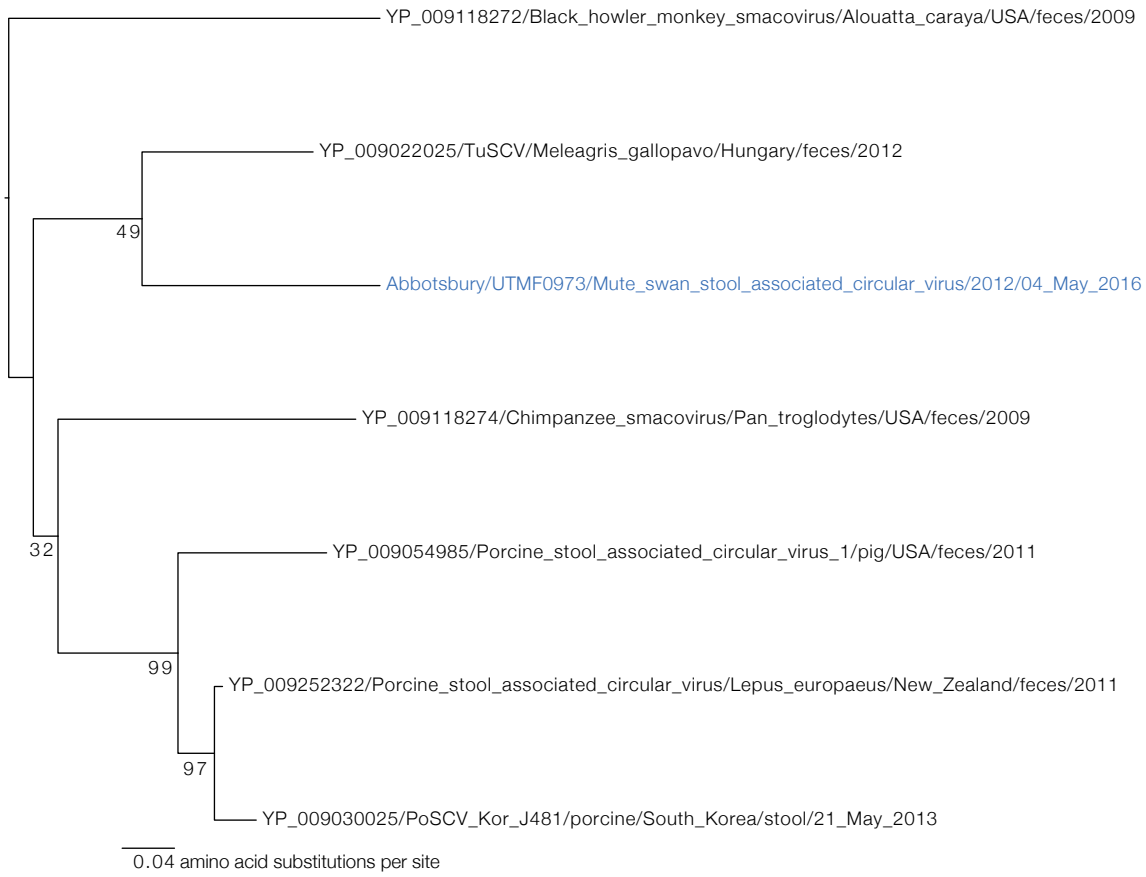

Figure 4.13: Maximum likelihood tree of replicase region of stool-associated circular viruses. The tree was constructed from an amino acid level alignment of the conserved region of the polyprotein using substitution model VT+G. Bootstrap scores >50 are shown, and the putative new species is highlighted in blue.

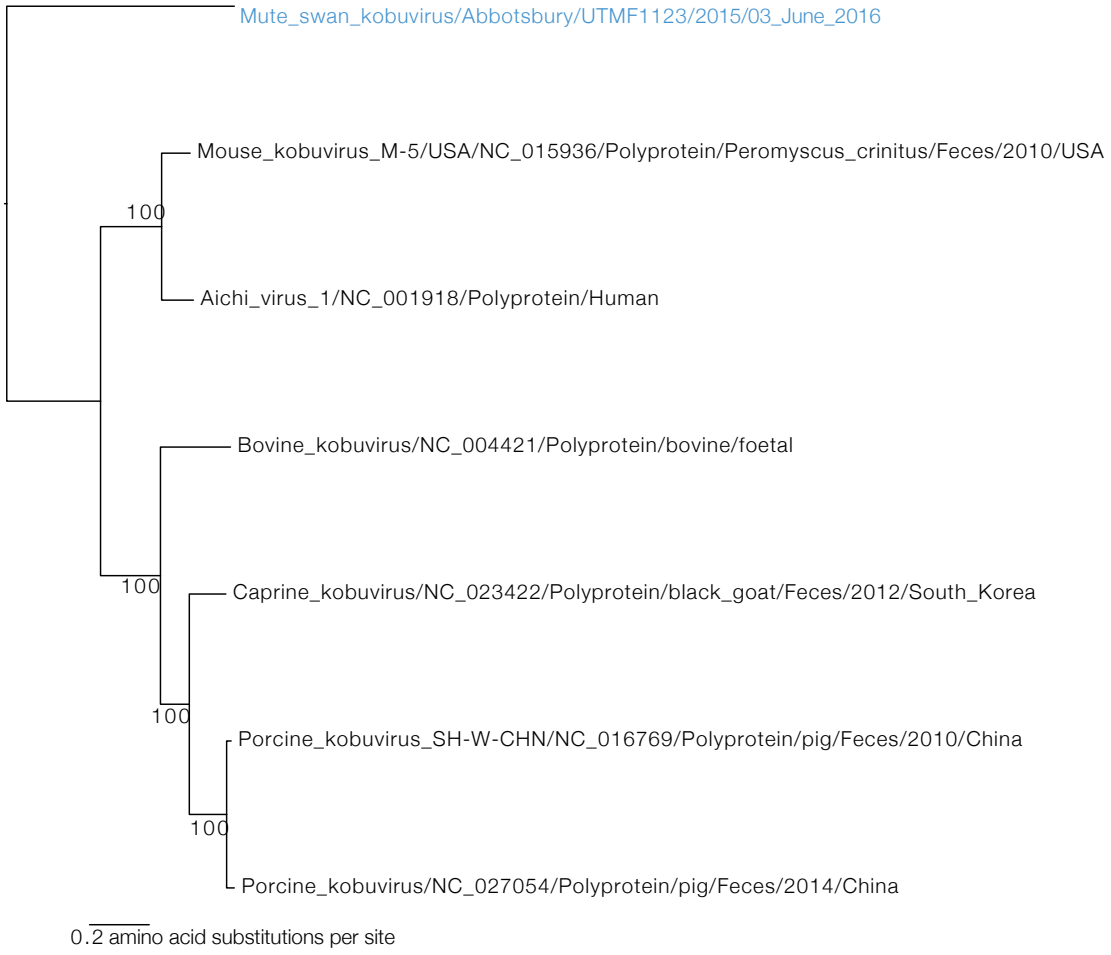

Figure 4.14: Maximum likelihood tree of *Kobuvirus* genera and putative novel related genus. The tree was constructed from an amino acid level alignment of the conserved region of the polyprotein using substitution model LG+G. Bootstrap scores >50 are shown, and the putative new genus is highlighted in blue.

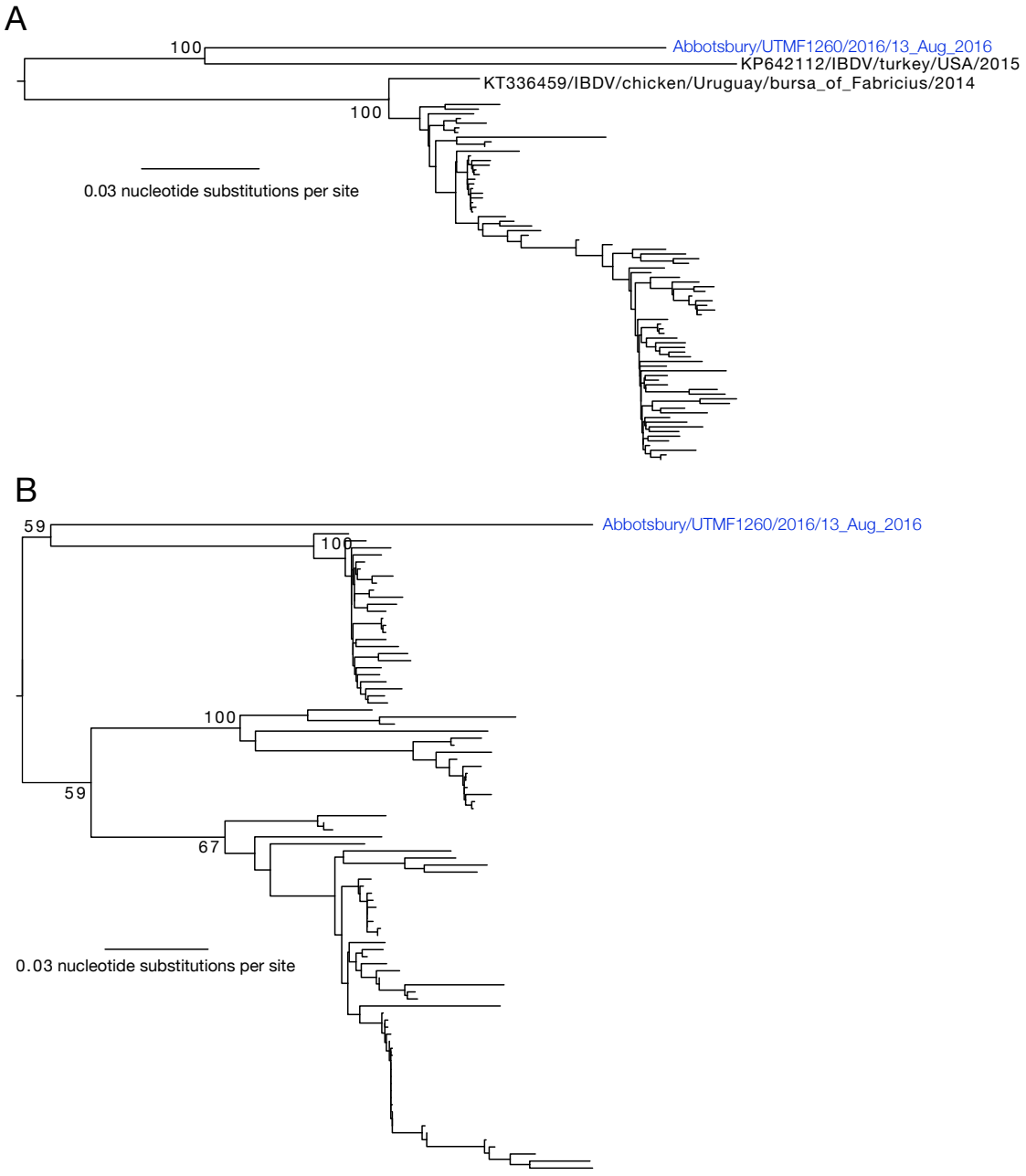

Figure 4.15: Midpoint rooted phylogenetic trees constructed from nucleotide-level alignments of the IBDV polyprotein segment (upper) and VP1 segment (lower). Both trees were constructed using the substitution model GTR+I+G. Bootstrap scores are only shown for major clades.

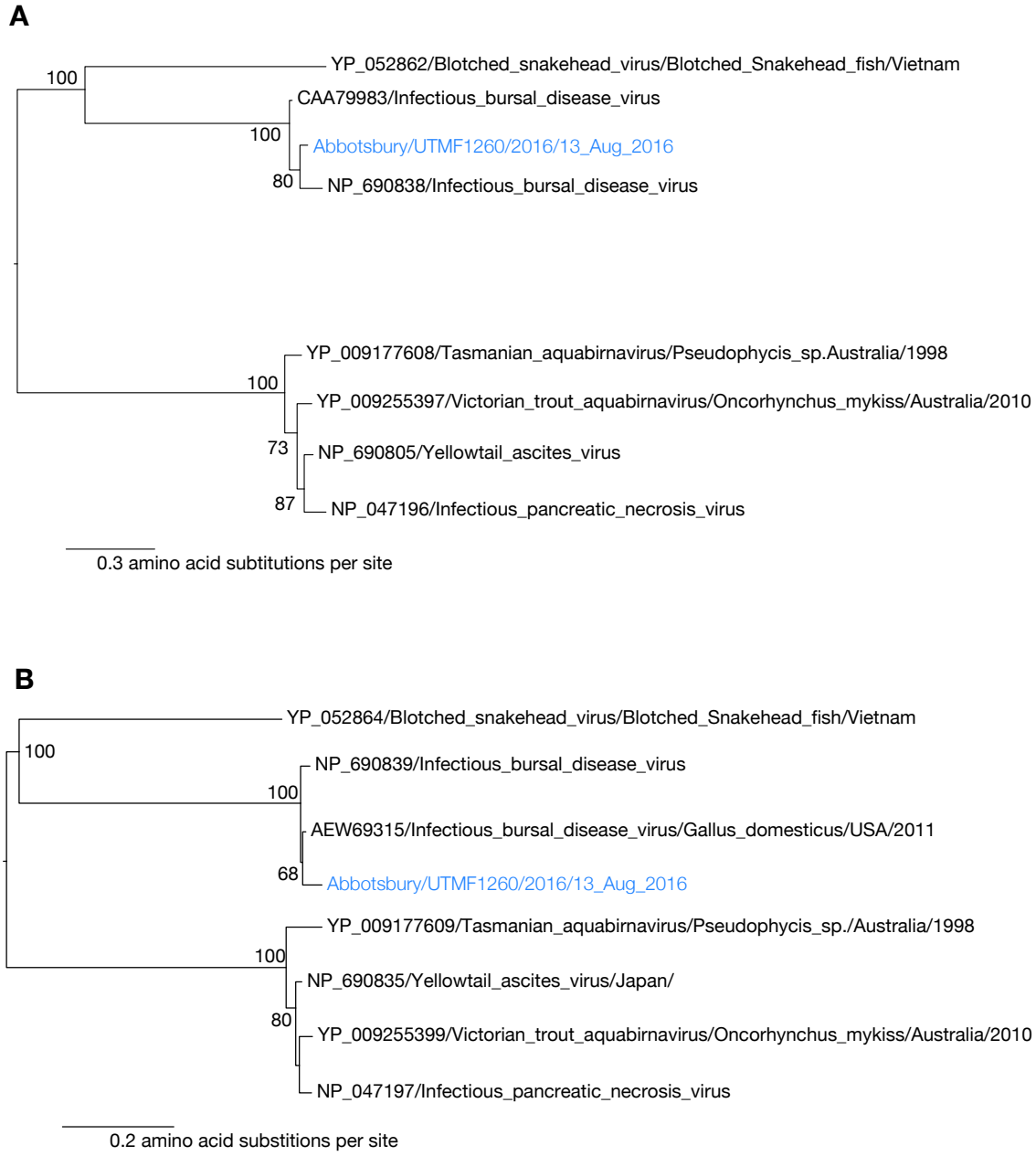

Figure 4.16: Midpoint rooted phylogenetic trees estimated from amino-acid-level alignments of the IBDV polyprotein segment (upper) and VP1 segment (lower). Both trees were estimated using the substitution model LG+G. Bootstrap scores > 50 are shown.

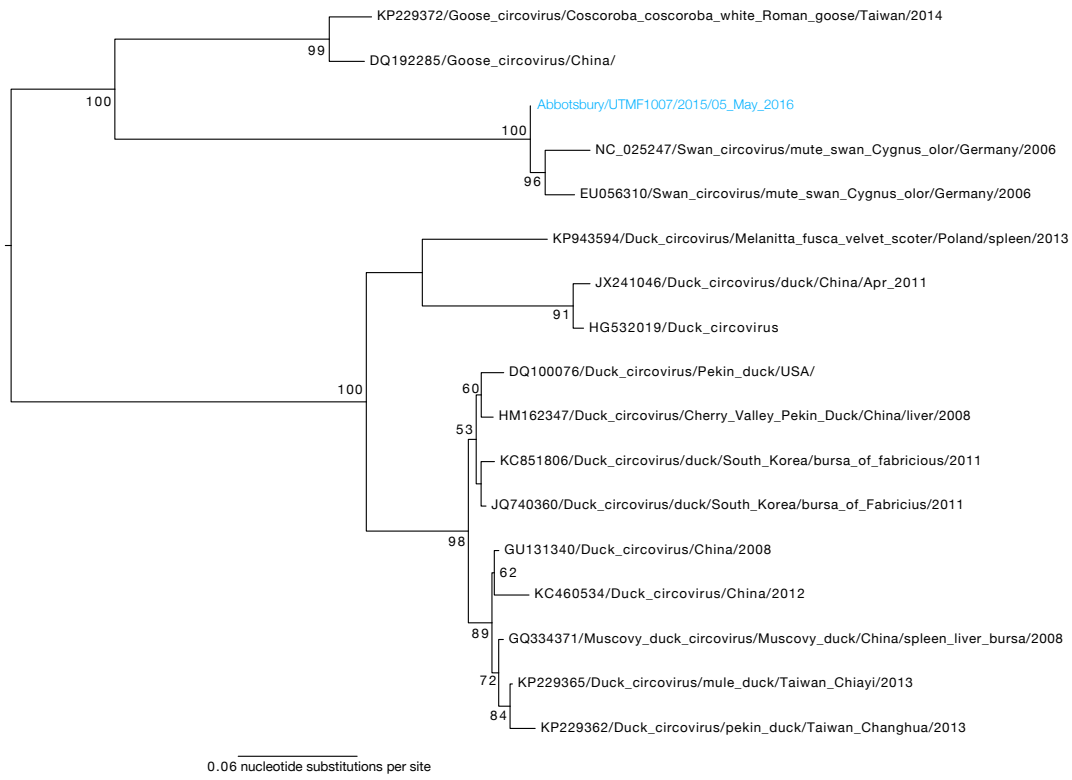

Figure 4.17: Midpoint rooted phylogenetic trees estimated from an alignment of avian circovirus NS1 sequences, including all sequenced isolates from the species *Swan circovirus* and a selection of duck and goose circoviruses. The tree was estimated using substitution model TrNef+G. Bootstrap scores > 50 are shown. A sequence from Abbotsbury is in blue.

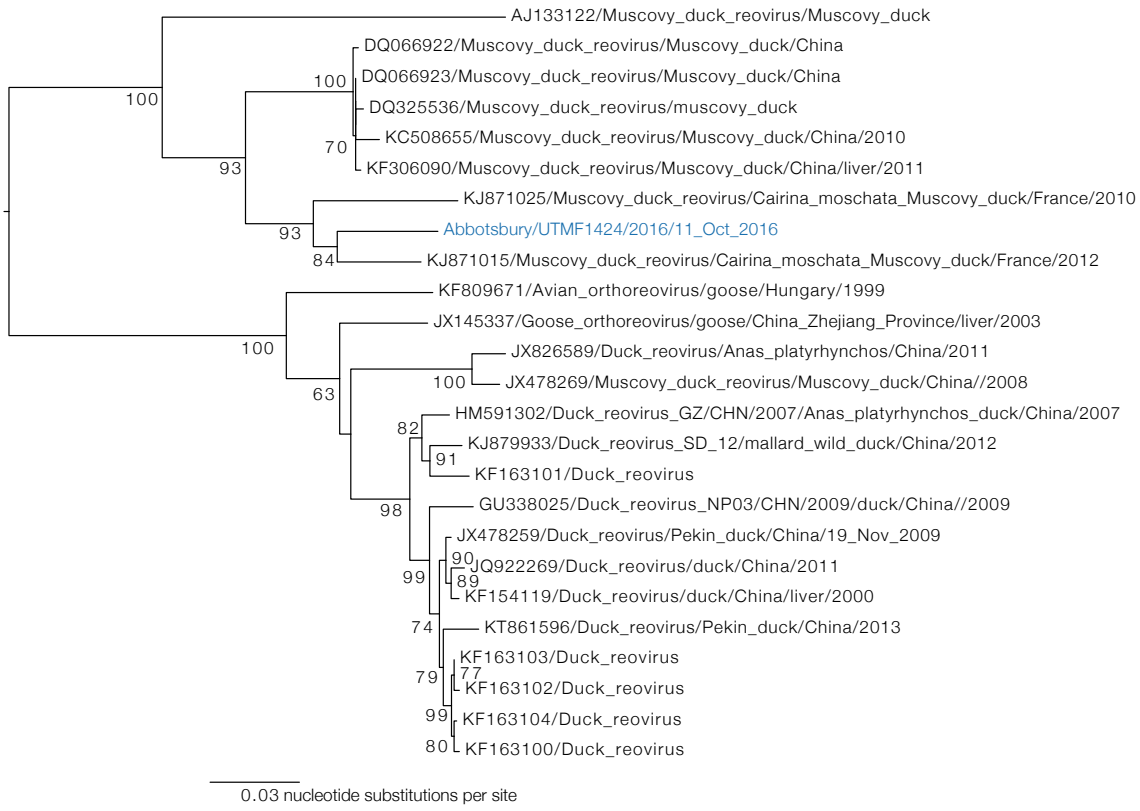

Figure 4.18: Midpoint rooted phylogenetic trees estimated from an alignment of avian orthoreovirus sigma NS segment, including the clade of sequences within which the Abbotsbury sequence falls. The tree was estimated using substitution model TrNef+G. Bootstrap scores > 50 are shown. A sequence from Abbotsbury is in blue.

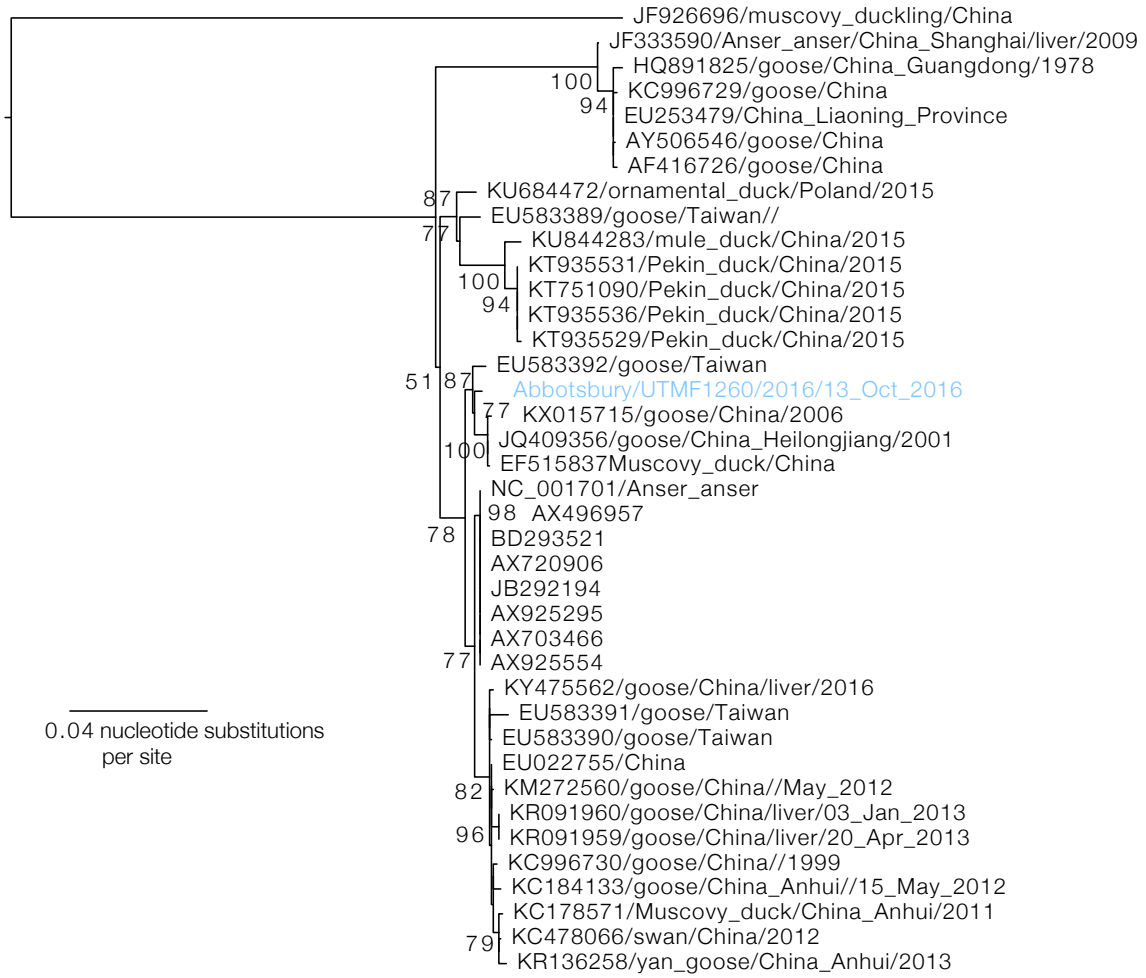

Figure 4.19: Midpoint rooted phylogenetic trees of goose parvovirus NS1. The tree was estimated using the HKY substitution model. Bootstrap scores > 50 are shown. A sequence from Abbotsbury is in blue.

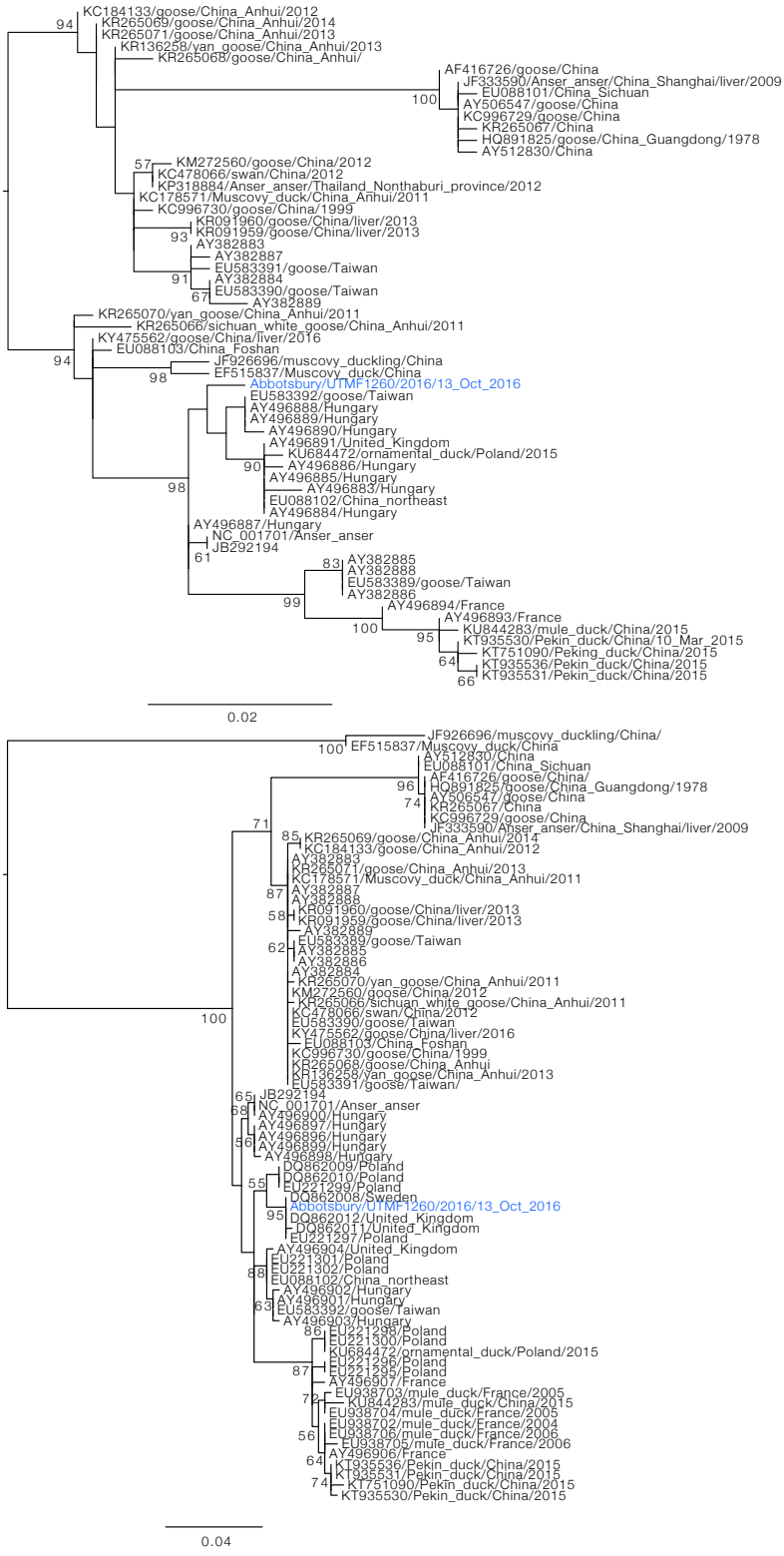

Figure 4.20: Midpoint rooted phylogenetic trees estimated from alignments of two different partial regions of goose parvovirus VP1, trimmed to regions that are commonly represented among partial sequences in GenBank. The trees were constructed using the K80+G substitution models (upper) and TPM1+G (lower). Bootstrap scores > 50 are shown. A sequence from Abbotsbury is in blue.

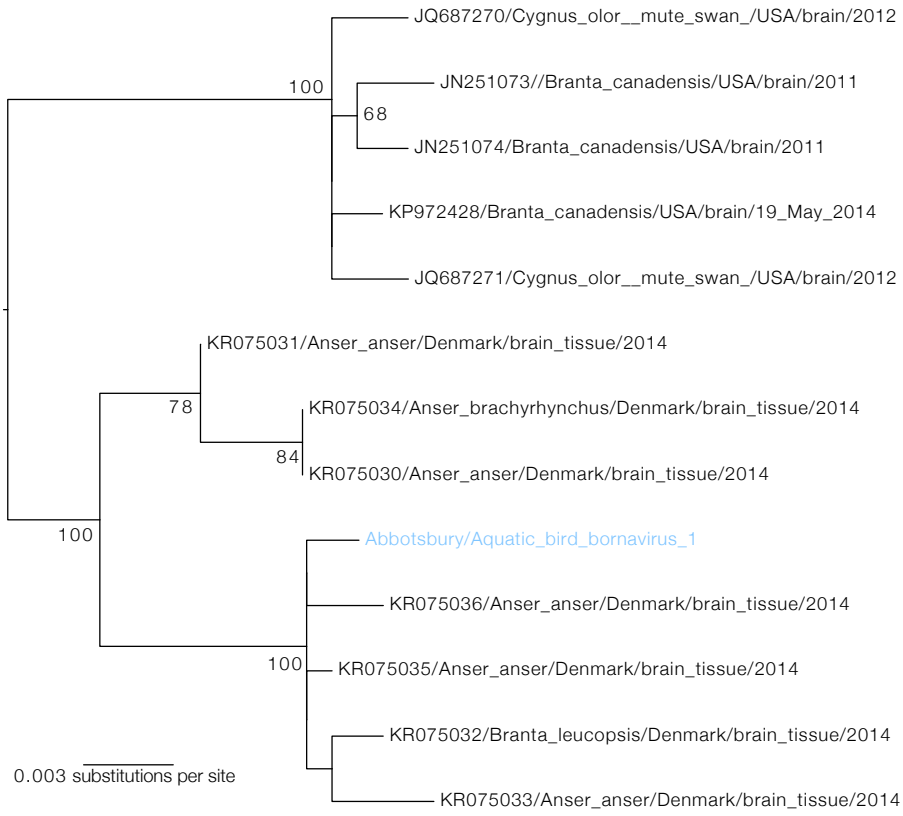

Figure 4.21: Midpoint rooted phylogenetic trees estimated from nucleotide alignments of aquatic bird bornavirus 1 trimmed to base 612-1177 of bornavirus KP972428. The tree was estimated using the HKY substitution model. Bootstrap scores > 50 are shown. A sequence from Abbotsbury is in blue.

4.4 Results: Impact of Host Factors on Viral Carriage and Abundance

4.4.1 Effect of age on viral carriage and abundance

The distribution of viruses in a population might be expected to correlate strongly with age due to the acquisition of immune responses to viruses encountered throughout the lifespan with increasing age. Furthermore, the introduction of immunologically naïve individuals to a population is expected to be strongly seasonal for many populations of birds due to the introduction of susceptible individuals into a transmission network through migration and breeding, which is highly seasonal in birds. In these results, I use metaviromic data to explore how age is associated with differences in viral prevalence and abundance, and explore the seasonal profiles of infection of two viral groups, gammacoronaviruses and astroviruses, which showed distinct intra-population diversity in the previous results section.

4.4.1.1 Comparison of Illumina read data and PCR data

In order to use metaviromic data to infer differences in viral carriage between samples from different birds, it was first necessary to demonstrate that viral reads detected within this metaviromic dataset are at least a relatively accurate reflection of the presence, absence or quanitity of a virus in a sample. To check this, Illumina read counts matching to particular viruses were compared to detection of that virus using qPCR and standard PCR followed by gel electrophoresis. Comparisons were made for two different viruses found in this study, astrovirus and megrivirus.

The qPCR cycle threshold (Ct) correlates log-linearly with Illumina read count per million reads for both viruses tested (Figure 4.22). Cycle threshold is therefore a significant predictor of the log of Illumina read numbers for both assays. Coefficients and p-values generated through application of linear models for samples with Ct less than or equal to 36 or more than 1 read per million are -0.50 ($p<0.005$) and -0.62

($p<0.05$) for megrivirus (see 4.2.8 for details of samples that were tested). Too few samples had astrovirus reads >1 read per million for robust statistical analysis, but the correlation between Ct and log of read numbers is reasonable, with coefficients of -0.56 ($p<0.005$). This suggests that read counts obtained through sequencing are an adequate reflection of quantity of virus for each assay, such that higher read counts mapping to a particular virus should broadly reflect higher abundance of that virus in the sample, and that epidemiological analysis based on measures of read abundance is valid for this dataset.

4.4.1.2 Total viral read count

Assuming that increased age in birds is associated with increased probability of having previously encountered an infection, and hence increased probaility of having acquired an immune response to that infection, it was hypothesised that samples from younger birds generate a higher proportion of reads that match to viruses that infect vertebrates than older birds. Samples from the youngest birds (hatch-year and year after hatch) were found to contain a higher proportion of reads that mapped to vertebrate viruses than samples from older birds (Figure 4.24). This is true when normalisation of reads is conducted against total reads (Wilcoxon rank sum test, $p < 5 \times 10^5$, median proportion is 7.9 times higher in younger birds) and against total viral reads ($p< 5 \times 10^4$, median proportion is 4.0 times higher in younger birds). This correlation remains significant even when hatch-year birds are excluded from the dataset, thus suggesting that it is not driven by differences between hatch-year birds and older birds, such as diet or geographic range.

Several datasets were explored to check that this result was robust. The above analysis of viral read abundance in birds of different age focuses on the relative abundance of reads that match to viruses that are believed to *exclusively* infect vertebrates, and excludes reads matching to viruses that might infect multiple hosts, such as vector-borne viruses. Furthermore, the above analysis did not make any attempt to remove specific reads that might represent false positives from the dataset prior

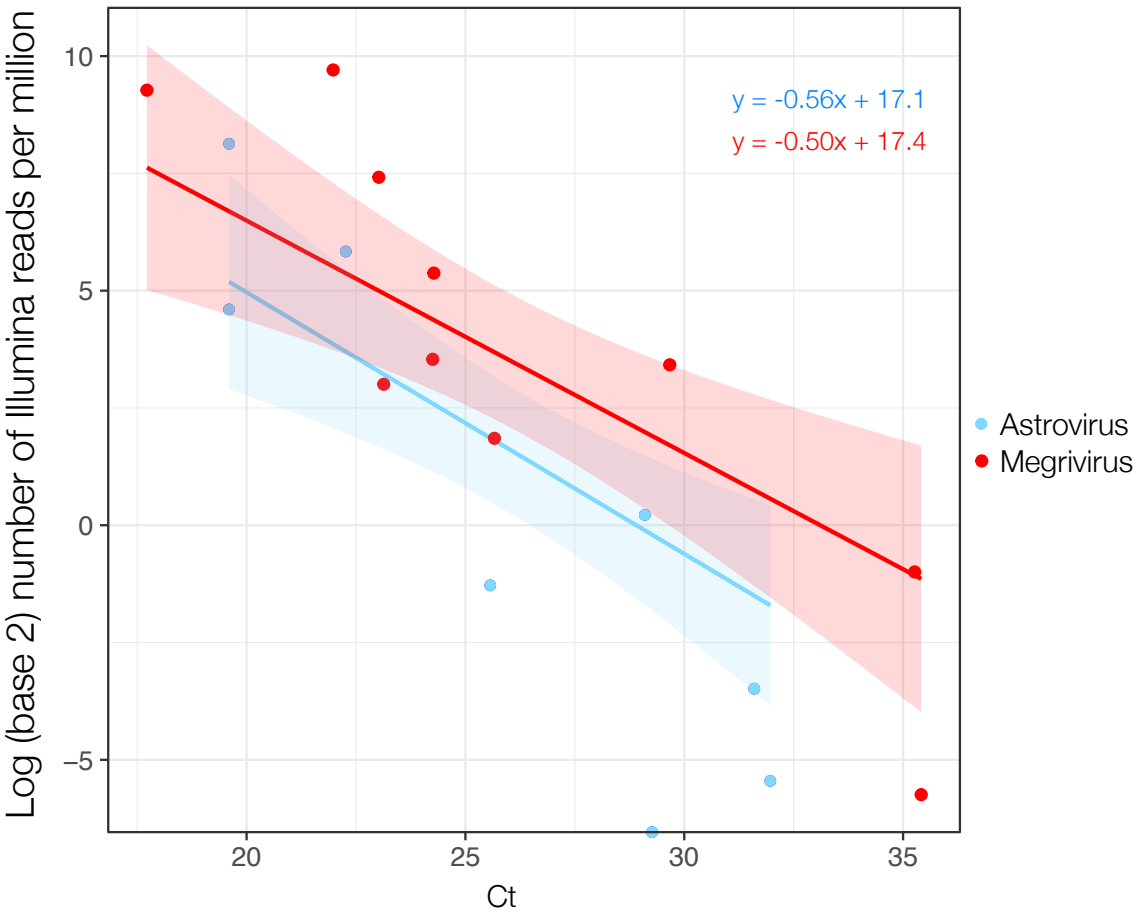

Figure 4.22: qPCR Cts compared to Illumina read counts for megrivirus assay (blue) and astrovirus assay (red), shown for all samples with Ct < 36. Equations derived from linear modelling are given in the top right corner. 95% confidence intervals for the linear model fit are shown by coloured ribbons.

to analysis, including reads matching to herpesvirus (herpesviruses co-opt a number host genes, so host nucleic acid can sometimes show a significant BLAST match to parts of the herpesvirus genome) and to retroviruses (that might represent the presence of endogenous retroviruses in the host genome). A significant effect of bird age on vertebrate read abundance remains even when, (a), herpesviruses and retroviruses are removed, (b), vector-borne viruses (defined as viruses with both a vertebrate and invertebrate host) are included with vertebrate viral reads, and, (c), when vertebrate viruses are defined as viruses that are believed to infect vertebrates but may also infect any other host taxonomic group. The effect of sequencing batch was also considered. As evident in Figure C.1, samples sequenced in the first of two Illumina sequencing runs that contributed to this dataset (broadly those samples from May 2016) had lower total read counts than those samples sequenced in later batched. However, excluding samples from May 2016 again makes no difference to the significance of association of increased vertebrate viral abundance in samples from younger birds.

To control for different numbers of samples taken from birds of each age-group in each month, Wilcoxon rank sum tests were conducted for the above analysis on 100 randomly downsampled datasets. In each dataset, the number of older age-group birds sampled in each month was downsampled to match the number of younger age-group birds sampled in that month. Younger birds had proportionally more viral reads mapping to vertebrate viruses in all datasets suggesting that the previous conclusion was robust, with the median proportion in younger birds ranging between 5.2 to 16.7 times more than in older birds for all datasets.

4.4.1.3 Age-specific variation in viral abundance and prevalence of different viral taxonomic groups

The overall observation of a slight increase in diversity of genera with age may be primarily driven by age-specific differences in infection with a small number of different taxa. Following primary viral infections that trigger the aquisition of an immune memory for that virus in the host, secondary infections with related viral strains that

share a portion of the previously observed antigenic diversity might be more rapidly prevented by the host immune system. Related sets of viruses that exhibit limited antigenic diversity, or single strains of viruses that are unable to evolve novel antigenic diversity, might be expercted to be primarily constrained to circulation within younger birds that have had little or no previous exposure to relevant viral antigens. On the other hand, viruses that can rapidly evolve novel antigenic diversity within a single host (such as HIV) might be expected to show little structuring with age, unless age is independently associated with behaviours that are key to the transmission route of that virus (such as breeding behaviour and sexual transmission).

To explore differences in viral read abundance and presence/absence by age, the R package metagenomeSeq [331] was used to assess the effect of avian age on virus detection at a number of different taxonomic levels, with sampling occasion as a covariate. Several families were detected in which bird age had a significant effect on abundance of that virus (all results in Table C.7). Reads from the viral families *Astroviridae*, *Picornaviridae* and *Adenoviridae* were 3.7 to 7.1 times more abundant in younger birds (hatch year and year-after-hatch) than in older birds (two or more years after hatching) (log2-fold changes; 1.89, 2.83 and 2.79 respectively, adjusted p-values $< 5\text{x}10^4$ for all genera) (Figure 4.25). Viral reads from the family *Bornaviridae* were 18.8 times more abundant in older birds (log2-fold change of younger compared to older; -4.23, adjusted p-value $< 5\text{x}10^{10}$). Considering only presence or absence of the viral family rather than read abundance, adenoviruses are significantly more likely to be present in younger birds (odds ratio 2.1, $p< 0.05$) and bornaviruses are absent from the younger birds (Figure 4.25).

Analyses of this kind are probably less robust at lower taxonomic levels (for example, genus) than at higher taxonomic levels (for example, family). This is because a single virus, particularly a novel virus, might have highest BLAST matches to a single viral *family* in all different regions of the genome, but might have highest BLAST matches to two (or more) different viral *species* (or genera) in different regions of the genome. However, in my analysis of the associations of bird age with

read abundance for different taxonomic groups, I analyse only those viral taxonomic groups that exclusively infect vertebrates, and therefore an entire viral family can be removed from testing because a single species or genus in that family is catalogued as infecting a non-vertebrate host. Testing different genera is therefore important here to detect associations that might not have been detected at the family level, although the aforementioned caveat should be kept in mind.

Patterns that were detected at the family level are supported by different reoccurence of those patterns across different genera of adenoviruses, picornaviruses and bornavirus. At the genus level, reads mapping to two parvovirus genera appear to be 4 times more abundant in younger birds. Analyses of presence or absence of the viral genera support the idea that bornaviruses and adenoviruses have different prevalences in birds of different age, and suggest that avian orthoreovirus, several picornaviruses and a parvovirus might also show differences in age-related prevalence. Full results are given in Tables C.8 and C.9.

Again, for the same reason detailed above, my results presented at the species level should be interpreted cautiously. Interestingly, several species for which several genomes were assembled in full or in part show differences in prevalence and/or abundance with age. At the species-level, avian orthoreovirus and the mute swan picornavirus reported here are notable for occurring more frequently in younger than older birds (respectively; odds ratios 3.45 and 5.19, adjusted p-values < 0.05 and $5\text{x}10^{6}$) (Figure 4.26), and the aquatic bird bornavirus 1 is again notable for being more prevalent and abundant in older birds.

4.4.1.4 Seasonality of viral infection for *Coronaviridae* and *Astroviridae*

Seasonality of infection in wild birds might derive from seasonal peaks associated with the hatching of the swans studied here, or seasonal effects of viral amplification in populations of wild birds that congregate at the site for overwintering and leave again to breeding grounds in spring.

There is substantial seasonal variation in the detected prevalence of coronaviruses

in my study population (Figures 4.27 and 4.29). As explored in Section 4.4.1.3, this variation is not explained simply by higher prevalences in the youngest birds (hatch-year 2016), which were sampled more frequently in later time points. (Table C.8).

There also appears to be a seasonal trend in prevalence of infection with an astrovirus in my study population (Figures 4.28 and 4.29). It is possible that this seasonal trend is driven by an increased viral prevalence in younger birds at later time points.

4.4.1.5 Stability of the viral repertoire in birds

Two different samples were sequenced for five birds in the population, with time between samples varying from one hour to six months. From these 5 resampled birds, there was no evidence of significant stability in infection status with different viral genera. Whilst this is not surprising for samples separated in time, because acute viral infections wouldn't be detected upon serial sampling, this is true even for samples collected within one hour of each other (2014 female bird on Figure 4.30), suggesting that the metaviromic data here is best interpreted at a population-level scale rather than at an individual level. Because two aliquots of the same sample were not independently processed, it is impossible to tell from this data whether the lack of consistency between samples from the same bird is due to genuine differences between different fecal samples from the same bird, or whether it reflects differences in the efficiency of processing different samples in the laboratory.

Figure 4.23: Heatmap of read abundance by viral species for viruses believed to exclusively infect vertebrates. Darker blues indicate greater read counts. Read counts are log(10) transformed and CSS normalized. The horizontal axis represents samples from 119 different birds, and is not sorted by age or sex. Note that viruses that appears < 6 times in the population, such as avian influenza virus, are not included in this heatmap, as these rarer viruses are removed from the analysis during CSS normalisation (see Section 4.2.10).

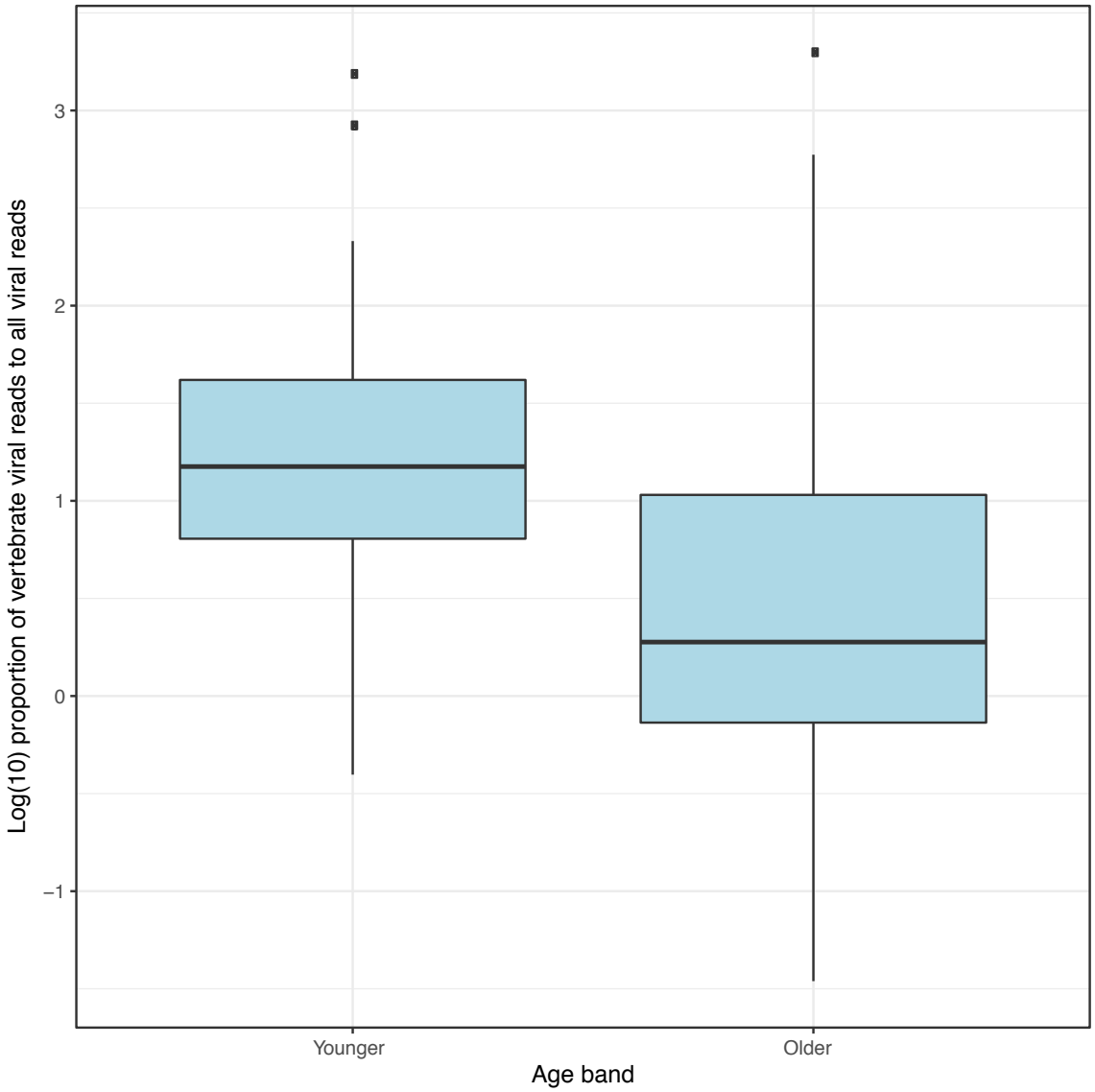

Figure 4.24: Log(10) proportion of viral reads to total reads for birds aged 2 or younger, and birds aged 3 or more.

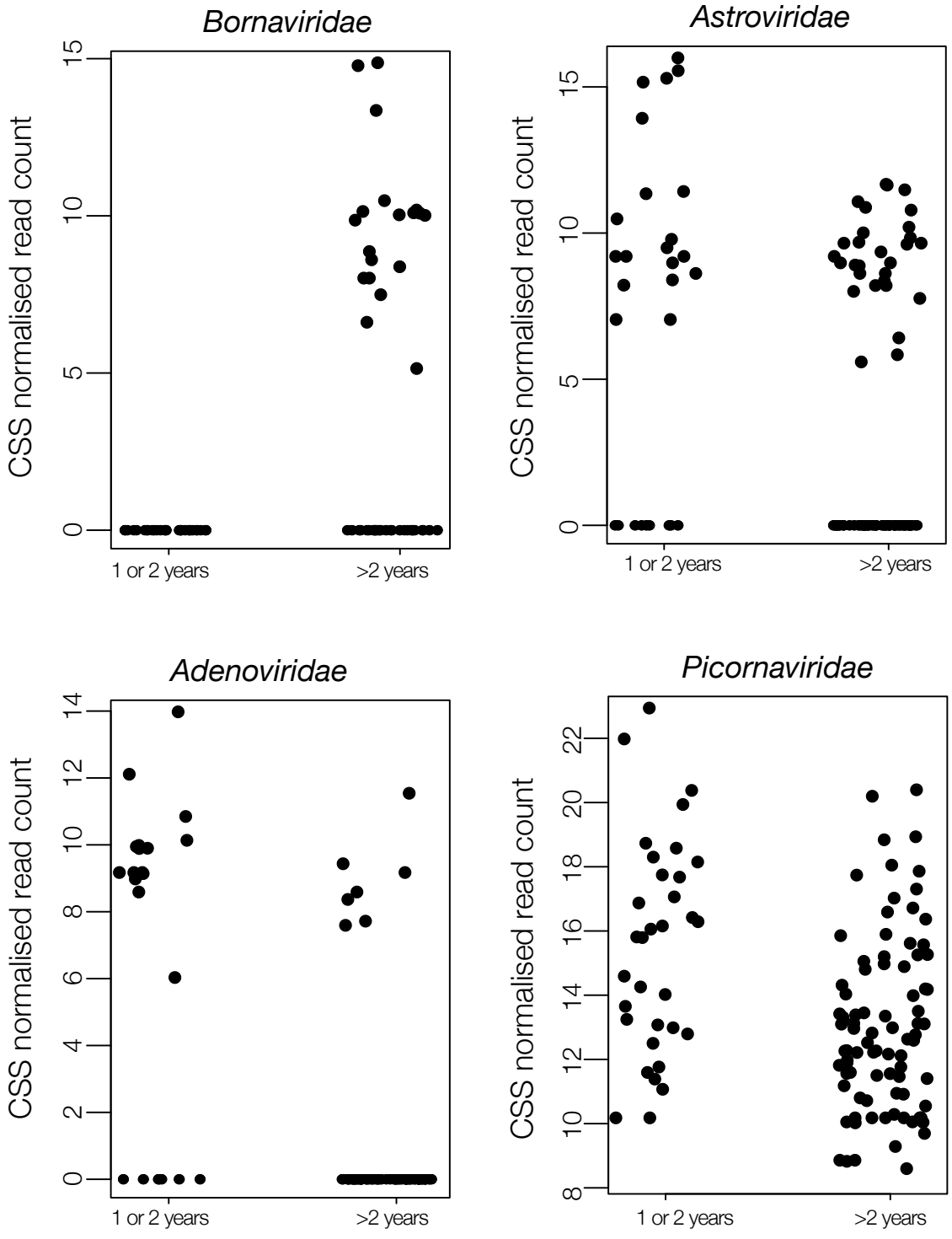

Figure 4.25: CSS normalised counts of viral families with differential read abundance in birds of different ages

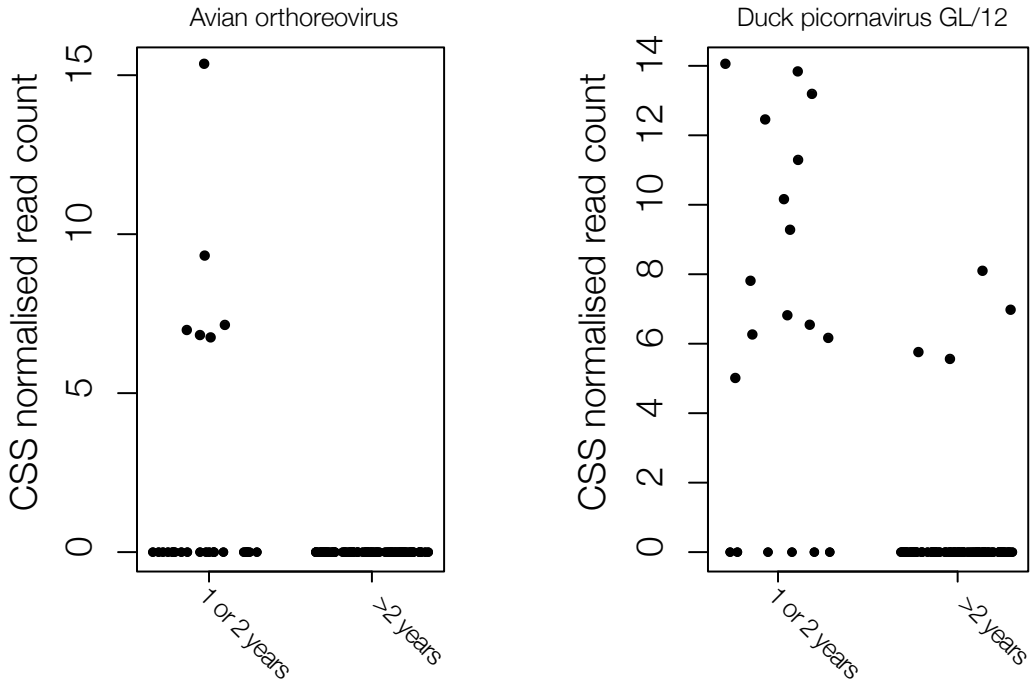

Figure 4.26: CSS normalised counts of viral species with differential read abundance in birds of different ages

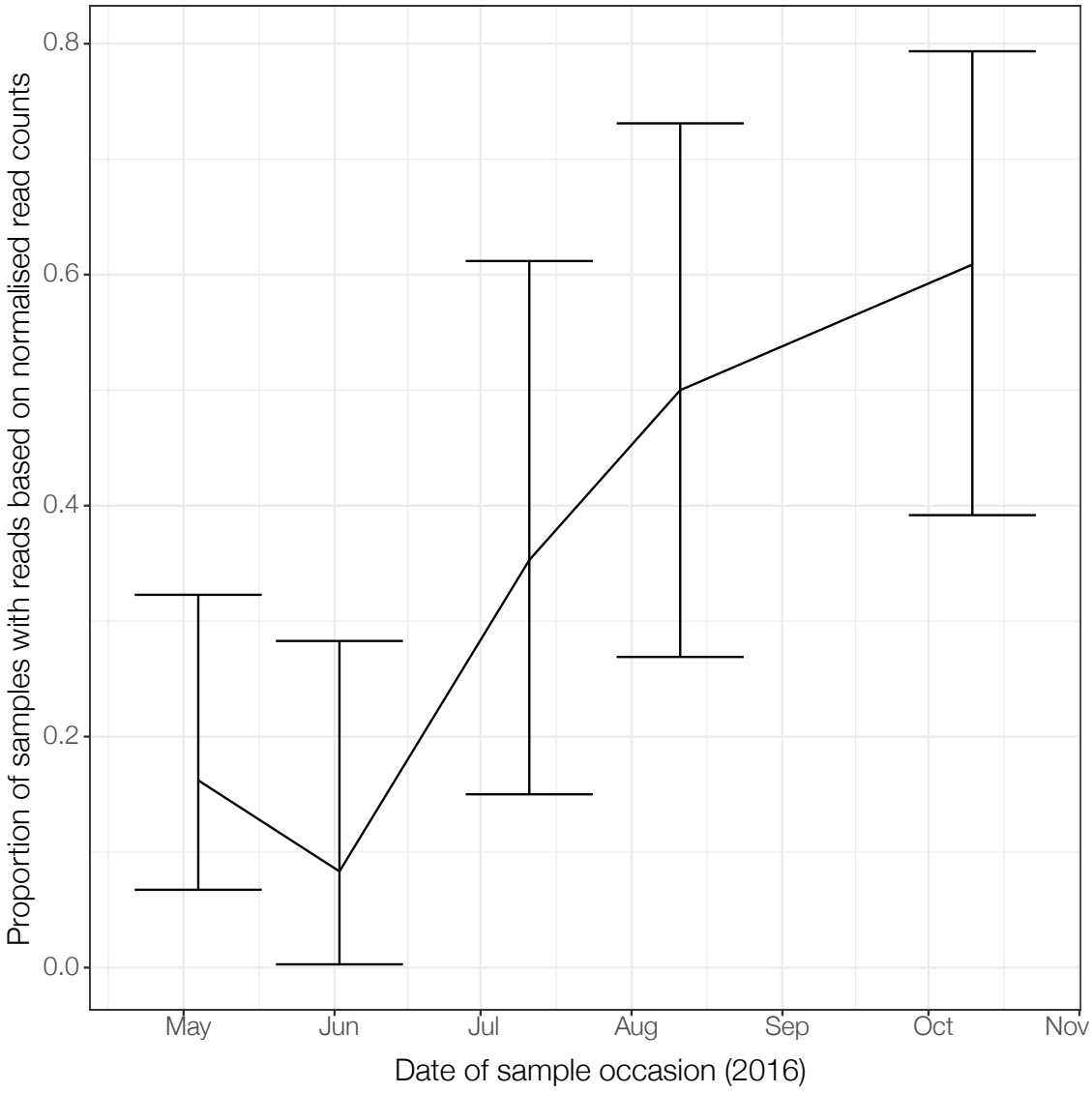

Figure 4.27: Proportion of samples with some evidence of coronavirus infection, based on positive normalised read counts with best BLASTP E-value <0.01 to any member of the family *Coronaviridae*. 95% confidence intervals were produced using the adjusted Wald method.

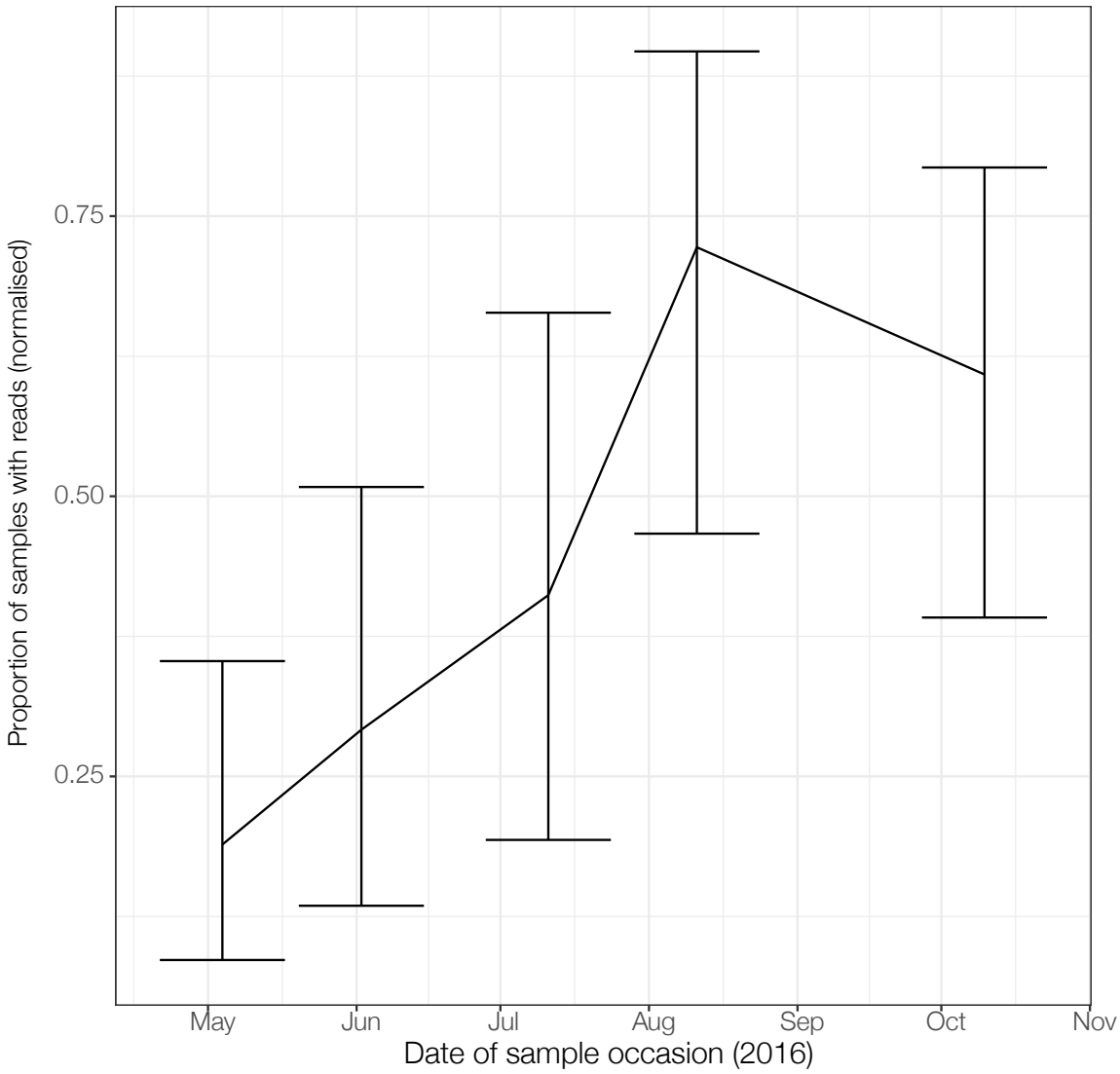

Figure 4.28: Proportion of samples with some evidence of astrovirus infection, based on positive normalised read counts with best BLASTP E-value <0.01 to any member of the family *Astroviridae*. 95% confidence intervals were produced using the adjusted Wald method.

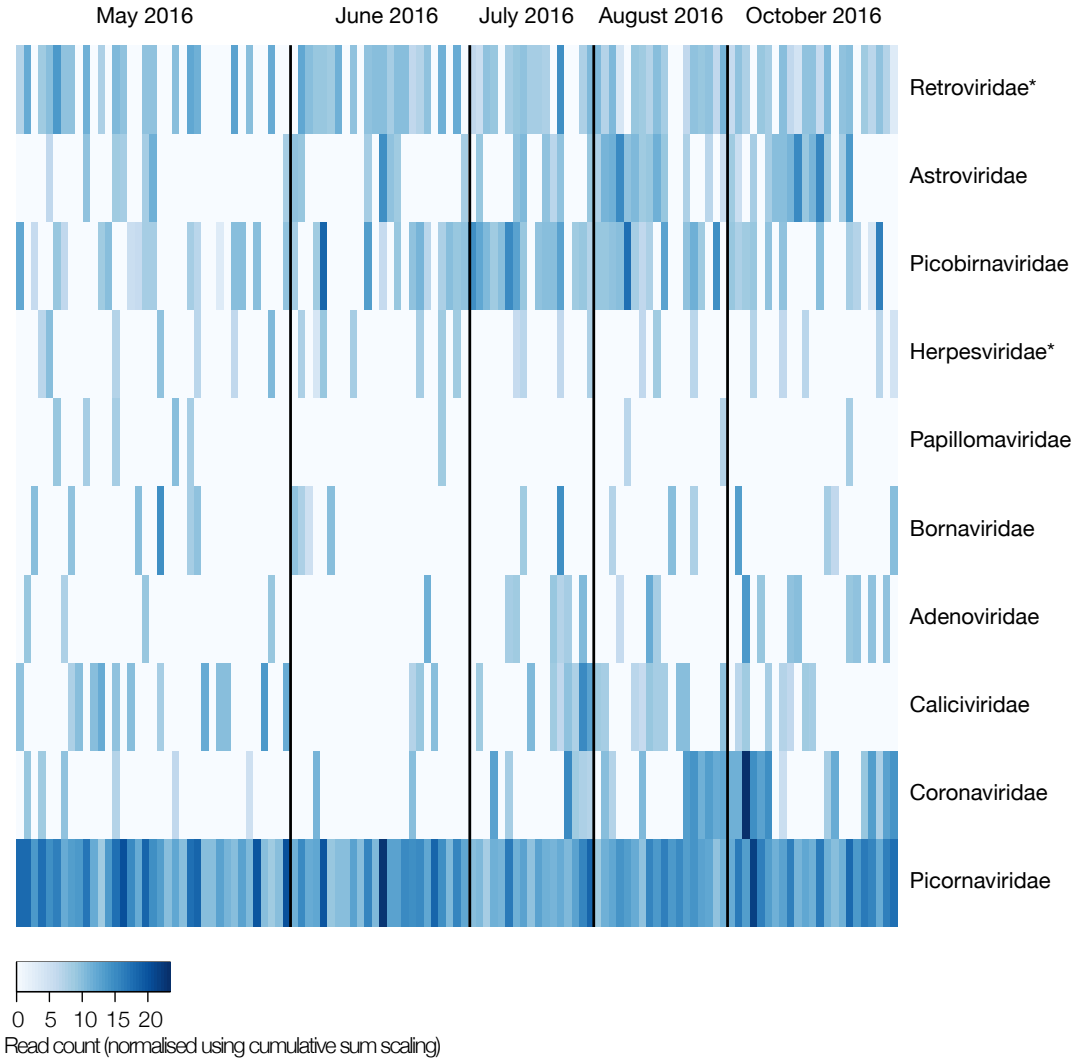

Figure 4.29: Heatmap showing normalised and log-transformed read counts with best BLASTP E-value <0.01 to any viral family believed to exclusively infect vertebrate hosts. Darker blues indicate higher read counts. Note that many of the reads mapping to retroviruses likely represent the presence of endogenous retroviruses in the (as yet-unsequenced) mute swan genome rather than free-virus. Similarly, all or almost all of the reads mapping to herpesviruses probably represent false positives due to similarities between many genes of herpes viruses and their hosts.

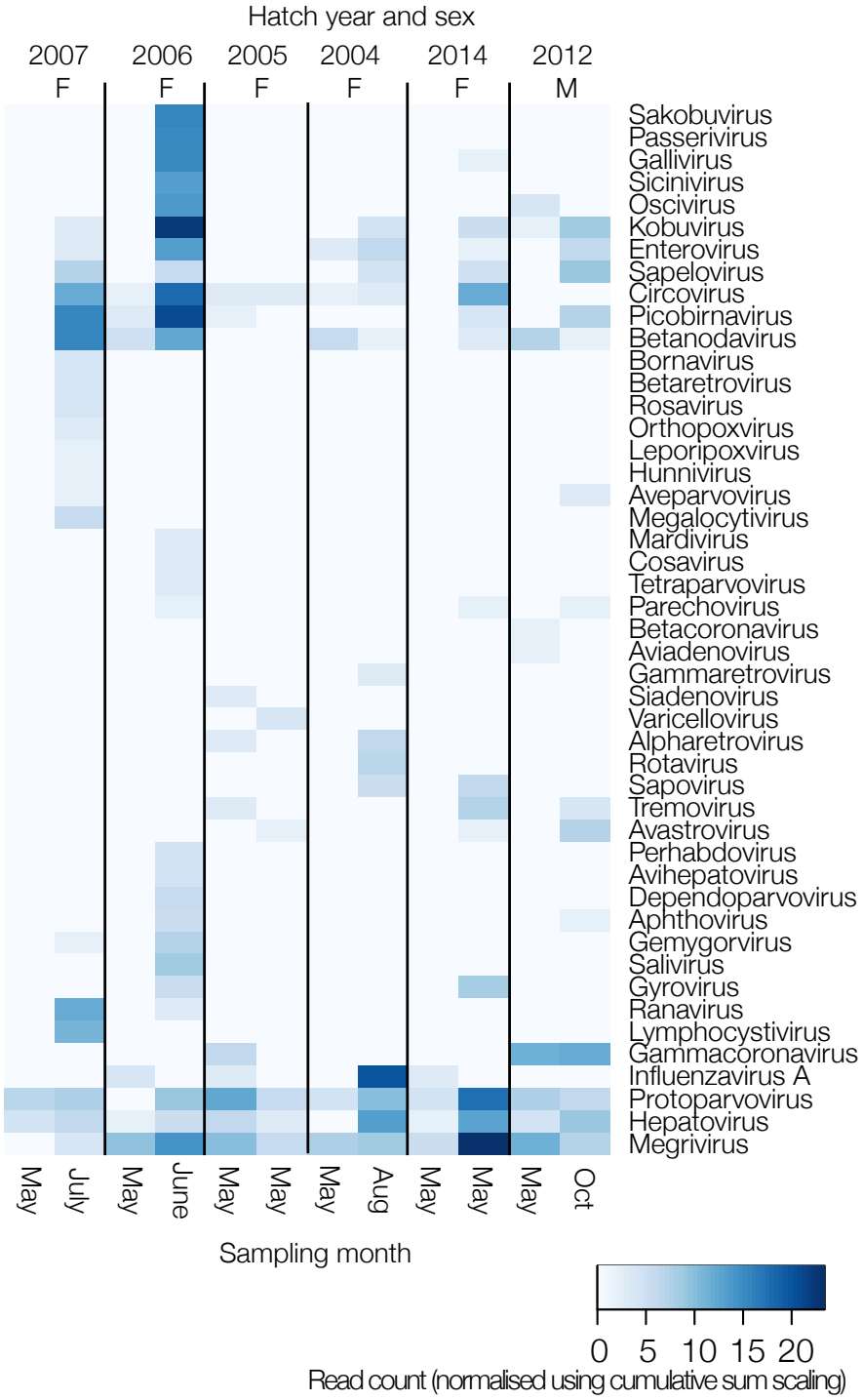

Figure 4.30: Heatmap showing normalised and log-transformed read counts with best BLASTP E-value <0.01 to any viral genus believed to exclusively infect vertebrate hosts. Darker blues indicate higher read counts. Sampling month, age and sex are given for each sample. Samples from the same bird are delineated by black lines.

4.5 Discussion

Here, I sequence the metavirome of 119 wild, mute swans (*Cygnus olor*), the largest metaviromic study of a wild avian population to date. To date, only nine viruses have been identified in swans, of which only six are believed to occur commonly [43, 97, 333, 159, 493, 157, 250, 293, 387, 185]. I report the whole genomes of a further nine new species of virus that likely infect swans. All species characterised here appear to be closely related to other viruses found in birds, with the exceptions of mute swan stool-associated virus (which probably infects a gut nematode based on the previously suggested host for its closest related viruses, porcine stool-associated virus [386]) and avian kobuvirus. Whilst I believe that most viral genomes described here are therefore likely to truly infect swans, I cannot rule out that these viruses are of dietary or environmental origin. For RNA viruses, base composition and dinucleotide frequencies have been used to estimate the likely host group (mammal, insect or plant) [386]. Similar approaches could be used to extend or corroborate the findings here, with the caveat that I could not entirely rule out environmental contamination from viruses shed by other vertebrates into lagoon water consumed during feeding.

Metaviromic studies have almost never been used to explore viral epidemiological dynamics in the wild. Immunological and epidemiological theory predicts that younger birds might be generally more susceptible to viruses that can exhibit constrained patterns of antigenic diversity, due to them lacking protective immunity that is subsequently acquired throughout the lifespan following single or repeated exposure. I demonstrate that younger birds tend to have statistically significantly higher numbers of viral reads that map to vertebrate viruses than older birds, and that this correlation may be driven by viruses from the families *Astroviridae*, *Picornaviridae* and *Adenoviridae*, all of which show greater abundance in young birds compared to older birds. I believe that this is the first time that a strong effect of age on the abundance of different viruses has been shown for any wildlife species using metaviromic techniques. This correlation remains true even when birds <1 year are removed from

the analysis, so is unlikely to be driven by dietary or behavioural differences between very young cygnets and older birds (including the inability of young cygnets to drink salty water, unlike adult swans).

The specific viral families implicated in contributing to this association have also been reported in similar studies of farmed pigs [375], including a higher abundance of picornavirus in 12 day old piglets compared to older piglets or sows and some evidence for a higher abundance of mamastrovirus (family: *Astroviridae*) sequences in piglets than in sows. Viral species from all three families have also been previously shown by specific PCR to be more prevalent in younger birds than older birds, which supports that the associations shown here are genuine. Specifically, age-related differences in prevalence have been demonstrated for astroviruses in turkeys, chickens and pigeons, [11, 103, 227, 503, 508]. Age difference in the prevalence of pigeon-adenovirus 2 have been noted for pigeons [438], and Group I adenoviruses are excreted in higher titres and for longer in juvenile birds than adults [285]. Four new species of *Picornaviridae* were assembled here, of which only one species (*Mute swan picornavirus*) showed significantly greater prevalence in younger birds than older birds. Although several of the most closely related viruses to the species *Mute swan picornavirus* are not well-characterized epidemiologically [33, 468], viruses from the related species *Duck hepatitis A virus* typically affect younger birds more than older birds [313]. No association with age was found for the novel sapelovirus, kobuvirus or megrivirus.

Viral prevalence was found to be different between younger and older birds for two taxonomic groups; the species *Avian orthoreovirus* and the familiy *Bornaviridae*. In concordance with previous findings in other avian species, *Avian orthoreovirus* was found to be more prevalent in younger birds [319, 496]. My data suggests that reads mapping to bornaviruses are significantly more abundant in older than younger birds. Few studies have investigated the effect of age on bornavirus infection, with only one study reporting no known effect of age on infection probability [97]. Assuming that bornavirus is primarily transmitted via the faecal-oral route [221] and therefore assuming that all birds would be equally exposed due to sharing of the same water

sources, it is difficult to understand why this virus might be more prevalent in older than younger birds and further studies of the diversity of the virus and prevalence of existing immune responses to the virus might be required to explore this finding in greater detail.

The finding that viral abundance is generally lower in young birds is consistent with previous suggestions that the youngest age group of birds is more susceptible to death as a result of highly pathogenic AIV infection [352], and that mortality within the population generally decreases with age up to to age of 3 years [283]. In combination, these findings hint that birds in the population accrue long-term immune responses following exposure to viruses that are protective against the (presumably) non-pathogenic viruses explored here in clinically healthy birds, and other pathogenic viruses such as AIV [352]. This conjecture is explored further in Chapters 5 and 6.

Phylogenetic analyses of the novel viruses observed here suggest that the viruses in swans tend to group with other bird viruses. To an extent, this probably represents bias during assembly, in which I directed greater effort at assembly of viral reads that appeared to be related to viruses with avian hosts than reads that appeared to be related to viruses from other hosts. However, the preferential grouping of these novel viruses with viruses that have been previously identified in waterfowl to the exlusion of viruses in galliform birds is interesting. For all the viruses detected here, sampling of birds from different orders and more intensive sampling of other species of bird at Abbotsbury would be useful to detect whether the patterns of preferential phylogenetic grouping of viruses from related hosts reflects genuine co-divergence of viruses with their hosts over long evolutionary time, or, alternatively, whether cross-species transmission events are not evident here simply because many new viral species have not yet been sought in most avian host species. A whole genome gammacoronavirus was sequenced here that appears to form a clade with other partial sequences of coronaviruses infecting geese and partial sequences of a further swan coronavirus found here. Whilst the phylogenetic intermixing of swan and geese coronaviruses reported here might be evidence for cross-species transmission

within waterfowl, the clustering of the swan gammacoronavirus found here with other coronaviruses infecting Anserinae birds, and not with viruses infecting Anatidae, lends support to suggestions that many avian coronaviruses have broadly co-evolved with their hosts [65].

Both coronaviruses and astroviruses in this dataset show clear evidence of seasonality, with a larger proportion of samples showing presence of both viral groups in autumn compared to early summer. Whilst prevalence varies by species, the proportion of samples tested here that show some evidence of coronavirus infection is at the higher end of what is typically reported for wild waterfowl [65, 103, 298, 485, 486]. An increase in prevalence in the autumn has been shown previously for both coronaviruses and astroviruses [485]. This finding, in combination with limited phylogenetic evidence for cross-species transmission of the novel coronavirus species described here (WatCV1), hints that the transmission of WatCV1 to swans in the population might be more likely when large numbers of birds of different species were visiting the Fleet Lagoon site, although the importance of migration for the local amplification and spread of WatCV1 should be further explored.

As in any shotgun metagenomic study, the viral diversity identified here is highly biased by published viral sequences and likely only represents a tiny fraction of the viruses truly infecting these animals. Bacteria and fungi respectively contain the universally conserved sequences regions 16S and ITS, so the presence of one of these conserved motifs in an assembled contig can be used to identify candidates for new bacterial and fungal species. However, viruses lack a universally conserved region, and must therefore be identified by genetic similarity to previously known viruses. Even in metagenomic studies in which viral particles are purified, 40-90% of sequence data from metagenomic studies are unalignable to any existing reference sequence [229]. Where sequences can be identified as viral, assembly into longer contigs can be made extremely difficult by the presence of quasispecies or several closely related species that can result in chimeras [402]. Here, I detected large numbers of viral reads with high E-value scores to hepatovirus, parvovirus and picobirnavirus that I was unable

to assemble reliably because of the very high diversity within the sequence data, and I expect that there are large numbers of viral reads that cannot be identified as viral because divergence to existing species is too great.

Appropriately normalising metagenomic data to accurately compare viremia across samples is challenging. Several existing normalisation approaches are commonly used, including total read count normalisation to reads per million per taxa. However, such normalisation can be heavily biased by a small number of high-count taxa [51, 102, 331]. In an attempt to avoid common normalisation biases, I used the R package metagenomeSeq to normalise the data presented here. This package attempts to avoid common normalisation biases through a normalisation approach in which raw counts are divided by the cumulative sum of counts up to a certain percentile, chosen based on the data [331]. A normalisation approach that is more intuitive, simpler to implement and more resistant to unexpected bias could include spiking each sample with a known quantity of an internal positive viral control prior to extraction of viral nucleic acids. This internal positive control could include a low quantity of a set of different viruses that were not expected to be present in the sample (for example, avian samples could be spiked with well-characterised strains of human viruses). Normalisation could subsequently be conducted for each sample against the number of sequencing reads found in that sample that are derived from the internal positive control. Including a range of different viruses in the internal positive control set could help to check that certain viral types were not being artifically removed during library preparation: for example, to confirm that both enveloped and non-enveloped viruses, or both RNA and DNA viruses, were maintained throughout the sequencing process at expected input proportions. Such an approach has been used previously in metagenomic papers and as a validation for effective viral extraction prior to RT-qPCR assays [150, 68, 378], but should be adopted more widely in metaviromic studies.

As well as identifying nine new species of virus that likely infect birds, I demonstrate that factors affecting viral prevalence, such as age, can be successfully explored

in wildlife populations using metagenomic sequencing data and broadly without specific PCRs- an approach that is still remarkably underutilised in studies of animal disease. The 50,000 fold decrease in the cost of sequencing driven by developments in high-throughput sequencing in the mid-2000s has dramatically expanded the realistic applications for sequencing in the field of wildlife disease. Recent developments in metagenomics using portable sequencing methods [44, 61, 153, 379] offer huge opportunities for circumventing logistical difficulties of transporting biological field-samples that may contain pathogens over long distances back to the laboratory, such that metagenomic studies of any wildlife population in the world is becoming increasingly within reach. Large-scale metagenomic studies that are supported by even the simplest of demographic data have the potential to revolutionise how we can study wildlife diseases, allowing joint exploration of population-level dynamics and viral diversity.

Chapter 5

Antibody responses to avian influenza viruses in wild birds broaden with age

5.1 Introduction

The ability to rapidly respond to viral infection via recognition of previously encountered antigens is a critical part of the host adaptive immune response. Viruses such as influenza A virus and HIV-1 effectively evade host immune systems because of their capacity to evolve novel antigenic variants. Existing immunity within a host population drives the emergence and spread of new antigenic strains by selecting for viruses with novel antigenic sites that avoid host immune recognition. Understanding the landscape of immunity within a population is consequently fundamental to understanding the epidemiological dynamics of antigenically-diverse pathogens.

Influenza A viral dynamics in human and avian hosts are fundamentally different. In humans, existing viruses are typically replaced by antigenically different strains belonging to the same virus subtype. Replacement by a strain of a different subtype ("antigenic shift") occurs only every 10-50 years [350]. The return of an antigenically similar strain also occurs comparatively rarely: antigenic similarities noted between

the H1N1 strains circulating in 1918 and 2009 were remarkable [501, 473] because strains separated by more than 2-8 years are usually considered to be antigenically distinct [400, 345, 226]. In bird populations, many genetically diverse subtypes can coexist (e.g., [240]). Modelling studies have investigated whether differences in influenza virus dynamics between species of different lifespans are in part a consequence of the faster rate at which immunologically naïve hosts are replenished in shorter-lived species compared to long-lived species [367, 484]. Such studies assume that immunological memory to a specific antigen can last for the lifetime of the host and thus the breadth of response to different antigenic strains increases as an individual ages. Whilst this seems probable for influenza A viruses in humans [160, 509, 395, 56], little is known about the acquisition and retention of immunity to influenza A viruses in wild birds, which form the primary reservoir of avian influenza A viruses (AIV) [471, 314]. Understanding the accumulation of adaptive immunity in wild birds is important for understanding the ecology of AIV viral prevalence and transmission dynamics, including predicting in which species novel strains are most likely to emerge.

Much of our knowledge about the acquisition of adaptive immunity to AIV in birds has been derived from experimental inoculation of immunologically naïve, domesticated waterfowl (e.g., [199, 75, 131, 457]). It has been particularly challenging to quantify changes in acquired immunity with age, because the exact ages of wild-caught birds are usually unknown or because lifespans of many domestic bird species that harbour AIV are too short for meaningful patterns to be discerned. An experimental study of wild-caught gulls reared in captivity showed that exposure to AIV results in increased protection against that strain for at least 1 year [457], suggesting that continual exposure to AIVs could result in better protection with increased age. However, whilst experimental studies provide important insights under controlled conditions they may not adequately describe long-term acquisition of immunity in wild populations that are exposed to a diverse range of AIV subtypes over prolonged periods and to other sources of mortality.

Previous studies have indicated that acquired adaptive immunity may shape the

observed incidence of AIV in wild birds of different ages. For example, juvenile birds from many Anseriforme species (including ducks, geese and swans) are less likely to carry antibodies against AIV than adult birds [182, 181, 179, 222, 487, 117, 352, 236]. Juveniles also have higher viral prevalence than adult birds (e.g., [471, 182, 224, 452]), suggesting that immunologically-naïve juveniles are more susceptible to infection or shed virus for longer than older birds. At the population level, seasonal peaks in viral prevalence have been observed following hatching, and attributed to the immunological naïvety of the unfledged birds [452]. Immunologically naïve birds that are challenged by an AIV have similar shedding patterns and probability of seroconversion regardless of bird age [457]. It is therefore likely that the age-related patterns of seroprevalence and viral prevalence observed in the wild result from birds gaining immunity to AIV with continual exposure throughout the lifespan, rather than changes in immune function specifically resulting from ageing.

Studies of adaptive immunity in wild bird populations may generate data with limited resolution, for two reasons. First, with few exceptions [117, 352], information on the exact age of the wild birds being studied is unavailable and thus age must be reduced to a binary variable (adult versus juvenile). Second, immune responses to AIV in wild birds are often characterised using the presence or absence of antibodies to the AIV nucleoprotein (NP), rather than to a specific haemagglutinin (HA) or neuraminidase (NA) type. The high sequence conservation of NP among AIV strains means that a single NP-antibody test can easily identify whether a bird has been infected with AIV, but cannot distinguish among acquired immune responses to specific subtypes.

It is not known whether studies that use NP-antibody tests are failing to detect subtype-specific variation in immunity among birds of different ages. Haemagglutinin (HA) is the most abundant of the two AIV surface proteins, and antibodies raised against HA are central to the adaptive immune defence [422]. These antibodies neutralise the virus by binding to the HA protein, preventing virus attachment and cell entry. Experimental and natural infections of waterfowl have shown that infection

with a given HA type can induce protective immune responses to that same HA type (homosubtypic immunity) and may also cause weaker immunity to different HA types [199, 75, 132, 76, 201, 443, 57] (heterosubtypic immunity). Haemagglutination inhibition (HI) titre is often correlated with the strength of protection against AIV when vaccine and challenge antigens are similar [430, 232, 429, 431]. If birds form a long-term immunological memory of encountered HA antigens, then birds exposed to many strains may be protected against a wider range of viruses than birds that have only ever been exposed to a few strains (e.g. younger birds). Whether the breadth of responses to different AIV HA types increases with age has not been investigated [367, 483].

The Abbotsbury Swannery in Dorset, UK, harbours a semi-habituated population of wild mute swans (*Cygnus olor*), a long-lived species that has been subject to long-term study at the site [339, 338]. All swans born into the colony are marked after hatching, so birth dates are known for most individuals. Birds are vent-sexed at hatching and by breeding behaviour. Birds that immigrate into the population are aged by plumage where possible and sexed during regular ringing events. The population is thus suitable for testing hypotheses about age-related acquisition of immunity. In winter 2007/2008, an outbreak of highly pathogenic avian influenza (HPAI) H5 occurred in the population [93]. During the outbreak, there appeared to be a predominance of young birds among the H5N1-positive dead birds [352]. Although this trend was not statistically robust due to small sample sizes, evidence from experimental challenge studies indicates that birds with prior exposure to AIV survive infection with pathogenic AIV and shed virus for shorter periods than immunologically naïve birds [76, 201, 24]. Since younger, immunologically naïve birds were more affected in the 2008 H5N1 outbreak, I hypothesise that antibody responses to related LPAI viruses were lower in younger than older birds prior to that event.

Here, I report the first investigation of the age structure of immunity to different AIV subtypes in a wild bird population for which bird ages are known. Blood samples from this study population were obtained in 2007 and 2008, either side of the H5N1

HPAI outbreak. To explore the pattern of immunity to AIV, I undertook statistical analysis on data from HI assays for five HA antigens (belonging to three HA types) to determine whether antibody responses to specific HA types and/or the breadth of antibody responses are associated with age in the population. I consider whether a lack of pre-existing immunity to different LPAI subtypes could have resulted in the raised mortality among young birds during the H5N1 outbreak in the winter of 2007/2008.

5.2 Materials and Methods

5.2.1 Study population

The Abbotsbury Swannery in Dorset, UK (50.6537°N, 2.6028°W), harbours a semi-habituated population of wild mute swans (*Cygnus olor*) that has been subject to long-term study [339, 338]. Following hatching, a small proportion of breeding pairs and cygnets are placed for 4 months in pens. All other swans have freedom of movement and mix naturally with other species. Survival rates of first-year and adult birds are similar to those across the UK [337, 284], but survival rates of birds in their second and third years are slightly higher at Abbotsbury [339]. Overall longevity is similar to elsewhere in the UK [284]. The population is unusually large and varies seasonally (in July 2007 population size was approximately 900-1000). Approximately 150 breeding pairs nested at Abbotsbury in 2007 and 2008. Supplementary food is provided from spring to autumn. A fence around the site reduces terrestrial predation.

5.2.2 Population sampling

Blood samples were collected from swans at Abbotsbury during July 2007 and August 2008 (UK Home Office licenses PPL 80/1944 and PPL 30/2572, respectively), as described previously [352]. Where a bird was sampled in both years (eleven birds total), a single sample was excluded randomly so that birds were not repeated across

years in the dataset, leaving 63 samples from 2007 and 95 samples from 2008. Birds with repeated sampling were considered in a separate analysis. Age and/or sex was unknown for a small proportion of birds (3.2% of birds had unknown age only, 8.9% had unknown sex only, and 1.9% had both unknown age and sex) and these birds were removed from analyses where appropriate. Sampling was intentionally slightly biased towards older birds, and the age-structure sampled across both years was similar (Figure D.1). Seropositivity prevalences reported here should therefore not be interpreted as if the sample was random.

5.2.3 Serological assays

HA antigens from 5 different viruses were chosen for HI assays. Antigen choice was motivated by an interest in protection against mortality from HPAI H5N1 infection following previous exposure with LPAI strains. Because protective immunity acts more effectively within than between HA groups, viruses from HA Clade 1 (the clade that includes H5 viruses) were chosen. Within this clade, HA types were chosen that had been experimentally associated with reducing the severity of subsequent HPAI H5N1 infections (H5 and H9 [76, 57, 210, 329]) or were antigenically close to H5 viruses (H6). H6 viruses were among the most common viruses in European waterfowl around the sampling period of 2007/2008, particularly in geese and swans [224, 296, 21]. LPAI H5 viruses were very common, comprising 4.9 8.4% of LPAI infections in European wild birds [296, 71]. H9 viruses were rarer than H5 or H6, but were not uncommon during this period in several species and locations [236, 296, 21] and were interesting because of their association with protection against H5N1 HPAI infection.

The Abbotsbury swans had not been screened for AIV prior to 2007/2008, so it was unknown which AIVs, other than HPAI H5N1, had circulated in the population. Antigens were therefore selected that were broadly cross-reactive with several modern strains of the same HA type (identified during years of serological testing at the

UK Animal and Plant Health Agency). The chosen antigens were H5N1 A/mute swan/England/26-20/2008 (the HPAI outbreak strain at Abbotsbury), H5N2 A/ostrich/Denmark/72420/1996, H6N8 A/ostrich/South Africa/946/1998, H9N2 A/turkey/USA Wisconsin/1966, and H9N9 A/knot/England/SV497/2002. Many of these antigens are commonly used in European Influenza Reference Laboratories, and the assays are correspondingly well-established.

HI assays were conducted to determine the presence of antibodies for specific HA types according to standard methods [50]. Sera were pre-treated with receptor destroying enzyme (RDE) (1 volume of serum to 5 volumes of RDE) before heat-inactivation at 56C for 30 minutes. Inactivated sera were treated with chicken red blood cells (RBCs). Serial two-fold dilutions of the sera were prepared from a starting dilution of 1:10. HI assays were performed using a 1% suspension of chicken RBCs in PBS. Due to low volume of some samples, all 5 assays were performed only on 61 and 88 of the 2007 and 2008 samples, respectively. Four HA units of antigen per 25ul were added to the serial dilutions. Haemagglutination titres of $\geq$20 were considered seropositive, based on the OIE recommended cut-off of 1/16 [312].

5.2.4 Statistical analyses

Cumulative link models were used to assess the effects of age on breadth of the antibody response. "Breadth" was calculated as the total number of subtypes a bird was seropositive for. To ensure that the specific choice of HA antigen included had no significant effect on the estimated relationship between breadth and age, I calculated breadth of responses based on the presence or absence of a serological response to (a) H5N2, H6N8, H9N2 (henceforth referred to as dataset A), (b) to either strain belonging to the same HA type (dataset B), (c) to both strains belonging to the same HA type (dataset C), (d) H5N1, H6N8 and H9N9 (dataset D), and (e) H5N2, H6N8 and a response to both H9N2 and H9N9 (dataset E) (Table 5.1). Age, sex, and sample year were included as fixed factors in the statistical model. Likelihood ratio tests of

cumulative link models were used to compare models with and without interaction terms between all pairs of factors. Interaction terms were included if a model with the interaction term had a probability of <0.05 of being a better explanatory model than one without the interaction, based on likelihood ratio tests. Two-sided tests were used to generate all p-values.

Table 5.1: Datasets used to calculate breadth of response

Dataset name	H5N1	H5N2	H6N8	H9N2	H9N9
Dataset A		✓	✓	✓	
Dataset B	Either H5 type		✓	Either H9 type	
Dataset B	Both H5 types		✓	Both H9 types	
Dataset D	✓		✓		✓
Dataset E		✓	✓	✓	✓

Generalised linear models with a gamma distributed response were used to assess the correlation between breadth of response in a sample and the result of the AIV NP-ELISA test on that sample (data for the latter obtained from [352]). The effect of age, sex and sample year on the NP-ELISA result was also tested.

Generalised linear models with a binomial response and logit link were used to determine whether the probability of being seropositive for each individual subtype increased with age. Age, sex and sample year were included as fixed factors, and interaction terms were tested using chi-squared tests of the difference in deviance between the two models. An interaction term was included if the model with the interaction term had a probability < 0.05 of being a better explanatory model than the model without the interaction.

Samples that tested seropositive ($\geq$20) for each antigen were separated into two groups; those with titres <40, and $\geq$40. generalised linear models with a binomial

response and logit link were used to determine whether the magnitude of positive titres varied with age. Sex and sample year were included as additive predictors. Binning titres into two categories was considered more appropriate than using raw titres in ordinal regression models, as there were few titres of $\geq$80 for all tested antigens except H9N9.

To analyse whether immune responses were stable across consecutive years, analyses were conducted on eleven birds sampled in both 2007 and 2008. Wilcoxon signed-rank tests were conducted on titres from each assay, and separately on the overall breadth of response, to determine whether there was an increase or decrease by year. McNemar tests were used to determine whether seropositivity in the population for each HI assay increased or decreased over the year. Ordinal logistic regression was used to test whether an increasing breadth of response between years was associated with a decrease in NP-ELISA result.

5.3 Results

5.3.1 Seroprevalence varies by subtype

HI assays were conducted for 5 AIV antigens belonging to 3 HA types. Seroprevalence in the (non-random) sample of birds varied according to subtype, with H5N2 being most seroprevalent in both sample years, H5N1 least seroprevalent in 2007 and H6N8 least seroprevalent in 2008 (Table 5.2, Figure 5.1 panel A).

Table 5.2: Population seroprevalence (number birds positive) for each test antigen.

Sampling date (birds sampled)	H5N1	H5N2	H6N8	H9N2	H9N9
July 2007 (61)	6.56 (n=4)	39.3 (n=24)	8.20 (n=5)	21.3 (n=13)	21.3 (n=13)
Aug 2008 (88)	15.9 (n=14)	35.2 (n=31)	6.82 (n=6)	23.9 (n=21)	29.5 (n=26)

5.3.2 Subtype breadth increases with age

As birds age they are increasingly more likely to respond to a wider variety of HA types (Figure 5.1 panel B). Due to the possibility of correlated responses to multiple H9 or H5 types, I explored a variety of dataset combinations (see Section 5.2.4 and (Table 5.1)). Likelihood ratio tests of cumulative link models suggested that inclusion of interaction terms did not significantly improve the model, so only the main effects of sex, age and sample year were included. In all of the tested datasets A-E except dataset C (see Section 5.2.4 and Table 5.1) older birds were significantly more likely to have a broader response to different AIV HA types than younger birds (Table D.1 for p-values and coefficients of the cumulative link model). When breadth is calculated as responses to H9N2, H5N2 and H6N8 subtypes (dataset A), the probability that a bird exhibits responses to X or more HA types increases by a factor of 1.1 with every extra year of age ($p < 0.05$; see Figure 5.2). Male birds were more likely to have narrower responses to AIV than female birds in all datasets, but this was only significant for a model where breadth was calculated from either dataset A or E.

The breadth of response observed in a sample was correlated with the raw result of the NP-ELISA for that sample, such that a larger breadth of response was associated

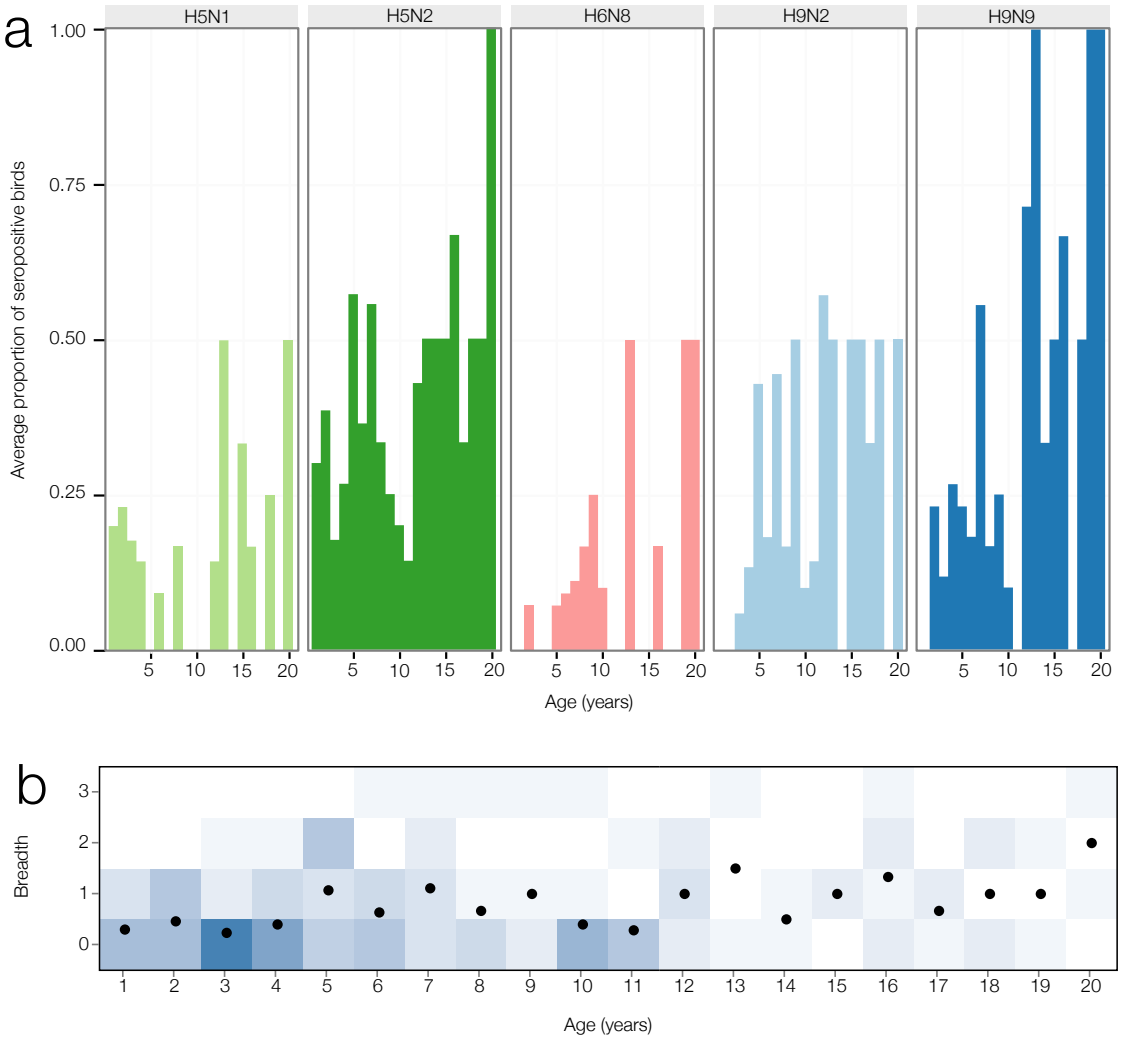

Figure 5.1: Immune responses of birds in the population. A) Proportion of birds of different ages that respond to each strain. B) Breadth of immune response (measured as the number of serotypes responded to from dataset A, see Section 5.2.4) for birds of different ages and Table 5.1 for description of datasets. The darker blue the square, the more birds of each age respond to each number of subtypes. The mean number of subtypes responded to at each age group is plotted as a black dot.

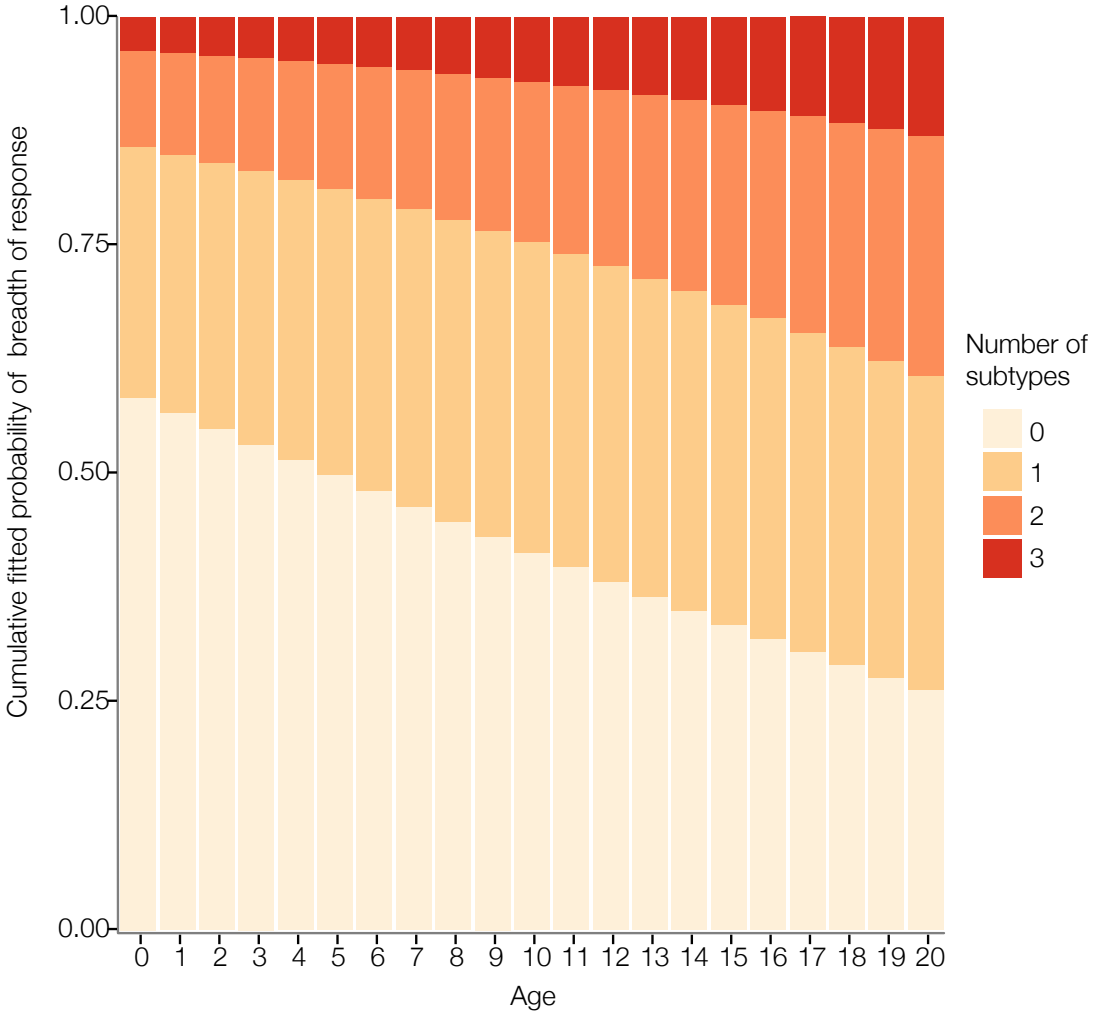

Figure 5.2: Cumulative fitted probabilities of effect of age on number of strains to which a swan exhibits a response (breadth of response) as determined by cumulative link models. Sex and sample year were included in the model, but do not significantly affect the probability of having a response, so for simplicity, data are shown for female birds in 2007. Breadth of response details the number of subtypes birds respond to (Dataset A, including H9N2, H5N2 and H6N8).

with lower ELISA values (note- ELISA values < 0.5 are considered to be positive for serological responses to AIV NP, and values ≥ 0.5 are considered to be negative). This result was robust to the way in which breadth was calculated (Table D.4). Age also correlated with NP-ELISA score, which is unsurprising as age is collinear with breadth of response in the data (Table D.5).

5.3.3 Effect of age, sex and sample year on each individual subtype

Generalised linear models were used to determine whether age was associated with responses to each individual subtype. Age, sex and sample year were included as fixed factors. For H5N1, H6N8 and H9N9 and H9N2, no interaction terms significantly increased the predictive power of the model. An interaction effect between sample year and age was included for H5N2 (D.2).

When individual subtypes were considered, age was found to be a significant predictor of response to H9N2, H9N9 and H5N2 subtypes ($p < 0.05$, $p < 0.0005$ and $p < 0.05$, respectively; Figure 5.3) (Table D.2). For every extra year of age, the odds of responding to H9N2 increased by 1.10 times, for H9N9, 1.14 times, and for H5N2, 1.17 times. Males were significantly less likely to be seropositive for H9N2 (odds 0.37, $p<0.05$). Whilst males were also slightly less likely to be seropositive for H9N9, H6N8 and H5N2, these effects were not significant.

Notably, antibody responses to H5N1 were significantly higher in 2008 than in 2007 (Table D.2). The sampling year significantly affected the relationship between age and response to subtype H5N2. In 2007, older birds were more likely to respond to H5N2 than younger birds, whereas in 2008 there was no significant relationship with age (Figure 5.4). When an interaction term between sampling year and age was fitted to the H5N1 data, I still did not observe a significant interaction effect. Age and sex had no significant relationship on the probability of response to H5N1. Sample year did not affect the probability of response to H9N2 or H9N9, and age,

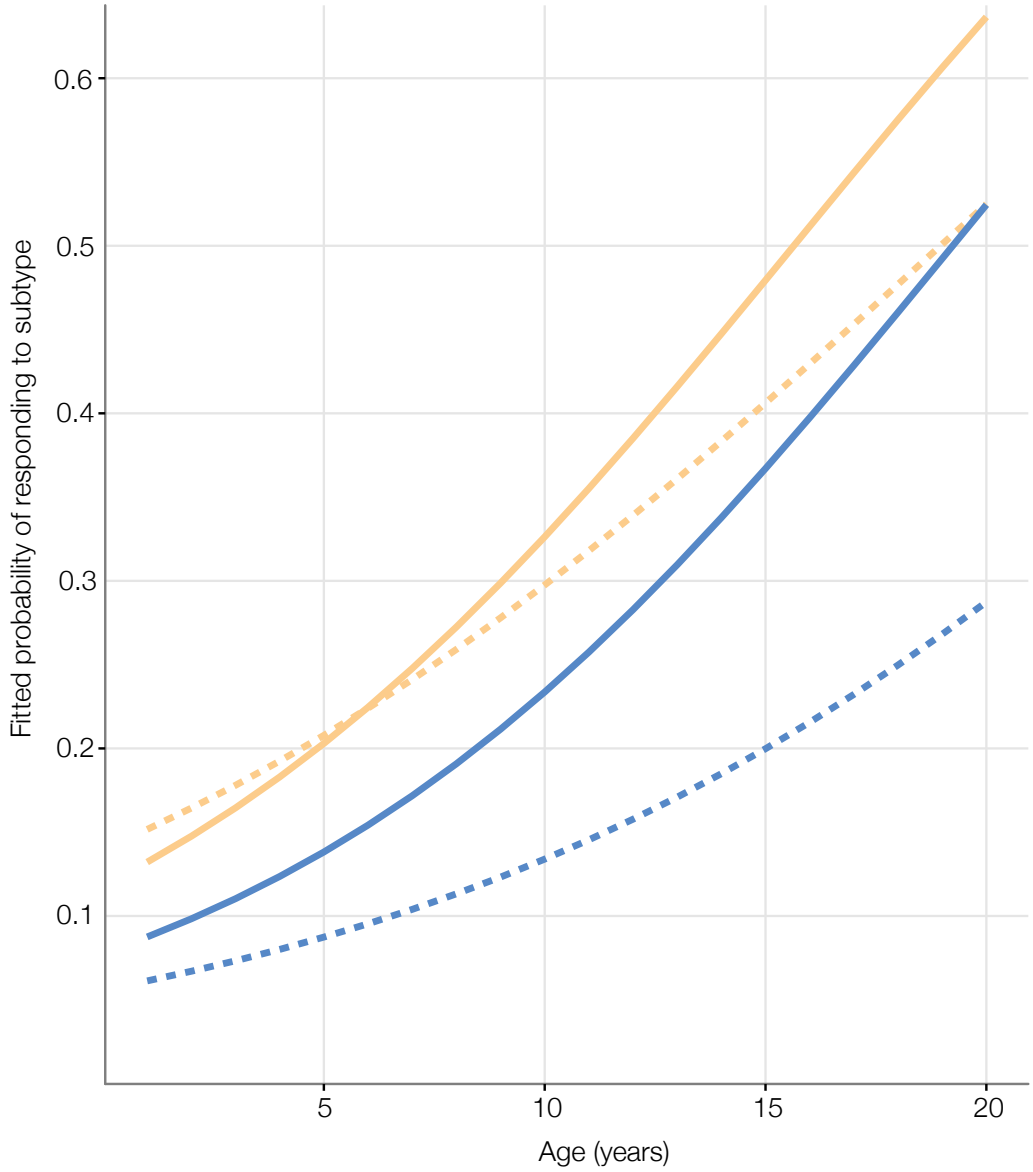

Figure 5.3: Fitted probability of effect of age and sex (females; pale orange, males; dark blue) on response to H9N9 (solid line) and H9N2 (dashed line) as determined by GLM with logit link. Sample year was included in the model, but does not significantly affect the probability of having a response, so for simplicity, data are shown for 2007.

sample year, and sex had no significant effect on the probability of response to H6N8. For seropositive samples from all tested antigens, there was no effect of age, sex or sample year on the magnitude of the titre (D.2).

5.3.4 Longitudinal sampling in birds from both years

Data were available from both 2007 and 2008 for eleven birds. Responses to each HA antigen were not stable between years (Figure D.2), but McNemar tests did not detect a significant increase or decrease in the number of seropositive birds for each assay by year ($p > 0.05$ for all tests). Wilcoxon signed-rank tests indicated that raw titres for every HA type, and also for breadth of response, showed no significant increase or decrease by year ($p > 0.05$ for all tests). The failure to detect a population-level effect of increasing breadth of response with increasing age between 2007 and 2008 was likely due to (i) the small sample size of 11 birds and (ii) the small estimated effect of age on breadth of response, which only becomes apparent across multiple years. A decrease in NP ELISA values between years was associated with an increase in response breadth, but this was not significant (datasets A-E, Tables D.3 and D.2).

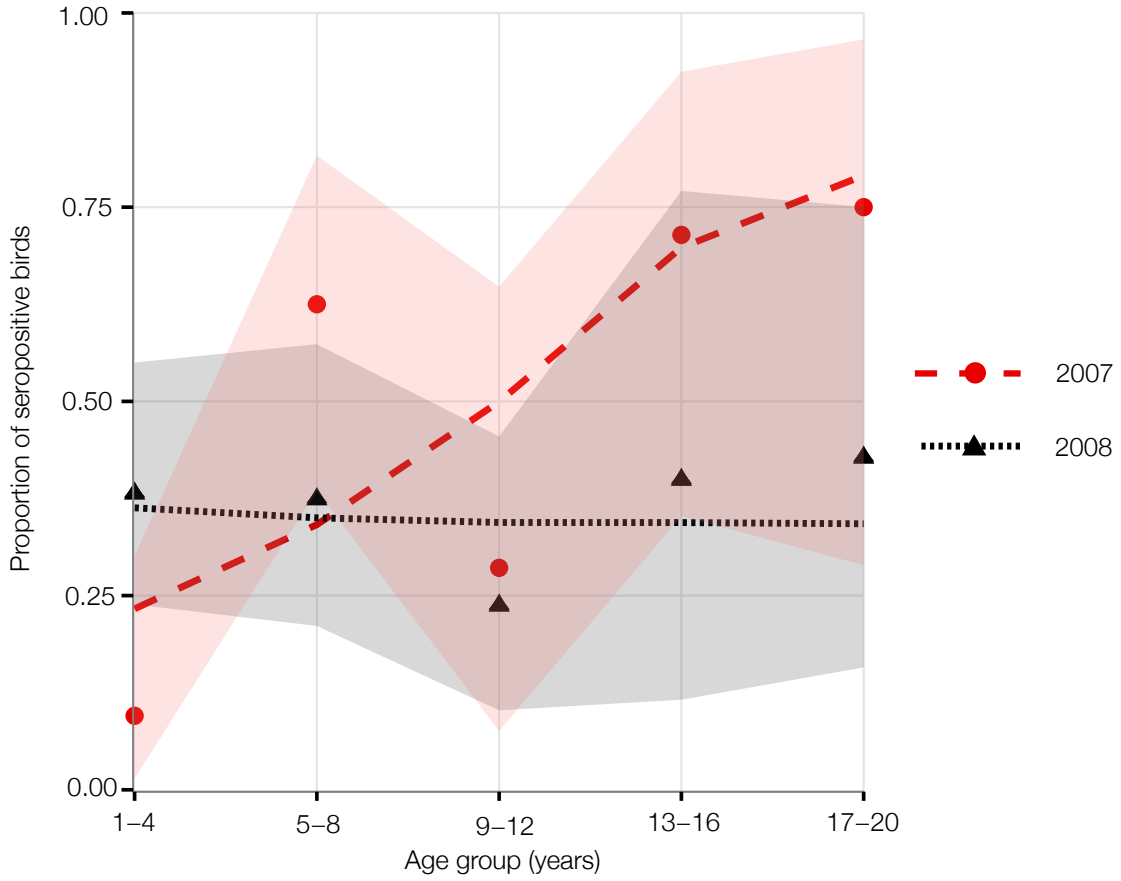

Figure 5.4: Proportion of birds observed to respond to H5N2 in 5 age groups for each sample year (dots, with 95% confidence intervals for proportions shaded (2007; pale red, 2008; dark grey). Lines represent the mean proportion of birds in each age category expected to respond to H5N2, based on the fitted probability of effect of age and sample year (2007; red, wide-dash, 2008; black, narrow-dash) on response to H5N2 as determined by GLM with logit link (D.2). As the inclusion of sex in the model does not significantly affect the probability of having a response, for simplicity, expected number of birds is based on the fitted probability estimated for a model without a sex term.

5.4 Discussion

Studies of AIV seroprevalence have shown that "adult" birds (typically individuals 1 or 2 years after hatch-year, depending on species) are more likely than "juvenile" birds (sampled during, or one year after, hatch-year) to harbour antibodies targeted at AIV NP [182, 181, 179, 222, 487, 117, 352, 236]. Here I show that not all birds defined as seropositive for AIV by NP antibody presence have equivalent levels of protection against AIV. Instead, older birds are increasingly more likely to produce antibodies directed at a wider range of HA types. This data suggest that mute swans have a long-term immune memory, possibly similar to that exhibited by humans in which memory B-cells produced during a primary infection are reactivated following further infections of a related HA type [160, 509, 395, 56]. The ability of wild waterfowl to form life-long immune memory in response to naturally encountered infections supports key assumptions made in modelling studies of AIV infection dynamics [367, 483].

The pattern observed here could be generated by at least two non-exclusive mechanisms. In the first scenario, a high proportion of birds is exposed to AIV each year, but younger birds are less likely to form long-term immune memory to the strains than older birds. However, whilst increases in the ability to form immune memory with age have not been investigated, such gains in immune function seem improbable given that birds appear to undergo immune-senescence as they age [308, 292, 318, 242]. In the second scenario, a low proportion of birds is exposed to different AIV HA types every year, but all age groups are equally likely to form a long-term immune response. The prevalence of AIV in healthy, wild mute swan populations is low (typically <3% [314, 236, 296, 21, 399]), with most infections observed in juveniles (6% prevalence) [236]. Experimental evidence in mallards (Anas platyrhynchos) suggests that birds are infected with approximately half of all HA types (6/11) circulating in the population within the first two years of life [443]. On average, the AIV prevalence in swans is thought to be considerably lower than the prevalence found in wild mallards [314], so it is plausible that the number of infections with different HA types per swan per year

is also lower than in mallards. If so, swans would encounter a relatively low number of HA types each year and would form a long-term HA immune response to at least a proportion of these viruses. This scenario is compatible with the observations in my study population of high seropositivity for AIV NP [352] and the accumulation of HA type specific responses throughout life.

Several previous studies of AIV seroprevalence have found that male birds are less likely than females to be serologically positive for AIV [182, 487, 236, 452]. However, other studies have shown the opposite trend [117]. This data suggest that male birds consistently had lower levels of NP antibodies and a slightly narrower response than female birds, but the significance of this result varied depending on which strains were used to calculate breadth. It has been suggested that sex-specific seroprevalence might result from different breeding behaviours of each sex [117] or inherent differences in immune response or antibody persistence [487, 308, 292]. I find no evidence to support the hypothesis that age affects the relationship between sex and seroprevalence, which might be expected if there were an effect of breeding status D.2.

Amongst those birds that did exhibit an antibody response, I found no evidence for an effect of age on the magnitude of HA antibody titre, suggesting that the responses of younger birds to the tested antigens are as strong as those of older birds. Hence any age-related differences in immunity derive only from the presence of absence of specific responses. It is unclear whether increasing seropositivity with age corresponds to increasing protection against the effects of AIV infection. As well as humoral immunity, cellular immunity may contribute to faster recovery and reduced viral shedding following infection, via the recognition of AIV-infected cells by cytotoxic T-lymphocytes (CTLs). Although little work has characterised CTL responses in birds, it has been shown ex vivo that chicken CTLs can target AIV NP and HA [393, 158]. Recognition of the conserved nucleoprotein could result in cross-protection among AIV viruses, including those with antigenically distinct HA types [383, 204]. In my study, I defined seropositive birds as those with titres $\geq$20. This is substantially lower than the threshold recommended by the OIE for vaccine-induced

protection against mortality ($\geq$32) or reduced viral shedding ($\geq$128) [312]. However, several studies suggest that protection may be afforded by lower titres than those specified by OIE, including a threshold of $\leq$20 for a good probability of protection against mortality from HPAI infection, and of $\geq$40 for prevention of viral shedding in most infected birds [232, 429, 431]. Secondary infections with LPAI viruses show significantly reduced shedding even when the original antibody HI response is weak (<20) [199]. If protection at low titres is caused by humoral immunity, I would expect that increasing seropositivity with age would result in increasing protection against the effects of AIV infection. Conversely, if protection at low HI titres results from immune cross-protection against conserved proteins, there may be little protective benefit of an increasing number of weak responses against a larger number of HA types.

Interestingly, this data suggest that older birds testing seropositive for previous infection by NP-ELISA typically have stronger responses in that NP-ELISA assay than seropositive younger birds. Anti-NP antibodies are non-neutralising but have been shown to contribute to protective cross-subtype immunity in mice [237, 54]. If NP antibodies are found to contribute to protective immunity in birds, the age-correlated differences in NP antibody levels observed here may result in differences in immune protection against AIV with age. Research into the relative importance of humoral and cellular responses against AIV when HA antibody responses are low is needed to understand how age, immune protection and serological breadth are related.

The natural outbreak of HPAI H5N1 in the Abbotsbury population in winter of 2007/2008 enables us to consider whether age specific immunity may have had an effect on, or been affected by, the dynamics of that outbreak. The only previously identified case affecting wild birds in the UK involved a single whooper swan (*Cygnus cygnus*) in Scotland in 2006 [4]. It is therefore very unlikely birds at Abbotsbury had encountered AIV from the HPAI H5N1 lineage prior to the start of the outbreak in late 2007. Consequently, the low levels of reactivity to HPAI H5N1 that was

observed in sera sampled in summer 2007 (6% of birds) may indicate a low level of protective immunity in the population as a result of previously encountered LPAI viruses. Interestingly, I observed a significant population-level increase in HPAI H5 antibody prevalence after the outbreak. This is consistent with a scenario in which a proportion of birds developed antibody responses against the HPAI H5N1 virus following undetected infections during the outbreak. Experimental work suggests that waterfowl previously exposed to an LPAI virus are more protected against the effects of subsequent exposure with H5N1 [132, 76, 24, 329]. This data suggests that older birds were more likely to have such pre-existing antibodies than younger birds. I speculate that this could have caused the higher mortality observed in younger birds (<3 years old in January 2008) during the 2008 HPAI H5N1 outbreak [352].

There are several caveats to my study. First, the samples used here were collected exclusively during July and August, and hence I cannot determine if seasonality affects any of the relationships that I report here. Active infection rates appear to fluctuate seasonally in wild waterfowl and some evidence suggests that serological responses may also vary seasonally [182]. Secondly, whilst my results are robust to a choice of H5 and H9 antigens, the choice of strains and HA types may have affected the results. Despite focusing on only 5 strains I found a significant increase in breadth of humoral response with age using just 3 HA types. Consequently I suspect my results are conservative and that the age-dependent effect would be stronger if more antigens had been tested. However, I cannot rule out that testing using a larger panel of viruses isolated over a long time in the test population might affect the results. Thirdly, the extent to which these results can be generalised to other long-lived, wild waterfowl is unknown. Swans at Abbotsbury may experience higher levels of AIV exposure than other mute swans due to the unusually large population size, and different environmental and nutritional stresses (see Section 5.2.1). Despite this, I think that it is unlikely that the capacity to form long-term immune memory would be qualitatively different in other mute swan populations or related species.

In summary, I find that older wild mute swans are more likely to have broader

antibody responses against AIV, and hence may be protected against a wider range of viruses. My data suggest that immune memory in this population, and perhaps also in other waterfowl populations or species, is long-lasting. The profile of existing immunity within a population almost certainly affects the demographic impact of new viral infections such as HPAI H5N1. Further, accumulation of antibody responses throughout life likely generates differences in the epidemiology of AIV among host species with different lifespans. A more complete understanding of influenza dynamics will require long term screening of active infection and serological immunity in other well-studied animal populations in nature.

Chapter 6

Epidemiology of highly pathogenic H5N8 avian influenza virus in a population of long-lived birds

6.1 Introduction

Since their emergence in 2010 [479, 515], highly pathogenic avian influenza (HPAI) clade 2.3.4.4 H5 viruses have spread rapidly across the globe, causing billions of dollars of associated damage [407]. The extensive geographic spread and high rate of reassortment of this clade has been unprecedented. Outbreaks have been detected widely across North America, Africa, Europe and Asia and reassortant strains involving at least 6 different neuraminidase (NA) types have been detected (N1, N2, N3, N5, N6, N8) [10, 245, 489, 515], and are sometimes also associated with novel internal gene segments.

Highly pathogenic avian influenza virus H5N8 has been introduced to Europe multiple times, with the first wave resulting in only twelve detected outbreaks in four countries during November to December 2014 (Germany, Italy, The Netherlands, the UK) [2]. The early wave of the outbreak and background on the virus is given in more detail in Section 3.1. In October 2016, H5N8 was reported for a second time

in Europe [120]. In contrast to the previous first wave, the second introduction was associated with high mortalities in wild birds [120]. The second wave resulted in a widespread European outbreak of long duration, with cases still being reported sporadically across Europe, as of September 2017. Birds from >50 different wild bird species have been found infected in EU countries [10]. Similar to earlier HPAI viruses such as H5N1 [46], observational [213] and experimental studies [23, 203] suggest that infections in some species cause high mortality, whilst in other species infection is often asymptomatic [26, 191, 193, 322, 458]. In total between October 2016 and June 2017 (i.e., during the second epidemic wave in Europe), 44% of all detection events in wild birds in EU countries were made in swans, of which almost half (22% of total detection events) were confirmed as mute swans (*Cygnus olor*) [10]. The extent to which this high detection rate in swans reflects increased suspectibility to the disease or simply reflects a detection bias towards larger, more visible and recognisable birds is unknown.

On 23rd December 2016, a mute swan (*Cygnus olor*) was found dead at Abbotsbury Swannery, Dorset, UK, which subsequently tested positive for H5N8. Mortality in the swan population escalated over the following weeks, generating to my knowledge the largest single outbreak of a highly pathogenic AIV ever recorded in wild birds in the UK. 182 mute swans died in total, with hatch-year birds being most affected. The outbreak at Abbotsbury is a unique event and scientifically valuable for several reasons. Firstly, the affected swans at Abbotsbury form a resident colony of approximately 800 birds that has been subject to intense ornithological study for over 60 years. Almost all birds that are hatched into the population are ringed upon hatching, and detailed demographic information such as sex, age and parentage is known for most birds. Secondly, this population of swans has been naturally infected by both HPAI H5N1 (in 2007/2008) [352] and, now, also HPAI H5N8 (in 2016/2017). To the best of my knowledge, the Abbotsbury mute swan population represents the only wild population in the world for which natural infections of H5N1 and H5N8 can be directly compared. I also believe that this is the only wild population affected by

H5N8 for which exact age of most birds that died in the outbreak is available. The population is therefore unusually suitable for exploring the effect of exact bird age on the epidemiology of H5N8.

For low pathogenic avian influenza (LPAI) viruses in wild birds, age is closely correlated with the probability of infection. In field studies of avian influenza, juvenile birds have a higher prevalence of low-pathogenic viruses than adult birds [181, 224, 452, 302, 138, 471], possibly as a result of acquisition of adaptive immunity during the lifetime [181, 182, 179, 222, 487, 117, 352, 236].

The extent to which age is correlated with protection against HPAI is typically extremely difficult to study in the wild. One study reported no obvious difference in detection rates of natural infection by HPAI H5N1 between adult and juvenile birds [392], based on haemagglutination assay detection of cultured viruses followed by RT-qPCR detection of H5 and N1 in samples. However, the numbers of juvenile birds tested was too small to draw significant conclusions. Laboratory studies suggest that both increased age and previous exposure to LPAI (which typically associates with increasing age) can contribute to protection against HPAI. For multiple different HPAI H5 viruses, experimental challenge of Pekin ducks (*Anas platyrhynchos domestica*) and Ruddy ducks (*Oxyura jamaicensis*) has shown that immunologically naïve younger ducks (less than or equal to 8 weeks of age) are more susceptible to severe effects of HPAI H5 infection than immunologically naïve older ducks [3, 270, 323, 324, 407]. Several models to explain this have been proposed, including the idea that as birds mature they can mount more effective immune responses to viral challenges, and that viral replication might be more effective in more mature host cells [323]. Previous exposure with an LPAI virus has also been associated with reduced morbidity upon subsequent challenge [24, 76, 200, 329]. Laboratory studies therefore suggest that increased bird age should be associated with decreased susceptibility to HPAI infection, both as a result of a maturing immune system (exclusively in juveniles) and also because of previous exposure to AIV (adults and juveniles). In practice this effect is difficult to measure in the wild, in part because detection of

HPAI infections is comparatively uncommon and data on ages are often unknown.

It is believed that wild bird movements drive the geographical spread of AIV H5N8. Genomic sequencing efforts and phylogenetic analysis have repeatedly demonstrated that H5N8 can be introduced to the same country multiple times in a short period of time [143, 169, 439] including a wide diversity of strains and reassortants. The Fleet Lagoon is an important wild bird reservoir on the south coast of England, with most overwintering avian visitors from a wide variety of species including teal, wigeon, shoveler and pochard ducks arriving from continental Europe [352].

Here, I explore the factors affecting avian mortality during the 2016/2017 H5N8 HPAI outbreak, including quantifying the effect of younger age, sex and last known weight on chance of death. To better understand the origins of the outbreak at Abbotsbury, and its relationship to other viruses in the UK and Europe, I undertook genomic sequencing of H5N8 positive samples from the outbreak. I estimate molecular-clock trees to date the origins of the outbreak at Abbotsbury, explore the time and route of introduction of this virus to the population and to place the virus within the known diversity of other outbreaks in Europe.

6.2 Materials and Methods

6.2.1 Site and sample collection

Abbotsbury Swannery harbours a large population of resident mute swans (*Cygnus olor*) for which detailed demographic data are known (for details of the site, see Chapter 4). Any dead swans found at the Swannery site are reported by staff to the Animal and Plant Health Agency (UK) and may be subjected to post-mortem and testing for notifiable diseases, particularly if multiple birds are observed to die within a short period. On 23rd December 2016, a dead swan was found at the site that subsequently tested positive for avian influenza H5N8. Four other birds that were tested prior to this during December 2016 all tested negative. From 23rd December

until the 1st January, a further 12 dead swans were found at the site, all of which were tested by cloacal and oropharangeal swabs for AIV. From 13th January 2017 until 31st January, all dead birds found at the site were tested for AIV. Two swabs were collected per bird, including one cloacal and one oropharangeal swab where possible. Swabs were stored at -80°C on site until shipping. Swannery staff patrolled the Swannery site at least once per day between late December 2016 and the beginning of February 2017, and patrolled the Fleet Lagoon at least once per day on weekdays and often also at weekends, so it is likely that most swans that died during the outbreak would have been collected. Date of finding and ring numbers (if any) were collected for all dead birds found during this period.

Throughout sampling of birds in this population and shipping of samples, personal protective equipment and biosecurity procedures were determined by partners at the Animal and Plant Health Agency. Although the risk of zoonotic transmission was thought to be extremely low, Public Health England staff checked daily with all workers at the Swannery site that they showed no signs of illness and Tamiflu was taken by all workers.

The number of birds of all sighted species on the Fleet Lagoon in November 2016, December 2016 and January 2017 were extracted from Wetland Bird Survey Data (Appendix E.1). Although a portion of the data is missing for the December 2016 survey, the survey presents a minimal likely count of the numbers of other birds present on the site during these months (Appendix E.1).

6.2.2 Estimation of LPAI AIV prevalence

To gauge the extent to which birds in this population are commonly exposed to LPAI viruses and whether this might be associated with patterns of H5N8 mortality in the population, I screened 866 faecal or cloacal swabs for AIV. These samples had been collected approximately once per month during April 2015 to March 2016 as part of preliminary work for Chapter 4, so whilst they do not immediately precede the

outbreak period they are useful for exploring typical AIV prevalence in the population. For faecal samples, birds were observed until they defecated, at which point a sterile swab was rolled over the faeces before being placed in a tube containing 1ml of Universal Transport Media (Copan). Note that the faecal *swabs* tested here therefore differ from the whole faecal samples described in Section 4.2.1. Samples were transported on dry ice and stored at -80°C until testing. Cloacal swabs were collected from birds that were over 1 year old, as identified by the plumage, specifically the clear colour change of the plumage from grey to white that occurs towards the end of the first year. Following consultation with the local Home Office inspector, University veterinarians and local ethical review board, this procedure of collecting cloacal swabs from adult swans as defined by the aforementioned change in plumage was deemed by the local Home Office inspector to be unlikely to cause significant distress, and was therefore not regulated under A(SP)A 1986. Training in collecting swabs was provided by a University veterinarian and a wildlife veterinarian from the Royal Society for the Prevention of Cruelty to Animals (RSPCA). Ethical permission for collecting cloacal swabs was granted by the University of Oxford Department of Zoology Animal Welfare Ethical Review Board committee.

6.2.3 Estimation of HPAI prevalence in healthy birds

To detect the presence of HPAI H5N8 in healthy birds in the population, I took cloacal swabs from live birds in the colony on 19th and 20th January 2017. Birds were caught with the help of several of the Swannery staff. 92 birds of known age over 1 year old were sampled.

6.2.4 Screening for positive samples

All HPAI diagnostic swabs were sent dry to the UK Animal and Plant Health Agency (Weybridge) for testing by the Animal and Plant Health Agency. Individual swabs were eluted into brain heart infusion broth containing antibiotics (1000IU penicillin

G, 10ug/ml amphotericin B, 1mg/ml gentamycin) and 140ul of swab-suspension was carried forward into nucleic acid extracton. For LPAI swabs from 2015-2016, 140ul of the Universal Transport Media used for swab collection was carried forwards into nucleic acid extraction. Nucleic acids were extracted using a Qiagen BioRobot Universal and a QiaAmp Viral kit with the inclusion of carrier RNA according to manufacturer's instructions. Negative and positive extraction controls were included. Reverse-transcription quantitative PCRs targeting the AIV matrix gene [301], H5 [397] and N8 were performed on extracted viral RNA. The matrix gene and H5 PCR protocols are published in detail elsewhere, but the N8 protocol followed by the Animal and Plant Health Agency was designed in response to the emergence of clade 2.3.4.4 during autumn 2014 and is as yet unpublished, so is described briefly here. The assay was designed by Prof Timm Harder (Friedrich Loeffler Institute, Germany). Primers and probes are given in Table E.2. qPCRs were performed using the Qiagen OneStep RT-PCR kit. 25ul reactions were performed using 2ul of extracted nucleic acid, 1ul of Qiagen OneStep RT-PCR Enzyme Mix, 3.75ul of Stratagene ROX reference dye (prediluted 1:500 with DEPC H_2O), 1ul of RNAsin and final concentrations of 1x Qiagen OneStep RT-PCR buffer, 0.8uM of each primer, 0.4uM of probe, 400uM of each dNTP and 3.75mM of $MgCl_2$. Thermocycling conditions involved reverse transcription for 30 minutes at 45 °C, followed by enzyme activation at 95 °C for 15 minutes, and then 40 cycles of denaturation at 95 °C for 15 seconds, primer annealing at 56 °C for 20 seconds, and extension at 72 °C for 30 seconds. Cycling was performed on a Stratagene Mx thermocycler. Samples <36 were considered positive.

All viruses detected during April 2015 to March 2016 were screened to test that they were not H5 or H7 subtypes in order to confirm that these viruses were not possible highly pathogenic viruses. The PCR protocols follow those of [397] and [398], so are not detailed further here. Several LPAI viruses from 2015 and 2016 were sequenced as part of metagenomic investigations in Chapter 4, so subtypes could also be confirmed for these viruses.

6.2.5 Amplification and sequencing of the H5N8 genome

RNA extracted by the Animal and Plant Health Agency was shipped to Oxford, and cDNA synthesis was performed on H5N8 positive samples using the Protoscript II First Strand cDNA Synthesis Kit (New England Biolabs). A multiplex primer scheme was designed to amplify the whole viral genome based the assumption that the strain at Abbotsbury would be similar to previously published sequences of H5N8 in Europe during late 2016 or early 2017 [354] and hence specific primers could be designed against these strains. The scheme consisted of 92 different primers designed to amplify overlapping amplicons of 400bp each, with an overlap length of 75bp between neighbouring amplicons (Appendix E.3). PCRs were conducted according to previously published methods [354]. 40 cycles of denaturation, annealing and extension were conducted. The negative controls used in cDNA synthesis and a PCR water control were included on each plate. PCR products were cleaned using 1x Ampure bead cleanups, and DNA concentration measured using a Qubit High Sensitivity dsDNA kit on a Qubit 3.0 fluorimeter. PCR products were standardised by concentration, and 90ng of total DNA per sequencing library was carried forwards into library preparation.

Amplified DNA and appropriate negative controls were sequenced in barcoded multiplexes of 6 - 8 samples per run on the MinION (Oxford Nanopore Technologies) using FLO-MIN106 flow cells. Library preparation was conducted using Ligation Sequencing 1D and Native Barcoding kits according to the manufacturer's instructions, but with the changes detailed in [354] (kits numbers; SQK-LSK108, EXP-NBD103). Libraries were loaded onto flow cells (FLO-MIN106) and sequencing was performed without basecalling for 48 hours using MinKnow 1.7.7.

Consensus sequences for each barcoded sample were generated following previously published methods [354]. Briefly, raw files were basecalled using Albacore 1.2.5 (Oxford Nanopore Technologies), demultiplexed and trimmed using Porechop, and then mapped with BWA to a reference genome (A/turkey/England/052131/2016;

GISAID Isolate ID EPI_ISL_239801). Nanopolish variant calling was applied to the assembly to detect single nucleotide variants to the reference genome [269]. All sites where the coverage was <20 and all primer binding sites were masked with Ns during generation of the consensus sequences.

To check that sequencing using this approach had not introduced errors into the consensus genomes, further specific PCRs were performed for 5 samples for a short section of the PB1 region for which the sequence was relvatively divergent compared to existing published sequences. PCR was conducting using Q5 High Fidelity Hot Start Polymerase according to the manufacturer's instructions and with 200uM final concentration of dNTPS and 0.5uM final concentration of each primer PB1_400_2L and PB1_400_3R (Table E.3). Thermocycling conditions were as follows; denaturation at 98°C for 30 seconds, followed by 40 cycles of denaturation (98°C, 7 seconds), primer annealing (68°C, 15 seconds) and extension (72°C, 25 seconds), followed by a final extension for 2 minutes at 72°C. Amplification of bands of the correct size was checking using agarose gel electrophoresis, and PCR products were cleaned following the ExoSAP protocol in Chapter 4 Section 4.2.5. The cleaned PCR products were directly Sanger sequenced from both ends of each amplicon (Source Bioscience). Chromatograms were trimmed to high-quality regions and aligned to the MinION consensus sequences to check for disparities between the two sequencing approaches.

6.2.6 Phylogenetics

All segments from all available Eurasian H5N8 sequences were downloaded from GiSAID (August 2017). Each segment was aligned with sequenced isolates from the Abbotsbury site using MUSCLE in Geneious 8.1.7, and the segments were manually trimmed to the same length as the Abbotsbury H5N8 sequences.

For all segments, simple Jukes-Cantor neighour-joining trees were conducted, and if several large clades were observed (presumably indicating reassorted internal genes), the major clade containing the Abbotsbury sequences was extracted (92-224 sequences

for each of the the 8 different segments). Sequences in this clade were used to etimate maximum likelihood trees. Appropriate substitution models were chosen using BIC calculations in jModelTest [349]. Maximum likelihood trees were estimated with Phyml [156], including 100 bootstrapped trees to evaluate statistical support for each branch.

In order to estimate the timing of the introduction of the outbreak to the site, a molecular clock phylogeny was estimated using BEAST [108]. Appropriate temporal signal was checked for the NA and HA alignments in TempEst [359]. Several phylogenetic models using different substitution models were constructed for each segment and compared using path-sampling/stepping stone sampling approaches implemented in BEAST [13, 14]. Starting with the simplest model of HKYG substitution model with strict clock and constant population size, I compared this model to a model of increasing complexity, and adopted the model with the highest support at each step. Sequentially, I compared and adopted all changes of; the lognormal relaxed clock in place of the strict clock model, SRD06 substitution model in place of the HKYG substiution model, and finally, the Bayesian skyline population model in place of the constant size model. Two independent runs of 10,000,000 step MCMC chains were performed for the best model with sampling ever 10,000 steps, and the first 10% discarded as burnin. The time to the most recent common ancestor was estimated using the 95% highest posterior density support interval for the Abbotsbury clade.

6.2.7 Epidemiological analysis

Epidemiological analysis was conducted to determine the distribution of mortality in the population by age.

In order to estimate the relative impact of the H5N8 outbreak in birds of different ages, it was necessary to estimate which birds would have likely been in the population in December 2016. Accurate censuses of the population are available from July 2015 and July 2017, and censuses of breeding birds and hatch-year birds are available

from October 2016. Information on ringing, ring-sightings and discovery date of any dead birds are available throughout 2016 and 2017. Further to the birds found dead during the H5N8 outbreak, the rings of >100 other birds seen on the site during the outbreak were also recorded. The December 2016 population was hence estimated in two ways. In the first dataset (A), the population was estimated by including all birds observed during July 2015 census, adding any birds sighted or ringed until January 2017, adding cygnets that survived to October 2016, and removing any birds known to have died prior to the first detection of H5N8. In the second dataset, (B), it was estimated from the censused population in July 2017, but adding any birds known to have died between 23rd December 2016 and July 2017 and adding any pre-2017 hatch-year birds sighted or ringed between December 2016 and July 2017. The datasets both estimate similar proportions of hatch-year birds in the population during the outbreak; 21% of the population (250 hatch year birds) in dataset A and 23% of the population (199 hatch year birds) in dataset B. Most results presented here are calculated only for dataset A.

Birds that died between 2nd January and 12th January 2017 were not tested for AIV [1]. However, if I could demonstrate a significant excess of mortality compared to the same period in previous years, it would seem reasonable to assume that the majority of these untested birds died of H5N8. To test this excess of mortality, recorded deaths during the outbreak period were compared to the same period in previous years (all alternate years from 1999-2015)[2]. Estimates of the December population size in previous years were calculated using dataset A, and the number of dead birds discovered in the 23rd December to 24th January period for each year

[1]Lack of sampling during this period was not a deliberate scientific decision. Sampling of a notifiable influenza outbreak for scientific purposes is unusual. Standard procedures to enable sampling by anyone from a non-governmental organisation were therefore, understandably, not in place prior to this outbreak. The period of missing sampling reflects the amount of time that it took to confirm the presence of AIV at the site following the Animal and Plant Health Agency's initial testing in early January, and the time taken to agree upon appropriate procedures and arrange logistics that would enable me to safely and legally conduct futher sampling in the population. Birds that died at the site between 2nd and 12th January were securely destroyed off-site before logistics and permission could be reasonably arranged.

[2]Specifically, in years in which a population census was conducted.

were also calculated. The crude mortality rates and the age-adjusted mortality rates for each year were calculated using 2016 as the reference population.

The proportion of birds of each age that died during the H5N8 outbreak was calculated, and a strong skew towards increased mortality in the youngest birds was noted. To test the significance of this skew, one-tailed Fisher's exact tests were performed to compare the number of dead birds to the number of live birds for birds of age <1 year and older birds. Generalised linear models with a binomial response (survived or died during the outbreak) were used to assess the impact of age and sex on likelihood of death. Analysis was performed first with ages group as <1 year (hatched in 2016) or 1 year or older (hatched before 2016), then repeated with ages grouped into <1 year, 1-2 years, 3-4 years and 5 or older.

To test whether the age skew was specific to H5N8-mortality or whether the skew is part of a normal winter mortality profile, the proportion of birds of each age group (separated by year for 0-7 years, and 8 or more) that died was calculated for the outbreak and for previous years (alternate years 1999-2015).

The reproduction number for the H5N8 outbreak was estimated using the R package R0 [310], which implements standard formulas linking the reproduction number to the exponential growth rate of an outbreak and a generation time distribution (reviewed in [463]), and also implements a maximum likelihood method to estimate the reproduction number. The implemented maximum likelihood method follows the method of [477]. In brief, the method does not require information on a pre-existing generation time distribution, but instead assumes that each index case causes a number of secondary cases that is Poisson distributed with a certain expected value, and attempts to maximise the log-likelihood of that expected value. Data on counts of carcasses at the Swannery were used as a proxy for H5N8 case counts, under the assumption that sampling effort did not change over long periods of time during the outbreak and that all bird deaths could be attributed to infection with H5N8 avian influenza. Fewer carcasses were collected on public holidays (such as Christmas) and on weekends (Figure 6.1). Spikes of sampling after these days suggest that carcasses

of birds that died on those days were still collected and recorded and hence on average the smoothed mortality profile of the outbreak would probably be reasonably correct. A distribution for the generation time was fit using the R package R0 [310] based on published experimental data on the time between experimental inoculation of geese or ducks with H5N8 isolated and subsequent infection of contact birds [321, 246, 424, 233, 203].

The majority of deaths occured in cygnets < 1 year old. To explore why some cygnets died when others survived, I tested whether the weight or exact age of a cygnet affected the probability of it dying during the outbreak. Exact hatch dates are known for all birds born into the population in 2016, and all cygnets are weighed upon ringing with adult leg rings in September or October. For all ringed cygnets, age (in days) was calculated for the day in which they were weighed, and a generalised linear model fitted to expected weight for a bird of that age and sex. The difference between the weight of the bird at ringing and its expected weight was calculated to measure how overweight or underweight a bird was compared to the average for its age and sex. Wilcoxon rank-sum tests were performed to test the association between the difference between expected weight and true weight at ringing and whether a bird died in the outbreak. Wilcoxon rank-sum tests were performed to test the association between age in days on 23rd December and whether a bird died in the outbreak.

Age-related differences in LPAI infections in birds were tested using Fisher's exact tests by grouping the swans into two age groups; up to 2 years old, or 3 years and older. Analyses were conducted on two different datasets, one considering positive samples to be those with RT-qPCR matrix protein assays of less than Ct 36, and one considering positive samples to be those with RT-qPCR matrix protein assays of up to Ct 38. All statistical analyses conducted here were performed in R.

6.3 Results

6.3.1 H5N8 outbreak

All swans found dead at the site from 6th December 2016 until the 1st January 2017, and from 13th January until 14th February 2017 were tested for H5N8 if the cloaca and/or oropharynx was present on the carcass. Three birds from other species that were found dead were also tested (one black-headed gull of unknown age (*Chroicocephalus ridibundus*), one adult coot (*Fulica atra*), and one adult Canada goose (*Branta canadensis*)). The first and last carcasses that tested positive were found on the site on 23rd December and 24th January 2017, respectively. In total, 182 dead swans were found during this time (henceforth referred to as the outbreak period), with most deaths occurring between 3rd-13th January 2017 (Figure 6.1). Of the dead birds, 23 were unringed or were of unknown ring-status due to missing legs as a result of predation, and 159 were ringed. Exact hatch years were known for 150 birds. Sex was known for 142 of these known-age birds.

The basic reproductive number (R0) of the outbreak was estimated from counts of recovered carcasses at the site and an estimated generation time distribution was calculated based on published experimental data (see Section 6.2.7 for details of methods and references). The best-fit distribution for the generation time distribution was a Weibull distribution with a mean of 2.9 days and a standard deviation of 1. The estimated R0 from the outbreak was calculated to be 2.02 (95% confidence intervals 1.75- 2.36) using the exponential growth method, and 1.59 (95% confidence intervals 1.22-2.04) using the maximum likelihood methods implemented in the R package R0 [310]. These methods are described in brief in Section 6.2.7.

Of the 92 live birds $>$ 1 year old that were tested in the population, only one bird (a male hatched in 2008) was positive for H5N8 AIV, with a Ct of 35.5 for the matrix protein assay, Ct 36.01 for H5 and Ct 34.1 for N8. This apparently healthy, H5N8 positive bird was observed again in the population in June 2017, so clearly survived the infection. I note that all samples from live birds were collected on 19th

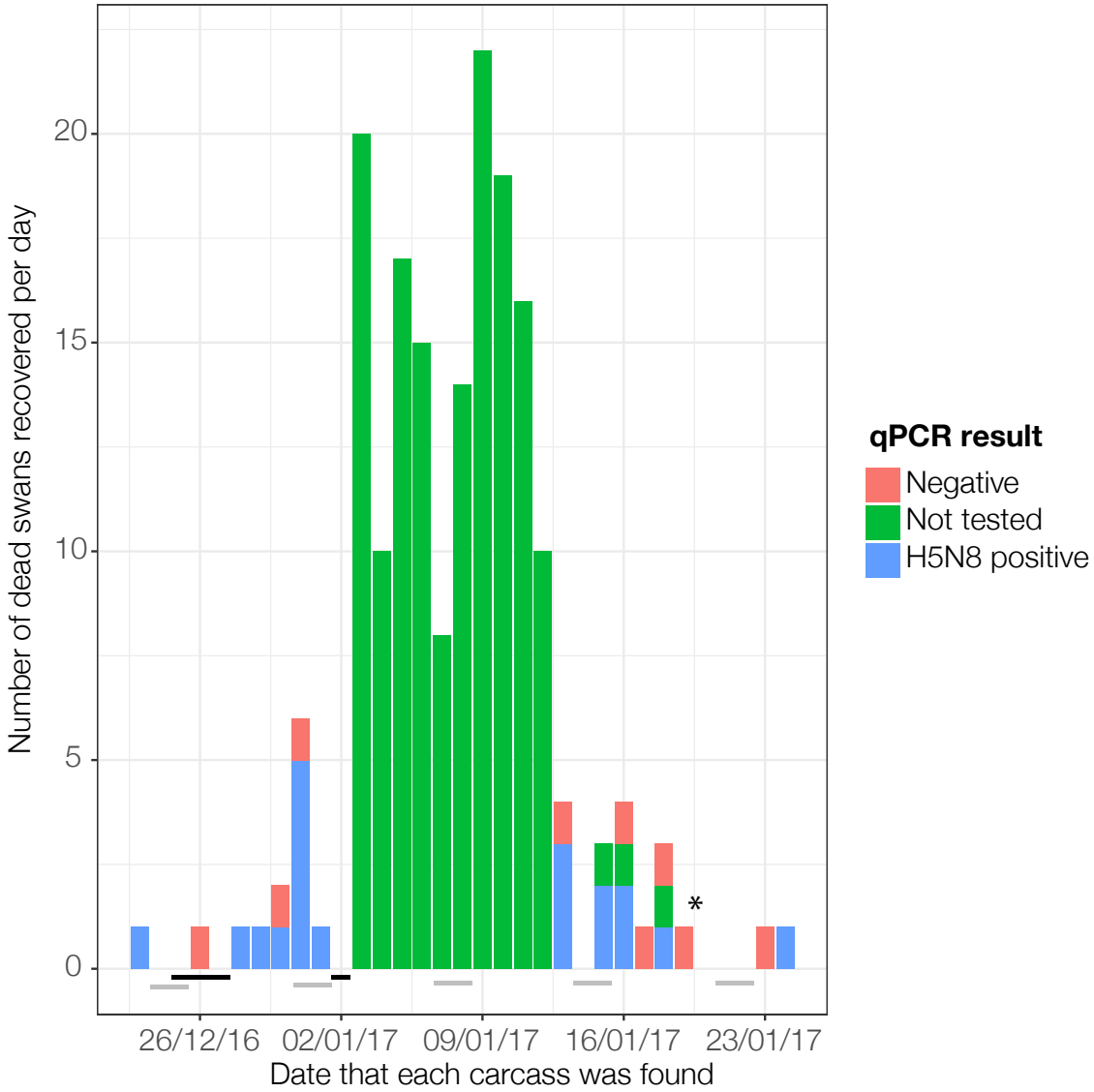

Figure 6.1: Number of carcasses found at Abbotsbury Swannery over time. Carcasses that were subsequently tested for AIV are coloured in blue (H5N8 positive) or red (H5N8 negative), and untested samples are coloured green. A star marks the date of testing of live birds in the population for AIV. Grey boxes towards the base of the figure indicate weekends, and black boxes indicate public holidays in England, both of which might have affected carcass collection effort.

and 20th January 2017, and therefore after almost all of the mortalities had passed (Figure 6.1). Whilst this finding confirms that there was little circulating H5N8 in the population by late January 2017, I suspect that the number of asymptomatic cases would have been much higher during the peak of the outbreak in early January.

The population in December 2016 was estimated to include between 1167 (dataset A, see Section 6.2.7 for population estimation details) and 841 ringed birds (dataset B). Based on observed counts of swans on the Fleet between November 2016 and January 2017, these numbers are both likely to be overestimates. Specifically, 985 swans were estimated to be on the Fleet in November 2016, but this count includes both ringed and unringed birds. Swans often move onto to Fleet Lagoon from other local areas during the summer moulting season, so estimating the population size from summer census populations might slightly overestimate the number of birds present in the winter. However, I have no reason to believe that the estimate of which birds were present in the population during winter 2016 would be significantly biased by age or sex. Even adopting the highest estimate of 1167 ringed swans in the population, at least 15% of birds at the site died during the outbreak period of 23rd December to 24th January. Several other avian species were present in high numbers at the site from November 2016 to January 2017 (Appendix E.1). No unusual mortality was noted in any of these species, although small numbers of dead birds in species with small body size or muted coloration may plausibly have been missed.

To my knowledge, the Abbotsbury population represents the only wild population in which both HPAI H5N1 and HPAI H5N8 have been detected. The number of carcasses found at the site was much higher during the H5N8 outbreak than during the H5N1 outbreak, despite very similar efforts to detect HPAI virus during both outbreak. Both outbreaks began in the same epidemiological weeks of 2007 and 2016 (Figure 6.2).

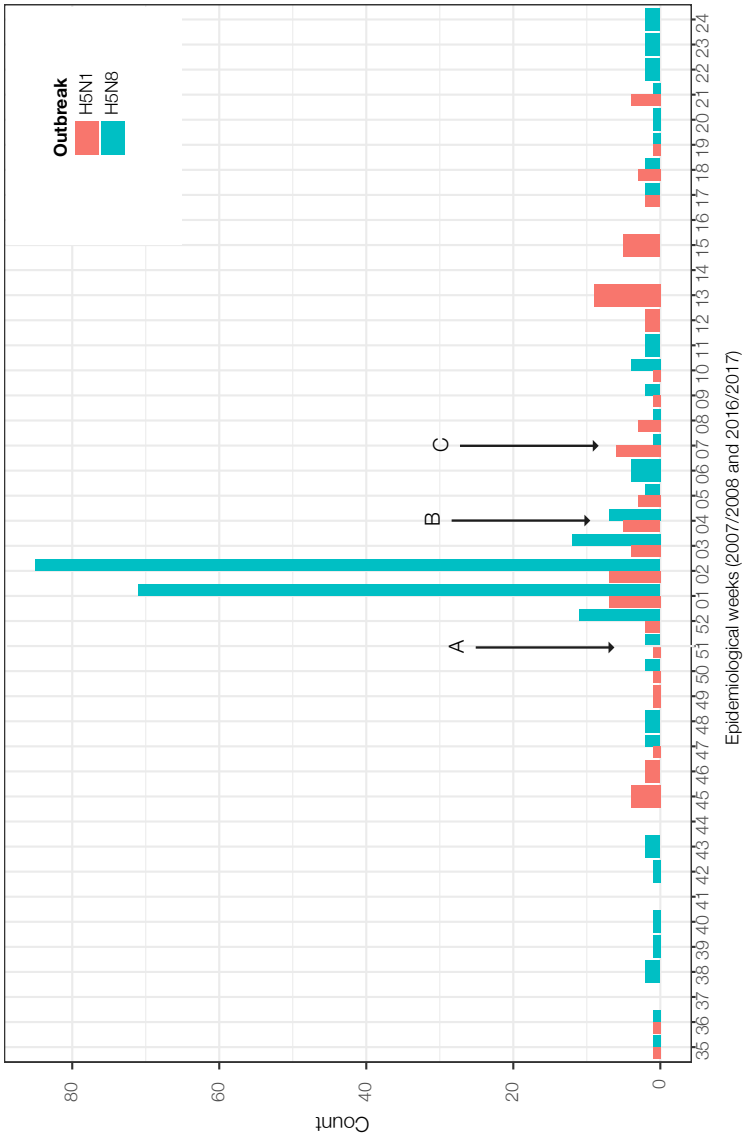

Figure 6.2: Comparison of mortalities during 2016-2017 (coloured blue, encompassing the HPAI H5N8 outbreak period) and 2007-2008 (coloured red, encompassing the HPAI H5N1 outbreak period). Arrow A marks the epidemiological week in which the first positive birds for H5N8 and H5N1 were found in the respective years. Arrow B marks the date of the last detected H5N1-positive bird. Arrow C marks the date of the last detected H5N8-positive bird. The count of carcasses is notably higher for H5N8 than H5N1, but the seasonal timing of infection is strikingly similar.

6.3.1.1 H5N8 HPAI caused high mortality in the population

Considering birds of known age only, approximately 16.8% - 22.0% of the estimated population died during the outbreak period. Birds that were found dead at the site between 2nd and 12th January 2017 inclusive were not tested for AIV (n=151), so it is unknown exactly how many of these deaths are attributable to AIV. To estimate how many birds that died during the outbreak period likely died of AIV, mortality rates were compared between 2016/2017 and the same period in previous years under the assumption that any elevated mortality rate could be most parsimoniously attributed to deaths from H5N8 infection. Both crude mortality rates and age-adjusted mortality rates (using 2016 as the reference year for the age structure) were calculated (Table 6.2). Age-adjusted mortality rates for the outbreak period in 2016/2017 were 168.2 deaths per 1000 birds (range, 157.8-178.5). Age-adjusted mortality rates for the same period in the previous 7 years ranged from 1.1 to 26.3 deaths per 1000 birds, with a median of 7.4 (Table 6.2). Hence, age-adjusted mortality in the H5N8 outbreak was on average 22.8 times the expected death rate of previous years and 6.4 times the highest mortality rate ever observed during the same period at this site (2015/2016). Based on the assumption that any increases in mortality during the outbreak period were a result of H5N8, it is expected that the majority of birds that died but were not tested would have died of AIV and not other natural causes. Based on the lowest, median and highest mortality rates observed at the site during the same period in previous years, I estimate that between 84% and 99% of birds that died during the H5N8 outbreak period would have died as a result of AIV infection, with a median estimate of 96% of birds (174 out of 182 birds).

Table 6.1: Estimated age-adjusted and crude-mortalities during outbreak period per year

Year[1]	Crude mortality	Adjusted mortality	Upper estimate	Lower estimate
2009	7.32	6.54	11.93	1.16
2010	1.25	1.08	1.71	0.46
2011	14.11	7.37	13.87	0.87
2012	10.84	8.94	10.56	7.32
2013	5.58	4.28	9.66	-1.10
2014	10.75	11.82	18.18	5.46
2015	31.38	26.31	36.64	15.99
2016	168.16	NA [2]	178.49	157.83

6.3.1.2 Mortality was skewed towards younger birds

The majority of birds that died during the H5N8 outbreak were hatch-year birds (Figure 6.3). Fifty-five percent of dead birds were <1 year old, representing 40.4% of the estimated hatch-year birds in the population at that time (Table 6.3). Increased mortality in hatch-year birds compared to the estimated population was confirmed statistically using two-sided Fisher's exact tests. Assuming that all birds were equally exposed to the virus, hatch-year birds died 8.2 times more frequently than birds of other ages (95% confidence interval of odds ratio; 5.5 - 12.3, p-value < 0.001). A generalised linear model was also performed to test the effect of age group and sex on probability of death. Hatch year birds were significantly more likely to die than

[1]Year is given as the start of the "outbreak period" (23rd December to 24th January) in each year. For example, the outbreak in December 2016 - January 2017 is referred to as "2016".

[2]Note that this value is not given, because the 2016 population is used as the reference population for the age structure for all years, and hence the adjusted mortality for this year is identical to the crude mortality estimate.

older birds, and sex did not have a significant effect. The significance of these results is retained when the dataset B population estimate is used (see Section 6.2.7). Using the alternative dataset B estimate of population size at the time of the outbreak (Section 6.2.7), I estimate that 50.8% of the hatch year age-group would have died, giving an odds ratio of 9.1.

Table 6.2: Estimated age-adjusted and crude-mortalities during outbreak period per year

Year[1]	Crude mortality	Adjusted mortality	Upper estimate	Lower estimate
2009	7.32	6.54	11.93	1.16
2010	1.25	1.08	1.71	0.46
2011	14.11	7.37	13.87	0.87
2012	10.84	8.94	10.56	7.32
2013	5.58	4.28	9.66	-1.10
2014	10.75	11.82	18.18	5.46
2015	31.38	26.31	36.64	15.99
2016	168.16	NA [2]	178.49	157.83

[1]Year is given as the start of the "outbreak period" (23rd December to 24th January) in each year. For example, the outbreak in December 2016 - January 2017 is referred to as "2016".

[2]Note that this value is not given, because the 2016 population is used as the reference population for the age structure for all years, and hence the adjusted mortality for this year is identical to the crude mortality estimate.

Table 6.3: Estimated proportion of birds that died by age

Age (years)	Estimated number of birds in the population	Number of birds that died	Proportion of the population that died
<1	250	101	40.4
1	178	12	6.7
2	138	11	8.0
3	99	7	7.1
4	60	5	8.3
5	37	5	13.5
6	39	4	10.3
7	21	1	4.8
8 or older	70	4	5.7
Unringed birds (age not recorded)	Unknown	32	Unknown

Although hatch year birds were more likely to be AIV positive than non-hatch year birds, the age-distribution of the AIV-positive tested swabs was not significantly different from the AIV-negative tested swabs (odds ratio for two-sided Fisher's exact test; 4.1, p-value 0.11). It is noted that swabs were only taken from a small number of birds (n=23), and only in the extreme temporal ends of the outbreak when few deaths were occuring. Consequently, this finding is not at odds with the above result based on total mortalities.

To rule out that the skew in mortality towards younger birds is not simply a

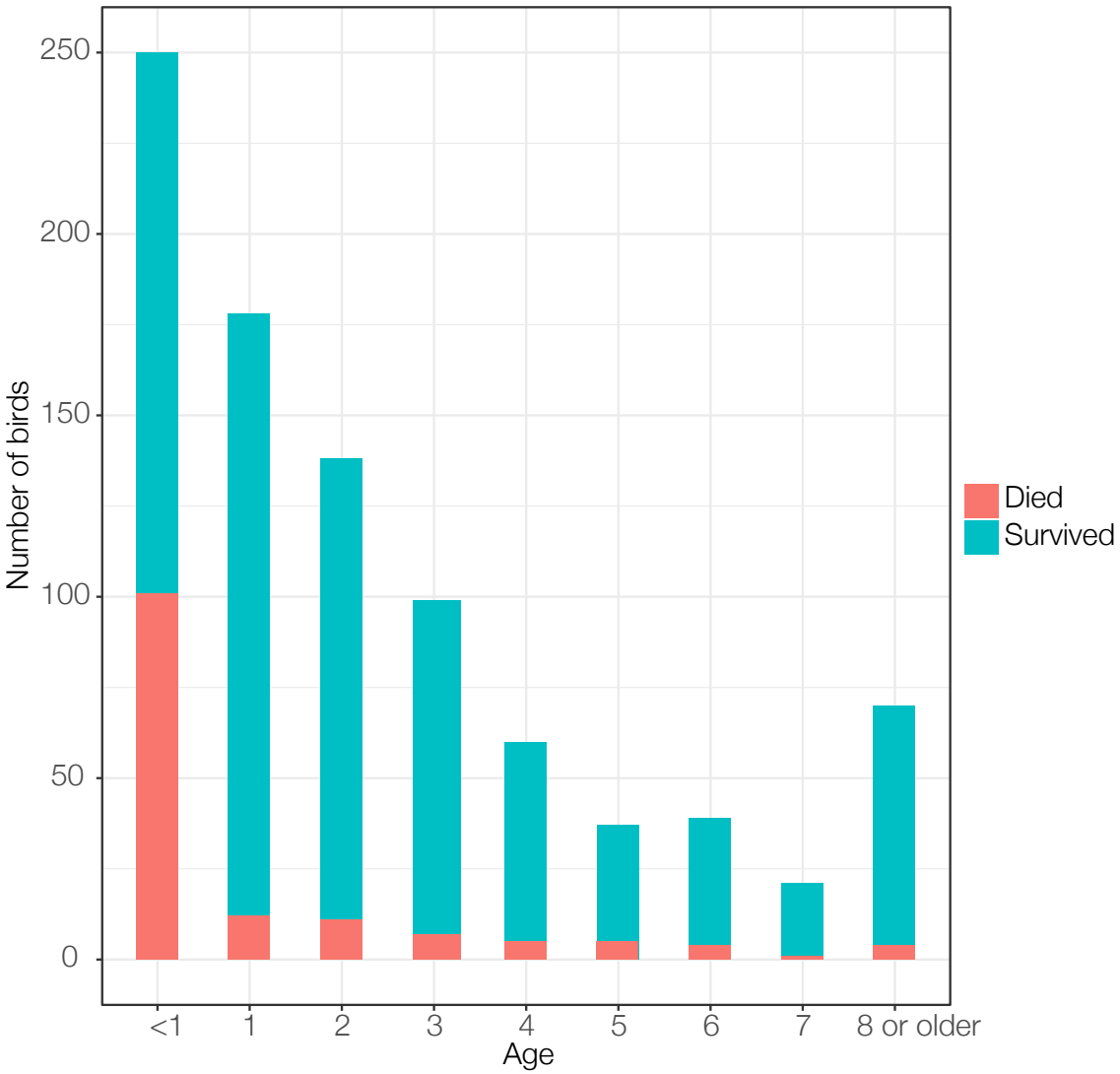

Figure 6.3: Estimated numbers of swans of each age in the population in December 2016 and number of each age that died. Birds that were estimated to be in the population but for which carcasses were not found are coloured blue, and those that are known to have died are coloured red.

standard feature of winter mortality in the population, age-specific mortality rates were compared between the 23rd December to 24th January in winter 2016/2017 to mortalities by age in the same period in previous years. Mortalities from previous years were combined to generate an average proportion of birds dying from each age-group, as individual death counts were extremely low per year (56 deaths in total recorded for this period between 2009/2010-2015/2016). The skew in the age-profile during this winter period appears distinct to the year of the H5N8 outbreak, and is not a factor of general mortality in the population during the winter (Figure 6.4).

6.3.1.3 There is no effect of weight on avian deaths amongst cygnets

On ringing days in September and October, hatch year cygnets weighed between 4.4kg and 12.2kg. At the start of the outbreak, cygnets were between 186 and 226 days old. Wilcoxon rank sum tests were conducted to determine whether birds that died in the outbreak had on average different ages or weights than birds that survived. There was no difference in average age of cygnets that died compared to cygnets that survived (Figure 6.5). This was true for analyses of raw weight at ringing (which do not take into account weight-differences caused by ringing birds up to a month apart or differences in weight caused by sex) (Figure 6.5), and for analyses based on difference in weight compared to the average expected weight for a bird of that age and sex. These results are robust using both estimation methods of which birds were believed to have been in the population in December 2016 (Section 6.2.7).

6.3.2 LPAI prevalence in the population

Samples from April 2015 to March 2016 were tested to investigate the typical prevalence of LPAI viral infection in the population. On average across the year 2015-2016, 1.3% of PCR tested samples were positive for AIV (11 out of 866) (PCR detection methods are included in Section 6.2.4). However, most PCR positive samples were collected on the same sample occasion (July 2015), such that the prevalence on this

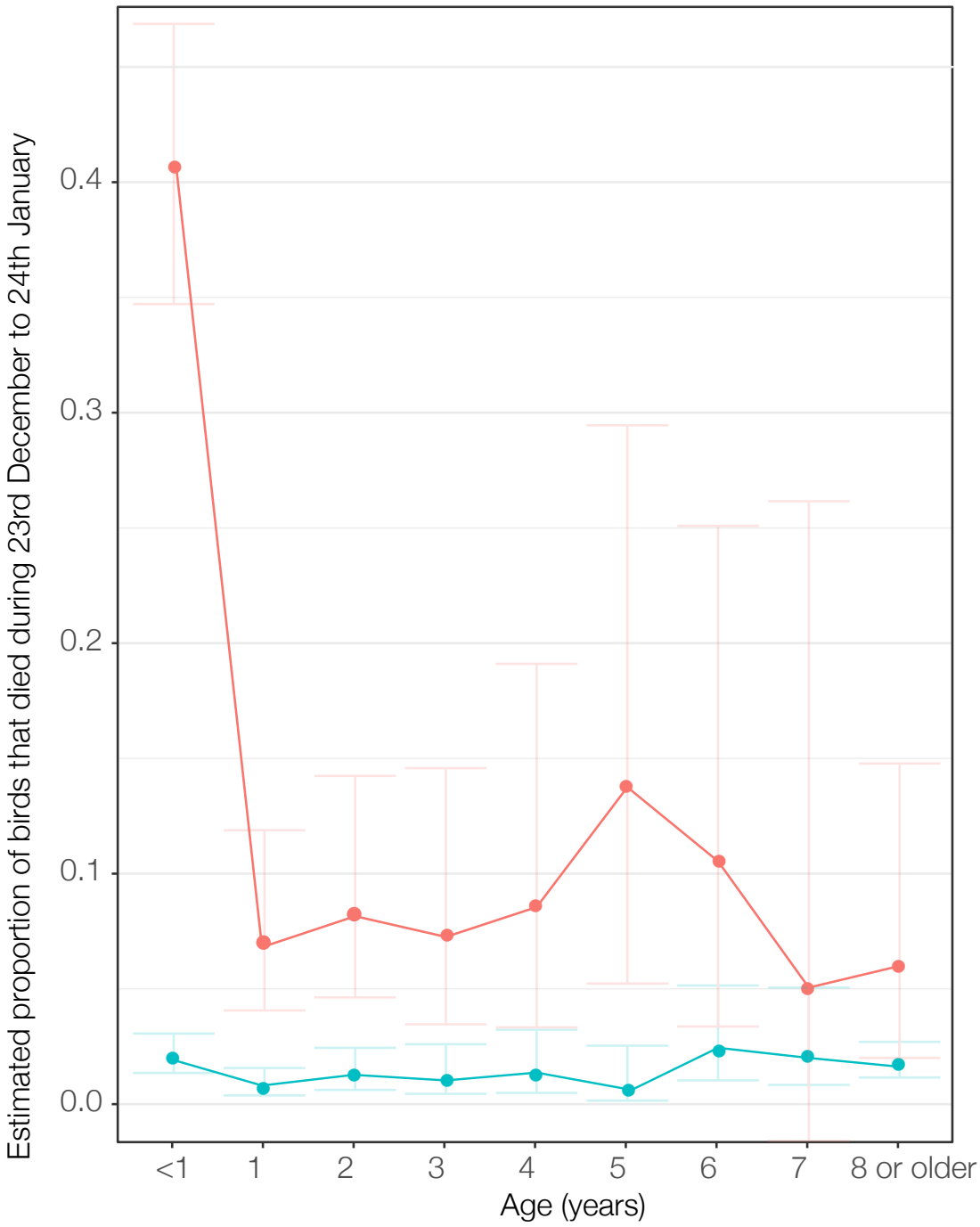

Figure 6.4: Proportion of birds of each age that died during 23rd December to 23rd January in 2016-2017 (red) and averaged deaths in the same period for previous years (blue). Wald-ajusted 95% confidence intervals are given in paler corresponding colours. Note the extreme increase in proportion of birds aged <1 that died during the period of known HPAI H5N8 circulation in 2016-2017.

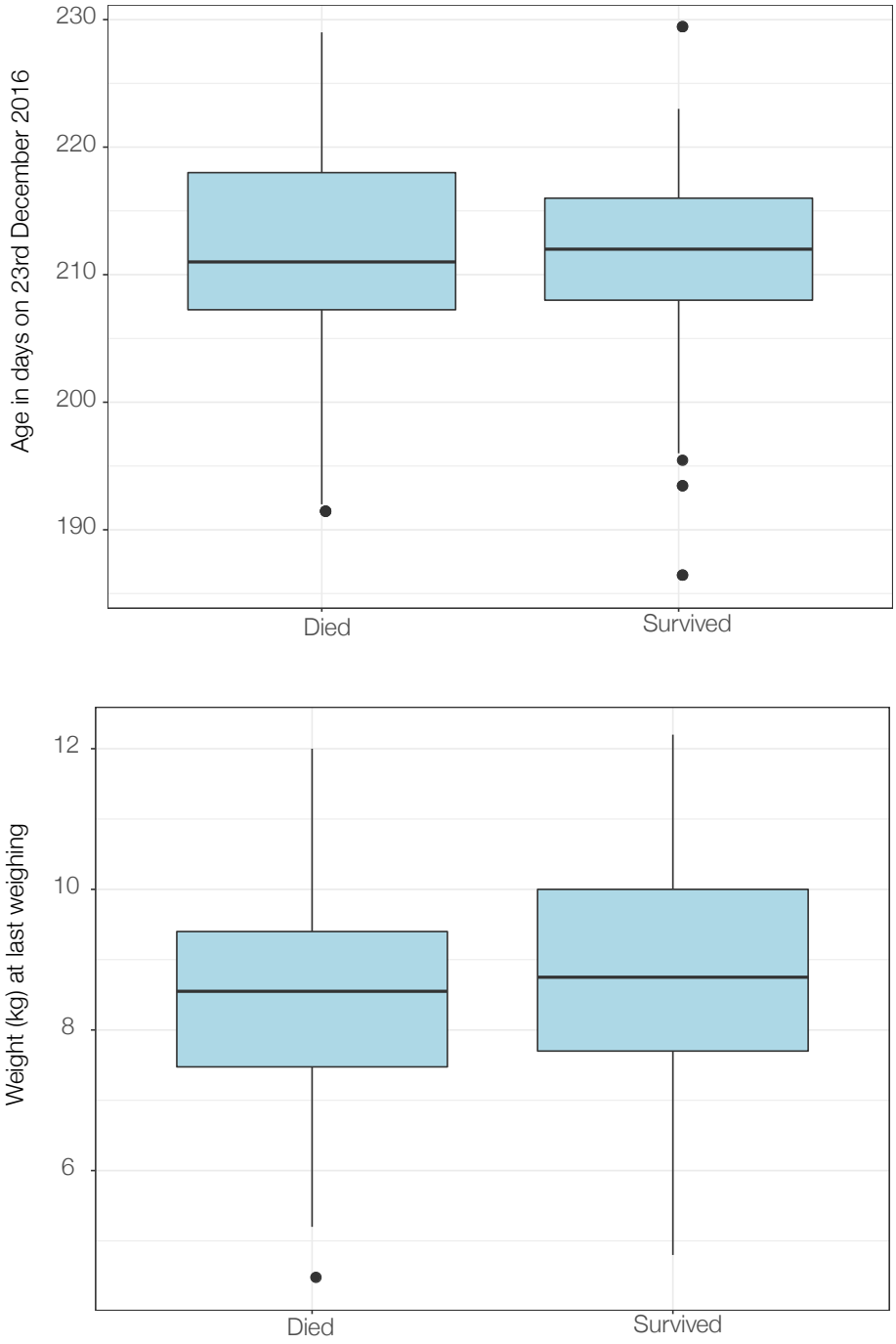

Figure 6.5: Boxplots showing no effect of exact age in days (upper) or last-known weight (lower) on cygnet survival of the outbreak in the population.

occasion was 9.9% (9 out of 91), and the prevalence in the rest of the year excluding this single occasion was therefore only 0.3%. Considering birds of known ages on this sampling occasion only, birds of 1 or 2 years of age are significantly more likely then older birds to have detectable AIV (Fisher's Exact test, $p<5x10^3$), with the mean proportion of AIV positive birds in each group being 20% (n=25) and 1.6% (n=61) respectively. No hatch year birds were sampled in July 2015, so I cannot quantify how the youngest age-group was affected by AIV on this occasion.

The effect of age on LPAI carriage is not significant when LPAI prevalence across the whole year in considered, although, as noted above, only very small numbers of birds were positive in total. Inclusion or exlusion of weak PCR positive samples (Cts 36-38) did not strongly affect either of these results, although proportions of positive birds increase substantially when weak positives were incuded, with 36% of birds aged 1 or 2 and 3.2% of birds aged over 2 being positive or weak-positives on the July sample occasion (Figure 6.7).

Although the numbers of infected birds are small, it was noticed that there was a peak in both LPAI carriage and HPAI mortality in birds that were 5 years old at the time of infection. Swans typically breed for the first time between 3 and 6 years old [60], and it was considered plausible that the high energetic requirement of breeding could have an effect on the likelihood of later AIV carriage. The effect of breeding status on HPAI mortality was explored, showing that there was no significant effect of breeding status on whether birds >2 years old died in the outbreak (Figure 6.6).

No specific effort was made to confirm the subtypes of the AIV positive samples obtained during April 2015 to March 2016. One sample from July 2015 that was sequenced along with samples included in Chapter 4 could be confirmed as H3N8 (see Section 4.2 for details of this sequencing method). I also detected the presence of H9N3 in one sample (a female hatched in 2004) in August 2016; however, metagenomics is not as sensitive as specific qPCR and therefore I cannot estimate AIV prevalence from the metagenomic data included in Chapter 4. Specific qPCR tests for different subtypes and the presence and absence of AIV should be conducted

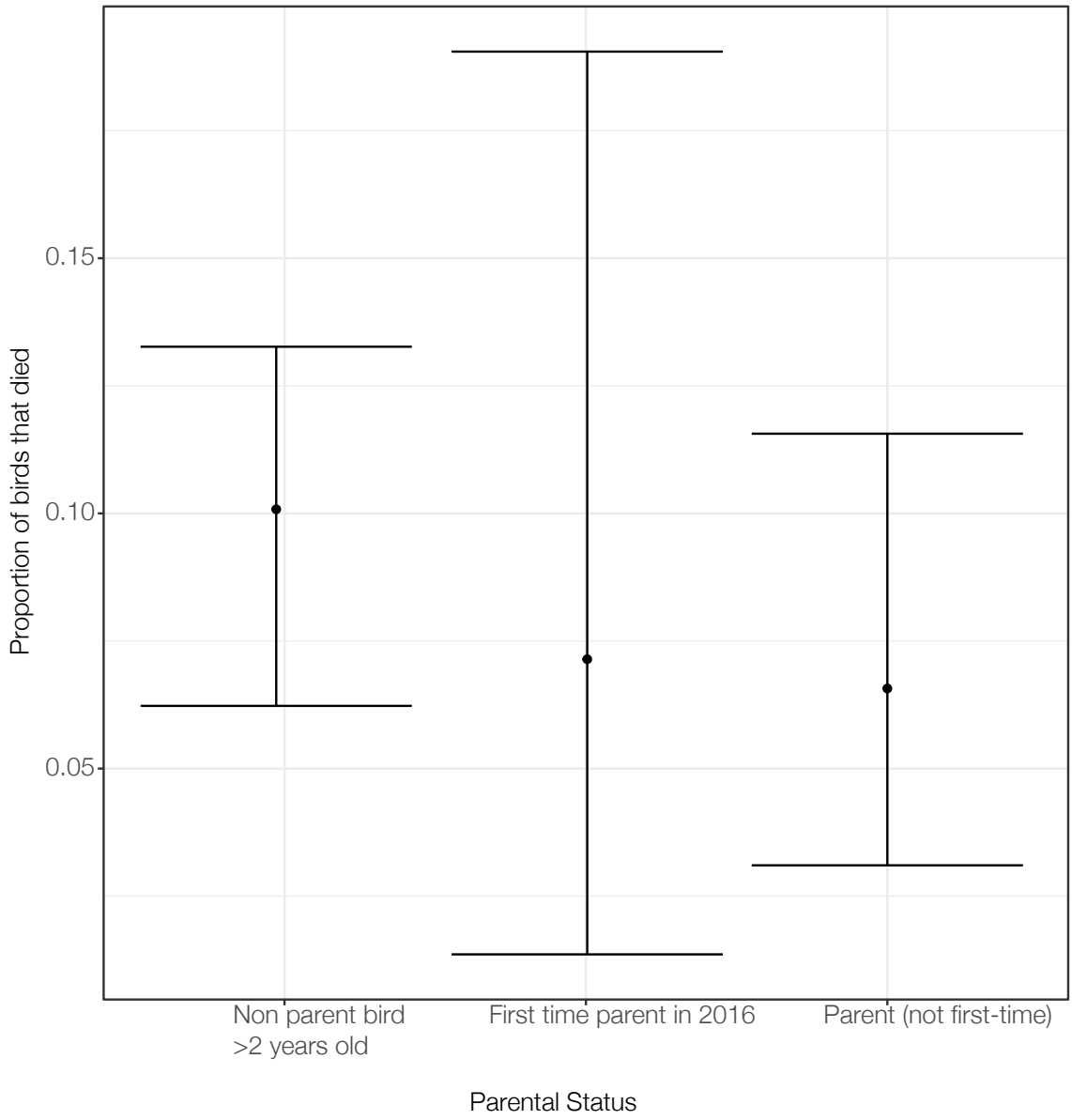

Figure 6.6: Proportion of birds that died of HPAI AIV and their parental status- classed as bird that had never been parents, first-time parents in 2016, and had been a parent before 2016. 95% confidence intervals were calculated using the adjusted Wald method. Only birds > 2 years old were considered, as few birds are capable of breeding before their third year.

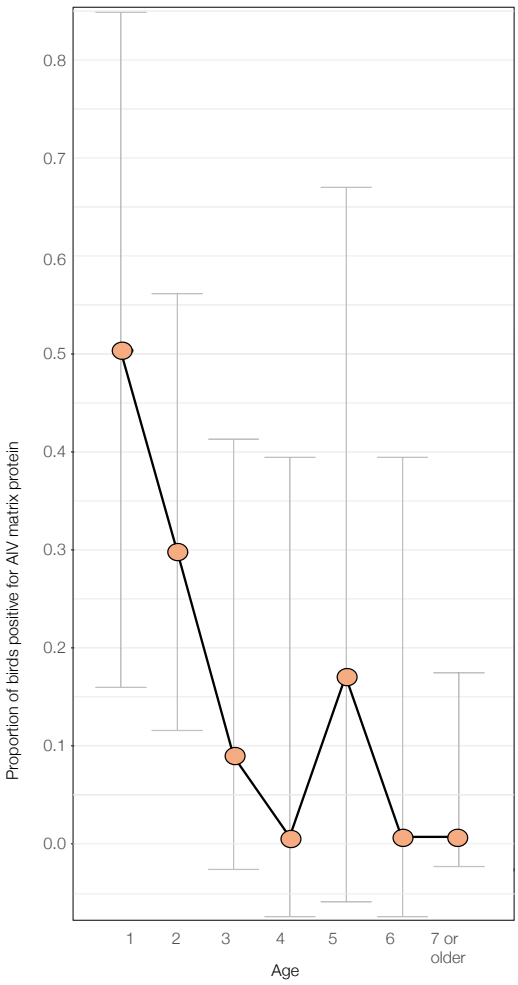

Figure 6.7: Proportion of birds of different ages sampled on 21st July 2015 that tested PCR positive for LPAI influenza. Wald-adjusted 95% confidence intervals are given in grey. Here, PCR positives include all birds with qPCR Ct values < 38. 91 birds were tested, of which none were hatch-year birds. Note that this trend is only apparent on this sample occasion, when a high proportion of birds were AIV positive, and is not apparent if all data from all other sampling points across the year are included, when very few birds were AIV positive (Section 6.3.2)

on the samples featured in Chapter 4 to quantify LPAI prevalence and subtypes in late 2016.

6.3.3 Molecular genetics and phylogenetics of HPAI H5N8

Cts of samples that were PCR positive for HPAI H5N8 were compared by swab type (oropharangeal or cloacal) for birds from which both swabs were PCR positive. Oropharangeal swabs have significantly lower Cts than cloacal swabs, with mean Cts of 27.4 and 30.5, respectively ($p = 0.039$). Dead hatch-year birds are not statistically more likely to have a significantly higher or lower Ct than dead older birds for either swab type.

Sequencing was conducted on samples from twelve AIV positive birds, with the lowest Ct sample chosen for each bird, regardless of sample-type. These samples cover both ends of the outbreak period, including six birds that were found positive at Abbotsbury up to and including 1st January 2017, and six birds that were found positive from 12th to 24th January 2017. Near-whole or partial sequences were generated for the coding regions of all segments for all samples (sequenced proportion of each coding region for each sample are provided in Table E.4).

Regression of root-to-tip genetic distance against sampling date in TempEst [359] suggested that there was sufficient temporal signal in a dataset of HA and NA sequences from 2010-2017 to estimate trees using molecular clock models (Figure 6.8). R^2 values for both segments were 0.91.

Molecular clock analyses of HA and NA evolution were conducted using BEAST in order to estimate the introduction date of the virus into the population (Figures 6.9, 6.10). Posterior density support distributions for the most recent common ancestor of the Abbotsbury outbreak clade were very similar for estimates produced from both the HA and NA segments (Figure E.1). Considering support from both segments together, the mean and median estimated dates of the most recent common ancestor of the outbreak sequences was November 19th 2017 and November 21st

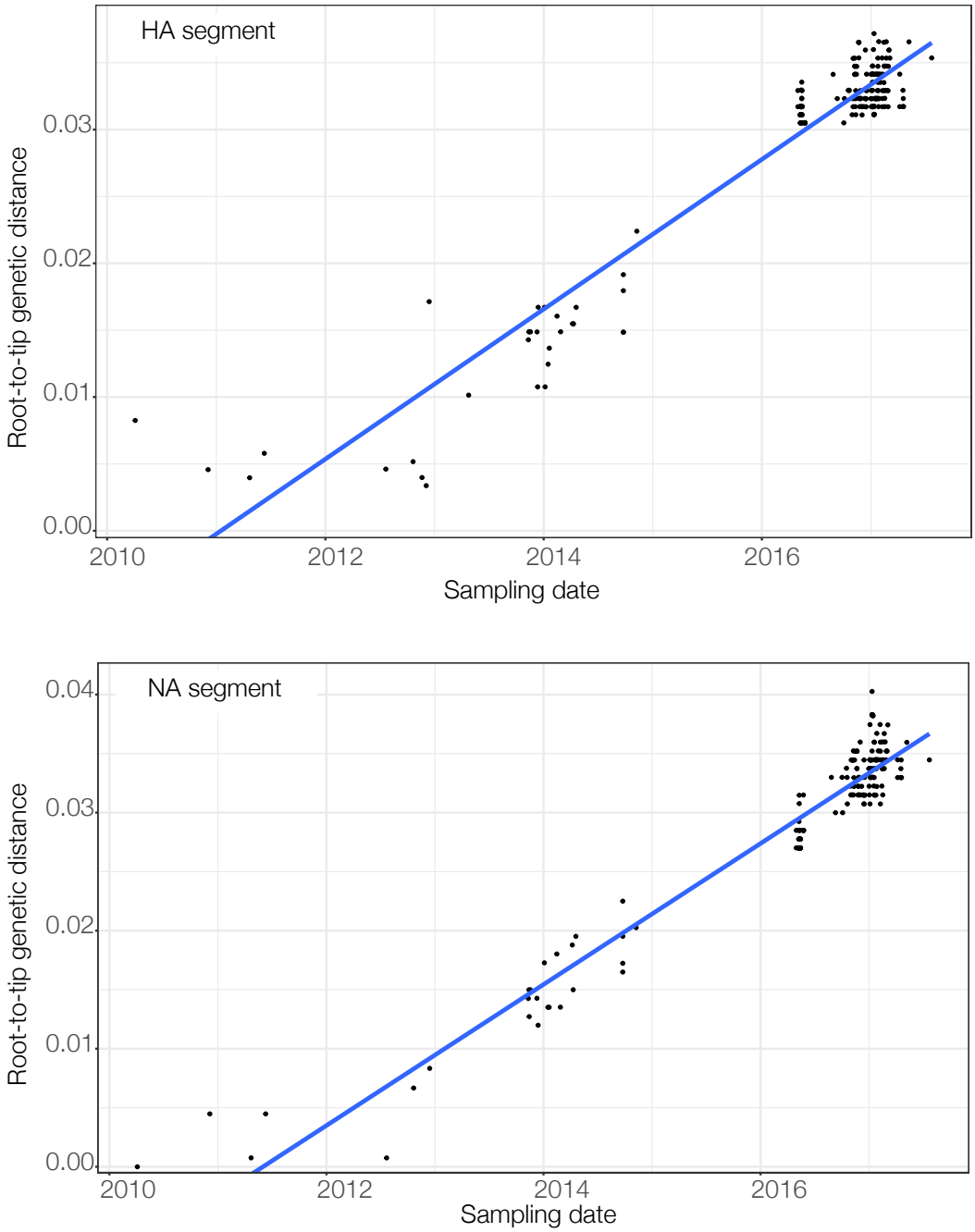

Figure 6.8: Root-to-tip regression of genetic distance against sampling date for HA and NA.

2017, with 95% HPD interval between October 18th 2017 and December 16th 2017. Virus genomes from Abbotsbury did not cluster with other samples from the UK in trees estimated from either the HA or NA segment. Viruses sampled in Russia and India were immediated basal to the Abbotsbury outbreak clade. Posterior support for the specific placement of the Abbotsbury sequence clade amongst neighbouring sequencing was poor, with posterior support values < 0.7.

Maximum likelihood phylogenetic trees estimated for each internal segment showed that all samples from Abbotsbury formed a single, well-supported monophyletic clade, as would be consistent with a single introduction to the site. In concordance with the molecular clock trees estimated from the HA and NA segments, virus genomes from Abbotsbury did not cluster with other samples from the UK, and instead more commonly formed clades with viruses sampled in Russia, Eastern Europe (for example, Hungary) and northwest China. (Figures 6.11, 6.12, 6.13, 6.14, 6.15, 6.16)

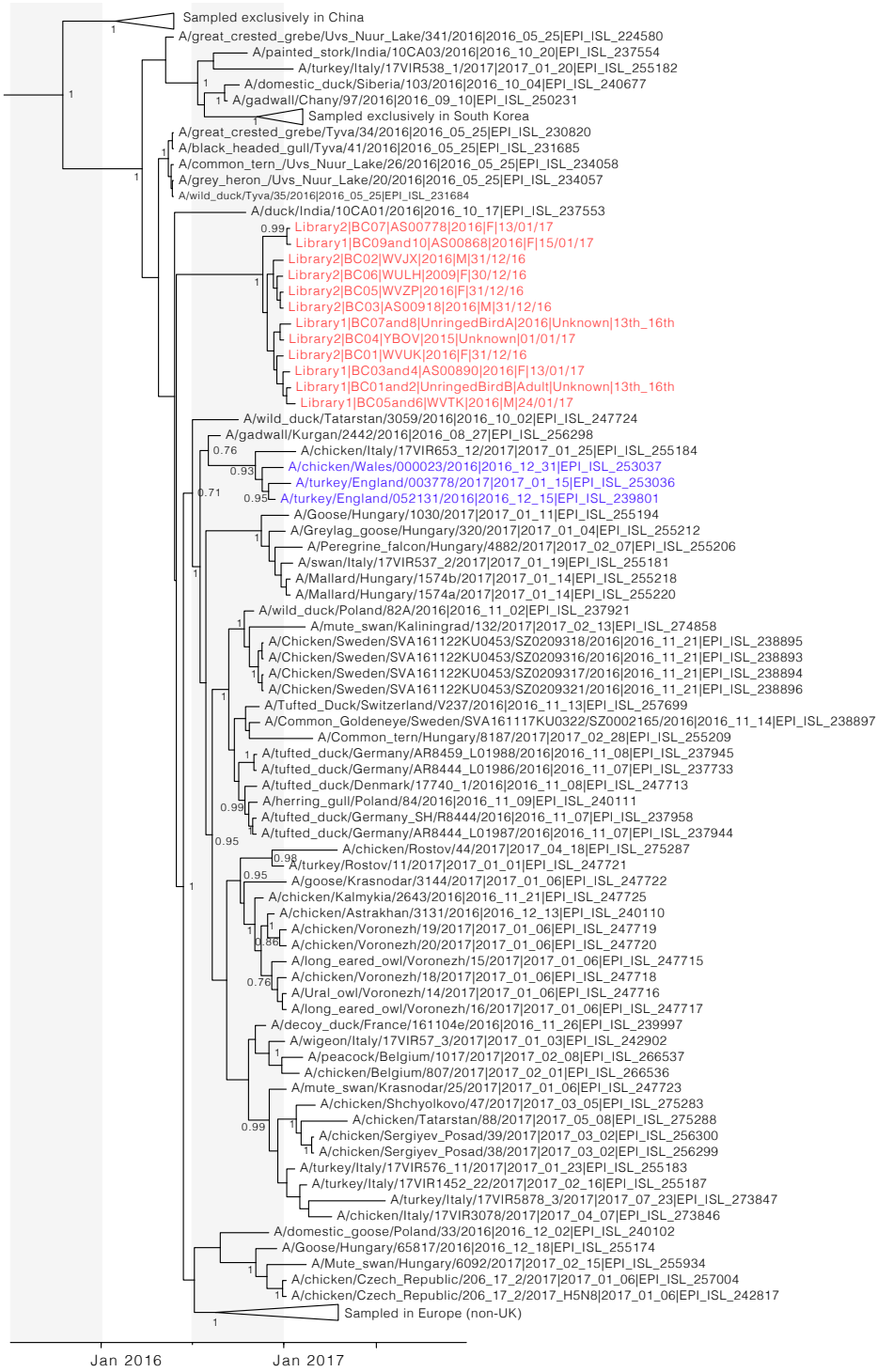

Figure 6.9: Maximum clade credibility tree estimated for HA segment. Posterior support is given at nodes where >0.7. Bars represent highest posterior density intervals for node times. Sequences from the outbreak at Abbotsbury are in red, and sequences from elsewhere in the UK are in blue. Infections sampled prior to January 2016 were included during phylogenetic estimation to provide temporal signal, but are not shown here for clarity of the figure. Several major clades are collapsed for clarity of the figure.

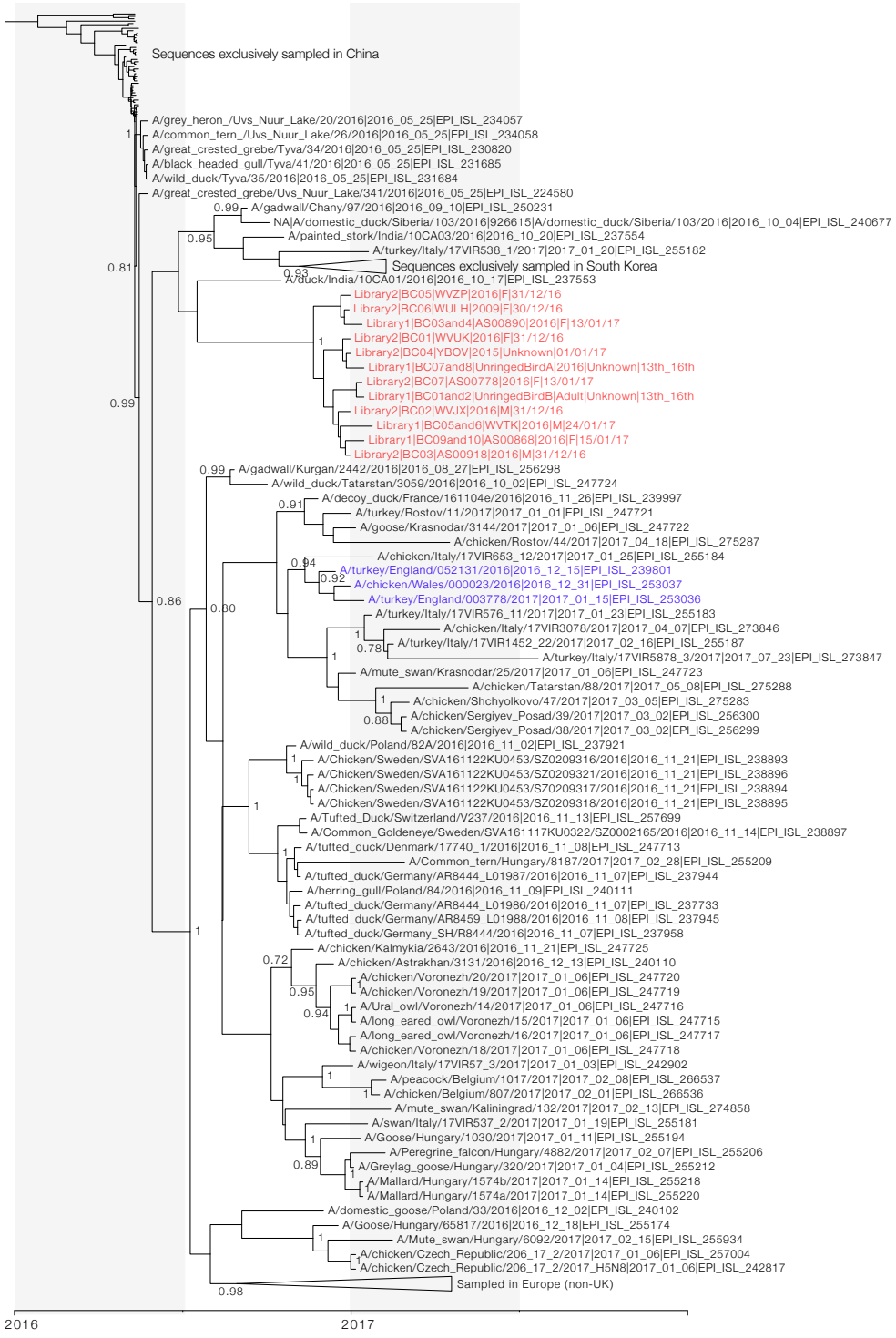

Figure 6.10: Maximum clade credibility tree estimated for NA segment. Posterior support is given at nodes where >0.7. Bars represent highest posterior density intervals for node times. Sequences from the outbreak at Abbotsbury are in red, and sequences from elsewhere in the UK are in blue. Infections sampled prior to January 2016 were included during phylogenetic estimation to provide temporal signal, but are not shown here for clarity of the figure. Several major clades are collapsed for clarity of the figure.

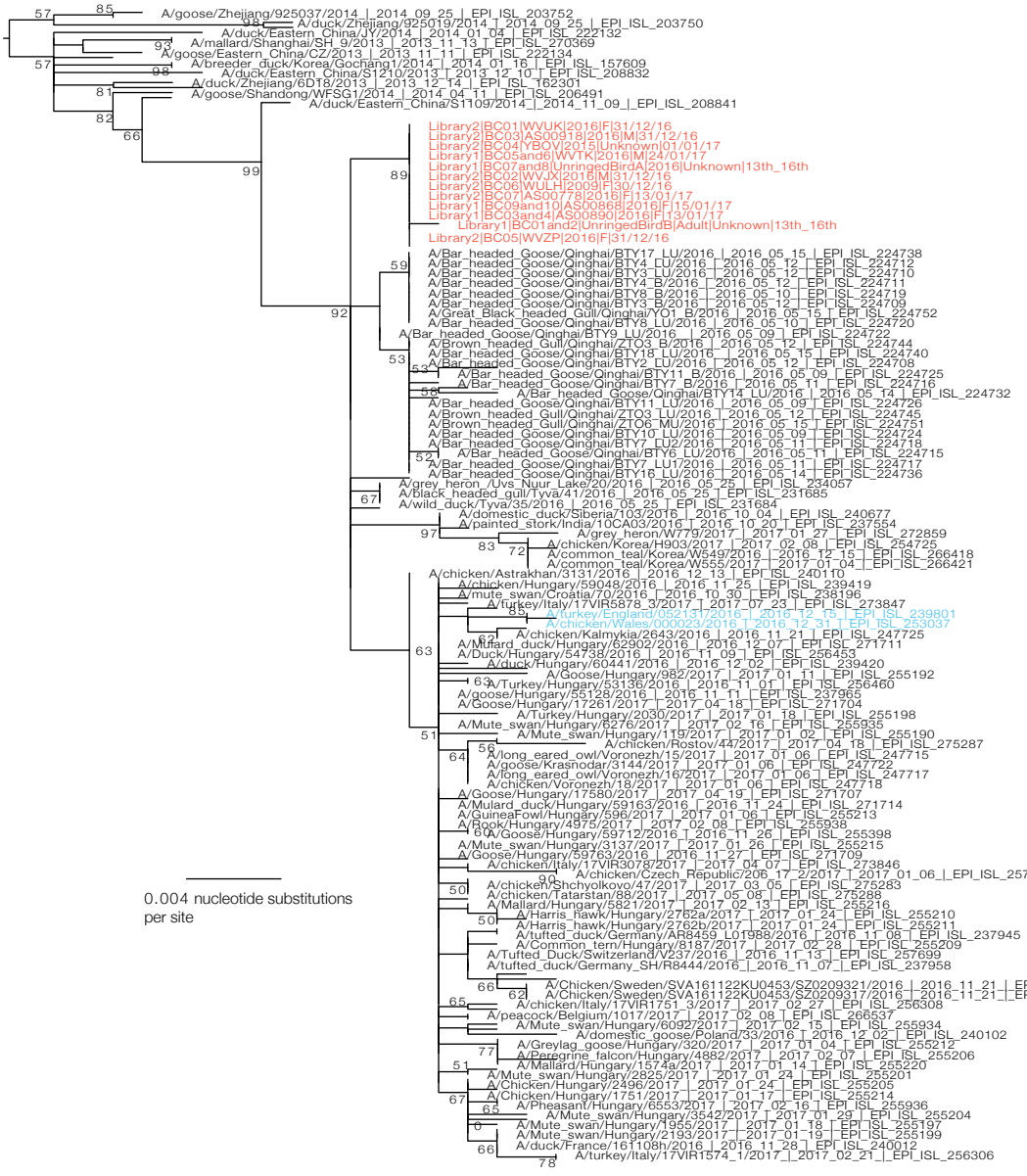

Figure 6.11: Maximum likelihood tree of NS segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model TPM1uf+I. Every alternate tip-label has been removed for clarity of illustration, except in the Abbotsbury clade which has been expanded to accommodate all tip labels. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighour-joining trees. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

Figure 6.12: Maximum likelihood tree of NP segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model HKY+I. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighour-joining trees. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

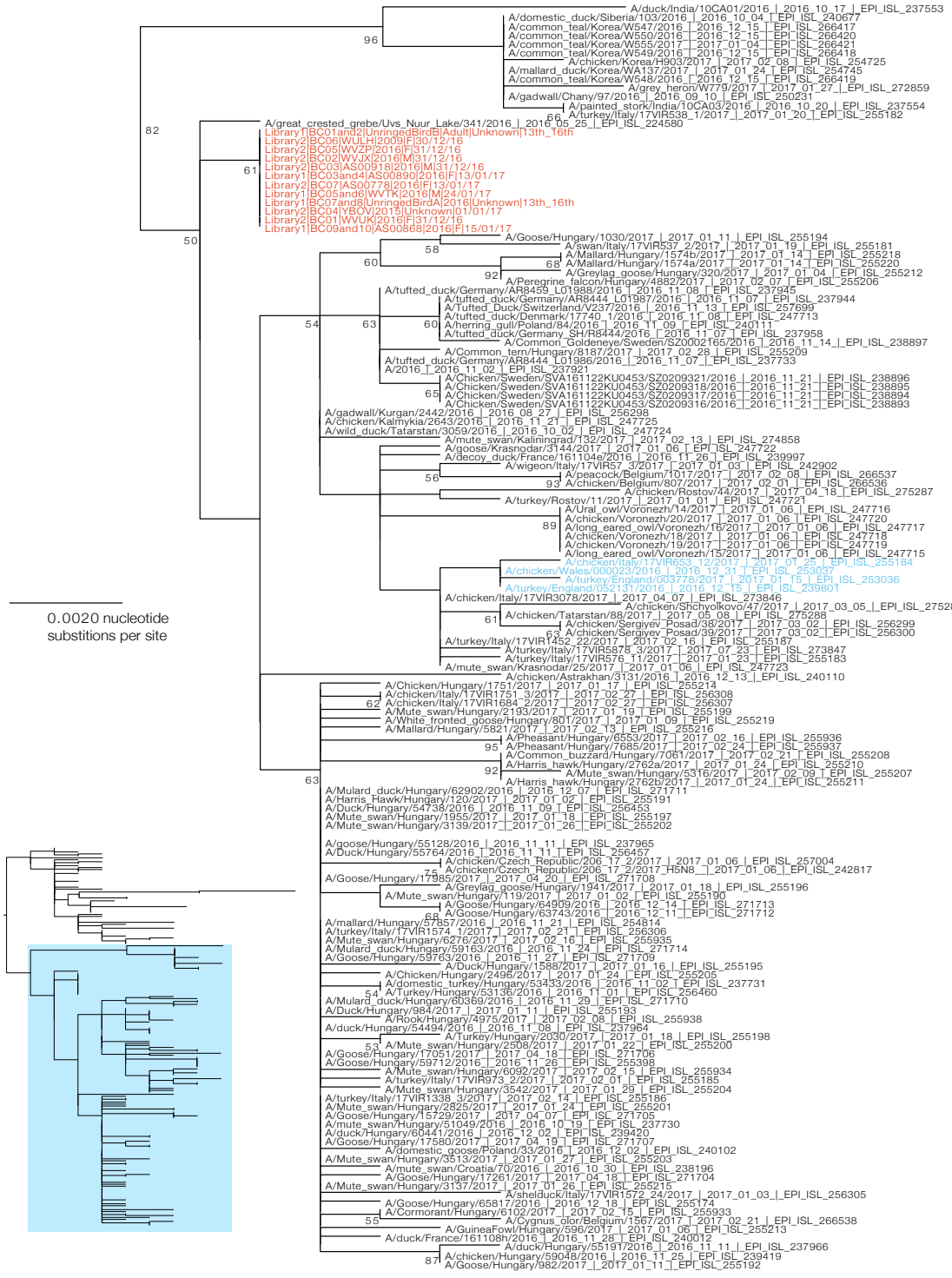

Figure 6.13: Maximum likelihood tree of MP segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model K80+G. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighbour-joining trees. The lower tree is an expansion of the well-supported clade highlighted in blue in the upper tree. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

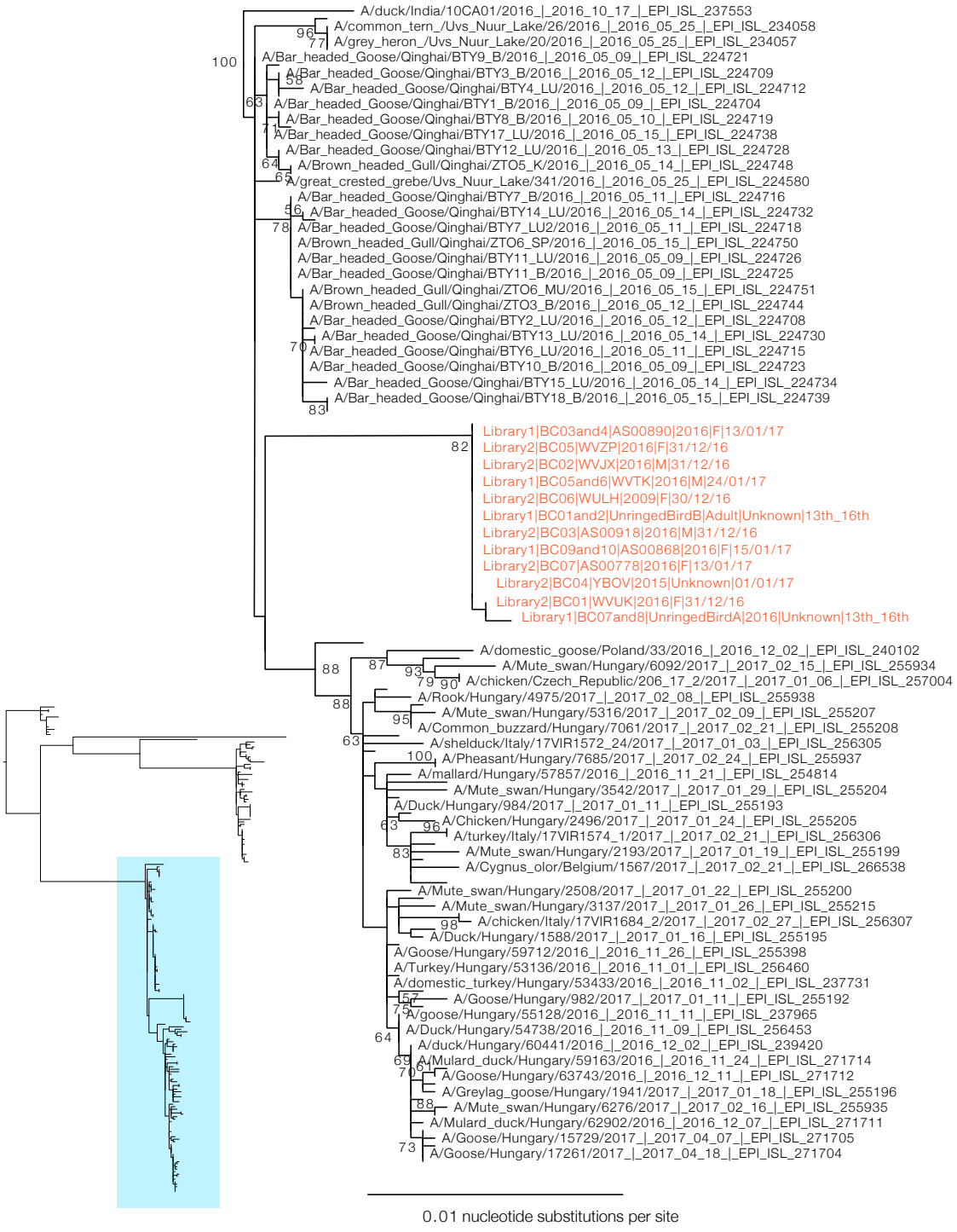

Figure 6.14: Maximum likelihood tree of PA segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model TVM+G. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighour-joining trees. The main tree is an expansion of the well-supported clade highlighted in blue in the smaller tree. Every alternate tip-label has been removed for clarity of illustration, except in the Abbotsbury clade which has been expanded to accommodate all tip labels. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

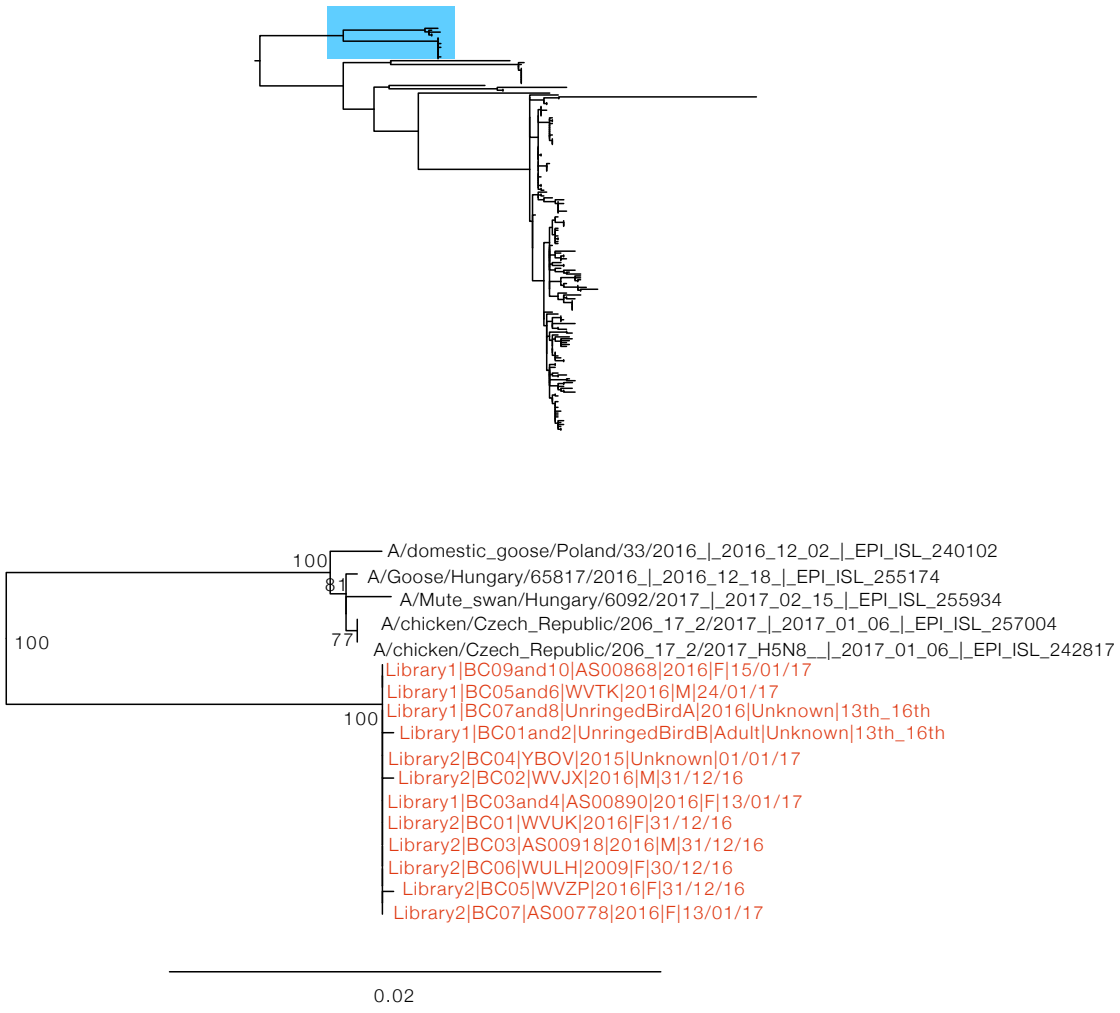

Figure 6.15: Maximum likelihood tree of PB1 segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model GTR+G. The lower tree is an expansion of the well-supported clade highlighted in blue in the upper tree. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighour-joining trees. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

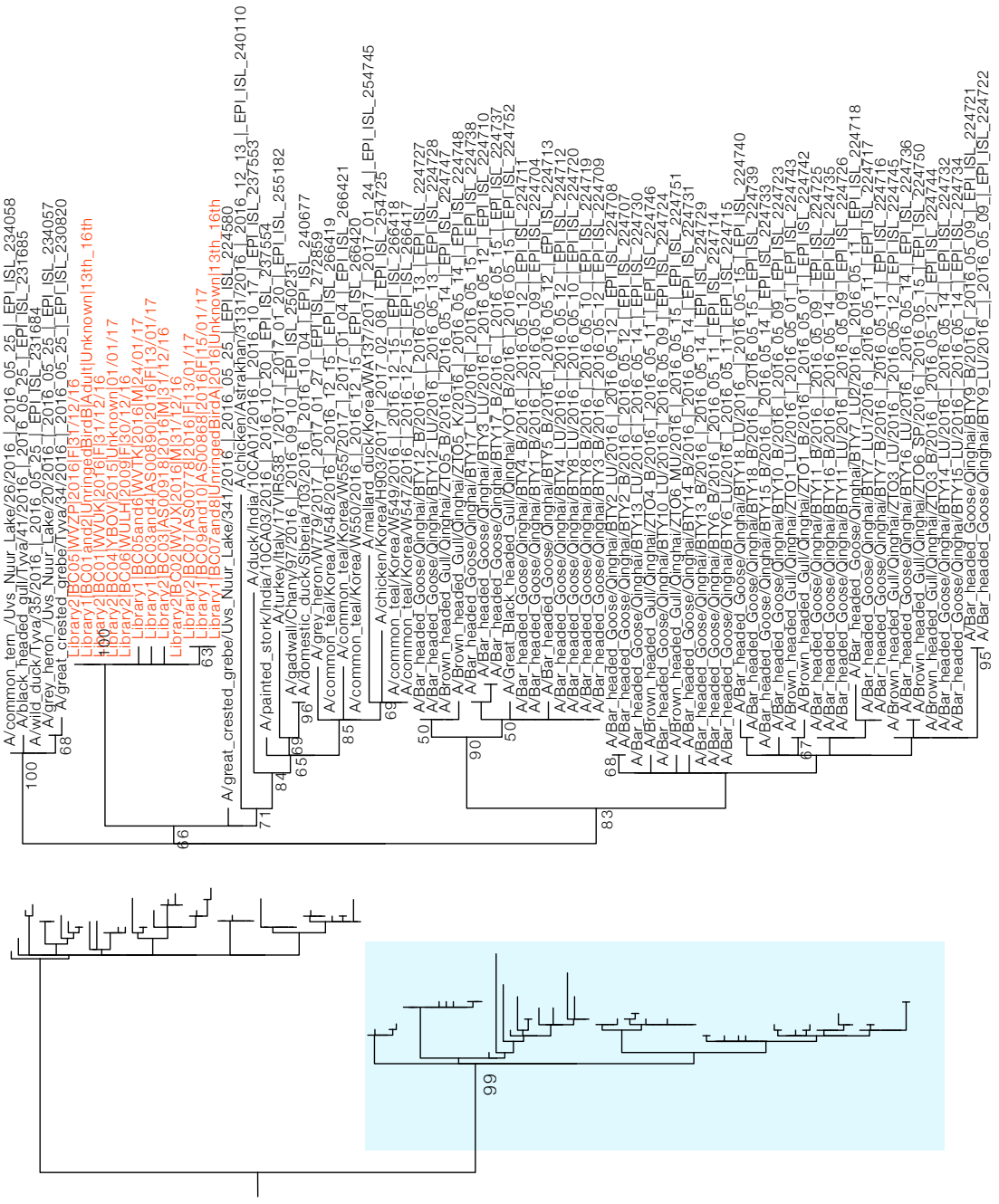

Figure 6.16: Maximum likelihood tree of PB2 segment of the H5N8 clade containing Abbotsbury sequences. The tree was constructed using substitution model TMP1uf+G. The right hand tree is an expansion of the well-supported clade highlighted in blue in the left hand tree. The tree is outgroup rooted based on the relationship of this clade to other major clades (not shown here, and presumed to represent reassortant internal genes) observed during estimation of neighour-joining trees. Node labels show bootstrap score > 50. Sequences from Abbotsbury are coloured red.

6.4 Discussion

Although only a small proportion of carcasses was tested for AIV, the mortality at the site during the confirmed period of H5N8 circulation was extremely high in comparison to previous years and had no other obvious etiology. It is therefore extremely likely that most deaths observed on the site during this period were the result of infection with avian influenza H5N8. During the period when no testing took place, many birds on the site showed symptoms that have been previously associated with highly pathogenic avian influenza, including general signs of illness (severe lack of co-ordination and lethargy) and more specific symptoms such as compulsive swimming in circles [209]; a behaviour not seen in this swan colony since the H5N1 outbreak in 2007/2008 (personal communication with Abbotsbury Swannery swanherds).

There are certain striking similarities between this HPAI H5N8 outbreak in 2016/2017 and HPAI H5N1 in 2007/2008 [352]. The estimated epidemiological weeks of introduction of both viruses are almost identical based on either the first detection of positive birds (same epidemiological week), or based on phylogenetic reconstruction of the most recent common ancestor (for H5N8 and H5N1, the most recent common ancestors of all outbreaks strains were both estimated to occur in 3rd week of November of the respective years, with credible interval between mid October to late or mid December [352]). During the 2007/2008 H5N1 outbreak several population-level trends were observed but were not significant due to low numbers of H5N1-attributable deaths, including higher detection of HPAI in swans than in any other species and higher mortalities in younger birds [352]. The mortality observed in the Abbotsbury swan population during the H5N8 outbreak was substantially higher than that seen during the 2007/2008 outbreak of H5N1, providing greater statistical power to confirm these findings.

Despite thousands of birds of many different species being present at the site in December and January, the only birds on the Fleet in which AIV was detected were mute swans. An adult coot, black-headed gull and adult Canada goose found

dead on the site were tested for AIV, and all three had no detectable AIV. Several staff members performed rigorous patrols of the site every day during the outbreak and a fence surrounding the land-boundary of the main Abbotsbury Swannery site prevented carcasses being removed or entirely destroyed by terrestrial scavengers. Predation outside of the fenced Swannery site around the Fleet Lagoon probably occured more frequently. Whilst carcasses from other species might be more difficult to detect as a result of smaller size, greater predation of carcasses or more muted coloration, it is unlikely that significant mortality of the same degree in any other of the other common species at the site would have been missed. Across Europe, detection events of H5N8 in swans represented almost half of detection events for which the specific species was recorded [10]. A high proportion (6-7%) of all swans tested for HPAI AIV were positive during the HPAI H5N1 outbreaks of 2006, and swans were considered to be highly susceptible to this outbreak [47, 167, 201, 437]. Differences in infectivity of recent clade 2.3.4.4 H5 viruses have been noted between several different species [321, 322, 94, 23]. Although I cannot rule out some bias in detection, the evidence from Abbotsbury suggests that the swans there (including adult swans) were more susceptible to HPAI AIV H5N8 than other birds and were also susceptible to the HPAI H5N1 outbreaks of 2007/2008, and that swans might therefore make a good sentinel species for HPAI AIV.

Younger birds were observed to die more than older birds in this outbreak. Whilst this result is based on estimation of the birds that were probably present in the population at the time of the outbreak, the result is robust to two different methods of estimating population size. It has been previously suggested that this trend could be because the youngest swans lack immune protection, as a result of little or no previous exposure to other low-pathogenic strains of avian influenza [352, 169]. Here, the population-level prevalence of LPAI virus infection in this population was calculated to be on average 1.2% (30 days of sampling conducted between April 2015 and March 2016). This is within the range of prevalences that are generally reported for healthy, wild mute swans (typically <3%, [21, 236, 296, 314, 394, 399, 425]). Assuming that

such a low prevalence is typical, it is possible that many of the youngest age group birds might have experienced few or no infections with avian influenza virus prior to the HPAI H5N8 outbreak in 2016/2017, resulting in a lack of immune protection as a result of previous exposure. However, I cannot rule out that age-related differences in mortality in this population were not caused or compounded by immunological immaturity. The main lymphoid organs in birds are the bursa of Fabricius and the thymus. In chickens and pigeons, structural changes in these organs can occur up 6 months of age [66]. The youngest cohort of swans studied here were 6-8 months old when HPAI H5N8 was first detected at the site. Whilst immune-maturation in swans is not well-studied, swans have a relatively long lifespan and a slower life-history compared to chickens and pigeons, so it might be expected that immunological maturation would proceed differently (for example, first breeding typically occurs at age 3-5 [59], compared to approximately age 1 for pigeons [168]). It is hence plausible that the youngest cohort of swans here was still immunologically immature compared to the older birds.

There is significant evidence from active surveillance and experimental studies that many species of birds that are infected with clade 2.3.4.4 HPAI viruses can be clinically mildly symptomatic or asymptomatic, including Pekin ducks, wigeons, Canada geese and mallards [191, 321]. I cannot estimate the number of asymptomatic, viral-shedding birds that might have been present at the site during the main period of the outbreak. Whilst at least one swan in the population appears to have survived a low-viremia HPAI H5N8 infection, live swans were tested only at the very end of the outbreak and I hence suspect that the number of asymptomatically infected birds would have been higher during the main peak of the outbreak.

Based on the short temporal span of the outbreak and phylogenetic clustering of all the sequences from Abbotsbury and phylogenetic distance to other samples from the UK, this outbreak represents one of at least two different introductions to the UK in late 2016. Interestingly, the genomic sequences from sampled viruses at Abbotsbury typically do not cluster with viruses sampled in Western Europe,

including other viruses sampled in the UK, and instead appear most closely related to viruses sampled in Russia, eastern Europe and northeastern China. The Fleet Lagoon (on which Abbotsbury Swannery is located) is an important winter migratory site for thousands of wild birds (Table E.1), including Brent geese (*Branta bernicla*) and Wigeon (*Anas penelope*) that migrate to the south coast of the UK for overwintering from Russia and northern Europe [420]. Migration of birds to the Fleet Lagoon from eastern Europe and Russia appears a plausible explanation for the clustering of these viruses with viruses from those locations. The estimated time of the most recent common ancestor of the Abbotsbury clade is mid November 2016, and the credible interval span until December 16th 2016; one week prior to the first detection of a dead birds with HPAI H5N8 at Abbotsbury on 23rd December 2016. The most plausible explanation for this appararent short period of cryptic circulation prior to detection at the site is that the virus was circulating in other species of bird at Abbotsbury prior to its incursion in the Abbotsbury flock, as has been previously suggested [352]. These findings should be further explored with more detailed phylogeographic study, as has been done for the HA segment of earlier clade 2.3.4.4 H5Nx samples [439].

The placement of the Abbotsbury sequences outside of the main clade of European sequences is unexpected, but is supported by separate phylogenetic estimations from all segments (including both maximum likelihood estimation (internal gene segments) and Bayesian estimation (antigenic protein segments)). The sequencing methods used here are relatively new. An avian influenza virus has been previously sequenced using MinION technology, and comparisons with data from Illumina suggested in that study that AIV consensuses produced through MinION sequencing should be accurate [467]. I chose to use a multiplex PCR method for amplification because, unlike standard methods of amplification of AIV using conserved segment ends (e.g., [171]), this method does not require full length RNA template and therefore should be more sensitive, and furthermore the mehtods produces dsDNA fragments of near-identical lengths that, in theory, should result in more even coverage across the eight segments during MinION sequencing than would be typical following a univeral priming

method. Here, I sequenced a short amplicon of the PB1 segment using Sanger sequencing that appeared to confirm that MinION sequencing was not inducing errors in my data. Although the multiplex method used here has been applied successfully to viruses including Zika virus [126], Yellow fever virus, and other flaviviruses, it has not yet been validated for AIV or other segmented viruses. Whilst I do not believe that the sequences included here are wrong, work validating the application of MinION sequencing and this multiplex method to segmented, rapidly evolving viruses such as AIV is a clear priority, especially because of the important possible applications of sequencing outbreaks in real-time at infected sites.

Here, I have demonstrated that younger swans appear to be significantly more susceptible to death following HPAI H5N8 infection than older swans, a pattern that was hinted at but not confirmed during the H5N1 outbreak of 2007/2008. Within the youngest age-group, there is no obvious effect of sex, last-known weight or exact age in days on the probability of death. I note that the swans appeared to be more heavily affected than other commonly occurring species at this site; a pattern that follows general reporting of H5N8-associated mortalities across Europe [10]. Mute swans are typically non-migratory, strongly territorial and not gregarious; behaviour that would presumably result in swans being poor conveyors of the virus. The apparent differences in species' susceptibility raises the interesting question of whether H5N8 might be specifically adapted to cause mild infection in other species that might represent its primary reservoir, and therefore the high virulence observed in swans is simply a blind consequence of specific viral adaptation to a different, primary host species. I again highlight the importance of detailed reporting of avian age and species during HPAI outbreaks. The fundamental biological processes behind the patterns that I observe here must be explored in greater detail by studies combining ornithological, virological and immunological expertise, if we are to fully understand the ecology of highly pathogenic avian influenza in the wild.

Chapter 7

Discussion and Conclusions

In this discussion, I highlight the key results of each chapter of this thesis and discuss links between the chapters. I also consider limitations or implications of my research that have not been discussed in previous sections, and, where appropriate, suggest some specific avenues for future research that would help to extend the findings presented here. I conclude with a more general discussion of possible future directions for the field as a whole.

7.1 Key Results, Implications and Future Direction of each Chapter

7.1.1 Chapter 2: discovery of a novel polyomavirus

Chapter 2 describes the detection and genomic characterisation of a new polyomavirus from the species *Meles meles polyomavirus 1*, in a European badger (*Meles meles*). I performed phylogenetic analysis of all polyomavirus genomes that were available at the time of publication and explored the evolution of the *Polyomaviridae*. Whilst there is significant association between host and viral phylogenies, which strongly suggests that host-viral co-divergence has played a role in the evolution of the *Polyomaviridae*, I demonstrated that this host-viral co-divergence has been interrupted

by other evolutionary processes that may include recombination and/or rate heterogeneity among different species.

Since publication of Chapter 2, *Meles meles polyomavirus 1* has been incorporated into the curated list of the accepted species by the International Committee on Taxonomy of Viruses [346]. Discovery of divergent new polyomavirus species in a guitarfish and a seabass, and detection of polyomavirus fragments in arthropods and reptiles suggest that the family is probably significantly more diverse and has a far longer evolutionary history than is currently known [50, 101, 335]. Particularly for non-human viruses, there remain significant gaps in our knowledge of the fundamental biology and ecology of these viruses, including transmission routes, their potential for cross-species transmission and their ability to cause disease.

7.1.2 Chapter 3: early spread of HPAI H5N8 in Asia

Chapter 3 considered the spread of the AIV HPAI H5N8 (H5 clade 2.3.4.4), later described in Chapter 6 at a local scale. In Chapter 3 I investigated the transmission and persistence of HPAI H5N8 during its early circulation in the Republic of Korea, the second country to report the strain. I suggested that the virus was most likely introduced and then reintroduced to the country by the migration of wild waterfowl. I suggested that long-term persistence of the virus in the country was primarily mediated by cryptic transmission within domestic ducks.

A major conclusion of the work of Chapter 3 is that current methods of phylogeographic estimation can be misleading if interpreted naïvely. During discrete geographic trait estimation, geographic locations along branches are constrained to be one of the discrete geographic locations attributed to each sequence. It is therefore challenging to accurately reconstruct the exit of a virus to an unsampled area and subsequent re-entry of that virus using discrete geographic phylogeography. Continuous phylogeographic models, in which viral movement is modelled according to co-ordinates such as longitude/latitude using random walk processes, may be better

able to reconstruct movement in unsampled areas under certain conditions. However, even continous phylogeographic models can struggle to reconstruct viral movement that takes take place outside of the geographic area that is bounded by the sampled locations. Because of this, any viral movement associated with host migration between sampled and unsampled areas may be missed, as was demonstrated in Chapter 3. Naïve interpretation of phylogeographic reconstruction without an appreciation for the limits of the utilised phylogeographic modelling approach, and hypotheses regarding important ecological niches of the viral host (including patterns of host migration or seasonal changes in naive host abundance), should be avoided.

Sparse sampling in certain geographic areas has the potential to cause innaccurate estimation of the geographic movement of viruses, including when BSSVS is used to infer the most important routes of viral migration. In circumstances where some geographic locations are associated with only a few sequences, phylogeographic reconstruction using BSSVS can overinflate the importance of links between sparsely sampled locations and other locations. Geographic migration corridors between these locations may appear to be highly supported by Bayes factors during BSSVS because they are statistically necessary to link sampled locations, whilst at the same time being relatively unimportant in the overall pattern of viral transmission, or missing other important transmission routes to sparesly sampled locations that have not been detected due to low levels of sampling. The phylogeographic reconstruction conducted in Chapter 3 appears to be relatively robust to sampling, in that the site of entrance to the Republic of Korea, and importance of viral migration along the western coast is confirmed using several different geographic grouping of locations and whilst suppressing sparsely sampled locations (Figure B.2). However, in general, BSSVS should be used cautiously or avoided where sampling is sparse at certain locations. As such, the Markov jump scores that are presented in Figure 3.4 should be given more credence than the Bayes factors that are also presented.

Since publishing Chapter 3, the HPAI clade 2.3.4.4 H5 viral clade has spread rapidly to an unprecedented geographic extent. Clearly, there is significant scope

for a more extensive phylogeographic study that analyses viral spread at both the global and continental levels that surpasses previously published work [439]. This is important for exploring transmission routes of HPAI viruses as a whole, as well as for resolving the origins of this particular H5N8 outbreak at Abbotsbury.

7.1.3 Chapter 4: exploring the faecal viral diversity of mute swans

As exemplified in Chapter 2, metaviromic data is appropriate for discovery and identification of new viral species, and hence assists exploration of long-term evolutionary patterns. The potential of metaviromics for exploring viral epidemiological dynamics in natural populations is almost entirely untapped. The acknowledgement of this potential led to the study in Chapter 4. In Chapter 4, I conducted a study of faecal viral diversity in a population of wild mute swans (*Cygnus olor*). To my knowledge, this represents the largest metaviromic study in a wild bird population to date. I showed that epidemiological variation of viral abundance according to age and season can be detected using metaviromic data alone. I demonstrated a novel effect: known vertebrate viruses make up a larger proportion of all sequencing reads (and a larger proportion of all known viral reads) in young swans <2 years old, than in older swans. I suggested that this higher proportion of viral reads might be particularly driven by astroviruses, picornaviruses and adenoviruses. I described nine complete or near-complete new viral genomes, doubling the number of viruses known to infect mute swans. I also demonstrated the existence of other new and previously described viruses in the population.

Part of Chapter 4 involved classifying viral reads identified in the viral sequencing dataset to specific host groups, by annotating each read with the host of the closest related viral species, as identified by BLAST matching. Although broadly useful for identifying the most likely host group, the strict host-viral co-divergence that is assumed by this classification approach (i.e., assuming that any viruses infecting a

vertebrate would tend to be more similar to another vertebrate virus than to an invertebrate virus) is obviously not always reliable. For example, the first identified polyomavirus viral reads recently found in arthropods [50] would have been designated as more likely to be vertebrate viral reads under this classification scheme, based on the fact that all previous polyomaviruses have been found in vertebrates. Furthermore, it relies on the correct host-classification of previously identified viruses, which is notoriously difficult for viruses identified in faeces. Achieving any degree of sensitivity with sequence-composition based methods for host-assignment methods would probably require *de-novo* assembly of longer contigs prior to processing, but it would be interesting to explore whether my dataset of putative viral reads could be ascribed to particular host groups using analyses of sequence-composition, as has been done for other datasets [205].

The viral enrichment laboratory protocol used in Chapter 4 was chosen to be as unbiased as possible with respect to particular types of virus. However, a consequence of deliberately not enriching virions by size or nucleic acid type is that non-viral material makes up an extremely large proportion of the dataset, and hence large amounts of redundant sequence data must be generated to achieve sufficient numbers of viral reads. The viral sequence data presented here represent only a small fraction of the sequencing data. Only 7% of the reads in my dataset have BLAST matches to viruses in the `nr` database, and of these only about 1.5% are viruses that can be identified through BLAST matching to be most closely related to a species that exclusively infects vertebrates. The remaining 93% are presumably comprised of divergent viruses and host material, commensal gut bacteria and contaminating environmental or dietary material that I do not consider here. In future studies, it should be explored whether immediately removing supernatant from centrifuged samples in the field prior to freezing could help reduce cellular nucleic acid further by reducing cell lysis and consequent release of cellular nucleic acids that can occur because of freezing-thawing. Whilst the protocol implemented here attempts to remove bacterial cells through centrifugation, which may have had some impact on the range of bacte-

ria represented in the data [72], it may still be possible to use my dataset to explore the development of the mute swan gut bacterial microbiome with age, the stability of the microbiome through age or season, and contribute to genetic investigations of the role of wild birds in the spread of specific bacteria to humans, such as *Campylobacter jejuni* [70].

Further sampling in this population or other populations would substantially extend the results of Chapter 4. For viruses such as the waterbird coronavirus 1 and the novel astroviruses sequenced here, sparse sampling in other species and a poor understanding of their evolutionary rates mean that it is difficult to assess whether phylogenetic clustering of swan viruses is a result of circulation of strains primarily within swans, or whether it is simply the result of missing information about the circulation of these strains in other host species.

A previous study of rhesus macaques followed similar reports in humans [175, 295] in suggesting that the faecal viromes of individuals in a single location are more similar to each other than individuals in other locations [9]. The sharing of viruses such as avian influenza A virus between Anseriformes that share the same water-body is actually surprisingly poorly evidenced, but occurs easily in experimental conditions [1] and presumably must occur regularly in the wild to generate phylogenetic patterns that do not show distinct clustering by host (for example, [257]). An interesting follow-up study would therefore involve focusing on sampling from two Anseriforme species across two different locations, in order to consider whether observed viral diversity was site-specific or species-specific, with the expectation that there may be a significant site-specific effect for sharing of pathogens transmitted via the faecal-oral route in waterfowl.

The study in Chapter 4 did not investigate vertical transmission of viruses in the population from adults to newly hatched birds, in part because collecting faecal samples from extremely young birds is difficult without distressing the parent bird, and also because the volume and consistency of faeces from new-hatched cygnets are so small and so liquid as to make sterile collection of the sample impractical. This

population would, however, be ideal to explore vertical transmission of blood-borne viruses, and their persistence throughout the lifetime, by sampling birds throughout their lives at the well-established series of catches that is currently undertaken in the population (first within 24 hours of hatching, then 6 months of age, and then every two years), and long-term records that link cygnet rings with those of parent birds. Many birds living in the colony are related by descent, with genealogies often reaching 3 or 4 generations amongst living birds.

The historical development of technologies for viral detection through the last 100 years has demonstrated that changing technologies reveal unexplored biases in earlier methods and introduce their own problems. Filtration accidentally removed giant viruses that were larger than it was believed a virus could be, cell-culture could not detect viruses that could not be easily grown in the laboratory and, currently, computational approaches cannot detect viruses with genomes that are significantly divergent to other known viruses. The swan virome dataset contains high numbers of picornavirus and parvovirus reads that I was unable to assemble confidently because I was wary of creating false chimaeras of closely related viruses. This poses an obvious question: to what extent are known viruses generated in the metagenomic era biased towards "easy-to-assemble" viruses? Difficult-to-assemble viruses might include viruses that occur at low viral load, segmented or multipartite viruses, viruses that co-exist with other closely related viral strains, or those that have significant intra-host structural diversity. If known viruses are biased against these types of virus, presumably that has the possibility to significantly skew our understanding of the most common types of life-history strategy that viruses follow. Comparing the number and proportion of viral genomes identified through *de novo* assembly of short-read data to those identified through mapping short-reads to (significantly less accurate) long-read data generated using the Nanopore MinION platform might help explore whether limitations of *de novo* assembly are significantly biasing our knowledge of viral diversity.

7.1.4 Chapters 5 and 6: pre-existing immune responses to AIV in birds and mortality following HPAI viral infection

Chapter 5 analysed the accumulation of humoral immune responses to different subtypes of avian influenza virus (AIV) that results from continued exposure to different subtypes throughout the lifespan of mute swans. Extensive previous literature has demonstrated that juvenile birds are more likely to harbour antibodies directed against AIV nucleoprotein than younger birds, but the acquisition of responses to different haemagglutinin (HA) subtypes is poorly characterised, despite HA being the primary antigen for AIV. Furthermore, most existing studies have been unable to accurately determine the age of tested adult birds, so it has been unclear whether birds continue to gain immunity to different subtypes throughout the adult lifespan. I showed that as birds age, they continue to gain antibody responses to a broader range of HA subtypes, presumably a result of continuous exposure to different subtypes throughout their lives.

Chapter 6 presented data from an outbreak of HA clade 2.3.4.4 AIV H5N8 that circulated in the Abbotsbury mute swan population in winter 2016/2017 and caused significant mortality. Estimated mortality rates in birds of different ages during the outbreak provides significant evidence that the youngest age-group birds (hatch-year 2016, and therefore <1 year old at the time of the outbreak) were significantly more likely to die of HPAI H5N8 than older birds. Limited data on AIV prevalence during 2015-2016 suggests that AIV infection is actually relatively rare in this population, and is in line with estimates produced by other studies that study swans in the wild. Patterns of viral introduction to the population suggest that the virus was not circulating for long periods prior to the outbreak in swans, and that its introduction was one of several viral introductions to the UK.

Chapters 5 and 6 are very complementary, both focusing on avian influenza virus in my mute swan study population. The limitations and insights that can be gleaned

from these two chapters are best considered together. In Chapter 5 I questioned the extent to which data from the Abbotsbury mute swan population could be generalised to other swans, or, indeed other wild birds. The population density at Abbotsbury is unusually high for a wild mute swan population, probably in part encouraged by a long-history of supplementary feeding at the site. Host community size and population density is known to be an extremely important driver of the dynamics of infectious diseases, including influenza virus [272, 279, 325]. If increased host population densities significantly affect the dynamics of low pathogenic avian influenza (LPAI) occurrence in the Abbotsbury population, then the profile of humoral immunity in the population would be expected to be different to that seen in most mute swans. In Chapter 6, I present evidence that the birds at Abbotsbury are rarely infected with LPAI viruses, a finding that is consistent with reporting of prevalence in swans at other sites where population density is significantly lower [21, 236, 296, 314, 399]. Consequently, the patterns of immunity observed at Abbotsbury in Chapter 5 may be generalisable to birds beyond this unique population.

Chapter 6 throws into question the significance of the patterns presented in Chapter 5 for actually impacting on bird health. I present evidence that suggests that swans at Abbotsbury are only rarely infected with LPAI AIV throughout the year, that birds continue to gain humoral responses to different encountered subtypes throughout their lives, and that during HPAI H5N8 exposure, one-year old birds have mortality rates that are more similar to those of older birds than they are to those of hatch-year birds. I currently cannot rule out the hypothesis that the youngest birds have higher mortality rates because they are physiologically less able to mount an effective immune response or because their generally smaller body mass left them less able to survive the immobilising course of the infection without starving to death. However, if we assume temporarily that mortality risk is driven primarily by acquisition of humoral immune responses to previously encountered LPAI infection, then at least two possible explanations emerge that can reconcile the patterns in Chapters 5 (the gradual, long-term acquisition of humoral immunity throughout life) and 6 (a step-

change in expected mortality upon exposure to HPAI H5N8). Firstly, let us suppose that protective immunity to HPAI H5N8 is only affected by exposure to a small number of very specific LPAI subtypes or strains. In this case, protective immunity would require the circulation of at least one of these specific LPAI strains in late 2015 or very early 2016, i.e., after the hatching of all birds in the population except the youngest cohort at the time of the HPAI H5N8 outbreak in December 2016. Given the apparent low annual rate of LPAI circulation in the population, this would seem unlikely. Alternatively, let us suppose that protective immunity can be generated by exposure to a wide range of LPAI subtypes or strains, or a combination of a small number of these. In this case, it is possible that only a very small number of previously encountered LPAI infections would be required to induce protective immunity to HPAI H5N8 AIV, and that having additional responses probably does not have a strong effect on protection against HPAI H5N8 AIV associated mortality. Under this scenario, the critical threshold for previous LPAI infections for protection against mortality appears to be related to either the number or diversity of AIV infections that swans in this population encounter in their first 6-18 months of life. Distinguishing between these two explanations is of critical importance for understanding the evolution of influenza viruses generally, and specifically for providing empirical evidence to support theoretical suggestions that antigenic evolution of influenza might primarily be driven by immunity against epitopes that can have limited structural variability [362].

It may be possible to explore these scenarios, in part, by using the faecal samples collected from the Abbotsbury population during 2016. There are approximately 600 samples, collected between 4th May 2016 and 25th November 2016, that have not yet been tested for AIV. Many of the birds that died during the HPAI H5N8 outbreak, or are known to have survived, were sampled during this period. Testing of these samples for AIV, along with approximately 400 samples that have been collected after the end of the outbreak may give a more robust picture of the prevalence and diversity of LPAI virus in the population, and might perhaps reveal an association

between specific previous infections and survival patterns during the H5N8 outbreak of 2016/2017.

HPAI H5N8 AIV is still being sporadically detected in Europe as of October 2017, so understanding the landscape of immunity to the virus in wild birds is important for attempts to predict the infection dynamics of this virus in the near future, including quantifying the level of risk to wildlife and to the poultry industry. There is currently no clear picture of patterns of immunity to HPAI H5N8 in wild birds, including the duration for which humoral immune responses might be sustained, or what proportion of wild birds might have already survived asymptomatic infection with HPAI H5N8. Additionally, the theory of "original antigenic sin" suggests that a strong humoral immune memory is generated to the first encountered strain, and that weaker immune responses are generated to subsequent infections with similar strains, because the host primarily relies on the original antibody repertoire, even if the binding between the existing antibodies and novel antigens is comparatively poor [141, 455]. This theory suggests that the age at which HPAI H5N8 is first encountered (and, correspondingly, the approximate number of previously encountered LPAI infections) might be associated with the specificity of any antibody responses upon re-challenge with HPAI H5N8 in species where AIV infection is relatively rare. It would be interesting to study existing immunological responses to HPAI H5N8, and to other previously detected LPAI viruses in the population in order to explore how long immunity lasts to specific viruses following known infections.

Phylogenetic analysis of viral genomes obtained from birds that died during the H5N8 outbreak at Abbotsbury showed that the viruses were not sufficiently diverse to attempt to resolve intra-population transmission routes. The low diversity in the population was likely caused by rapid viral spread through the population. Within a single host, the diversity of influenza virus populations is high. Reconstructing the intra-population transmission route from patterns of minority variants that are present within individual hosts presents a complementary approach to studying transmission over short time scales, even when consensus sequences from each host in that

transmission pathway are identical [409]. This approach has been applied to reconstructing transmission routes of influenza viruses in horses [184], birds [409], pigs [299] and dogs [170]. In general it is believed that younger birds tend to shed higher quantities of virus than older birds [452], which might mean that infected younger birds are more important donors of virus to new birds than older birds. Determining the route of transmission of avian influenza amongst the Abbotsbury swans could be interesting for both HPAI H5N8 and well-sampled LPAI viruses identified in the population, especially if an age-based directionality of transmission could be observed. Although the MinION data obtained in Chapter 6 may not currently be appropriate for such an analysis because the high error rate prevents accurate identification of minority variants, it would be interesting to re-sequence the samples from the infected birds using a low error rate technique (such as Illumina sequencing) to investigate the pattern of minority variants and attempt to reconstructe viral transmission within the Abbotsbury swan population.

The results of Chapter 5 and Chapter 6 support the findings of experimental viral challenge studies, which show that HPAI AIVs exhibit species-specific differences in virulence and infectivity [321, 322, 94, 23]. Wild waterfowl of different species exhibit a huge spectrum of sociality, population size, migration patterns, previous exposure to LPAI AIV, and feeding behavior. This raises the interesting question of whether HPAI AIV viruses might be adapted to infect and transmit among a primary reservoir comprised of a limited number of bird species that best facilitate viral transmission because of their particular behavioural or migratory patterns. For example, viral adaptation to gregarious hosts that mix with other birds at high density, to waterfowl hosts that consume contaminated water during feeding, and/or to hosts that regularly move between different areas, might facilitate the transmission of HPAI AIV. A virus might adapt to cause low virulence in species that represent the primary host reservoir for a number of different reasons, including if low virulence allows greater host movement and hence facilitates viral transmission. Transmission to a host species that does not represent the key reservoir might therefore result in

infections of higher virulence because the virus has not adapted to transmission in that host. Under this hypothesis, the high virulence infections observed in swans might indicate that swans are generally not as important for further transmission of the virus as birds that exhibit low-virulence infections: perhaps because they are usually non-migratory and territorial birds, the native host range does not include parts of Asia where the HPAI H5N8 virus appears to have emerged, and the absolute number of mute swans is low in comparison to certain other species.

Surprisingly, there has been limited investigation to date of the association of wild bird ecology and the detection of HPAI AIV in wild birds. Currently, studies of cross-species transmission and species-specific adaptation of HPAI viruses typically group all species of wild waterfowl together, in opposition to poultry, mammals or other wild birds such as gulls. However, this may be failing to account for epidemiologically important ecological differences between waterfowl species. Collating and analyzing currently available data on species-specific AIV prevalence is an important future avenue for understanding the epidemiological dynamics that we observe.

7.2 General Implications and Future Directions for the Study of Viral Epidemiological Dynamics

Inter-disciplinary studies that investigate viral epidemiological dynamics in animal populations that have been subjected to long-term studies of host behaviour and demography offer enormous potential for improving our understanding of viral transmission in wildlife. I show this potential here with the inclusion of three chapters that greatly benefit from the existence of continuous, long-term ornithological records of a population of wild mute swans. Such an approach has already occasionally been used to explore the movement of specific viruses in wildlife populations (for example, Fountain-Jones et al., (2017)[139]) but should be adopted more widely, and, as costs

of sequencing continue to reduce, could include more frequent analysis of metaviromic data. New technologies that assist the remote tracking of wildlife could also be incorporated into studies of how host movement affects viral transmission, or, conversely, how viral infection affects host movement. This might include more frequent use of GPS tracking or activity tracking of individuals to detect frequency of movement or geographic dispersal, depth sensing, video-monitoring [374, 444, 450], or, for highly visible animals such as mute swans, regular drone-based or satellite photography of geographic location. Investigation of larger-scale geographic spread might be possible through consideration of seasonal changes based on existing databases of variables that affect whether an animal can live in a particular ecosystem, such as temperature or vegetation [307]. Weather surveillance radar data (for example, NEXRAD in the USA) often routinely detects the movement of birds, and is currently being used to achieve a more extensive quantitative picture of the timing and route of avian migration [180]. Movement of birds between radar stations might help provide a more quantitative picture of the temporal and geographic patterns in avian migratory flyways, that could be incorporated into network or phylodynamics models. Creatively exploring other global datasets that could be used to track animal movement on a global scale will help us to better understand the widespread movement of viruses. Recently developed computational tools such as SERAPHIM [96] offer flexible approaches for analysing the importance of environmental variables to viral transmission modelled within continuous space using phylogenetic estimation approaches. Such computational tools could be extended to evaluate the importance of host geographic movement, or contact between hosts, on phylogenetic reconstructions.

The importance of host ecology to viral transmission logically means that surveillance of pathogens should include detailed metadata, which, at a minimum, includes host species, date, location (ideally longitude and latitude) and approximate host age. Whilst existing information is useful, the availability of precise longitude-latitude data would have allowed a more detailed model of transmission between locations

within South Korea in Chapter 3. Data on wild bird outbreak reporting on the HPAI H5N8 outbreak featured here suggests that 9% of all outbreaks had no species associated with them, and a total of 38% of all outbreaks were associated with "ducks", "geese""or "swans" [10]. Higher-level classifications are useful to an extent, but hide the huge variation in migration strategy, life-history or feeding behavior that exist in all of those groups. Technological advances, including computational image-based identification of species, could easily be incorporated into existing reporting systems to allow better ecological annotation of genomic or epidemiological data without putting significantly greater pressure on key government, scientific or veterinary workers.

Real-time, portable genome sequencing has the potential to revolutionise the way in which viruses are studied. Previous projects that sequenced human viruses such as Ebola virus and Zika virus have already demonstrated the utility of taking sequencing machines to the site of ongoing outbreaks [355], or to clinics that have vitally important samples but might lack the infrastructure to conduct rapid and large-scale viral sequencing and which cannot easily ship samples elsewhere due to legal or logistical restrictions [126, 128, 125]. There is no reason that the same methodology that is currently being applied in humans could not also be used to study viruses in animals, including in geographically remote animal populations. Excitingly, real-time data production could also be used to modulate research questions in real time. Sample collection is non-trivial in terms of researcher time and cost. In principle, real-time sequencing could be used to generate metaviromic information on novel pathogens in a small set sample of samples in real-time in the field, and the data generated fast enough to subsequently inform how further sample collection should be targeted during a single fieldtrip. For example, in Chapter 4, initial identification of an orthoreovirus exclusively in the youngest birds in the population might have merited a change to sample collection that would subsequently focus primarily on very young birds. This approach of real-time feedback between data-generation and sampling would be particularly useful for investigating viruses that only circulate rarely or for short periods in a specific population. The potential consequences of adopting such

an approach to real-time-data driven sampling could be huge, allowing far greater amounts of information to be gained from the same numbers of samples.

In this thesis, I contribute to our existing knowledge of viral diversity and distribution, and explore some of the host ecological drivers that lead to transmission dynamics in nature. More broadly, my research is motivated by the premise that viral epidemiological processes and long-term evolution of viral diversity should be explored by studying viral distribution and transmission in a natural ecological setting, in which heterogeneous contact patterns between hosts, competition between different viruses, and pre-existing immunological responses are the norm rather than the exception. This approach does not devalue laboratory and modelling studies, but rather underlines the notion that research that aims to understand the transmission of viruses must surpass traditional boundaries between the medical and natural sciences and draw on existing knowledge about the natural ecology of hosts. Creative approaches to integrating innovations from multiple disciplines will be fundamental to furthering our knowledge of the epidemiology and diversity of viruses.

Appendix A

Appendix to Chapter 2

A.1 Tables

Table A.1: Tissue samples and PCR positivity for presence of polyomavirus using VP1 specific assay for 11 animals

	Animal[1]											
Organ type	1	2	3	4	5	6	7	8	9	10	11	Number positive
Salivary gland	-	-	-	+	-	+	-	-	-	-	-	2 (18)
Spleen	-	-	-	-	-	-	-	-	-	+	+	2 (18)
Liver	+	-	+	+	-	-	+	-	-	+	+	6 (55)
Kidney	-	-	-	+	-	-	-	-	-	-	-	1 (9)
Mediastinal lymph node	+	+	-	-	-	+	+	-	-	X	-	4 (40)
Bronchial lymph node	+	-	+	+	-	+	-	-	-	+	-	5 (45)
Blood	+	-	-	-	-	-	-	-	-	-	-	1 (9)
Faeces	-	-	-	-	-	-	-	-	-	-	-	0 (0)

[1]Separate animals are represented by different numbers. +: tested negative, -: tested positive, X: not tested.

Table A.2: Tissue samples and PCR positivity for presence of polyomavirus using NCCR specific assay for 11 animals

	Animal[1]										
Organ type	1	2	3	4	5	6	7	8	9	10	11
Salivary gland	X	X	X	+	X	+	X	X	X	X	X
Spleen	X	X	X	X	X	X	X	X	X	+	+
Liver	+	X	+	+	X	X	+	X	X	-	+
Kidney	X	X	X	-	X	X	X	X	X	X	X
Mediastinal lymph node	-	+	X	X	X	+	+	X	X	X	X
Bronchial lymph node	-	X	-	-	X	+	X	X	X	+	X
Blood	+	X	X	X	X	X	X	X	X	X	X
Faeces	X	X	X	X	X	X	X	X	X	X	X

[1]Separate animals are represented by different numbers. +: tested negative, -: tested positive, X: not tested.

Table A.3: Host group categories for BaTS analysis

Number	Accession number and name	Group 1 (order)	Group 2 (family)	Group 3 (order) [1]
1	AB588640 Mastomys polyomavirus	Rodentia	Muridae	Rodent
2	AB767294 Yellow baboon polyomavirus 1	Primate	Cercopithecidae	Old world monkey
3	AB767295 Yellow baboon polyomavirus 2	Primate	Cercopithecidae	Old world monkey
4	AB767297 Vervet monkey polyomavirus 3	Primate	Cercopithecidae	Old world monkey
5	AB767299 Vervet monkey polyomavirus 2	Primate	Cercopithecidae	Old world monkey
6	KP644239 Meles meles polyomavirus 1 UK isolate	Carnivora	Mustelidae	Carnivora
7	HQ385746 Pan troglodytes verus polyomavirus 1a isolate 6444	Primate	Hominidae	Ape
8	HQ385748 Pan troglodytes verus polyomavirus 2a isolate 6512	Primate	Hominidae	Ape
9	HQ385752 Gorilla gorilla gorilla polyomavirus 1 isolate 5766	Primate	Hominidae	Ape
10	JQ178241 Raccoon polyomavirus strain R45	Carnivora	Procyonidae	Carnivora
11	JQ958886 Bat polyomavirus isolate A1055	Chiroptera	Phyllostomidae	New world bat
12	JQ958887 Bat polyomavirus isolate R104	Chiroptera	Phyllostomidae	New world bat
13	JQ958888 Bat polyomavirus isolate B0454	Chiroptera	Phyllostomidae	New world bat
14	JQ958889 Bat polyomavirus isolate C1109	Chiroptera	Phyllostomidae	New world bat
15	JQ958890 Bat polyomavirus isolate A504	Chiroptera	Phyllostomidae	New world bat
16	JQ958891 Bat polyomavirus isolate R266	Chiroptera	Mormoopidae	New world bat
17	JQ958892 Bat polyomavirus isolate AT7	Chiroptera	Phyllostomidae	New world bat
18	JQ958893 Bat polyomavirus isolate B1130	Chiroptera	Molossidae	New world bat
19	JX159985 Cercopithecus erythrotis polyomavirus 1	Primate	Cercopithecidae	Old world monkey
20	JX159989 Saimiri sciureus polyomavirus 1	Primate	Cebidae	New world monkey
21	JX259273 MX polyomavirus isolate UC MXPyV 1	Primate	Hominidae	Human
22	KC594077 Dolphin polyomavirus 1 isolate DPyV 1	Cetacea	Delphinidae	Ungulate
23	NC001442 Bovine polyomavirus	Artiodactyla	Bovidae	Ungulate
24	NC 001505 Murine pneumotropic virus	Rodentia	Muridae	Rodent
25	NC 001515 Murine polyomavirus	Rodentia	Muridae	Rodent
26	NC 001538 BK polyomavirus	Primate	Hominidae	Human

Continued
on next page

Number	Accession number and name	Group 1 (order)	Group 2 (family)	Group 3 (order) [1]
27	NC 001663 Hamster polyomavirus	Rodentia	Cricetidae	Rodent
28	NC 001669 Simian virus 40	Primate	Cercopithecidae	Old world monkey
29	NC 001699 JC polyomavirus	Primate	Hominidae	Human
30	NC 004763 African green monkey polyomavirus	Primate	Cercopithecidae	Old world monkey
31	NC 007611 Simian virus 12	Primate	Cercopithecidae	Old world monkey
32	NC 009238 KI polyomavirus Stockholm 60	Primate	Hominidae	Human
33	NC 009539 WU Polyomavirus	Primate	Hominidae	Human
34	NC 009951 Squirrel monkey polyomavirus	Primate	Cebidae	New world monkey
35	NC 010277 Merkel cell polyomavirus	Primate	Hominidae	Human
36	NC 011310 Myotis polyomavirus VM 2008	Chiroptera	Vespertilionidae	New world bat
37	NC 013439 Orangutan polyomavirus	Primate	Hominidae	Ape
38	NC 013796 California sea lion polyomavirus 1	Carnivora	Otariidae	Carnivora
39	NC 014361 Trichodysplasia spinulosa associated polyomavirus	Primate	Hominidae	Human
40	NC 014406 Polyomavirus HPyV6	Primate	Hominidae	Human
41	NC 014407 Polyomavirus HPyV7	Primate	Hominidae	Human
42	NC 014743 Chimpanzee polyomavirus	Primate	Hominidae	Ape
43	NC 015150 Human polyomavirus 9	Primate	Hominidae	Human
44	NC 017982 Equine polyomavirus	Perissodactyla	Equidae	Ungulate
45	NC 018102 MW polyomavirus	Primate	Hominidae	Human
46	NC 019844 Vervet monkey polyomavirus 1	Primate	Cercopithecidae	Old world monkey
47	NC 019850 Piliocolobus rufomitratus polyomavirus 1	Primate	Cercopithecidae	Old world monkey
48	NC 019851 Macaca fascicularis polyomavirus 1	Primate	Cercopithecidae	Old world monkey
49	NC 019853 Ateles paniscus polyomavirus 1	Primate	Atelidae	New world monkey
50	NC 019854 Cebus albifrons polyomavirus 1	Primate	Cebidae	New world monkey
51	NC 019855 Pan troglodytes verus polyomavirus 3	Primate	Hominidae	Ape
52	NC 019856 Pan troglodytes verus polyomavirus 4	Primate	Hominidae	Ape
53	NC 019857 Pan troglodytes verus polyomavirus 5	Primate	Hominidae	Ape
54	NC 019858 Pan troglodytes schweinfurthii polyomavirus 2	Primate	Hominidae	Ape

Continued
on next page

Number	Accession number and name	Group 1 (order)	Group 2 (family)	Group 3 (order) [1]
55	NC 020065 Chaerephon polyomavirus 1 isolate KY397	Chiroptera	Molossidae	Old world bat
56	NC 020066 Otomops polyomavirus 2 isolate KY156	Chiroptera	Molossidae	Old world bat
57	NC 020067 Cardioderma polyomavirus isolate KY336	Chiroptera	Megadermatidae	Old world bat
58	NC 020068 Eidolon polyomavirus 1 isolate KY270	Chiroptera	Pteropodidae	Old world bat
59	NC 020069 Miniopterus polyomavirus isolate KY369	Chiroptera	Vespertilionidae	Old world bat
60	NC 020070 Pteronotus polyomavirus isolate GTM203	Chiroptera	Mormoopidae	New world bat
61	NC 020071 Otomops polyomavirus 1 isolate KY157	Chiroptera	Molossidae	Old world bat
62	NC 020106 STL polyomavirus strain MA138	Primate	Hominidae	Human
63	NC 020890 Human polyomavirus 12 strain hu1403	Primate	Hominidae	Human
64	NC 022519 African elephant polyomavirus 1	Proboscidea	Elephantidae	Proboscidea

[1]Bats split by old/new world and primates split into old world monkeys, new world monkeys, and apes, and ungulates grouped

Table A.4: Primers used in sequencing of French badger polyomavirus isolate

Primer name [1]	Sequence (5'-3')	Product size (bp)	Purpose
6814-s	TGAAGTGTTGTGTAATGAACCCT	260	NBL-7 cell line
6814-as	AGTTTAGCATTTTCATGGTGC		
6815-s	ACAGGCCCTAGAAAGCTTTGG	308	NBL-7 cell line
6815-as	GGTTAATTAGGTCTGGTCCCCA		
6816-s	ACTGGAGACAAAATATTGATGTGGG	345	NBL-7 cell line
6816-as	ATTTTTGGCTGGGTCTGGGT		
6817-s	ACTGGAGACAAAATATTGATGTGGG	450	NBL-7 cell line
6817-as	CCATTTTCATCAAGCAAAATGGT		
6894-s	TGAAAATGGAGTGGGACCAC	264	Diagnostic VP1 PCR
6894-as	TATACCCTGACCTCCTCCAC		
6895-s	TGGCCATACATCTGCATGGG	459	Diagnostic NCCR PCR
6895-as	GGCTGCTGCAAGTTCAAACA		
6942-s	GTTTCGGTTTGGCTAAGGGC	2017	Partial genome amp. 1st round
6942-as	CATGGGCTGGCTATCCACAT		
6943-s	AGCCATCTTCCCGCGTTATG	1974	Partial genome amp. 2nd round
6943-as	GGCTATCCACATCAGGGGC		
6944-s	GGGACATATACAGGAGGTGTCC	2190	Partial genome amp. 1st round
6944-as	AGAAACTGTGGAGGGAGGACT		
6945-s	CACCCCCAGTGCTGCAATT	2120	Partial genome amp. 2nd round
6945-as	ACCTGTGGTGTCCTGGGTAA		
6946-s	ACATCCTCTATTTTGTTTTCTAAGGCA	1470	Partial genome amp. 1st round
6946-as	TTCAGGCTCAGGCTTCATGT		
6947-s	TACCCAGGACACCACAGGTT	1390	Partial genome amp. 2nd round
6947-as	ATAACGCGGGAAGATGGCTG		
6962-s	TCTCTTTTCTCCACGTGGGA	751	Validation
6962-as	CAGGCCCTAGAAAGCTTTGGA		
6963-s	CCCTTCCCCCAAAGATGCAA	800	Validation
6963-as	AGTTTGGCTTCTCAGGCTGT		
6964-s	CCCAGACAAGGGTGGTGATG	752	Validation
6964-as	GGGCATACTTGTCCTGTACCT		
6965-s	TCCTGAGTAAAGCCCTAATATCT	757	Validation
6965-as	TCTTCAAATTTAGGTAACAGGCA		
6966-s	TGGAGGCAATATTTGGTCCCA	801	Validation
6966-as	TGGAAATCATCTTGGCCCCT		
6967-s	TGATCCCCAAATTCAAACAGATGT	781	Validation

Continued on next page

Primer name [1]	Sequence (5'-3')	Product size (bp)	Purpose
6967-as	GCCACCACAGGACACAGAAA		
6968-s	ATTTGGGGAGGGTCAGGTCT	759	Validation
6968-as	TGCTCAGCCTATTGTTGGACT		
6969-s	ACTGTAGCTGCAGTGGTTCC	611	Validation
6969-as	AAATATGGTGCCCCAGGTGG		
6991-s	ACAGATGTTTGTAAATGGAAGGA	749	Validation
6991-as	CACAGGACACAGAAAGAAATCT		
6992-s	CCCCCACTGCAAACATATGA	611	Validation
6992-as	CTGGCCAAAGACATGCTCCT		

[1] s=sense, as=antisense

Appendix B

Appendix to Chapter 3

B.1 Figures

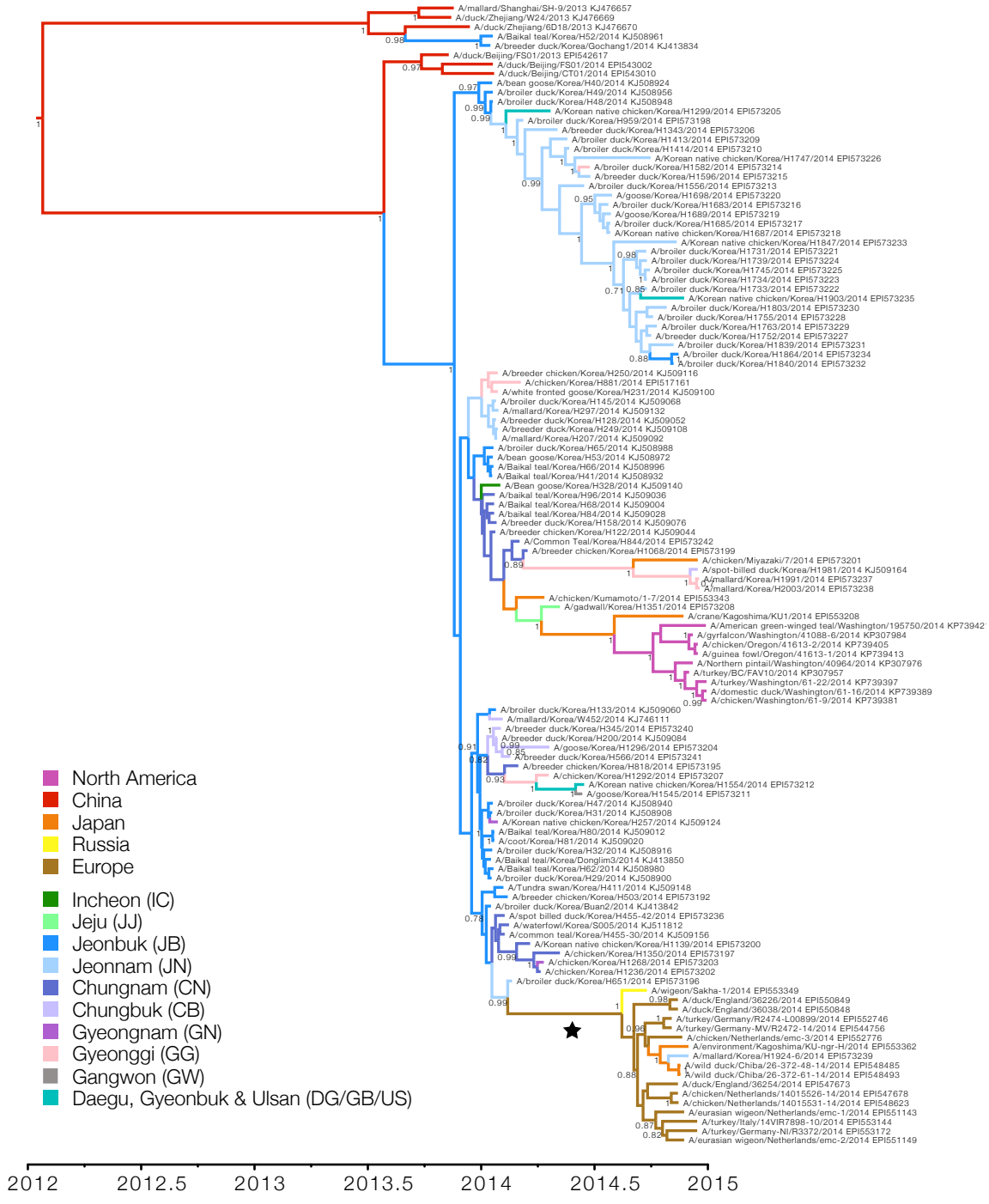

Figure B.1: Maximum clade credibility tree of avian influenza sequences for H5 clade 2.3.4.4. NA sequences were used to inform the structure of the starred Eurasian clade, where available. Tips are labelled with isolate name and accession number (refer to Supplementary Table 2 for accession number details). Posterior support is given at nodes where >0.7. Tree is equivalent to Figure 3.3. Note that whilst all branches are coloured according to the inferred ancestral location based on asymmetric trait reconstruction using BSSVS, branch locations in some parts of the tree may not accurately reflect the true dynamics of the virus. Refer to the main text for discussion.

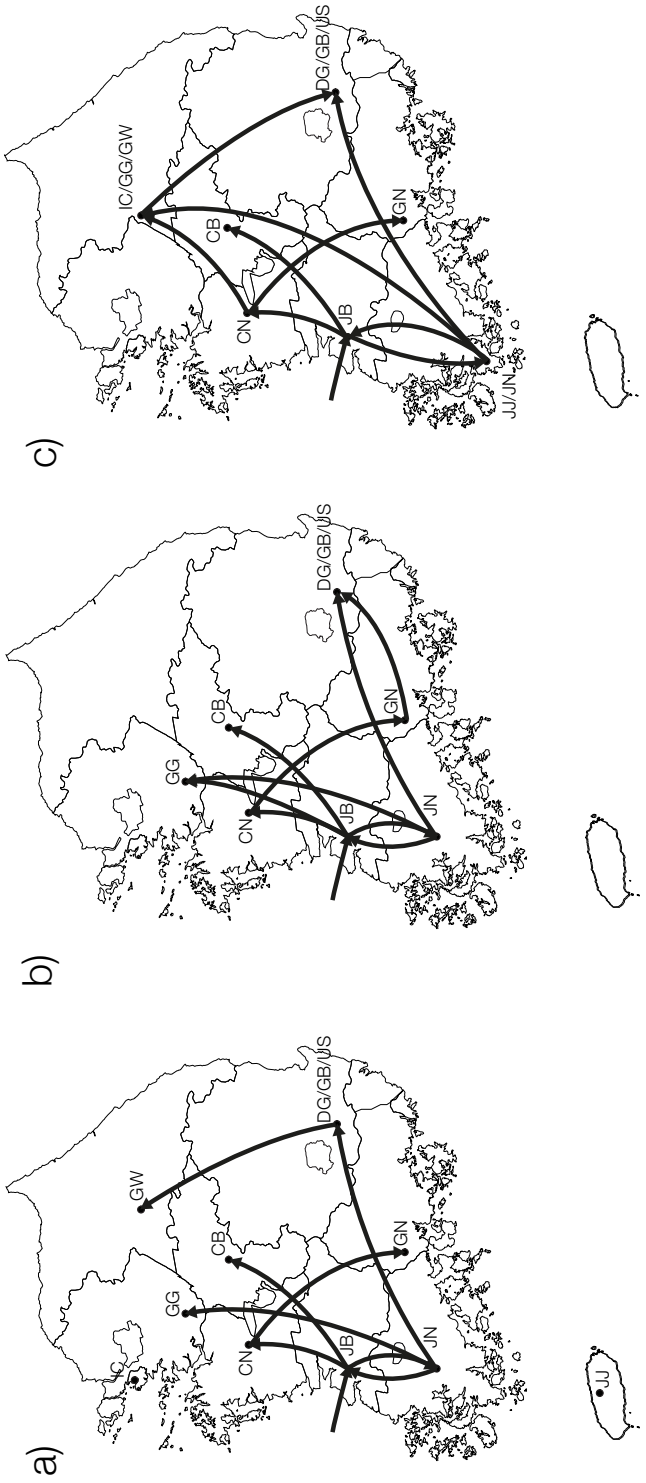

Figure B.2: Effect of different sampling and grouping of provinces with only one sequence on phylogeographic inference. Arrows show inferred epidemiological links where Bayes Factor support for rate indictors >10, as determined using SPREAD. a) Results as shown in Figure 3. b). Sequences isolated in provinces from which only one sequence is available have been removed from the alignment (except DG/GB/US, which have been grouped). c) Sequences isolated in provinces from which only one sequence is available have been grouped with geographically neighbouring provinces for phylogeographic inference.

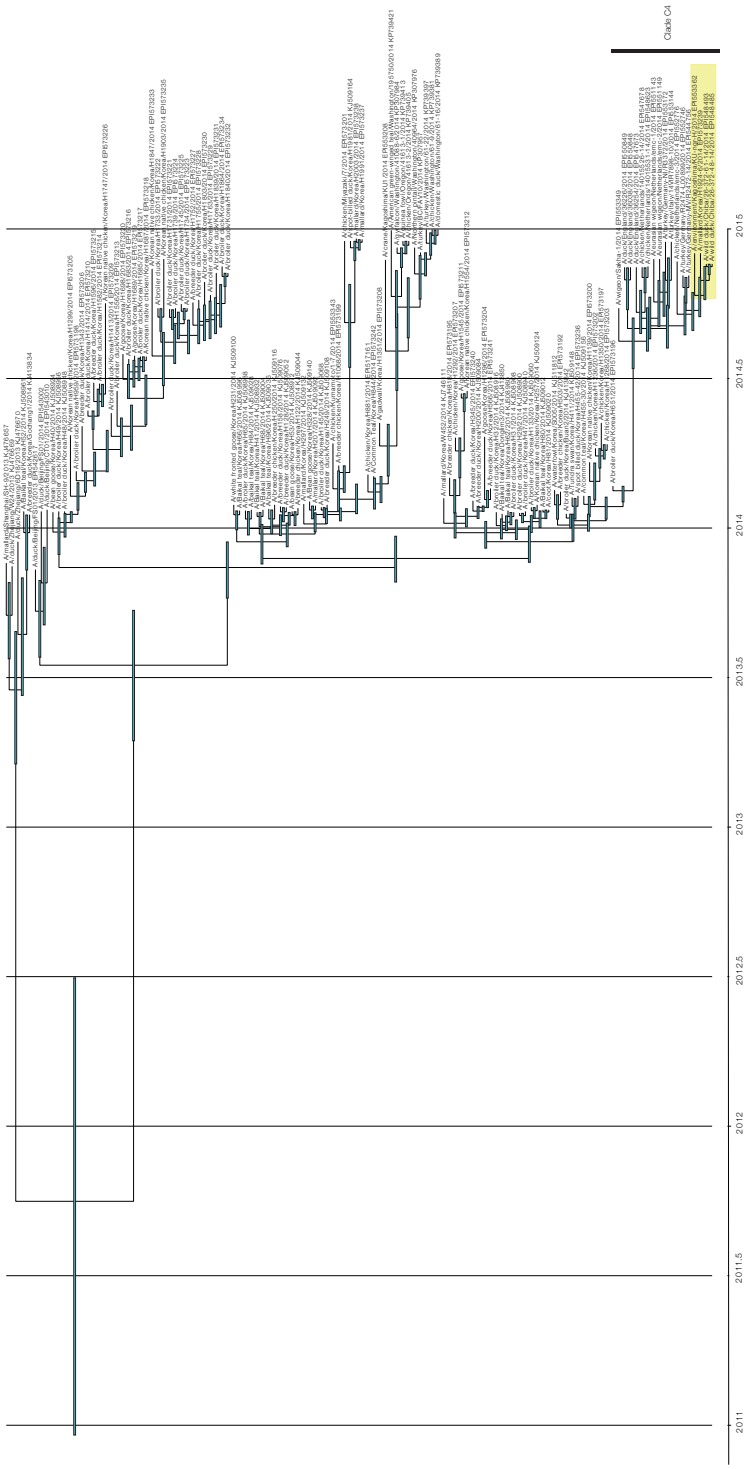

Figure B.3: Maximum clade credibility tree for reconstruction without phylogeographic model. Yellow box shows "Japan and Korea" clade in which monophyly is strongly supported (Table 3.1). Note that, although A/wigeon Sakha-1/2014 appears to be basal to clade C4 is this MCC tree, this position is not well supported by monophyly statistics (monophyly statistic 0.28), which indicates the Russia sequence often occurs internally the the C4 clade.

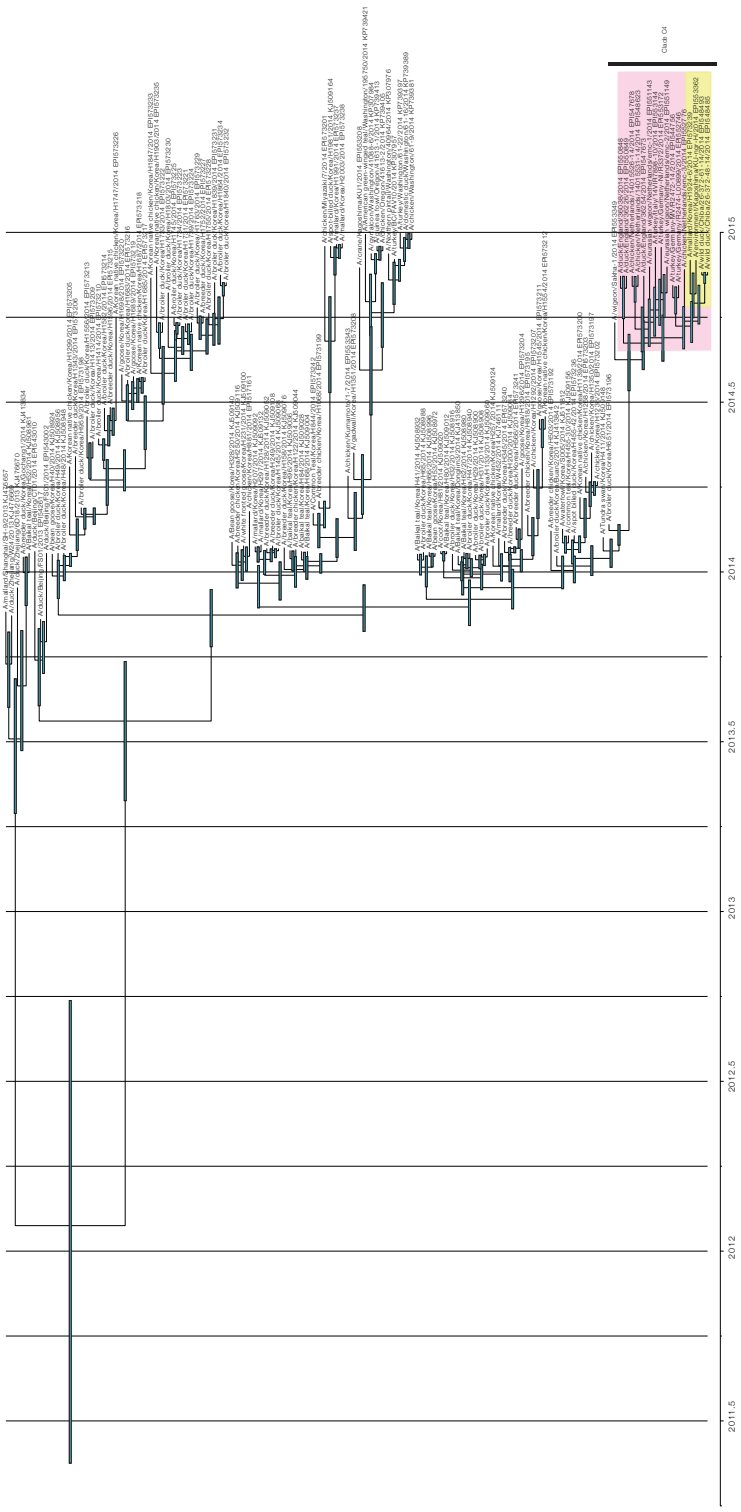

Figure B.4: Maximum clade credibility tree for reconstruction with phylogeographic model and without BSSVS. Yellow box shows "Japan and Korea" clade in which monophyly is strongly supported (Table 1). Pink box shows "Japan, Korea and Europe" clade (Table 3.1). Monophyly of this clade is only very weakly favoured by this model (monophyly statistic 0.55).

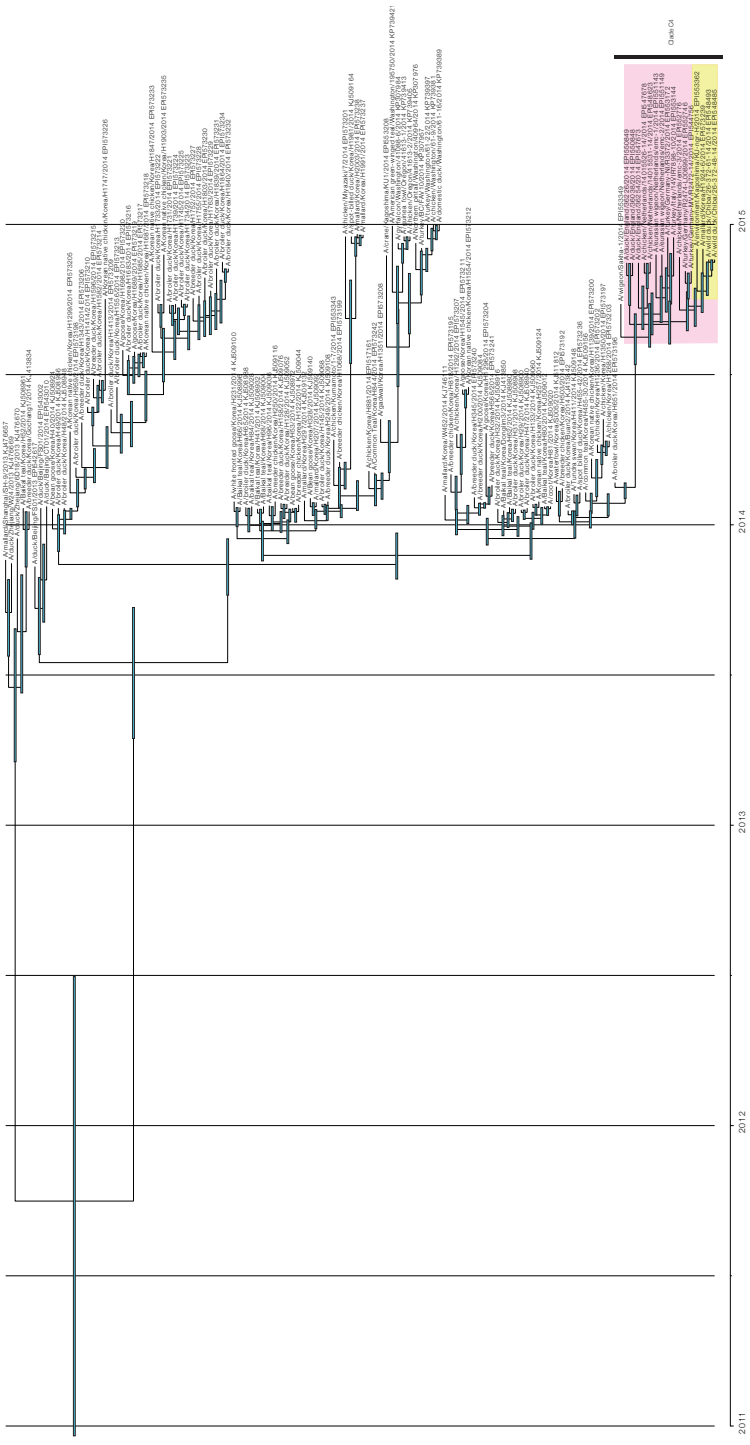

Figure B.5: Maximum clade credibility tree for reconstruction with phylogeographic model and with BSSVS. Yellow box shows "Japan and Korea" clade in which monophyly is strongly supported (Table 3.1). Pink box shows "Japan, Korea and Europe" clade (Table 3.1). The monophyly statistic for this clade (0.62) only very weakly supports consistent monophyly during tree estimation.

B.2 Tables

Table B.1: Positive samples and selected isolates from poultry in ROK.

	Month[1]												
Province	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Province totals
CB	4(1)	49(2)	4	1(1)									58(4)
CN	2(2)	9	15(3)	2(1)	1(1)								29(7)
DG						1(1)							1(1)
GB			2								1(1)		3(1)
GG	1(1)	14	5(1)	2(1)		1(1)						5	28(4)
GN	1(1)			1(1)								1	3(2)
GW						1(1)							1(1)
JB	30(11)	9(1)	5	3							3(2)		50(14)
JN	3 (3)	13(1)	19(1)	2	3(3)	3(2)	4(4)		11(5)	17(5)	3(2)	1(1)	79(28)
US				1(1)									1(1)
Month totals	41(19)	94(4)	50(5)	12(5)	4(4)	6(5)	4(4)	0(0)	11(5)	17(5)	7(5)	7(1)	253(63)

[1]From January 2014 to December 2014, H5N8 viruses (n=296 across wild birds (Table B.2) and poultry) were isolated from provinces of Korea (Chungbuk(CB), Chungnam(CN), Daegu(DG), Gangwon(GW), Gyeongbuk(GB), Gyeonggi(GG), Gyeongnam(GN), Incheon(IC), Jeju(JJ), Jeonnam(JN), Ulsan(US)) (text not in brackets). 56 strains were selected for sequencing across wild birds (Table B.2) and poultry on the basis of isolation time and location (text in brackets). At least one strain per province and four strains per month were included.

Table B.2: Positive samples and selected isolates from wild birds in ROK.

	Month[1]												
Province	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Province totals
CB		1(1)										2(1)	3(2)
CN	4(3)	7(3)	3(1)										14(7)
GG	3(1)		1	1.00								3(2)	8(3)
GW		1											1(0)
IC		1(1)											1(1)
JB	10(8)	2(1)	1										13(9)
JJ					1(1)								1(1)
JN	2(2)												2(2)
Month totals	19(14)	12(7)	5(1)	1	1(1)	0	0	0	0	0	0	5(3)	43(25)

[1]From January 2014 to December 2014, H5N8 viruses (n=296 across wild birds and poultry (Table B.1)) were isolated from provinces of Korea (Chungbuk(CB), Chungnam(CN), Daegu(DG), Gangwon(GW), Gyeongbuk(GB), Gyeonggi(GG), Gyeongnam(GN), Incheon(IC), Jeju(JJ), Jeonnam(JN), Ulsan(US)) (text not in brackets). 56 strains were selected for sequencing across wild birds and poultry (Table B.1) on the basis of isolation time and location (text in brackets). At least one strain per province and four strains per month were included.

Table B.3: Sequences and accession numbers used in phylogenetic analyses

Virus name	Host	Isolation date	Location	GISAID or GenBank accessions	GISAID originating lab [1]	GISAID submitting lab [1]
A/breeder duck/Korea/Gochang1/2014	Domestic duck	2014-01-16	Jeonbuk(JB)	KJ413834		
A/broiler duck/Korea/Buan2/2014	Domestic duck	2014-01-17	Jeonbuk(JB)	KJ413842		
A/Baikal teal/Korea/Donglim3/2014	Wild duck	2014-01-17	Jeonbuk(JB)	KJ413850		
A/Baikal teal/Korea/H52/2014	Wild duck	2014-01-20	Jeonbuk(JB)	KJ508961		
A/breeder duck/Korea/H200/2014	Domestic duck	2014-01-27	Chungbuk(CB)	KJ509084		
A/Bean goose/Korea/H328/2014	Wild goose	2014-02-01	Incheon(IC)	KJ509140		
A/Tundra swan/Korea/H411/2014	Wild other	2014-02-06	Jeonbuk(JB)	KJ509148		
A/common teal/Korea/H455-30/2014	Wild duck	2014-02-08	Chungnam(CN)	KJ509156		
A/breeder chicken/Korea/H503/2014	Domestic chicken	2014-02-13	Jeonbuk(JB)	EPI573192	QIA	QIA
A/broiler duck/Korea/H651/2014	Domestic duck	2014-02-20	Jeonnam(JN)	EPI573196	QIA	QIA
A/breeder chicken/Korea/H818/2014	Domestic chicken	2014-03-02	Chungnam(CN)	EPI573195	QIA	QIA
A/chicken/Korea/H881/2014	Domestic chicken	2014-03-06	Gyeonggi(GG)	EPI573197	QIA	QIA
A/broiler duck/Korea/H959/2014	Domestic duck	2014-03-10	Jeonnam(JN)	EPI573198	QIA	QIA
A/breeder chicken/Korea/H1068/2014	Domestic chicken	2014-03-17	Chungnam(CN)	EPI573199	QIA	QIA
A/Korean native chicken/Korea/H1139/2014	Domestic chicken	2014-03-21	Chungnam(CN)	EPI573200	QIA	QIA
A/chicken/Korea/H1236/2014	Domestic chicken	2014-04-07	Chungnam(CN)	EPI573201	QIA	QIA
A/chicken/Korea/H1268/2014	Domestic chicken	2014-04-12	Gyeongnam(GN)	EPI573202	QIA	QIA
A/chicken/Korea/H1292/2014	Domestic chicken	2014-04-20	Gyeonggi(GG)	EPI573203	QIA	QIA
A/goose/Korea/H1296/2014	Domestic goose	2014-04-21	Chungbuk(CB)	EPI573204	QIA	QIA
A/Korean native chicken/Korea/H1299/2014	Domestic chicken	2014-04-23	Ulsan(US)	EPI573205	QIA	QIA
A/breeder duck/Korea/H1343/2014	Domestic duck	2014-05-04	Jeonnam(JN)	EPI573206	QIA	QIA
A/chicken/Korea/H1350/2014	Domestic chicken	2014-05-08	Chungnam(CN)	EPI573207	QIA	QIA
A/gadwall/Korea/H1351/2014	Wild duck	2014-05-08	Jeju(JJ)	EPI573208	QIA	QIA
A/broiler duck/Korea/H1413/2014	Domestic duck	2014-05-20	Jeonnam(JN)	EPI573209	QIA	QIA
A/broiler duck/Korea/H1414/2014	Domestic duck	2014-05-23	Jeonnam(JN)	EPI573210	QIA	QIA
A/goose/Korea/H1545/2014	Domestic goose	2014-06-13	Gangwon(GW)	EPI573211	QIA	QIA
A/Korean native chicken/Korea/H1554/2014	Domestic chicken	2014-06-16	Daegu(DG)	EPI573212	QIA	QIA
A/broiler duck/Korea/H1556/2014	Domestic duck	2014-06-16	Jeonnam(JN)	EPI573213	QIA	QIA
A/broiler duck/Korea/H1582/2014	Domestic duck	2014-06-25	Gyeonggi(GG)	EPI573214	QIA	QIA
A/breeder duck/Korea/H1596/2014	Domestic duck	2014-06-26	Jeonnam(JN)	EPI573215	QIA	QIA
A/broiler duck/Korea/H1683/2014	Domestic duck	2014-07-25	Jeonnam(JN)	EPI573216	QIA	QIA
A/broiler duck/Korea/H1685/2014	Domestic duck	2014-07-28	Jeonnam(JN)	EPI573217	QIA	QIA
A/Korean native chicken/Korea/H1687/2014	Domestic chicken	2014-07-28	Jeonnam(JN)	EPI573218	QIA	QIA
A/goose/Korea/H1689/2014	Domestic goose	2014-07-29	Jeonnam(JN)	EPI573219	QIA	QIA
A/goose/Korea/H1698/2014	Domestic goose	2014-07-31	Jeonnam(JN)	EPI573220	QIA	QIA
A/broiler duck/Korea/H1731/2014	Domestic duck	2014-09-24	Jeonnam(JN)	EPI573221	QIA	QIA
A/broiler duck/Korea/H1733/2014	Domestic duck	2014-09-25	Jeonnam(JN)	EPI573222	QIA	QIA
A/broiler duck/Korea/H1734/2014	Domestic duck	2014-09-25	Jeonnam(JN)	EPI573223	QIA	QIA

Continued on next page

Virus name	Host	Isolation date	Location	GISAID or GenBank accessions	GISAID originating lab [1]	GISAID submitting lab
A/broiler duck/Korea/H1739/2014	Domestic duck	2014-09-25	Jeonnam(JN)	EPI573224	QIA	QIA
A/broiler duck/Korea/H1745/2014	Domestic duck	2014-09-30	Jeonnam(JN)	EPI573225	QIA	QIA
A/Korean native chicken/Korea/H1747/2014	Domestic chicken	2014-10-01	Jeonnam(JN)	EPI573226	QIA	QIA
A/breeder duck/Korea/H1752/2014	Domestic duck	2014-10-02	Jeonnam(JN)	EPI573227	QIA	QIA
A/broiler duck/Korea/H1755/2014	Domestic duck	2014-10-05	Jeonnam(JN)	EPI573228	QIA	QIA
A/broiler duck/Korea/H1763/2014	Domestic duck	2014-10-11	Jeonnam(JN)	EPI573229	QIA	QIA
A/broiler duck/Korea/H1803/2014	Domestic duck	2014-10-27	Jeonnam(JN)	EPI573230	QIA	QIA
A/broiler duck/Korea/H1839/2014	Domestic duck	2014-11-07	Jeonnam(JN)	EPI573231	QIA	QIA
A/broiler duck/Korea/H1840/2014	Domestic duck	2014-11-07	Jeonbuk(JB)	EPI573232	QIA	QIA
A/Korean native chicken/Korea/H1847/2014	Domestic chicken	2014-11-12	Jeonnam(JN)	EPI573233	QIA	QIA
A/broiler duck/Korea/H1864/2014	Domestic duck	2014-11-15	Jeonbuk(JB)	EPI573234	QIA	QIA
A/Korean native chicken/Korea/H1903/2014	Domestic chicken	2014-11-24	Gyeongbuk(GB)	EPI573235	QIA	QIA
A/spot-billed duck/Korea/H1981/2014	Wild duck	2014-12-16	Chungbuk(CB)	EPI573236	QIA	QIA
A/mallard/Korea/H1991/2014	Wild duck	2014-12-18	Gyeonggi(GG)	EPI573237	QIA	QIA
A/mallard/Korea/H2003/2014	Wild duck	2014-12-19	Gyeonggi(GG)	EPI573238	QIA	QIA
A/mallard/Korea/H1924-6/2014	Domestic duck	2014-12-01	Jeonnam(JN)	EPI573239	QIA	QIA
A/broiler duck/Korea/H29/2014	Domestic duck	2014-01-18	Jeonbuk(JB)	KJ508900		
A/broiler duck/Korea/H31/2014	Domestic duck	2014-01-19	Jeonbuk(JB)	KJ508908		
A/broiler duck/Korea/H32/2014	Domestic duck	2014-01-19	Jeonbuk(JB)	KJ508916		
A/bean goose/Korea/H40/2014	Domestic goose	2014-01-19	Jeonbuk(JB)	KJ508924		
A/Baikal teal/Korea/H41/2014	Wild duck	2014-01-19	Jeonbuk(JB)	KJ508932		
A/broiler duck/Korea/H47/2014	Domestic duck	2014-01-20	Jeonbuk(JB)	KJ508940		
A/broiler duck/Korea/H48/2014	Domestic duck	2014-01-20	Jeonbuk(JB)	KJ508948		
A/broiler duck/Korea/H49/2014	Domestic duck	2014-01-20	Jeonbuk(JB)	KJ508956		
A/bean goose/Korea/H53/2014	Wild goose	2014-01-20	Jeonbuk(JB)	KJ508972		
A/Baikal teal/Korea/H62/2014	Wild duck	2014-01-21	Jeonbuk(JB)	KJ508980		
A/broiler duck/Korea/H65/2014	Domestic duck	2014-01-21	Jeonbuk(JB)	KJ508988		
A/Baikal teal/Korea/H66/2014	Wild duck	2014-01-21	Jeonbuk(JB)	KJ508996		
A/Baikal teal/Korea/H68/2014	Wild duck	2014-01-22	Chungnam(CN)	KJ509004		
A/Baikal teal/Korea/H80/2014	Wild duck	2014-01-22	Jeonbuk(JB)	KJ509012		
A/coot/Korea/H81/2014	Wild other	2014-01-22	Jeonbuk(JB)	KJ509020		
A/baikal teal/Korea/H84/2014	Wild duck	2014-01-22	Chungnam(CN)	KJ509028		
A/baikal teal/Korea/H96/2014	Wild duck	2014-01-23	Chungnam(CN)	KJ509036		
A/breeder chicken/Korea/H122/2014	Domestic chicken	2014-01-24	Chungnam(CN)	KJ509044		
A/breeder duck/Korea/H128/2014	Domestic duck	2014-01-25	Jeonnam(JN)	KJ509052		
A/broiler duck/Korea/H133/2014	Domestic duck	2014-01-25	Jeonbuk(JB)	KJ509060		
A/broiler duck/Korea/H145/2014	Domestic duck	2014-01-25	Jeonnam(JN)	KJ509068		
A/breeder duck/Korea/H158/2014	Domestic duck	2014-01-26	Chungnam(CN)	KJ509076		
A/mallard/Korea/H207/2014	Wild duck	2014-01-27	Jeonnam(JN)	KJ509092		
A/white fronted goose/Korea/H231/2014	Wild goose	2014-01-28	Gyeonggi(GG)	KJ509100		
A/breeder duck/Korea/H249/2014	Domestic duck	2014-01-28	Jeonnam(JN)	KJ509108		

Continued on next page

Virus name	Host	Isolation date	Location	GISAID or GenBank accessions	GISAID originating lab [1]	GISAID submitting lab
A/breeder chicken/Korea/H250/2014	Domestic chicken	2014-01-28	Gyeonggi(GG)	KJ509116		
A/Korean native chicken/Korea/H257/2014	Domestic chicken	2014-01-28	Gyeongnam(GN)	KJ509124		
A/mallard/Korea/H297/2014	Wild duck	2014-01-29	Jeonnam(JN)	KJ509132		
A/mallard/Korea/W452/2014	Wild duck	2014-02-05	Chungbuk(CB), Chungcheongbuk-do, (Mihocheon River)	KJ746111		
A/spot billed duck/Korea/H455-42/2014	Wild duck	2014-02-08	Chungnam(CN)	KJ509164		
A/waterfowl/Korea/S005/2014	Wild other	2014-02-15	Chungnam(CN), Pungse River	KJ511812		
A/duck/Beijing/FS01/2013	Domestic duck	2013-11-10	China, Beijing Municipality/ Fangshan District	EPI542617	IoM	IoM
A/duck/Zhejiang/W24/2013	Domestic duck	2013-11-14	China/ Zhejiang	KJ476669		
A/mallard/Shanghai/SH-9/2013	Wild duck	2013-11-18	China, Shanghai Municipality	KJ476657		
A/duck/Zhejiang/6D18/2013	Domestic duck	2013-12-14	China/ Zhejiang	KJ476670		
A/duck/Beijing/FS01/2014	Domestic duck	2014-01-20	China, Beijing Municipality/ Fangshan District	EPI543002	IoM	IoM
A/duck/Beijing/CT01/2014	Domestic duck	2014-01-22	China, Beijing Municipality/ Shunyi District	EPI543010	IoM	IoM
A/chicken/Kumamoto/1-7/2014	Domestic chicken	2014-04-13	Kumamoto (Japan)	EPI517161	NIAH	NAFRO
A/wigeon/Sakha-1/2014	Wild duck	2014-09-25	Russian Federation, Sakha (Yakutia) Republic	EPI553349 EPI553350	RRVBVT	RRVBVT
A/turkey/Germany-MV/R2472-14/2014	Domestic turkey	2014-11-04	Germany, Mecklenburg-Vorpommern	EPI544756 EPI544759		FLI
A/turkey/Germany/R2474-L00899/2014	Domestic turkey	2014-11-04	Germany, Mecklenburg-Vorpommern	EPI552746 EPI552748		FLI
A/chicken/Netherlands/14015526-14/2014	Domestic chicken	2014-11-14	Netherlands, Provincie Utrecht, Gemeente Oudewater	EPI547678 EPI547683	CVI	CVI
A/duck/England/36254/2014	Domestic duck	2014-11-14	East Riding of Yorkshire (England)	EPI547673 EPI547675	APHA	APHA
A/duck/England/36038/2014	Domestic duck	2014-11-14	East Riding of Yorkshire (England)	EPI550848 EPI558007	APHA	APHA
A/duck/England/36226/2014	Domestic duck	2014-11-14	East Riding of Yorkshire (England)	EPI550849 EPI558000	APHA	APHA
A/chicken/Netherlands/14015531-14/2014	Domestic chicken	2014-11-15	Netherlands, Provincie Utrecht, Gemeente Oudewater	EPI548623 EPI548626	CVI	CVI
A/wild duck/Chiba/26-372-48-14/2014	Wild duck	2014-11-18	Chiba (Japan)	EPI548485 EPI548487	NIAH	NIAH
A/wild duck/Chiba/26-372-61-14/2014	Wild duck	2014-11-18	Chiba (Japan)	EPI548493 EPI548495	NIAH	NIAH
A/chicken/Netherlands/emc-3/2014	Domestic chicken	2014-11-21	Netherlands, South Holland, Gemeente Nieuwkoop	EPI552776 EPI552778	EMC	EMC
A/crane/Kagoshima/KU1/2014	Wild other	2014-11-23	Kagoshima (Japan)	EPI553208	KU	KU

Continued on next page

Virus name	Host	Isolation date	Location	GISAID or GenBank accessions	GISAID originating lab [1]	GISAID submitting lab
A/eurasian wigeon/Netherlands/emc-2/2014	Wild duck	2014-11-24	Netherlands, Provincie Utrecht, Gemeente Woerden. Between Kamerik and Kockengen	EPI551149 EPI551150	EMC	EMC
A/eurasian wigeon/Netherlands/emc-1/2014	Wild duck	2014-11-24	Netherlands, Provincie Utrecht, Gemeente Woerden. Between Kamerik and Kockengen	EPI551143 EPI551144	EMC	EMC
A/environment/Kagoshima/KU-ngr-H/2014	Wild other*	2014-12-01	Kagoshima (Japan)	EPI553362 EPI553364	KU	KU
A/turkey/Italy/14VIR7898-10/2014	Domestic turkey	2014-12-15	Italy, Veneto, Province of Rovigo	EPI553144 EPI555068	IZS	IZS
A/turkey/Germany-NI/R3372/2014	Domestic turkey	2014-12-15	Lower Saxony (Germany)	EPI553172 EPI553152		FLI
A/chicken/Miyazaki/7/2014	Domestic chicken	2014-12-16	Miyazaki (Japan)	EPI553343		NIAH
A/breeder duck/Korea/H345/2014	Domestic duck	2014-02-02	Chungbuk (CB)	EPI573240	QIA	QIA
A/breeder duck/Korea/H566/2014	Domestic duck	2014-02-17	Chungbuk (CB)	EPI573241	QIA	QIA
A/Common Teal/Korea/H844/2014	Wild duck	2014-03-04	Chungnam (CN)	EPI573242	QIA	QIA
A/American green-winged teal/Washington/195750/2014	Wild duck	2014-12-29	Washington (USA)	KP739421		
A/Northern pintail/Washington/40964/2014	Wild duck	2014-12-08	Washington (USA)	KP307976		
A/chicken/Oregon/41613-2/2014	Domestic chicken	2014-12-16	Oregon (USA)	KP739405		
A/guinea fowl/Oregon/41613-1/2014	Domestic other	2014-12-16	Oregon (USA)	KP739413		
A/gyrfalcon/Washington/41088-6/2014	Domestic other	2014-12-08	Washington (USA)	KP307984		
A/turkey/BC/FAV10/2014	Domestic turkey	2014-12-02	British Columbia (Canada)	KP307957		
A/domestic duck/Washington/61-16/2014	Domestic duck	2014-12-30	Washington (USA)	KP739389		
A/chicken/Washington/61-9/2014	Domestic chicken	2014-12-30	Washington (USA)	KP739381		
A/turkey/Washington/61-22/2014	Domestic turkey	2014-12-30	Washington (USA)	KP739397		

[1]QIA: Animal and Plant Quarantine Agency (QIA), Republic of Korea, IoM: Institute of Microbiology, Chinese Academy of Science, CVI: Central Veterinary Institute, Netherlands, RRVBVT: State Research Center of Virology and Biotechnology Vector, Russian Federation, NIAH: National Institute of Animal Health, Japan, APHA: Animal and Plant Health Agency (APHA), UK, EMC: Erasmus Medical Center, Netherlands, IZS: Istituto Zooprofilattico Sperimentale Delle Venezie, Italy, KU: Kagoshima University, Japan, NAFRO: National Agriculture and Food Reseach Organization, Japan, FLI: Friedrich-Loeffler-Institut, Germany

Appendix C

Appendix to Chapter 4

C.1 Figures

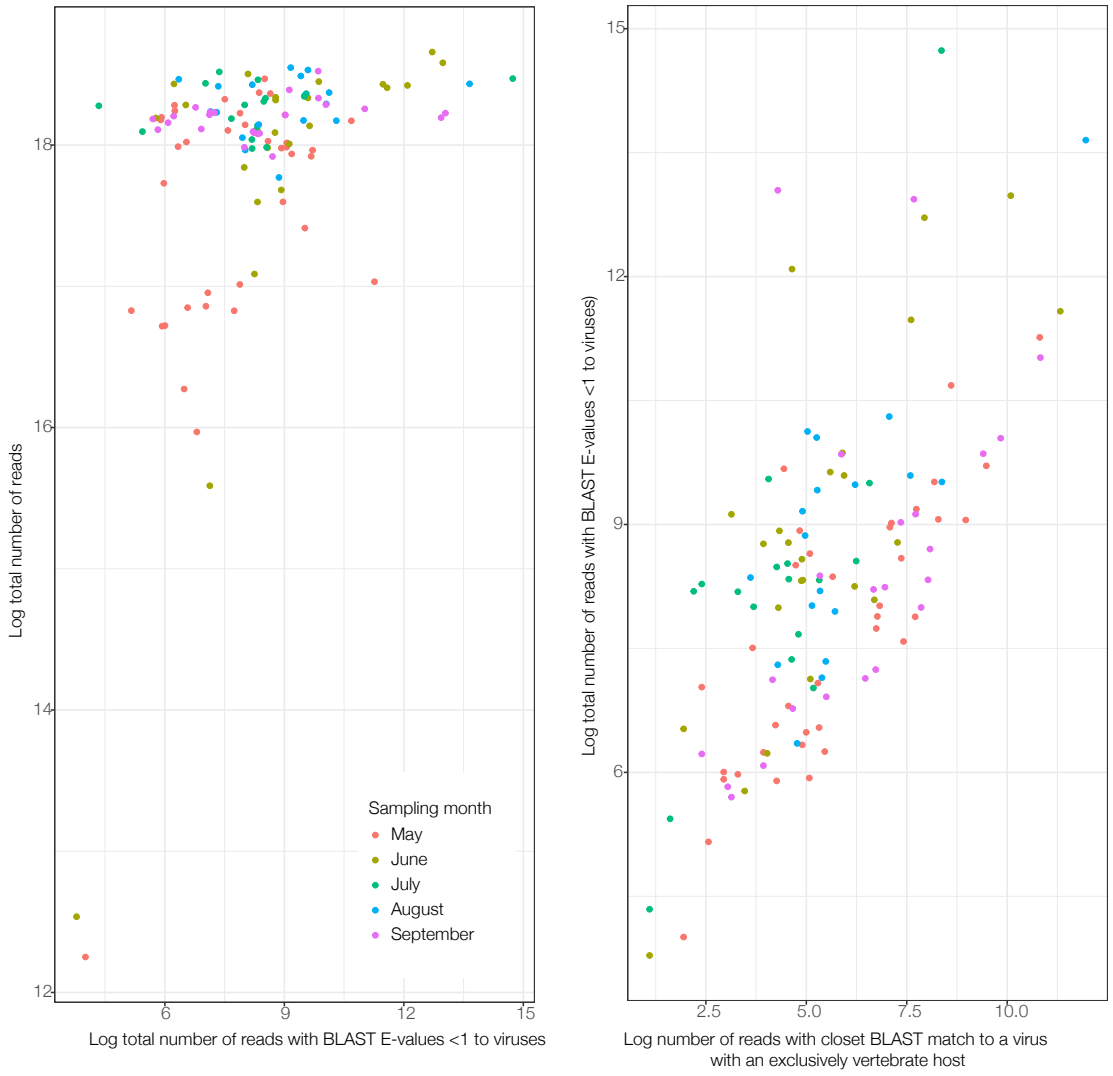

Figure C.1: Log-transformed counts per sample of sequencing reads obtained, reads that were similar to known viruses based on BLAST matching, and reads matching to viruses that exclusively occur in vertebrates. Colours represent sampling visits. Note that the lower number of sample reads obtained for a proportion of the May 2016 samples reflects the fact that these samples were sequenced in a different Illumina run than the remaining samples, that in general obtained fewer reads.

C.2 Tables

Table C.1: Number of times that each uniquely identifiable bird was sampled at Abbotsbury per sample type

Times sampled	Number of birds	Sample type
1	292	Whole feces
2	90	Whole feces
3	26	Whole feces
4	12	Whole feces
5	5	Whole feces
6	1	Whole feces
1	108	Swabs of feces
2	68	Swabs of feces
3	30	Swabs of feces
4	17	Swabs of feces
5	9	Swabs of feces
6	2	Swabs of feces
7	1	Swabs of feces
1	153	Either sample type
2	103	Either sample type
3	66	Either sample type
4	44	Either sample type
5	27	Either sample type
6	21	Either sample type
7	8	Either sample type
8	2	Either sample type
9	1	Either sample type
10	2	Either sample type

Table C.2: Number of samples collected from uniquely identifiable birds per visit

Date	Number of samples	Sample type
2015-04-16	21	Swab of feces
2015-05-14	23	Swab of feces
2015-05-31	29	Swab of feces
2015-06-25	37	Swab of feces
2015-07-24	47	Swab of feces
2015-09-01	54	Swab of feces
2015-10-10	41	Swab of feces
2015-11-24	42	Swab of feces
2015-12-15	72	Swab of feces
2016-01-26	43	Swab of feces
2016-03-08	57	Swab of feces
2016-05-03	85	Whole feces
2016-06-01	96	Whole feces
2016-07-10	72	Whole feces
2016-08-10	85	Whole feces
2016-10-09	91	Whole feces
2016-11-21	52	Whole feces
2017-03-21	92	Whole feces
2017-06-01	56	Whole feces

Table C.3: Whole genome PCR primers for novel viral genomes resequenced using Sanger sequencing

Target sample	Forward primer name	Forward primer sequence (5' to 3')	Reverse primer name	Reverse primer sequence (5' to 3')	Annealing temp °C
UTMF1260	AvibirnaA_F	ACGTGGCTACTAGGGGTGAT	AvibirnaA_R	GGTCACCGTTGTCTGACCAT	68
UTMF1260	AvibirnaB_F	TGGATACTCGCCCCAGAAGA	AvibirnaB_R	TTGTGGAGTTTCTCGGCCTC	68
UTMF1260	AvibirnaC_F	ATTCTCTTGACGGCGCAGAA	AvibirnaC_R	CTTCAGCACACAGTCGAGGT	68
UTMF1260	Avibirna_E_F	ACTCGCCAAATTCCTCAGGG	Avibirna_E_R	GTGCTGTCTCGTGCAGTTTG	68
UTMF1341	CalicivirusA_F	TGATGCACCCATCCAACCTC	CalicivirusA_R	AACAGCACAGAAGGAGACCG	68
UTMF1341	CalicivirusB_F	ATCTGCAGGTCAACCACTGG	CalicivirusB_R	ACAGGCAAGCAATCTACCCC	68
UTMF1341	CalicivirusC_F	GGCACTAGACTTGGCACCAT	CalicivirusC_R	GGGTGCGACAAGTGGAGTAT	68
UTMF1342	CalicivirusD_F	CGTTGTCTGTTTGGCCACTG	CalicivirusD_R	CTGCGACATTCTGAATGCCG	68
UTMF1343	CalicivirusE_F	GTTTTGTGACACACCTCGGC	CalicivirusE_R	GCCCTTTTGAGGCACAACAG	68
UTMF1344	CalicivirusF_F	GCTTTAAGGCTGGCAAACCC	CalicivirusF_R	ACGGATTGCCAGTCAGGAAG	68
UTMF1345	CalicivirusG_F	ACGGCACTGGTCATTTCCTT	CalicivirusG_R	CTCCTCAATGGTGGTGTCCC	68
UTMF1346	CalicivirusH_F	CCGTCTATGCCTCACAACCA	CalicivirusH_R	GCCAACCCAAACAGGCATTT	68

Continued on next page

Target sample	Forward primer name	Forward primer sequence (5' to 3')	Reverse primer name	Reverse primer sequence (5' to 3')	Annealing temp °C
UTMF0973	TuSCV_A_F	TCTATCATGACGCGG AGTGC	TuSCV_A_R	GCTCACCAACACACT ACCGA	68
UTMF0973	TUSCV_B_F	TCTTCTCCAGCGACA CCAAC	TuSCV_B_R	GCACTCCGCGTCATG ATAGA	68
UTMF0973	TUSCV_B_F	TCTTCTCCAGCGACA CCAAC	TuSCV_A_R	GCTCACCAACACACT ACCGA	68

Table C.4: Primers for targeted sequencing PCRs

Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')	Annealing temp (°C)	Extension time (secs)
Coronavirus996_F	ACGAAGCGCAAT GTAATGCC	Coronavirus996_R	AAGGGCAACATA ACGCTCCA	67	30
Astrovirus1076_F	CTGCTAYGGTGA TGATAGRCTCTTG	Astrovirus1076_R	ATCCATTCCCCAC CAACCTG	67	35
Astrovirus1096_F	TGAAAATTAGGGATA CACCCGATG	Astrovirus1096_R	ATCCATATTCCA CCAACCCT	64	35
Astrovirus1076_F	CTGCTAYGGTGA TGATAGRCTCTTG	Astrovirus1096_R	ATCCATATTCCA CCAACCCT	64	35
Adenovirus1347_F	CATCGCATGAGGG TGTCGTA	Adenovirus1347_R	GTTACCACACGG CTCTCCTC	68	60

Table C.5: Primers for SYBR green qPCRs

Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')	Annealing temp (°C)
SYBRmegrivirus_F	ATGGTTGCAGCGTTTAAGGA	SYBRmegrivirus_R	CCACCAGCTCTCACTGAACA	66
SYBRastrovirus_F	AGCGGCCTGTGTTCTGTTAA	SYBRastrovirus_R	CCCGCCTATCCGAAGTTTCA	68

Table C.6: Primers for whole genome multiplex PCRs

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
UTMF1260	Astrovirus_1_LEFT	CACTGCGCACATTAA GGGATGA	Astrovirus_1_RIGHT	AGGCACGTTGTGTA GAGGTGAT
UTMF1260	Astrovirus_2_LEFT	GCTGTGTGCTTGTTA ATGTCTCCT	Astrovirus_2_RIGHT	TTCATAAAACGCCAG CCTGTCC
UTMF1260	Astrovirus_3_LEFT	CGCTAATCAGTGTAA GAAAGGGATAGG	Astrovirus_3_RIGHT	AATATGTTTCTGCCC ACGGGTG
UTMF1260	Astrovirus_4_LEFT	ACCCTGGAACAAGTC TGCAAGA	Astrovirus_4_RIGHT	GTGCCCCTCATTCTC TGAATGC
UTMF1260	Astrovirus_5_LEFT	AGCGTAAGAAAGTGT GCAGATGG	Astrovirus_5_RIGHT	GAATGGAGTCCAACC AACCTGC
UTMF1260	Astrovirus_6_LEFT	TCAGAATGATCCTGT GTACTGATCCT	Astrovirus_6_RIGHT	TGCCTACTATTTCGG TCCTCCC
UTMF1260	Astrovirus_7_LEFT	AGGAAGTCAGAGACT ATTTGGATGCT	Astrovirus_7_RIGHT	CAATGAGCCGTCTGT CCCATTC
UTMF1260	Astrovirus_8_LEFT	ACATGGTGGAGATTA AAGTTACATATCTCT	Astrovirus_8_RIGHT	GGTGCACGAAAGCTA ACCCATT
UTMF1260	Astrovirus_9_LEFT	CGGACAATTGAGACT GACAGGC	Astrovirus_9_RIGHT	GGTGATTATCTTGCC GGTGAGC
UTMF1348	Sapelovirus_1_LEFT	TGATGGAACCTACAT GCACAACA	Sapelovirus_1_RIGHT	CTCCTCGATCCCTTT GGGTGAT
UTMF1348	Sapelovirus_2_LEFT	TGAACCAGACACTAA AGAAGAGGAGA	Sapelovirus_2_RIGHT	TACCATCAACCTCTG CCTCTGG
UTMF1348	Sapelovirus_3_LEFT	AGTGTTTGGTCAAAT GTGCCAGT	Sapelovirus_3_RIGHT	GGGATGAACCAACAT TGTCTGCA
UTMF1348	Sapelovirus_4_LEFT	AGGCTTACAGTCGTC ATTATCATTTACTG	Sapelovirus_4_RIGHT	GGCCTTGCATAGACG GAAAACA

Continued on next page

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
UTMF1348	Sapelovirus_5_LEFT	ACAACACAGTATAATGGCTTTAATGGATT	Sapelovirus_5_RIGHT	GTTGCTTTCTGAGTGTACCGCA
UTMF1348	Sapelovirus_6_LEFT	GCAAGTTTACAGAGTCTAGTGTTGTCA	Sapelovirus_6_RIGHT	TCTTTGGATCATCAACGCTGCG
UTMF1348	Sapelovirus_7_LEFT	TCCAAGGAAAGAATGGATGTGTTGT	Sapelovirus_7_RIGHT	GTGCTGGTCTCTCTGGTGATCT
UTMF1348	Sapelovirus_8_LEFT	GTGGTGGTCTTATCTGCAAGGC	Sapelovirus_8_RIGHT	TACCAGATGGCATCCCACCTTC
UTMF1348	Sapelovirus_9_LEFT	ATTCCACCAAGAGGAGAACCCC	Sapelovirus_9_RIGHT	CAGGAATGGTAAAAGCCCGTCC
UTMF0980	Duckpicorna_1_LEFT	TTCCACAAAAGATTCCCCTGGC	Duckpicorna_1_RIGHT	AGCTGCCACGCCAATATTCAAA
UTMF0980	Duckpicorna_2_LEFT	TCCATCTTCAATACTGGGACTTGGT	Duckpicorna_2_RIGHT	CCAAATGTGGGTTGTCCGAGAG
UTMF0980	Duckpicorna_3_LEFT	TGTCAGTGCGTGGATGTATTTTGA	Duckpicorna_3_RIGHT	AATATCCTTGTCCCATGGGCCA
UTMF0980	Duckpicorna_4_LEFT	TCTTCATGCTTTCCCGCCAAAC	Duckpicorna_4_RIGHT	TGCCAGGAGAGTGATTGTCCAT
UTMF0980	Duckpicorna_5_LEFT	GCCGGCTAATTGGGTCAGCTAA	Duckpicorna_5_RIGHT	CTGTGCGGCATTTCACAATGTG
UTMF0980	Duckpicorna_6_LEFT	TCAAGTGTTTTGAGTGGATTCGGA	Duckpicorna_6_RIGHT	CTCTTGGGCGGATTAGGGAAGA
UTMF0980	Duckpicorna_7_LEFT	TGAGCTAGATTCAAGGGACGCC	Duckpicorna_7_RIGHT	GCACAGGGGACTCCTCATACTT
UTMF0980	Duckpicorna_8_LEFT	TAGCTGGAAATCGCAACATGGG	Duckpicorna_8_RIGHT	CCAGCAAGTTACACACCGTGTT
UTMF0980	Duckpicorna_9_LEFT	ACTGTTCAATGACCATGATCGCTC	Duckpicorna_9_RIGHT	CGGCAGGTGTTTTACGTGAACT

Continued on next page

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
UTMF1123	Passeriviru_1_LEFT	TAAAGTCGACCCTCA CCTAGCC	Passeriviru_1_RIGHT	TTGAGTGGGGCTGAT CCATCAT
UTMF1123	Passeriviru_2_LEFT	CTCGGCTCATAGATT GATTGGCT	Passeriviru_2_RIGHT	AGTGTGTGTTGTACA TCGCAGC
UTMF1123	Passeriviru_3_LEFT	TGGAAACAACCAATG GAGAATCCTG	Passeriviru_3_RIGHT	GCCATGTCAACCGTC TTCCAAT
UTMF1123	Passeriviru_4_LEFT	TGAGCCACCTACCAC TATGCAA	Passeriviru_4_RIGHT	AAATCTCCTCCAGCC TGTCTCC
UTMF1123	Passeriviru_5_LEFT	AACCCCATTCTTTGC CTTTGGT	Passeriviru_5_RIGHT	AGGCTCAGCTTCTTC TACCACA
UTMF1123	Passeriviru_6_LEFT	TGCTGACTATACTCC TGATGTTGTCT	Passeriviru_6_RIGHT	CGTTGCAGGTCCATC AGGAGTA
UTMF1123	Passeriviru_7_LEFT	AATAACACATCTGTT CGGGCCC	Passeriviru_7_RIGHT	GCCATCCCTGTTGAC CTTCCTA
UTMF1123	Passeriviru_8_LEFT	ATAAGGACATTCGCC GCTTCCT	Passeriviru_8_RIGHT	ACCAGTTCACTTCAG GGTCACA
UTMF1123	Passeriviru_9_LEFT	ATTGAAGCTGCCCCT ATGCATG	Passeriviru_9_RIGHT	TAGTTGCCGGGTTTT ACTCCCA
UTMF1123	Megrivirus_1_LEFT	CGGTATTCTAGACTC TCCCTGTCA	Megrivirus_1_RIGHT	CGGTTGAAGTCTGTG GAGCAAG
UTMF1123	Megrivirus_2_LEFT	GTTCCAAACCCATGG TCTCTACC	Megrivirus_2_RIGHT	TCCTGGGTCTAAAGT CTGCAGG
UTMF1123	Megrivirus_3_LEFT	GGCCCTATCTATCTT CACTGGCT	Megrivirus_3_RIGHT	CCCTTGTTGAGAAGA CTGTGGC
UTMF1123	Megrivirus_4_LEFT	ATGTTGTGGTTGGAT ATCGCCC	Megrivirus_4_RIGHT	AGCGTTGCCTTCCAT TACAGTG
UTMF1123	Megrivirus_5_LEFT	ATTCTTGAAACTGAG CGCTCGG	Megrivirus_5_RIGHT	TGAGTGTTGATCCGT ACCCCTG

Continued on next page

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
UTMF1123	Megrivirus_6_LEFT	TCTCTCTTGAGGAGG AATTCGATGA	Megrivirus_6_RIGHT	CTCCTGCTTTCATCG ACGTGTC
UTMF1123	Megrivirus_7_LEFT	AGCCAATACTTCTGC AGTTGCC	Megrivirus_7_RIGHT	ACAGCCTGGTCCACC ATGAATA
UTMF1123	Megrivirus_8_LEFT	CAACCCGCAGCTCTT CTCTCTA	Megrivirus_8_RIGHT	CACAATGTGGGTCAA GACAGGG
UTMF1123	Megrivirus_9_LEFT	CCGTAAAGGAGATGA AGTCTGGC	Megrivirus_9_RIGHT	CCTCTTCATGTGTGT CCTTGGC
UTMF1123	Megrivirus_10_LEFT	CGCTTCTTTCTGTCT GTCAAGGA	Megrivirus_10_RIGHT	CGCTCAAACATGGGT CCAAAGA
UTMF1123	Megrivirus_11_LEFT	CTTCTTAAAGGATGA GCTTAGGCCC	Megrivirus_11_RIGHT	TGCCCATCACATGAA GTTTGCA
UTMF1123	Megrivirus_12_LEFT	CCGGAAACATCTACT ATTTATGATGTCACT	Megrivirus_12_RIGHT	AAGGGCAAGGGGACT GATAGAC
UTMF1260	PigeonParvo_A_F	GCCTGCTTATCCTTT GCAGC	PigeonParvo_A_R	TGCAGCATTCTTCCC ACCAT
UTMF1260	PigeonParvo_B_F	GCACTCTGCAAAATG GTAGGC	PigeonParvo_B_R	AAGTCCTGGTGGACA TGCTTT
UTMF1260	PigeonParvo_C_F	CGGGTATTTCTTCTC TCTCTCCTG	PigeonParvo_C_R	GGTCTCCATTTGCTG GCTGT
UTMF1260	PigeonParvo_D_F	CAACTCCAGGTTGCT GGCTA	PigeonParvo_D_R	AGGGTTCCATTGGGC ACTTT
UTMF1388	Coronavirus_1_LEFT	AGGACCTTATAGATC ACATTGTTACTGATT	Coronavirus_1_RIGHT	ACTGTCTCGTTGTCA CCACTCA
UTMF1388	Coronavirus_2_LEFT	CTGGTAAGGGTGTGG CTTTTGT	Coronavirus_2_RIGHT	GAACGACACCGCATT TAACAACA
UTMF1388	Coronavirus_3_LEFT	CAAAATTCATTTGCA GAACATCTTTGGT	Coronavirus_3_RIGHT	TCGCACAAAACACAA CAAAGCC

Continued on next page

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
UTMF1388	Coronavirus_4_LEFT	CCTGGTGGGTCTTGG GTGTATA	Coronavirus_4_RIGHT	GGCGAATCTGAGTGG AAACTGG
UTMF1388	Coronavirus_5_LEFT	GGCAGATCTTCAATA TGCTGACCC	Coronavirus_5_RIGHT	GGACCAATTCCACTC ACACTCAC
UTMF1388	Coronavirus_6_LEFT	GGAAATTCAGCTCTA GGAAGTTCAGAG	Coronavirus_6_RIGHT	AAAACCTAAGGGTGT GCCGTTC
UTMF1388	Coronavirus_7_LEFT	TGTCCAGTACCTCCT GATAAAACTGT	Coronavirus_7_RIGHT	GTCCATGTGCTGTAC CCTCGAT
UTMF1388	Adenovirus1347_F	CATCGCATGAGGGTG TCGTA	Adenovirus1347_R	GTTACCACACGGCTC TCCTC
	Bornavirus_1_LEFT	TAAAAGATGCCACCC AAACGCC	Bornavirus_1_RIGHT	TACTCCAGTAGAATG CCGCAGA
	Bornavirus_10_LEFT	TAAGGCCGCGAGATA TGTCTGC	Bornavirus_10_RIGHT	CCGCATAGAGGCAAC CAAATGT
	Bornavirus_11_LEFT	CTGGGTATGTAAACA TAATTAGTACCTGGT	Bornavirus_11_RIGHT	TCCCTTTACCTGCCA GCGAATA
	Bornavirus_12_LEFT	GGCAAATATAGACCT CTATTTAGCATCGG	Bornavirus_12_RIGHT	CTCAGAGGGAATGTC CAGCAGA
	Bornavirus_13_LEFT	GCCCAATTACTCAAA ATGGTCATAAGG	Bornavirus_13_RIGHT	ACACAATGACAATGA CGATAACCACA
	Bornavirus_2_LEFT	AGTGTCGTCAGAGCT TAAGTCTGA	Bornavirus_2_RIGHT	CGCGGATGAAGCATA AAGTAGGT
	Bornavirus_3_LEFT	CAAGTAGAGACAATC CAGGCTATTCA	Bornavirus_3_RIGHT	GGTAACTAACCTCGC ATGAAATGACA
	Bornavirus_4_LEFT	CTGACATCCGCGTTT CAAGCTC	Bornavirus_4_RIGHT	TGACGCACTGGTTAA ACCTGAAC
	Bornavirus_5_LEFT	AGAGATCTCTCACAG ATGGAATATTTAGTC	Bornavirus_5_RIGHT	TGCAAAATTCAACTT CTGTACCAATTAATG

Continued on next page

Sample	Forward primer	Sequence (5'-3')	Reverse primer	Sequence (5'-3')
	Bornavirus_6_LEFT	TGGTGGAATACTATA TCTCATCACACTTT	Bornavirus_6_RIGHT	TGGGATTGAGGCGCA AAAGTTT
	Bornavirus_7_LEFT	TGTCATTTACTAATT GCTTCTGTAGTCCA	Bornavirus_7_RIGHT	CTACGGAGCCCTTGT AAAACGG
	Bornavirus_8_LEFT	ACCGGAAGAAACATG GTGATCAAT	Bornavirus_8_RIGHT	ATTGCTTGAGACTGC CTGGAAC
	Bornavirus_9_LEFT	GTTCATTATACAAGC ATTCACGATTAGCA	Bornavirus_9_RIGHT	CGACTCTCTAACACT ACATCTTCCAA

Table C.7: Viral read abundance between birds of different age per taxonomical unit

Taxonomic unit	Age younger[1]	p-values	Adjusted p-values [2]	Taxonomic level
Abbotsbury bornavirus (Aquatic bird bornavirus 1)	-1.84	0.00	0.00	Abbotsbury species
Abbotsbury Swan stool associated circular virus	-0.40	0.01	0.03	Abbotsbury species
Abbotsbury swan picornavirus	0.70	0.00	0.00	Abbotsbury species
Bornaviridae	-4.24	0.00	0.00	Family
Astroviridae	1.89	0.00	0.00	Family
Adenoviridae	2.79	0.00	0.00	Family
Picornaviridae	2.83	0.00	0.00	Family
Bornavirus	-4.20	0.00	0.00	Genus
Avisivirus	1.91	0.00	0.00	Genus
Protoparvovirus	2.01	0.00	0.01	Genus
Aveparvovirus	2.09	0.00	0.00	Genus
Siadenovirus	2.49	0.00	0.00	Genus
Avihepatovirus	2.53	0.00	0.00	Genus
Porcine picobirnavirus	-3.67	0.00	0.00	Species
Po-Circo-like virus 41	-2.05	0.00	0.00	Species
Avisivirus Pf-CHK1/AsV	1.51	0.00	0.00	Species
Turkey parvovirus 1078	2.35	0.00	0.00	Species
unidentified adenovirus	2.89	0.00	0.00	Species
Duck picornavirus GL/12	3.22	0.00	0.00	Species

[1]Log2-fold changes of occuring in birds <3 years compared to >2 years
[2]Adjusted for multiple testing of different OTUs

Table C.8: Viral read presence/absence differences birds of different age per taxonomical unit

Taxonomic unit	Odds ratio	Lower estimate	Upper estimate	p-values	Adjusted p-values [1]	TaxonomicLevel
Adenoviridae	9.17	3.01	30.83	1.20e-05	1.20e-04	Family
Bornaviridae	0.00	0.00	0.47	1.56e-03	7.81e-03	Family
Aviadenovirus	8.65	1.90	54.39	1.49e-03	8.43e-03	Genus
Avihepatovirus	9.99	2.64	47.28	1.09e-04	2.35e-03	Genus
Avisivirus	13.08	2.40	134.37	5.09e-04	3.44e-03	Genus
Bocaparvovirus	10.12	2.30	62.71	4.53e-04	3.44e-03	Genus
Bornavirus	0.00	0.00	0.47	1.56e-03	8.43e-03	Genus
Cosavirus	3.84	1.34	11.21	7.92e-03	3.57e-02	Genus
Enterovirus	4.03	1.58	11.11	1.88e-03	9.22e-03	Genus
Mastadenovirus	13.08	2.40	134.37	5.09e-04	3.44e-03	Genus
Orthoreovirus	Inf	3.43	Inf	3.19e-04	3.44e-03	Genus
Parechovirus	9.99	2.64	47.28	1.09e-04	2.35e-03	Genus
Siadenovirus	11.71	2.73	71.63	1.31e-04	2.35e-03	Genus
Teschovirus	13.08	2.40	134.37	5.09e-04	3.44e-03	Genus
Avian orthoreovirus	Inf	3.43	Inf	3.19e-04	6.51e-03	Species
Avisivirus Pf-CHK1/AsV	Inf	3.43	Inf	3.19e-04	6.51e-03	Species
Bat picornavirus 1	4.94	1.77	14.25	8.29e-04	9.39e-03	Species
Bat picornavirus 3	4.28	1.47	12.82	3.37e-03	2.65e-02	Species
Bovine picornavirus	3.95	1.42	11.19	4.72e-03	3.21e-02	Species
Canine picornavirus	4.47	1.63	12.61	2.21e-03	1.88e-02	Species
Chicken picornavirus 4	3.25	1.30	8.28	7.76e-03	4.40e-02	Species
Duck picornavirus GL/12	12.94	3.53	60.32	8.00e-06	8.16e-04	Species
Enterovirus A	5.46	1.86	16.80	9.37e-04	9.56e-03	Species
Enterovirus C	7.94	2.69	25.29	2.85e-05	1.45e-03	Species
Enterovirus SEV-gx	9.12	1.52	97.57	5.65e-03	3.60e-02	Species

Continued on next page

Taxonomic unit	Odds ratio	Lower estimate	Upper estimate	p-values	Adjusted p-values [1]	TaxonomicLevel
Feline picornavirus	3.06	1.23	7.75	9.21e-03	4.95e-02	Species
Human parechovirus	7.29	1.53	46.81	4.62e-03	3.21e-02	Species
Porcine bocavirus	11.02	1.94	115.22	1.74e-03	1.62e-02	Species
Quail picornavirus QPV1/HUN/2010	3.60	1.31	10.01	6.38e-03	3.83e-02	Species
Rhinovirus A	4.76	1.80	12.94	6.95e-04	8.86e-03	Species
Rhinovirus C	4.58	1.77	12.15	5.79e-04	8.44e-03	Species
Teschovirus A	13.08	2.40	134.37	5.09e-04	8.44e-03	Species
unidentified adenovirus	26.34	3.28	1212.16	1.27e-04	4.33e-03	Species
Abbotsbury bornavirus (Aquatic bird bornavirus 1)	0.00	0.00	0.60	5.49e-03	4.67e-02	Abbotsbury_species
Abbotsbury swan picornavirus	18.61	5.19	86.02	1.06e-07	1.80e-06	Abbotsbury_species

[1]Adjusted for multiple testing of different OTUs

Table C.9: Unique features that occur only in one age group

Taxonomic Unit	Number of positive samples in birds >2 years	Number of positive samples in birds <2 years	Number of reads in birds >2 years	Number of reads in birds <3 years	TaxonomicLevel
Bornaviridae	19.00	0.00	970.00	0.00	Family
Bornavirus	19.00	0.00	970.00	0.00	Genus
Orthoreovirus	0.00	6.00	0.00	609.00	Genus
Avian orthoreovirus	0.00	6.00	0.00	609.00	Species

Appendix D

Appendix to Chapter 5

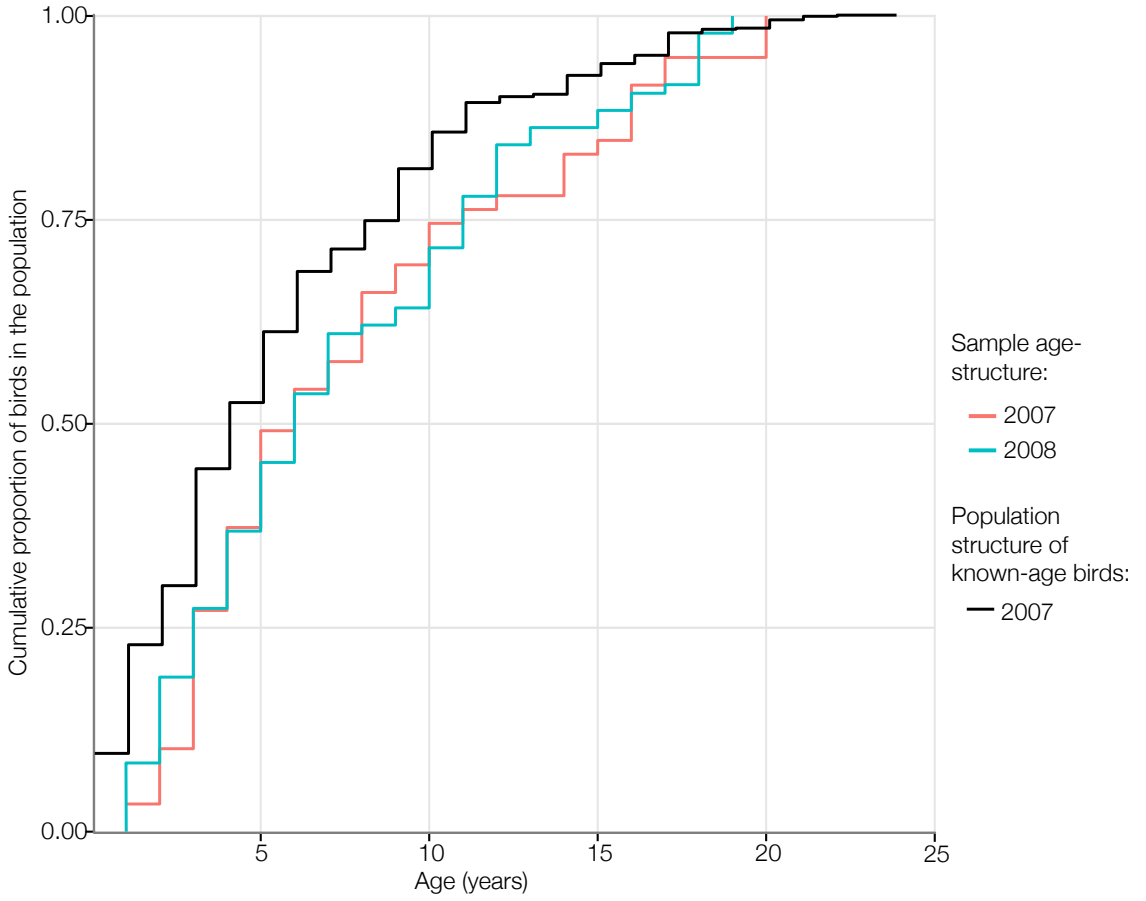

Figure D.1: Cumulative proportion of birds sampled during each sampling year (red and blue lines) and in the population (black lines). The population age-structure is based on birds with known ages caught or known to be alive (breeding birds) during a biennial catch of all adult birds on July 21st 2007. Because cygnets born during 2007 at Abbotsbury were not caught on this day, the age-structure also includes all Abbotsbury cygnets that were ringed after July 2007 (typically during late August to mid October). Cygnet mortality during late summer and prior to ringing may mean that the number of cygnets alive during July 2007 is slightly underestimated here. In total, approximately 80% of the birds believed to be in the population on July 21st 2007 had known ages. Birds of unknown ages typically entered the population as adults and could not be aged.

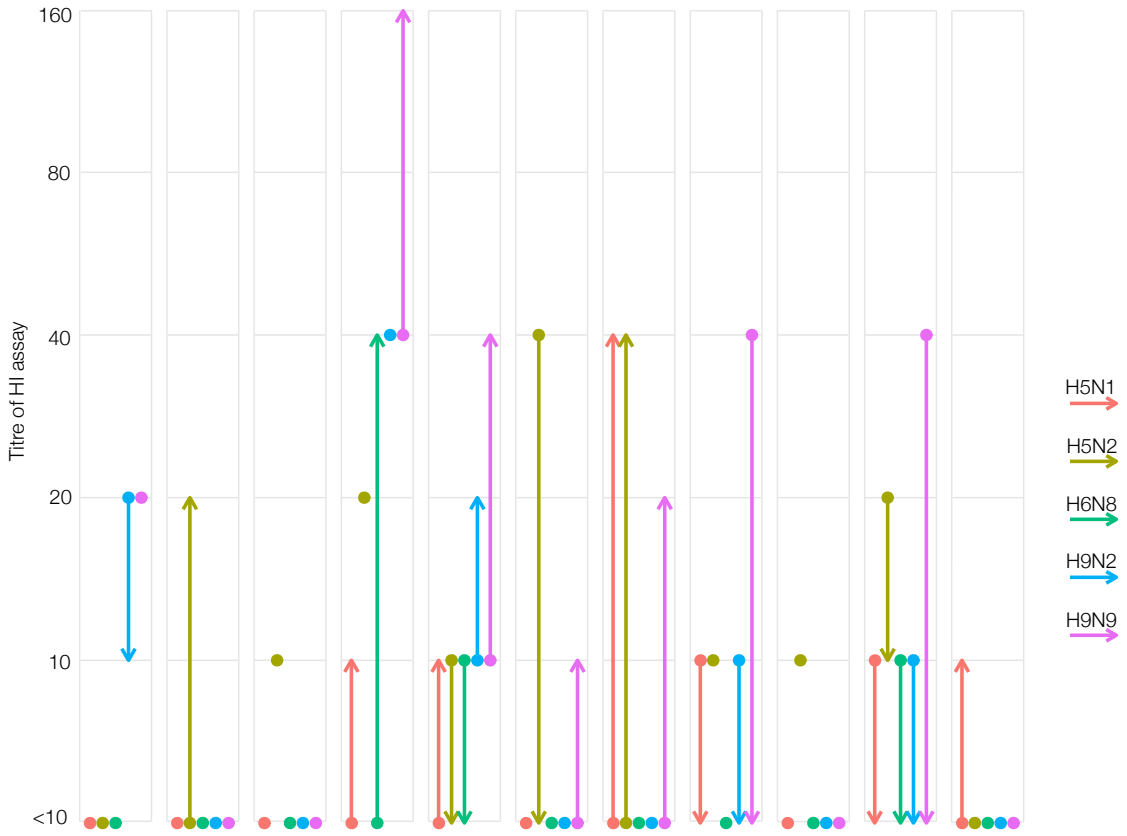

Figure D.2: Change in HI titre for each antigen (colours) between 2007 and 2008 for each of the eleven birds that were sampled in both years. Dots represent 2007 values, such that lone dots show not change in titre. Arrows show the change from 2007 to 2008. Many titres are not stable between years. Values less than or equal to 10 are considered seronegative in this study.

D.1 Tables

Table D.1: Generalised linear models for individual subtypes

Dataset	Age [1]	Sample year (2008)	Sex (M)
Dataset A (H5N2, H6N8, H9N2)	0.068 ($p<0.05$)	-0.033 ($p=0.927$)	-0.885 ($p<0.05$)
Dataset B (either H5, H6N8, either H9)	0.0790 ($p<0.05$)	0.114 ($p=0.750$)	-0.601 ($p=0.095$)
Dataset C (responds to both H5N1 and H5N2, H6N8, both H9N2 and H9N9)	0.062 ($p=0.105$)	0.599 ($p=0.184$)	-0.653 ($p=0.158$)
Dataset D (H5N1, H6N8, H9N9)	0.081 ($p<0.01$)	0.579 ($p=0.142$)	-0.157 ($p=0.688$)
Dataset E (H5N2, H6N8, both H9N2 and H9N9)	0.068 ($p<0.05$)	0.077 ($p=0.834$)	-0.751 ($p<0.05$)

[1] P-values for each of the coefficient estimates are given in brackets

Table D.2:

Model	Age [1]	Sample year (2008)	Sex (M)	Interaction effect
H5N1, Age+SampleYear+Sex	-0.0449 (p=0.434)	1.64 (p=0.0365)	0.380 (p=0.494)	
H5N2, Age*SampleYear+Sex	0.156 (p=0.0184)	1.13 (p=0.111)	-0.768 (p=0.0540)	-0.184 (p=0.0229)
H6N8, Age+ SampleYear+Sex	0.111 (p=0.0950)	-0.0539 (p= 0.944)	-1.51 (p=0.167)	
H9N2, Age+ SampleYear+Sex	0.0959 (p=0.0190)	0.115 (p=0.803)	-1.01 (p=0.0477)	
H9N9, Age+ SampleYear+Sex	0.129 (p= 0.000154)	0.294 (p=0.501)	-0.463 (p=0.303)	

[1] P-values for each of the coefficient estimates are given in brackets

Table D.3:

Dataset	Change in AIV IDEXX value between years (p-value) [1]
Dataset A (H5N2, H6N8, H9N2)	-12.657 (0.0859)
Dataset B (either H5, H6N8, either H9)	-46.28 (0.0246)
Dataset C (responds to both H5N1 and H5N2, H6N8, both H9N2 and H9N9)	-10.079 (0.14)
Dataset D (H5N1, H6N8, H9N9)	-34.98 (0.0192)
Dataset E (H5N2, H6N8, both H9N2 and H9N9)	-12.657 (0.0859)

[1]Note that negative coefficients indicate an association between decreasing raw results of NP ELISA (i.e., increased levels of NP antibody) and an increase in breadth of response to different HA types.

Table D.4:

Dataset	AIV IDEXX ELISA (dataset including all raw ELISA data)	AIV IDEXX ELISA (dataset including raw ELISA data <0.5 only) [1]
Dataset A (H5N2, H6N8, H9N2)	1.872 (6.6e-16)	1.353 (0.000397)
Dataset B (either H5, H6N8, either H9)	1.954 (4.62e-07)	1.585 (1.13e-05)
Dataset C (responds to both H5N1 and H5N2, H6N8, both H9N2 and H9N9)	2.411 (0.000262)	1.930 (0.000812)
Dataset D (H5N1, H6N8, H9N9)	2.665 (8.81e-06)	2.385 (1.05e-05)
Dataset E (H5N2, H6N8, both H9N2 and H9N9)	1.940 (1.22e-05)	1.418 (0.000515)

[1]Values <0.5 indicates the bird is seropositive for previous AIV infection by NP-ELISA. This dataset was used to check that the association was not being driven by high ELISA values occurring for birds with no breadth of response, and low ELISA values at breadths greater than 1. P-values are given in brackets next to the coefficient estimates.

Table D.5:

Dataset	Age	Sex (M)	Sample Year (2007)
Dataset F (including all raw ELISA data)	0.182 (0.00134)	-1.41 (0.00366)	1.48 (0.00219)
Dataset G (including raw ELISA data <0.5 only [1])	0.137 (0.0322)	-1.01 (0.0794)	-1.160 (0.0438)

[1]Values <0.5 indicates the bird is seropositive for previous AIV infection by NP-ELISA. This dataset was used to check that the association was not being driven by high ELISA values occurring for birds with no breadth of response, and low ELISA values at breadths greater than 1. P-values are given in brackets next to the coefficient estimates.

Appendix E

Appendix to Chapter 6

E.1 Figures

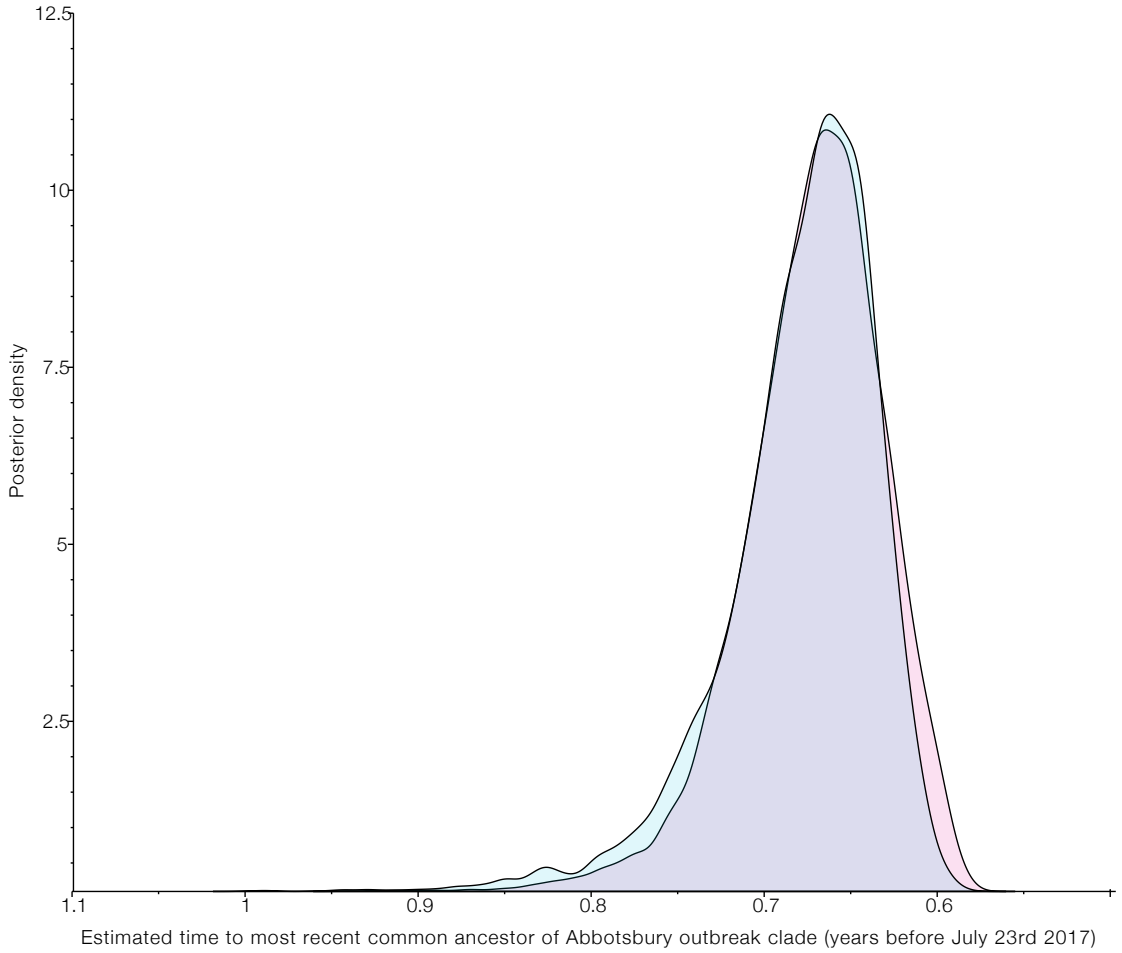

Figure E.1: Kernel density estimated distribution of posterior support for the time to most recent common ancestor of the Abbotsbury outbreak clade. The NA estimated distribution is shown in pink, and the HA estimated distribution is shown in blue.

E.2 Tables

Table E.1: Wild birds observed at site during Wetland Bird Surveys from November 2016 to January 2017

Bird species	November	December[1]	January
Brent Goose (Dark-bellied)	2222	2864	1174
Mute Swan	985	900	737
Wigeon	6537	874	1492
Coot	1115	735	780
Teal	1120	530	617
Mallard	415	511	302
Pochard	60	315	360
Tufted Duck	265	290	240
Black-headed Gull	192	148	87
Pintail	207	102	322
Common Gull	60	65	121
Herring Gull	71	61	82
Oystercatcher	15	58	58
Red-breasted Merganser	20	53	133
Canada Goose	829	47	25
Shoveler	91	38	72
Little Grebe	34	29	18
Great Crested Grebe	15	29	16
Mediterranean Gull	84	27	73
Snipe	1	24	8
Moorhen	14	22	30
Cormorant	17	20	16
Great Black-backed Gull	8	20	0
Bar-tailed Godwit	7	20	0
Lapwing	270	18	245
Little Egret	7	17	9
Turnstone	29	14	27
Brent Goose (Pale-bellied)	10	14	0
Redshank	31	12	14
Scaup	0	12	8
Shelduck	12	10	52
Goosander	0	8	0
Mallard (domestic)	3	4	3
Dunlin	41	3	8
Grey Heron	6	3	1
Long-tailed Duck	1	3	0
Goldeneye	0	3	2
Black Swan	2	2	2
Kingfisher	2	2	0

Continued on next page

Bird species	November	December[1]	January
Gadwall	0	2	3
Slavonian Grebe	0	2	0
Brent Goose (Black Brant)	0	1	0
Egyptian Goose	0	1	0
Golden Plover	0	1	0
Knot	0	1	0
Ruff	0	1	0
Grey Plover	7	0	1
Barnacle Goose	7	0	0
Curlew	2	0	0
Water Rail	2	0	0
Whimbrel	1	0	0
Black-necked Grebe	0	0	1
Common Scoter	0	0	0
Eider	0	0	0
Razorbill	0	0	0
Shag	0	0	0
Black-throated Diver	0	0	0
Great Northern Diver	0	0	0

[1]Note that some data is missing for December 2016, as the entirety of the Fleet Lagoon was not surveyed during this time. As such, the numbers for December 2016 represent minimal bird numbers only. Specifically, the missing data corresponds to the western part of the Fleet Lagoon between Rodden and Moonfleet, so represents perhaps a quarter of the possible data although the effect cannot be exactly quantified.

Table E.2: Primers used in N8 screening PCRs

Primer or probe	Primer or probe name	Sequence (5'-3')
Primer	IVA-N8-1296F	TCC ATG YTT TTG GGT TGA RAT GAT
Primer	IVA-N8-1423R	GCT CCA TCR TGC CAY GAC CA
Probe	IVA-N8-1354FAM	FAM-TCH AGY AGC TCC ATT GTR ATG TGT GGA GT-BHQ1

Table E.3: Primers used in whole-genome multiplex PCRs

Primer name	Sequence (5'-3')
HA_400_1_LEFT	GCAGGGGTTCACTCTGTCAAAA
HA_400_1_RIGHT	ATAGTCATTGAGGCTCCCTGGG
HA_400_2_LEFT	TGCGACGAATTCATCAGAGTGC
HA_400_2_RIGHT	TGTCTGCTCTTCTGCATTGTTGG
HA_400_3_LEFT	CAAAAAGAACGATGCATACCCAACAA
HA_400_3_RIGHT	GTGTTGCAGTGGCCATATTCCA
HA_400_4_LEFT	TTTCGAGAGTAATGGAAATTTCATTGCT
HA_400_4_RIGHT	CACTCCCCTGCTCATTGCTATG
HA_400_5_LEFT	AGAGGGCTGTTTGGGGCTATAG
HA_400_5_RIGHT	CGTTACCCAGCTCCTTTGCATT
HA_400_6_LEFT	TTCTCATGGAAAACGAGAGGACTCT
HA_400_6_RIGHT	TTCTGCACTGTAACGACCCATTG
MP_400_1_LEFT	AAGCAGGTAGATATTGAAAGATGAGTCT
MP_400_1_RIGHT	CACCGGTTGAGTAACTGAGTGC
MP_400_2_LEFT	TTAGGGTTTGTGTTCACGCTCAC
MP_400_2_RIGHT	GCCTTAGCTGTAGTACTGGCCA
MP_400_3_LEFT	TACAACAAGATGGGGACGGTGA
MP_400_3_RIGHT	GATCCCAATGATACTTGCGGCAA
MP_400_4_LEFT	TGCTAGTCAGGCTAGGCAGATG
MP_400_4_RIGHT	TCCAGCTCTATGTTGACAAAATGACC
NA_400_1_LEFT	GCAAAAGCAGGAGTTTAAAATGAATCCA
NA_400_1_RIGHT	CAGGTGAACAAGAGACGAAAGGC
NA_400_2_LEFT	AGGGCTTTGCACCTTTTTCCAA
NA_400_2_RIGHT	AACAACATCAGTAGGCACCCCT
NA_400_3_LEFT	GTCAGCAACAGCCTGTCATGAT
NA_400_3_RIGHT	TAGCACAGGCCTGTTAGTTCCC
NA_400_4_LEFT	AGATGTCAGCTTTAGTGGAGGACA
NA_400_4_RIGHT	ACCCACTGTATCCCGACCAATT
NA_400_5_LEFT	ATGGCGTAAAAGGTTTCGGGTTT
NA_400_5_RIGHT	GGAAGAATAGCTCCATCGTGCC
NP_400_1_LEFT	GCAGGGTAGATAATCACTCACTGAG
NP_400_1_RIGHT	GCAGTTGCGTCTTCTCCATTGT
NP_400_2_LEFT	AGGAGAGATGGGAAATGGGTGAG
NP_400_2_RIGHT	TCATTGCTCTTTGTGCTGCTGT
NP_400_3_LEFT	CTTCTGGAGAGGCGAGAATGGA
NP_400_3_RIGHT	GCAGAATGACATGCCATCCACA
NP_400_4_LEFT	TTCCGTCTGCTTCAAAACAGCC
NP_400_4_RIGHT	CCGCCATAATGGTCGCTCTTTC
NP_400_5_LEFT	TGGACTCCAGCACTCTTGAACT
NP_400_5_RIGHT	TCCTCTGCATTGTCTCCGAAGAA
NS_400_1_LEFT	GCAAAAGCAGGGTGACAAAAACAT
NS_400_1_RIGHT	AACCTGCCACTTTCTGTTTGGG

Continued on next page

Primer name	Sequence (5'-3')
NS_400_2_LEFT	TGACTGTTGCTTCAAGTCCGTC
NS_400_2_RIGHT	ACTTTGGAGGGAGTGAAGGTCTC
NS_400_3_LEFT	TGGGAGAAATCTCACCGTTACCTT
NS_400_3_RIGHT	AGCTGAAACGAGAAAGTTCTTATCTCTT
PA_400_1_LEFT	AGGTACTGATCCAAAATGGAAGACTT
PA_400_1_RIGHT	ATGTGAACTTCCCTTCGCGTTAC
PA_400_2_LEFT	ACTACAGGAGTCGAAAAGCCCA
PA_400_2_RIGHT	TTTTCAAGGCTGGAGAAGTTCGG
PA_400_3_LEFT	CCGAGAGAGGCGAAGAGACAAT
PA_400_3_RIGHT	TTTCACGATGTTGGGCTCTTTCC
PA_400_4_LEFT	AATTGAGCATTGAAGACCCGAGC
PA_400_4_RIGHT	TCTGTCAATTCGCAAGCCTTGTT
PA_400_5_LEFT	AGATGTTAGCGATCTAAGACAGTACGA
PA_400_5_RIGHT	CGGTGTCATTCCTCAAATGGGA
PA_400_6_LEFT	ACTTCCAACTGATTCCAATGATAAGCA
PA_400_6_RIGHT	GACTCGGCTTCAATCATGCTCTC
PA_400_7_LEFT	ATGTGAGAACCAATGGGACTTCC
PA_400_7_RIGHT	CAGGCACTCCTCAATTGCTTCA
PA_400_8_LEFT	AAGGACATGACCAAGGAATTCTTTGAAA
PA_400_8_RIGHT	GCAAATAGTAGCATTGCCACAACT
PB1_400_1_LEFT	GGATGTCAATCCGACTTTACTTTTCTT
PB1_400_1_RIGHT	CAGTCATAAGTCTGGCGACCTTG
PB1_400_2_LEFT	TTTCCTTGAAGAGTCCCACCCAG
PB1_400_2_RIGHT	CCGCCTCTTCAATTTGCCTCTT
PB1_400_3_LEFT	ACAATAGGGAAAAAGAAACAAAGGCTG
PB1_400_3_RIGHT	GGGGCAATGCTCAAGACATTTCT
PB1_400_4_LEFT	CCAAATGGAATGAGAATCAAAATCCTCG
PB1_400_4_RIGHT	GGAGTCCATCCCACCAGTATGT
PB1_400_5_LEFT	CCTGGAATGATGATGGGCATGTT
PB1_400_5_RIGHT	GTTGCTGGTCCAAGGTCATTGT
PB1_400_6_LEFT	CCAGCTTTGGAGTGTCTGGAATC
PB1_400_6_RIGHT	ACCACAGCATTGTTTACGGACT
PB1_400_7_LEFT	CTGCTTGAAGTGGGAGCTGATG
PB1_400_7_RIGHT	GAGCTCTTCAATGGTGGAACAGA
PB2_400_1_LEFT	CGAAAGCAGGTCAAATATATTCAATATGGA
PB2_400_1_RIGHT	TGAACAGGACCAAAGGTTCCATG
PB2_400_2_LEFT	GGTGGAATAGAAATGGGCCAACA
PB2_400_2_RIGHT	GCAGGTCCCTTGAGTCAAATGC
PB2_400_3_LEFT	TGGAGAGAGAACTGGTTCGCAA
PB2_400_3_RIGHT	ACAGACGACCCACTTGTCCTTT
PB2_400_4_LEFT	AACCCAACAGAAGAGCAAGCTG
PB2_400_4_RIGHT	CGCTGATTCGCTCTGTTGACAA
PB2_400_5_LEFT	CGCTGAAGCAATCATAGTGGCA
PB2_400_5_RIGHT	TCGTCAGTTTCTCTGTTCCCTGT
PB2_400_6_LEFT	AGAGAGTGGTCGTGAGCATTGA

Continued on next page

Primer name	Sequence (5'-3')
PB2_400_6_RIGHT	TGCATCCTACTCTGTTCCGGTG
PB2_400_7_LEFT	TTCCAACAGATGCGTGATGTACTG
PB2_400_7_RIGHT	ACGTCTCCTTGCCCTATCAACA
PB2_400_8_LEFT	AGGTTCAGGAATGAGAATACTTGTGAG
PB2_400_8_RIGHT	CGACACTAATTGATGGCCATCCG

Table E.4: Percentage of each genomic segment sequenced at >20x coverage

Library	Barcode	Bird	HatchYear	Sex	Death of date	HA % [1]	MP %	NA %	NP %	NS %	PA %	PB1 %	PB2 %
1	BC01/2	Unringed BirdB	Adult	Unknown	13th-16th January	98	99	76	80	97	64	97	99
1	BC03/4	AS00890	2016	F	13/01/17	83	99	76	80	97	64	97	99
1	BC05/6	WVTK	2016	M	24/01/17	83	99	76	73	97	64	86	99
1	BC07/8	Unringed BirdA	2016	Unknown	13th-16th January	83	99	76	55	97	48	94	99
1	BC09/10	AS00868	2016	F	15/01/17	98	99	76	80	97	48	86	99
2	BC01	WVUK	2016	F	31/12/17	98	99	96	98	97	88	97	99
2	BC02	WVJX	2016	M	31/12/17	98	99	96	98	97	88	97	99
2	BC03	AS00918	2016	M	31/12/17	98	99	76	97	97	64	97	99
2	BC04	YBOV	2015	Unknown	01/01/17	98	99	96	98	97	88	97	99
2	BC05	WVZP	2016	F	31/12/17	98	99	96	98	97	65	97	99
2	BC06	WULH	2009	F	30/12/17	98	99	76	88	97	64	97	99
2	BC07	AS00778	2016	F	13/01/17	98	99	76	80	97	48	97	99

[2]Percentage of coding region of corresponding segment that was sequenced at >20 nucleotide depth

Bibliography

[1] Achenbach, J. E., and Bowen, R. A. Transmission of Avian Influenza A Viruses among Species in an Artificial Barnyard. *PLoS ONE 6*, 3 (2011), e17643.

[2] Adlhoch, C., Gossner, C., Koch, G., Brown, I., Bouwstra, R., Verdonck, F., Penttinen, P., and Harder, T. Comparing introduction to Europe of highly pathogenic avian influenza viruses A(H5N8) in 2014 and A(H5N1) in 2005. *Euro Surveillance 19*, 50 (2014).

[3] Aiello, R., Beato, M. S., Mancin, M., Rigoni, M., Romero Tejeda, A., Maniero, S., Capua, I., and Terregino, C. Differences in the detection of highly pathogenic avian influenza H5N1 virus in feather samples from 4-week-old and 24-week-old infected Pekin ducks (Anas platyrhynchos var. domestica). *Veterinary Microbiology 165*, 3 (2013), 443–447.

[4] Alexander, D. J., Manvell, R. J., Irvine, R., Londt, B. Z., Cox, B., Ceeraz, V., Banks, J., and Browna, I. H. Overview of incursions of Asian H5N1 subtype highly pathogenic avian influenza virus into Great Britain, 2005-2008. *Avian Diseases 54*, 1 Suppl (2010), 194–200.

[5] Almeida, J. D. A Classification of Virus Particles Based on Morphology. *Canadian Medical Association Journal 89*, 16 (1963), 787–798.

[6] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology 215*, 3 (1990), 403–410.

[7] An, P., Sáenz Robles, M. T., and Pipas, J. M. Large T Antigens of Polyomaviruses: Amazing Molecular Machines. *Annual Review of Microbiology 66*, 1 (2012), 213–236.

[8] Andrews, S. FastQC: a quality control tool for high throughput sequence data., 2010.

[9] Anthony, S. J., Islam, A., Johnson, C., Navarrete-Macias, I., Liang, E., Jain, K., Hitchens, P. L., Che, X., Soloyvov, A., Hicks, A. L., Ojeda-Flores, R., Zambrana-Torrelio, C., Ulrich, W., Rostal, M. K., Petrosov, A., Garcia, J., Haider, N., Wolfe, N., Goldstein, T., Morse, S. S., Rahman, M., Epstein, J. H., Mazet, J. K., Daszak, P., and Lipkin, W. I. Non-random patterns in viral diversity. *Nature Communications 6* (2015), 8147.

[10] APHA. Situation assessment following detection and spread of H5 HPAI in EU Member States since October 2016.

[11] Awe, O. O., Kang, K.-i., Ibrahim, M., Ali, A., Elaish, M., Saif, Y. M., and Lee, C.-W. Age-Related Susceptibility of Turkeys to Enteric Viruses. *Avian Diseases 59*, 2 (2015), 207–212.

[12] Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., and Suchard, M. A. BEAGLE: an Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology* (2011), syr100.

[13] Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., and Alekseyenko, A. V. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* (2012), mss084.

[14] Baele, G., Suchard, M. A., Rambaut, A., and Lemey, P. Emerging Concepts of Data Integration in Pathogen Phylodynamics. *Systematic Biology 66*, 1 (2017), e47–e65.

[15] Bahl, J., Vijaykrishna, D., Holmes, E. C., Smith, G. J. D., and Guan, Y. Gene flow and competitive exclusion of avian influenza A virus in natural reservoir hosts. *Virology 390*, 2 (2009), 289–297.

[16] Baker, T. S., Drak, J., and Bina, M. The capsid of small papova viruses contains 72 pentameric capsomeres: direct evidence from cryo-electron-microscopy of simian virus 40. *Biophysical Journal 55*, 2 (1989), 243–253.

[17] Bande, F., Arshad, S. S., and Omar, A. R. Isolation and Metagenomic Identification of Avian Leukosis Virus Associated with Mortality in Broiler Chicken. *Advances in Virology 2016* (2016), 9058403.

[18] Banks, M., King, D. P., Daniells, C., Stagg, D. A., and Gavier-Widen, D. Partial characterization of a novel gammaherpesvirus isolated from a European badger (Meles meles). *Journal of General Virology 83*, 6 (2002), 1325–1330.

[19] Bartlett, M. S. Measles Periodicity and Community Size. *Journal of the Royal Statistical Society. Series A (General) 120*, 1 (1957), 48–70.

[20] Bartlett, M. S. The Critical Community Size for Measles in the United States. *Journal of the Royal Statistical Society. Series A (General) 123*, 1 (1960), 37–44.

[21] Baumer, A., Feldmann, J., Renzullo, S., Müller, M., Thür, B., and Hofmann, M. A. Epidemiology of Avian Influenza Virus in Wild Birds in Switzerland Between 2006 and 2009. *Avian Diseases 54*, 2 (2010), 875–884.

[22] Berger, J. R., Miller, C. S., Mootoor, Y., Avdiushko, S. A., Kryscio, R. J., and Zhu, H. JC Virus Detection in Bodily Fluids: Clues to Transmission. *Clinical Infectious Diseases 43*, 1 (2006).

[23] Berhane, Y., Kobasa, D., Embury-Hyatt, C., Pickering, B., Babiuk, S., Joseph, T., Bowes, V., Suderman, M., Leung, A., Cottam-Birt, C., Hisanaga, T., and Pasick, J. Pathobiological Characterization of a Novel Reassortant Highly Pathogenic H5N1 Virus Isolated in British Columbia, Canada, 2015. *Scientific Reports 6* (2016).

[24] Berhane, Y., Leith, M., Embury-Hyatt, C., Neufeld, J., Babiuk, S., Hisanaga, T., Kehler, H., Hooper-McGrevy, K., and Pasick, J. Studying possible cross-protection of Canada geese preexposed to North American low pathogenicity avian influenza virus strains (H3n8, H4N6, and H5N2) against an H5N1 highly pathogenic avian influenza challenge. *Avian Diseases 54*, 1 Suppl (2010), 548–554.

[25] Bermejo, M., Rodríguez-Teijeiro, J. D., Illera, G., Barroso, A., Vilà, C., and Walsh, P. D. Ebola Outbreak Killed 5000 Gorillas. *Science 314*, 5805 (2006), 1564–1564.

[26] Bevins, S. N., Dusek, R. J., White, C. L., Gidlewski, T., Bodenstein, B., Mansfield, K. G., DeBruyn, P., Kraege, D., Rowan, E., Gillin, C., Thomas, B., Chandler, S., Baroch, J., Schmit, B., Grady, M. J., Miller, R. S., Drew, M. L., Stopak, S., Zscheile, B., Bennett, J., Sengl, J., Brady, C., Ip, H. S., Spackman, E., Killian, M. L., Torchetti, M. K., Sleeman, J. M., and Deliberto, T. J. Widespread detection of highly pathogenic H5 influenza viruses in wild birds from the Pacific Flyway of the United States. *Scientific Reports 6* (2016).

[27] Bhatt, S., Lam, T. T., Lycett, S. J., Leigh Brown, A. J., Bowden, T. A., Holmes, E. C., Guan, Y., Wood, J. L. N., Brown, I. H., Kellam, P., Pybus, O. G., Brown, I., Brookes, S., Germundsson, A., Cook, A., Williamson, S., Essen, S., Garcon, F., Gunn, G., Sanchez, M., Marques, D., Wood, J., Tucker, D., McCrone, I., Gog, J., Saenz, R., Staff, M., Murcia, P., Barclay, W., Donnelly, C., Elderfield, R. A., Kellam, P., Baillie, G., Coulter, E., Wieland, B., Mastin, A., McCauley, J., Brown, A. L., Lycett, S., Woolhouse, M., Pybus, O., Bhatt, S., Hayward, A., Ishola, D., Archibald, A., Freeman, T., Charleston, B., LeFevre, E., Bailey, M., Inman, C., Stokes, C., Chang, K. C., Dunham, S., White, G., Nguyen-Van-Tam, J., and Enstone, J. The evolutionary dynamics of influenza A virus adaptation to mammalian hosts. *Philosophical Transactions of the Royal Society B: Biological Sciences 368*, 1614 (2013).

[28] Bielejec, F., Rambaut, A., Suchard, M. A., and Lemey, P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics 27*, 20 (2011), 2910–2912.

[29] Bofill-Mas, S., Formiga-Cruz, M., Clemente-Casares, P., Calafell, F., and Girones, R. Potential transmission of human polyomaviruses through the gastrointestinal tract after exposure to virions or viral DNA. *Journal of Virology 82*, 16 (2008), 82448244.

[30] Boldorini, R., Allegrini, S., Miglio, U., Paganotti, A., Cocca, N., Zaffaroni, M., Riboni, F., Monga, G., and Viscidi, R. Serological evidence of vertical transmission of JC and BK polyomaviruses in humans. *Journal of General Virology 92*, 5 (Sep 2011), 10441050.

[31] Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., and van der Noordaa, J. Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology 28*, 3 (1990), 495–503.

[32] Boros, Á., Nemes, C., Pankovics, P., Kapusinszky, B., Delwart, E., and Reuter, G. Identification and complete genome characterization of a novel picornavirus in turkey (Meleagris gallopavo). *Journal of General Virology 93*, Pt 10 (2012), 2171–2182.

[33] Boros, Á., Nemes, C., Pankovics, P., Kapusinszky, B., Delwart, E., and Reuter, G. Genetic characterization of a novel picornavirus in turkeys (Meleagris gallopavo) distinct from turkey galliviruses and megriviruses and distantly related to the members of the genus Avihepatovirus. *Journal of General Virology 94*, Pt 7 (2013), 1496–1509.

[34] Boros, Á., Pankovics, P., Adonyi, Á., Fenyvesi, H., Day, J. M., Phan, T. G., Delwart, E., and Reuter, G. A diarrheic chicken simultaneously co-infected with multiple picornaviruses: Complete genome analysis of avian picornaviruses representing up to six genera. *Virology 489* (2016), 63–74.

[35] Bosch, F. X., Garten, W., Klenk, H. D., and Rott, R. Proteolytic cleavage of influenza virus hemagglutinins: primary structure of the connecting peptide between HA1 and HA2 determines proteolytic cleavability and pathogenicity of Avian influenza viruses. *Virology 113*, 2 (1981), 725–735.

[36] Bouwstra, R., Heutink, R., Bossers, A., Harders, F., Koch, G., and Elbers, A. Full-Genome Sequence of Influenza A(H5N8) Virus in Poultry Linked to Sequences of Strains from Asia, the Netherlands, 2014. *Emerging Infectious Diseases 21*, 5 (2015), 872–874.

[37] Brachman, P. S. Epidemiology. In *Medical Microbiology*, S. Baron, Ed., 4th ed. University of Texas Medical Branch at Galveston, Galveston (TX), 1996.

[38] Brady, A., and Salzberg, S. PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nature Methods 8*, 5 (2011), 367.

[39] Brady, A., and Salzberg, S. L. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods 6*, 9 (2009), 673–676.

[40] Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. Diversity and population structure of a near-shore marine-sediment viral community. *Proceedings of the Royal Society B: Biological Sciences 271*, 1539 (2004), 565–574.

[41] Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology 185*, 20 (2003), 6220–6223.

[42] Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., and Rohwer, F. Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences 99*, 22 (2002), 14250–14255.

[43] Brenner, D., Larsen, R. S., Wack, R. F., Agnew, D., and Imai, D. Concurrent West Nile virus and mycobacterium avium infection in a black-necked swan (Cygnus melanocoryphus). *Journal of Zoo and Wildlife Medicine: Official Publication of the American Association of Zoo Veterinarians 38*, 2 (2007), 357–362.

[44] Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., and Franklin, R. B. MinION nanopore sequencing of environmental metagenomes: a synthetic approach. *GigaScience* (2017).

[45] Brown, J. D., Goekjian, G., Poulson, R., Valeika, S., and Stallknecht, D. E. Avian influenza virus in water: Infectivity is dependent on pH, salinity and temperature. *Veterinary Microbiology 136*, 12 (2009), 20–26.

[46] Brown, J. D., Stallknecht, D. E., Beck, J. R., Suarez, D. L., and Swayne, D. E. Susceptibility of North American ducks and gulls to H5N1 highly pathogenic avian influenza viruses. *Emerging Infectious Diseases 12*, 11 (2006), 1663–1670.

[47] Brown, J. D., Stallknecht, D. E., and Swayne, D. E. Experimental infection of swans and geese with highly pathogenic avian influenza virus (H5N1) of Asian lineage. *Emerging Infectious Diseases 14*, 1 (2008), 136–142.

[48] Brown, J. D., Swayne, D. E., Cooper, R. J., Burns, R. E., and Stallknecht, D. E. Persistence of H5 and H7 avian influenza viruses in water. *Avian Diseases 51*, 1 Suppl (2007), 285–289.

[49] Buchfink, B., Xie, C., and Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods 12*, 1 (2015), 59–60.

[50] Buck, C. B., Doorslaer, K. V., Peretti, A., Geoghegan, E. M., Tisza, M. J., An, P., Katz, J. P., Pipas, J. M., McBride, A. A., Camus, A. C., McDermott, A. J., Dill, J. A., Delwart, E., Ng, T. F. F., Farkas, K., Austin, C., Kraberger, S., Davison, W., Pastrana, D. V., and Varsani, A. The Ancient Evolutionary History of Polyomaviruses. *PLoS Pathogens 12*, 4 (2016), e1005574.

[51] Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics 11* (2010), 94.

[52] Campillo-Balderas, J. A., Lazcano, A., and Becerra, A. Viral Genome Size Distribution Does not Correlate with the Antiquity of the Host Lineages. *Frontiers in Ecology and Evolution 3* (2015).

[53] Cappelle, J., Zhao, D., Gilbert, M., Nelson, M. I., Newman, S. H., Takekawa, J. Y., Gaidet, N., Prosser, D. J., Liu, Y., Li, P., Shu, Y., and Xiao, X. Risks of avian influenza transmission in areas of intensive free-ranging duck production with wild waterfowl. *EcoHealth 11*, 1 (2014), 109–119.

[54] Carragher, D. M., Kaminski, D. A., Moquin, A., Hartson, L., and Randall, T. D. A Novel Role for Non-Neutralizing Antibodies against Nucleoprotein in Facilitating Resistance to Influenza Virus. *The Journal of Immunology 181*, 6 (2008), 4168–4176.

[55] Carter, J. J., Daugherty, M. D., Qi, X., Bheda-Malge, A., Wipf, G. C., Robinson, K., Roman, A., Malik, H. S., and Galloway, D. A. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. *Proceedings of the National Academy of Sciences 110*, 31 (2013), 12744–12749.

[56] CDC. Serum Cross-Reactive Antibody Response to a Novel Influenza A (H1N1) Virus After Vaccination with Seasonal Influenza Vaccine. *Morbidity and Mortality Weekly Report 58*, 19 (2009), 521–524.

[57] Chaise, C., Lalmanach, A.-C., Marty, H., Soubies, S. M., Croville, G., Loupias, J., Marc, D., Quéré, P., and Guérin, J.-L. Protection Patterns in Duck and Chicken after Homo- or Hetero-Subtypic Reinfections with H5 and H7 Low Pathogenicity Avian Influenza Viruses: A Comparative Study. *PLoS ONE 9*, 8 (2014), e105189.

[58] Charleston, M. A., and Robertson, D. L. Preferential Host Switching by Primate Lentiviruses Can Account for Phylogenetic Similarity with the Primate Phylogeny. *Systematic Biology 51*, 3 (2002), 528–535.

[59] Charmantier, A., Perrins, C., McCleery, R. H., and Sheldon, B. C. Age-Dependent Genetic Variance in a Life-History Trait in the Mute Swan. *Proceedings: Biological Sciences 273*, 1583 (2006), 225–232.

[60] Charmantier, A., Perrins, C., McCleery, R. H., and Sheldon, B. C. Quantitative genetics of age at reproduction in wild swans: Support for antagonistic pleiotropy models of senescence. *Proceedings of the National Academy of Sciences 103*, 17 (2006), 6587–6592.

[61] Check Hayden, E. Pint-sized DNA sequencer impresses first users. *Nature News 521*, 7550 (2015), 15.

[62] Chen, G.-Q., Zhuang, Q.-Y., Wang, K.-C., Liu, S., Shao, J.-Z., Jiang, W.-M., Hou, G.-Y., Li, J.-P., Yu, J.-M., Li, Y.-P., and Chen, J.-M. Identification and Survey of a Novel Avian Coronavirus in Ducks. *PLoS ONE 8*, 8 (2013), e72918.

[63] Chen, H., Smith, G. J. D., Li, K. S., Wang, J., Fan, X. H., Rayner, J. M., Vijaykrishna, D., Zhang, J. X., Zhang, L. J., Guo, C. T., Cheung, C. L., Xu, K. M., Duan, L., Huang, K., Qin, K., Leung, Y. H. C., Wu, W. L., Lu, H. R., Chen, Y., Xia, N. S., Naipospos, T. S. P., Yuen, K. Y., Hassan, S. S., Bahri, S., Nguyen, T. D., Webster, R. G., Peiris, J. S. M., and Guan, Y. Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *Proceedings of the National Academy of Sciences 103*, 8 (2006), 2845–2850.

[64] Chen, H. Y., Mascio, M. D., Perelson, A. S., Ho, D. D., and Zhang, L. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. *Proceedings of the National Academy of Sciences 104*, 48 (2007), 19079–19084.

[65] Chu, D. K. W., Leung, C. Y. H., Gilbert, M., Joyner, P. H., Ng, E. M., Tse, T. M., Guan, Y., Peiris, J. S. M., and Poon, L. L. M. Avian Coronavirus in Wild Aquatic Birds. *Journal of Virology 85*, 23 (2011), 12815–12820.

[66] Ciriaco, E., Píñera, P. P., Díaz-Esnal, B., and Laurà, R. Age-related changes in the avian primary lymphoid organs (thymus and bursa of Fabricius). *Microscopy Research and Technique 62*, 6 (2003), 482–487.

[67] Claas, E. C., Osterhaus, A. D., van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., Krauss, S., Shortridge, K. F., and Webster, R. G. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet 351*, 9101 (1998), 472–477.

[68] Cleland, A., Nettleton, P., Jarvis, L., and Simmonds, P. Use of bovine viral diarrhoea virus as an internal control for amplification of hepatitis C virus. *Vox sanguinis 77*, 8 (1999), 170–174.

[69] Cockrem, J. F. Timing of seasonal breeding in birds, with particular reference to New Zealand birds. *Reproduction, Fertility and Development 7*, 1 (1995), 1–19.

[70] Cody, A. J., McCarthy, N. D., Bray, J. E., Wimalarathna, H. M. L., Colles, F. M., Jansen van Rensburg, M. J., Dingle, K. E., Waldenström, J., and Maiden, M. C. J. Wild bird-associated Campylobacter jejuni isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environmental Microbiology Reports 7*, 5 (2015), 782–788.

[71] Community Reference Laboratory for Influenza. Annual Report on surveillance for avian influenza in wild birds in the EU. *Technical Report from the Community Reference Laboratory for Influenza* (2006).

[72] Conceição-Neto, N., Godinho, R., Álvares, F., Yinda, C. K., Deboutte, W., Zeller, M., Laenen, L., Heylen, E., Roque, S., Petrucci-Fonseca, F., Santos, N., Van Ranst, M., Mesquita, J. R., and Matthijnssens, J. Viral gut metagenomics of sympatric wild and domestic canids, and monitoring of viruses: Insights from an endangered wolf population. *Ecology and Evolution 7*, 12 (2017), 4135–4146.

[73] Conceição-Neto, N., Zeller, M., Lefrère, H., Bruyn, P. D., Beller, L., Deboutte, W., Yinda, C. K., Lavigne, R., Maes, P., Ranst, M. V., Heylen, E., and Matthijnssens, J. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports 5* (2015), srep16532.

[74] Conlan, A. J., and Grenfell, B. T. Seasonality and the persistence and invasion of measles. *Proceedings of the Royal Society B: Biological Sciences 274*, 1614 (2007), 1133–1141.

[75] Costa, T. P., Brown, J. D., Howerth, E. W., and Stallknecht, D. E. Effect of a prior exposure to a low pathogenic avian influenza virus in the outcome of a heterosubtypic low pathogenic avian influenza infection in mallards (Anas platyrhynchos). *Avian Diseases 54*, 4 (2010), 1286–1291.

[76] Costa, T. P., Brown, J. D., Howerth, E. W., Stallknecht, D. E., and Swayne, D. E. Homo- and Heterosubtypic Low Pathogenic Avian Influenza Exposure on H5N1 Highly Pathogenic Avian Influenza Virus Infection in Wood Ducks (Aix sponsa). *PLoS ONE 6*, 1 (2011), e15987.

[77] Cottam, E. M., Thébaud, G., Wadsworth, J., Gloster, J., Mansley, L., Paton, D. J., King, D. P., and Haydon, D. T. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proceedings of the Royal Society B: Biological Sciences 275*, 1637 (2008), 887–895.

[78] Cottam, E. M., Wadsworth, J., Shaw, A. E., Rowlands, R. J., Goatley, L., Maan, S., Maan, N. S., Mertens, P. P. C., Ebert, K., Li, Y., Ryan, E. D., Juleff, N., Ferris, N. P., Wilesmith, J. W., Haydon, D. T., King, D. P., Paton, D. J., and Knowles, N. J. Transmission Pathways of Foot-and-Mouth Disease Virus in the United Kingdom in 2007. *PLoS Pathogens 4*, 4 (2008), e1000050.

[79] Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., and van der Hoek, L. Full Genome Virus Detection in Fecal Samples Using Sensitive Nucleic Acid Preparation, Deep Sequencing, and a Novel Iterative Sequence Classification Algorithm. *PLoS ONE 9*, 4 (2014), e93269.

[80] Courcoul, A., Moyen, J.-L., Brugère, L., Faye, S., Hénault, S., Gares, H., and Boschiroli, M.-L. Estimation of Sensitivity and Specificity of Bacteriology, Histopathology and PCR for the Confirmatory Diagnosis of

Bovine Tuberculosis Using Latent Class Analysis. *PLoS ONE 9*, 3 (2014), e90334.

[81] Dalby, A. R., and Iqbal, M. The European and Japanese outbreaks of H5N8 derive from a single source population providing evidence for the dispersal along the long distance bird migratory flyways. *PeerJ 3* (2015).

[82] Dalziel, A. E., Delean, S., Heinrich, S., and Cassey, P. Persistence of Low Pathogenic Influenza A Virus in Water: A Systematic Review and Quantitative Meta-Analysis. *PLoS ONE 11*, 10 (2016), e0161929.

[83] Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., Ross, J. V., and Vernon, M. C. Networks and the Epidemiology of Infectious Disease. *Interdisciplinary Perspectives on Infectious Diseases 2011* (2011), 1–28. arXiv: 1011.5950.

[84] Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics 27*, 8 (2011), 1164–1165.

[85] Day, J. M., Ballard, L. L., Duke, M. V., Scheffler, B. E., and Zsak, L. Metagenomic analysis of the turkey gut RNA virus community. *Virology Journal 7* (2010), 313.

[86] Day, J. M., Oakley, B. B., Seal, B. S., and Zsak, L. Comparative analysis of the intestinal bacterial and RNA viral communities from sentinel birds placed on selected broiler chicken farms. *PLoS ONE 10*, 1 (2015), e0117210.

[87] Day, J. M., and Zsak, L. Molecular and phylogenetic analysis of a novel turkey-origin picobirnavirus. *Avian Diseases 58*, 1 (2014), 137–142.

[88] Day, J. M., and Zsak, L. Investigating Turkey Enteric Picornavirus and Its Association with Enteric Disease in Poults. *Avian Diseases 59*, 1 (2015), 138–142.

[89] de Bruijn, N. G. A combinatorial problem. *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 49*, 7 (1946), 758–764.

[90] De Gascun, C. F., and Carr, M. J. Human Polyomavirus Reactivation: Disease Pathogenesis and Treatment Approaches. *Clinical and Developmental Immunology 2013* (2013), 373579.

[91] de Vries, E., Guo, H., Dai, M., Rottier, P. J., van Kuppeveld, F. J., and de Haan, C. A. Rapid Emergence of Highly Pathogenic Avian Influenza Subtypes from a Subtype H5N1 Hemagglutinin Variant. *Emerging Infectious Diseases 21*, 5 (2015), 842–846.

[92] DeCaprio, J. A., and Garcea, R. L. A cornucopia of human polyomaviruses. *Nature Reviews Microbiology 11*, 4 (2013), 264–276.

[93] DEFRA. Highly Pathogenic Avian Influenza H5N1 in swans in Dorset. Version 2, released February 2008.

[94] DeJesus, E., Costa-Hurtado, M., Smith, D., Lee, D.-H., Spackman, E., Kapczynski, D. R., Torchetti, M. K., Killian, M. L., Suarez, D. L., Swayne, D. E., and Pantin-Jackwood, M. J. Changes in adaptation of H5N2 highly pathogenic avian influenza H5 clade 2.3.4.4 viruses in chickens and mallards. *Virology 499* (2016), 52–64.

[95] Dela Cruz, F. N., Giannitti, F., Li, L., Woods, L. W., Del Valle, L., Delwart, E., and Pesavento, P. A. Novel Polyomavirus associated with Brain Tumors in Free-Ranging Raccoons, Western United States. *Emerging Infectious Diseases 19*, 1 (2013), 77–84.

[96] Dellicour, S., Rose, R., Faria, N. R., Lemey, P., and Pybus, O. G. SERAPHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics 32*, 20 (2016), 3204–3206.

[97] Delnatte, P., Nagy, É., Ojkic, D., Leishman, D., Crawshaw, G., Elias, K., and Smith, D. A. Avian bornavirus in free-ranging waterfowl: prevalence of antibodies and cloacal shedding of viral RNA. *Journal of Wildlife Diseases 50*, 3 (2014), 512–523.

[98] Delport, W., Poon, A. F. Y., Frost, S. D. W., and Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics 26*, 19 (2010), 2455–2457.

[99] Denesvre, C., Dumarest, M., Rémy, S., Gourichon, D., and Eloit, M. Chicken skin virome analyzed by high-throughput sequencing shows a composition highly different from human skin. *Virus Genes 51*, 2 (2015), 209–216.

[100] Devaney, R., Trudgett, J., Trudgett, A., Meharg, C., and Smyth, V. A metagenomic comparison of endemic viruses from broiler chickens with runting-stunting syndrome and from normal birds. *Avian Pathology: Journal of the W.V.P.A 45*, 6 (2016), 616–629.

[101] Dill, J. A., Ng, T. F. F., and Camus, A. C. Complete Sequence of the Smallest Polyomavirus Genome, Giant Guitarfish (Rhynchobatus djiddensis) Polyomavirus 1. *Genome Announcements 4*, 3 (2016).

[102] Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., and Jaffrézic, F. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics 14*, 6 (2013), 671–683.

[103] Domańska-Blicharz, K., Bocian, Ł., Lisowska, A., Jacukowicz, A., Pikuła, A., and Minta, Z. Cross-sectional survey of selected enteric viruses

in Polish turkey flocks between 2008 and 2011. *BMC Veterinary Research 13* (2017).

[104] Donato, C., and Vijaykrishna, D. The Broad Host Range and Genetic Diversity of Mammalian and Avian Astroviruses. *Viruses 9*, 5 (2017).

[105] Donis, R. O., Smith, G. J., and World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization (WHO/OIE/FAO) H5 Evolution Working Group. Nomenclature updates resulting from the evolution of avian influenza A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013-2014. *Influenza and Other Respiratory Viruses* (2015), 1750–2659.

[106] dos Reis, M., Tamuri, A. U., Hay, A. J., and Goldstein, R. A. Charting the host adaptation of influenza viruses. *Molecular Biology and Evolution 28*, 6 (2011), 1755–1767.

[107] Dowell, S. F. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerging Infectious Diseases 7*, 3 (2001), 369–374.

[108] Drummond, A. J., and Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology 7* (2007), 214.

[109] Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution 29*, 8 (2012), 1969–1973.

[110] Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., Park, D. J., Ladner, J. T., Arias, A., Asogun, D., Bielejec, F., Caddy, S. L., Cotten, M., DAmbrozio, J., Dellicour, S., Di Caro, A., Diclaro, J. W., Duraffour, S., Elmore, M. J., Fakoli, L. S., Faye, O., Gilbert, M. L., Gevao, S. M., Gire, S., Gladden-Young, A., Gnirke, A., Goba, A., Grant, D. S., Haagmans, B. L., Hiscox, J. A., Jah, U., Kugelman, J. R., Liu, D., Lu, J., Malboeuf, C. M., Mate, S., Matthews, D. A., Matranga, C. B., Meredith, L. W., Qu, J., Quick, J., Pas, S. D., Phan, M. V. T., Pollakis, G., Reusken, C. B., Sanchez-Lockhart, M., Schaffner, S. F., Schieffelin, J. S., Sealfon, R. S., Simon-Loriere, E., Smits, S. L., Stoecker, K., Thorne, L., Tobin, E. A., Vandi, M. A., Watson, S. J., West, K., Whitmer, S., Wiley, M. R., Winnicki, S. M., Wohl, S., Wölfel, R., Yozwiak, N. L., Andersen, K. G., Blyden, S. O., Bolay, F., Carroll, M. W., Dahn, B., Diallo, B., Formenty, P., Fraser, C., Gao, G. F., Garry, R. F., Goodfellow, I., Günther, S., Happi, C. T., Holmes, E. C., Kargbo, B., Keïta, S., Kellam, P., Koopmans, M. P. G., Kuhn, J. H., Loman, N. J., Magassouba, N., Naidoo, D., Nichol, S. T., Nyenswah, T., Palacios, G., Pybus, O. G., Sabeti, P. C., Sall, A., Ströher, U., Wurie, I., Suchard, M. A., Lemey, P., and Rambaut, A. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature 544*, 7650 (2017), 309–315.

[111] Duffy, S., Shackelton, L. A., and Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics 9*, 4 (2008), 267–276.

[112] Duke-Sylvester, S. M., Bolzoni, L., and Real, L. A. Strong seasonality produces spatial asynchrony in the outbreak of infectious diseases. *Journal of the Royal Society Interface 8*, 59 (2011), 817–825.

[113] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research 32*, 5 (2004), 1792–1797.

[114] Edwards, C., Suchard, M., Lemey, P., Welch, J., Barnes, I., Fulton, T., Barnett, R., O'Connell, T., Coxon, P., Monaghan, N., Valdiosera, C., Lorenzen, E., Willerslev, E., Baryshnikov, G., Rambaut, A., Thomas, M., Bradley, D., and Shapiro, B. Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline. *Current Biology 21*, 15 (2011), 1251–1258.

[115] Edwards, R. A., and Rohwer, F. Viral metagenomics. *Nature Reviews Microbiology 3*, 6 (2005), 504–510.

[116] Ehlers, B., and Moens, U. Genome analysis of non-human primate polyomaviruses. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 26* (2014), 283–294.

[117] Ely, C. R., Hall, J. S., Schmutz, J. A., Pearce, J. M., Terenzi, J., Sedinger, J. S., and Ip, H. S. Evidence that Life History Characteristics of Wild Birds Influence Infection and Exposure to Influenza A Viruses. *PLoS ONE 8*, 3 (2013), e57614.

[118] Enders, J. F., Weller, T. H., and Robbins, F. C. Cultivation of the Lansing Strain of Poliomyelitis Virus in Cultures of Various Human Embryonic Tissues. *Science 109*, 2822 (1949), 85–87.

[119] Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., and Hayashi, M. Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Research 33*, 6 (2005), e65.

[120] European Centre for Disease Prevention and Control. Outbreaks of highly pathogenic avian influenza A(H5N8) in Europe- 18 November 2016.

[121] European Food Safety Authority. Highly pathogenic avian influenza A subtype H5N8. *EFSA Journal 12*, 12 (2014), 3941.

[122] Evans, G. L., Caller, L. G., Foster, V., and Crump, C. M. Anion homeostasis is important for non-lytic release of BK polyomavirus from infected cells. *Open Biology 5*, 8 (2015), 150041.

[123] Ewald, P. W. Guarding Against the Most Dangerous Emerging Pathogens: Insights from Evolutionary Biology. *Emerging Infectious Disease journal - CDC 2*, 4 (1996).

[124] Fan, S., Zhou, L., Wu, D., Gao, X., Pei, E., Wang, T., Gao, Y., and Xia, X. A novel highly pathogenic H5N8 avian influenza virus isolated from a wild duck in China. *Influenza and Other Respiratory Viruses 8*, 6 (2014), 646–653.

[125] Faria, N. R., Azevedo, R. d. S. d. S., Kraemer, M. U. G., Souza, R., Cunha, M. S., Hill, S. C., Thézé, J., Bonsall, M. B., Bowden, T. A., Rissanen, I., Rocco, I. M., Nogueira, J. S., Maeda, A. Y., Vasami, F. G. d. S., Macedo, F. L. d. L., Suzuki, A., Rodrigues, S. G., Cruz, A. C. R., Nunes, B. T., Medeiros, D. B. d. A., Rodrigues, D. S. G., Queiroz, A. L. N., Silva, E. V. P. d., Henriques, D. F., Rosa, E. S. T. d., Oliveira, C. S. d., Martins, L. C., Vasconcelos, H. B., Casseb, L. M. N., Simith, D. d. B., Messina, J. P., Abade, L., Lourenço, J., Alcantara, L. C. J., Lima, M. M. d., Giovanetti, M., Hay, S. I., Oliveira, R. S. d., Lemos, P. d. S., Oliveira, L. F. d., Lima, C. P. S. d., Silva, S. P. d., Vasconcelos, J. M. d., Franco, L., Cardoso, J. F., Vianez-Júnior, J. L. d. S. G., Mir, D., Bello, G., Delatorre, E., Khan, K., Creatore, M., Coelho, G. E., Oliveira, W. K. d., Tesh, R., Pybus, O. G., Nunes, M. R. T., and Vasconcelos, P. F. C. Zika virus in the Americas: Early epidemiological and genetic findings. *Science* (2016), aaf5036.

[126] Faria, N. R., Quick, J., Claro, I. M., Thézé, J., de Jesus, J. G., Giovanetti, M., Kraemer, M. U. G., Hill, S. C., Black, A., da Costa, A. C., Franco, L. C., Silva, S. P., Wu, C.-H., Raghwani, J., Cauchemez, S., du Plessis, L., Verotti, M. P., de Oliveira, W. K., Carmo, E. H., Coelho, G. E., Santelli, A. C. F. S., Vinhal, L. C., Henriques, C. M., Simpson, J. T., Loose, M., Andersen, K. G., Grubaugh, N. D., Somasekar, S., Chiu, C. Y., Muñoz-Medina, J. E., Gonzalez-Bonilla, C. R., Arias, C. F., Lewis-Ximenez, L. L., Baylis, S. A., Chieppe, A. O., Aguiar, S. F., Fernandes, C. A., Lemos, P. S., Nascimento, B. L. S., Monteiro, H. a. O., Siqueira, I. C., de Queiroz, M. G., de Souza, T. R., Bezerra, J. F., Lemos, M. R., Pereira, G. F., Loudal, D., Moura, L. C., Dhalia, R., França, R. F., Magalhães, T., Marques Jr, E. T., Jaenisch, T., Wallau, G. L., de Lima, M. C., Nascimento, V., de Cerqueira, E. M., de Lima, M. M., Mascarenhas, D. L., Neto, J. P. M., Levin, A. S., Tozetto-Mendoza, T. R., Fonseca, S. N., Mendes-Correa, M. C., Milagres, F. P., Segurado, A., Holmes, E. C., Rambaut, A., Bedford, T., Nunes, M. R. T., Sabino, E. C., Alcantara, L. C. J., Loman, N. J., and Pybus, O. G. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature 546*, 7658 (2017), 406–410.

[127] Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G., and Lemey, P. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science 346*, 6205 (2014), 56–61.

[128] Faria, N. R., Sabino, E. C., Nunes, M. R. T., Alcantara, L. C. J., Loman, N. J., and Pybus, O. G. Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine 8* (2016), 97.

[129] Fawaz, M., Vijayakumar, P., Mishra, A., Gandhale, P. N., Dutta, R., Kamble, N. M., Sudhakar, S. B., Roychoudhary, P., Kumar, H., Kulkarni, D. D., and Raut, A. A. Duck gut viral metagenome analysis captures snapshot of viral diversity. *Gut Pathogens 8* (2016).

[130] Felsenstein, J. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. *Systematic Zoology 27*, 4 (1978), 401–410.

[131] Fereidouni, S. R., Grund, C., Häuslaigner, R., Lange, E., Wilking, H., Harder, T. C., Beer, M., and Starick, E. Dynamics of specific antibody responses induced in mallards after infection by or immunization with low pathogenicity avian influenza viruses. *Avian Diseases 54*, 1 (2010), 79–85.

[132] Fereidouni, S. R., Starick, E., Beer, M., Wilking, H., Kalthoff, D., Grund, C., Häuslaigner, R., Breithaupt, A., Lange, E., and Harder, T. C. Highly Pathogenic Avian Influenza Virus Infection of Mallards with Homo- and Heterosubtypic Immunity Induced by Low Pathogenic Avian Influenza Viruses. *PLoS ONE 4*, 8 (2009), e6706.

[133] Fine, P. E., and Clarkson, J. A. Measles in England and Wales–I: An analysis of factors underlying seasonal patterns. *International Journal of Epidemiology 11*, 1 (1982), 5–14.

[134] Firth, C., Kitchen, A., Shapiro, B., Suchard, M. A., Holmes, E. C., and Rambaut, A. Using Time-Structured Data to Estimate Evolutionary Rates of Double-Stranded DNA Viruses. *Molecular Biology and Evolution 27*, 9 (2010), 2038–2051.

[135] Fish, E. N. The X-files in immunity: sex-based differences predispose immune responses. *Nature Reviews. Immunology 8*, 9 (2008), 737–744.

[136] Fouchier, R. A. M., Munster, V., Wallensten, A., Bestebroer, T. M., Herfst, S., Smith, D., Rimmelzwaan, G. F., Olsen, B., and Osterhaus, A. D. M. E. Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls. *Journal of Virology 79*, 5 (2005), 2814–2822.

[137] Fouchier, R. A. M., and Munster, V. J. Epidemiology of low pathogenic avian influenza viruses in wild birds. *Revue Scientifique Et Technique (International Office of Epizootics) 28*, 1 (2009), 49–58.

[138] Fouchier, R. A. M., Olsen, B., Bestebroer, T. M., Herfst, S., Kemp, L. v. d., Rimmelzwaan, G. F., and Osterhaus, A. D. M. E. Influenza A Virus Surveillance in Wild Birds in Northern Europe in 1999 and 2000. *Avian Diseases 47* (2003), 857–860.

[139] Fountain-Jones, N. M., Packer, C., Troyer, J. L., VanderWaal, K., Robinson, S., Jacquot, M., and Craft, M. E. Linking social and spatial networks to viral community phylogenetics reveal subtype specific transmission dynamics in African lions. *The Journal of Animal Ecology* (2017).

[140] Fourment, M., Darling, A. E., and Holmes, E. C. The impact of migratory flyways on the spread of avian influenza virus in North America. *BMC Evolutionary Biology 17* (2017), 118.

[141] Francis, Jr., T. On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society 104*, 6 (1960), 572–578.

[142] Fries, A. C., Nolting, J. M., Bowman, A. S., Lin, X., Halpin, R. A., Wester, E., Fedorova, N., Stockwell, T. B., Das, S. R., Dugan, V. G., Wentworth, D. E., Gibbs, H. L., and Slemons, R. D. Spread and Persistence of Influenza A Viruses in Waterfowl Hosts in the North American Mississippi Migratory Flyway. *Journal of Virology 89*, 10 (2015), 5371–5381.

[143] Fusaro, A., Monne, I., Mulatti, P., Zecchin, B., Bonfanti, L., Ormelli, S., Milani, A., Cecchettin, K., Lemey, P., Moreno, A., Massi, P., Dorotea, T., Marangon, S., and Terregino, C. Genetic Diversity of Highly Pathogenic Avian Influenza A(H5N8/H5N5) Viruses in Italy, 2016-17. *Emerging Infectious Diseases 23*, 9 (2017).

[144] Galvani, A. P. Epidemiology meets evolutionary ecology. *Trends in Ecology & Evolution 18*, 3 (2003), 132–139.

[145] Geoghegan, J. L., Duchêne, S., and Holmes, E. C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathogens 13*, 2 (2017), e1006215.

[146] Gerits, N., and Moens, U. Agnoprotein of mammalian polyomaviruses. *Virology 432*, 2 (2012), 316–326.

[147] Gilbert, J., Dahl, J., Riney, C., You, J., Cui, C., Holmes, R., Lencer, W., and Benjamin, T. Ganglioside GD1a restores infectibility to mouse cells lacking functional receptors for polyomavirus. *Journal of Virology 79*, 1 (2005), 615–618.

[148] Gilbert, M., Golding, N., Zhou, H., Wint, G. R. W., Robinson, T. P., Tatem, A. J., Lai, S., Zhou, S., Jiang, H., Guo, D., Huang, Z., Messina, J. P., Xiao, X., Linard, C., Van Boeckel, T. P., Martin, V., Bhatt, S., Gething, P. W., Farrar, J. J., Hay, S. I., and Yu, H. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nature Communications 5* (2014), 4116.

[149] Goldsmith, C. S., and Miller, S. E. Modern Uses of Electron Microscopy for Detection of Viruses. *Clinical Microbiology Reviews 22*, 4 (2009), 552–563.

[150] Graf, E. H., Simmon, K. E., Tardif, K. D., Hymas, W., Flygare, S., Eilbeck, K., Yandell, M., and Schalaberg, R. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *Journal of Clinical Microbiology* (Apr 2016).

[151] Gregor, I., Dröge, J., Schirmer, M., Quince, C., and McHardy, A. C. PhyloPythiaS+: a self-training method for the rapid reconstruction of low-ranking taxonomic bins from metagenomes. *PeerJ 4* (2016).

[152] Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science 303*, 5656 (2004), 327–332.

[153] Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J. M., Dodd, R., Mulembakani, P., Schneider, B. S., Muyembe-Tamfum, J.-J., Stramer, S. L., and Chiu, C. Y. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine 7* (2015), 99.

[154] Gribaldo, S., and Philippe, H. Ancient phylogenetic relationships. *Theoretical Population Biology 61*, 4 (2002), 391–408.

[155] Guan, Y., Shortridge, K. F., Krauss, S., and Webster, R. G. Molecular characterization of H9n2 influenza viruses: Were they the donors of the internal genes of H5N1 viruses in Hong Kong? *Proceedings of the National Academy of Sciences 96*, 16 (1999), 9363–9367.

[156] Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology 59*, 3 (2010), 307–321.

[157] Gyuranecz, M., Foster, J. T., Dán, Á., Ip, H. S., Egstad, K. F., Parker, P. G., Higashiguchi, J. M., Skinner, M. A., Höfle, U., Kreizinger, Z., Dorrestein, G. M., Solt, S., Sós, E., Kim, Y. J., Uhart, M., Pereda, A., González-Hein, G., Hidalgo, H., Blanco, J.-M., and Erdélyi, K. Worldwide Phylogenetic Relationship of Avian Poxviruses. *Journal of Virology 87*, 9 (2013), 4938–4951.

[158] Haghighi, H. R., Read, L. R., Haeryfar, S. M. M., Behboudi, S., and Sharif, S. Identification of a Dual-Specific T Cell Epitope of the Hemagglutinin Antigen of an H5 Avian Influenza Virus in Chickens. *PLoS ONE 4*, 11 (2009).

[159] Halami, M. Y., Nieper, H., Müller, H., and Johne, R. Detection of a novel circovirus in mute swans (Cygnus olor) by using nested broad-spectrum PCR. *Virus Research 132*, 1-2 (2008), 208–212.

[160] Hancock, K., Veguilla, V., Lu, X., Zhong, W., Butler, E. N., Sun, H., Liu, F., Dong, L., DeVos, J. R., Gargiullo, P. M., Brammer, T. L., Cox, N. J., Tumpey, T. M., and Katz, J. M. Cross-Reactive Antibody Responses to the 2009 Pandemic H1N1 Influenza Virus. *New England Journal of Medicine 361*, 20 (2009), 1945–1952.

[161] Hanna, A., Banks, J., Marston, D. A., Ellis, R. J., Brookes, S. M., and Brown, I. H. Genetic Characterization of Highly Pathogenic Avian Influenza (H5N8) Virus from Domestic Ducks, England, November 2014. *Emerging Infectious Diseases 21*, 5 (2015), 879–882.

[162] Hannigan, G. D., Meisel, J. S., Tyldsley, A. S., Zheng, Q., Hodkinson, B. P., SanMiguel, A. J., Minot, S., Bushman, F. D., and Grice, E. A. The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio 6*, 5 (2015).

[163] Harder, T., Maurer-Stroh, S., Pohlmann, A., Starick, E., Höreth-Böntgen, D., Albrecht, K., Pannwitz, G., Teifke, J., Gunalan, V., Lee, R. T., Sauter-Louis, C., Homeier, T., Staubach, C., Wolf, C., Strebelow, G., Höper, D., Grund, C., Conraths, F. J., Mettenleiter, T. C., and Beer, M. Influenza A(H5N8) Virus Similar to Strain in Korea Causing Highly Pathogenic Avian Influenza in Germany. *Emerging Infectious Diseases 21*, 5 (2015), 860–863.

[164] Hashizume, M., Armstrong, B., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., and Sack, D. A. Rotavirus infections and climate variability in Dhaka, Bangladesh: a time-series analysis. *Epidemiology and Infection 136*, 9 (2008), 1281–1289.

[165] Heather, J. M., and Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics 107*, 1 (2016), 1–8.

[166] Hénault, S., Karoui, C., and Boschiroli, M. L. A PCR-based method for tuberculosis detection in wildlife. *Developments in Biologicals 126* (2006), 123–132; discussion 325–326.

[167] Hesterberg, U., Harris, K., Stroud, D., Guberti, V., Busani, L., Pittman, M., Piazza, V., Cook, A., and Brown, I. Avian influenza surveillance in wild birds in the European Union in 2006. *Influenza and Other Respiratory Viruses 3*, 1 (2009), 1–14.

[168] Hetmański, T. Timing of Breeding in the Feral Pigeon Columba livia f. domestica in Słupsk (NW Poland). *Acta Ornithologica 39*, 2 (2004), 105–110.

[169] Hill, S. C., Lee, Y.-J., Song, B.-M., Kang, H.-M., Lee, E.-K., Hanna, A., Gilbert, M., Brown, I. H., and Pybus, O. G. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. *Infection, Genetics and Evolution 34* (2015), 267–277.

[170] Hoelzer, K., Murcia, P. R., Baillie, G. J., Wood, J. L. N., Metzger, S. M., Osterrieder, N., Dubovi, E. J., Holmes, E. C., and Parrish, C. R. Intrahost evolutionary dynamics of canine influenza virus in naive and partially immune dogs. *Journal of Virology 84*, 10 (2010), 5329–5335.

[171] Hoffmann, E., Stech, J., Guan, Y., Webster, R. G., and Perez, D. R. Universal primer set for the full-length amplification of all influenza A viruses. *Archives of Virology 146*, 12 (2001), 2275–2289.

[172] Holmes, D., and Austad, S. Declining Immunity with Age in the Wild. *Science of Aging Knowledge Environment 2004*, 21 (2004), pe22.

[173] Holmes, E. C. The phylogeography of human viruses. *Molecular Ecology 13*, 4 (2004), 745–756.

[174] Holmes, E. C. The Evolutionary Genetics of Emerging Viruses. *Annual Review of Ecology, Evolution, and Systematics 40*, 1 (2009), 353–372.

[175] Holtz, L. R., Cao, S., Zhao, G., Bauer, I. K., Denno, D. M., Klein, E. J., Antonio, M., Stine, O. C., Snelling, T. L., Kirkwood, C. D., and Wang, D. Geographic variation in the eukaryotic virome of human diarrhea. *Virology 0* (2014), 556–564.

[176] Hong, S.-B. Visitation Aspect of Mergansers (Mergus spp.) in the Nakdong Estuary, Busan, Korea. *Journal of Asia-Pacific Biodiversity 6*, 1 (2013), 83–89.

[177] Hong, S.-B., and Lee, I.-S. Visitation Aspect of Common Goldeneye (Bucephala clangula) in the Nakdonggang Estuary, Busan, Korea. *Journal of Korean Nature 5*, 4 (2012), 329–334.

[178] Honkavuori, K. S., Shivaprasad, H. L., Briese, T., Street, C., Hirschberg, D. L., Hutchison, S. K., and Lipkin, W. I. Novel Picornavirus in Turkey Poults with Hepatitis, California, USA. *Emerging Infectious Diseases 17*, 3 (2011), 480–487.

[179] Hoque, M. A., Burgess, G. W., Cheam, A. L., and Skerratt, L. F. Epidemiology of avian influenza in wild aquatic birds in a biosecurity hotspot, North Queensland, Australia. *Preventive Veterinary Medicine 118*, 1 (2015), 169–181.

[180] Horton, K. G., Van Doren, B. M., Stepanian, P. M., Farnsworth, A., and Kelly, J. F. Where in the air? Aerial habitat use of nocturnally migrating birds. *Biology Letters 12*, 11 (2016).

[181] Hoye, B. J., Fouchier, R. A. M., and Klaassen, M. Host behaviour and physiology underpin individual variation in avian influenza virus infection in migratory Bewick's swans. *Proceedings of the Royal Society B: Biological Sciences 279*, 1728 (2012), 529–534.

[182] Hoye, B. J., Munster, V. J., Nishiura, H., Fouchier, R. A. M., Madsen, J., and Klaassen, M. Reconstructing an annual cycle of interaction: natural infection and antibody dynamics to avian influenza along a migratory flyway. *Oikos 120*, 5 (2011), 748–755.

[183] Huang, Y., Wille, M., Dobbin, A., Walzthöni, N. M., Robertson, G. J., Ojkic, D., Whitney, H., and Lang, A. S. Genetic structure of avian influenza viruses from ducks of the Atlantic flyway of North America. *PLoS ONE 9*, 1 (2014), e86999.

[184] Hughes, J., Allen, R. C., Baguelin, M., Hampson, K., Baillie, G. J., Elton, D., Newton, J. R., Kellam, P., Wood, J. L. N., Holmes, E. C., and Murcia, P. R. Transmission of Equine Influenza Virus during an Outbreak Is Characterized by Frequent Mixed Infections and Loose Transmission Bottlenecks. *PLoS Pathogens 8*, 12 (2012).

[185] Hughes, L. A., Savage, C., Naylor, C., Bennett, M., Chantrey, J., and Jones, R. Genetically Diverse Coronaviruses in Wild Bird Populations of Northern England. *Emerging Infectious Diseases 15*, 7 (2009), 1091–1094.

[186] Hur, W.-H., Lee, W.-S., and Rhim, S.-J. Changes in bird community in artificial wetlands of Sihwa Lake, South Korea. *Korean Journal of Environmental Ecology* (2005).

[187] Hurwitz, B. L., and Sullivan, M. B. The Pacific Ocean Virome (POV): A marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE 8*, 2 (2013).

[188] Imperiale, M., and Major, E. Polyomaviruses. In *Fields Virology*, D. Knipe, P. Howley, D. Griffen, R. Lamb, R. Martin, B. Roizman, and S. Straus, Eds., 4th edition ed. Lippincott Williams and Wilkins, Philadelphia, 2007, pp. 2263–2298.

[189] Imperiale, M., and Major, E. Herpes simplex viruses. In *Fields Virology*, D. Knipe, P. Howley, D. Griffen, R. Lamb, R. Martin, B. Roizman, and S. Straus, Eds., 6th edition ed. Lippincott Williams and Wilkins, Philadelphia, 2013, pp. 1823–1897.

[190] Imperiale, M. J., and Jiang, M. Polyomavirus persistence. *Annual Review of Virology 3*, 1 (2016), 517532.

[191] Ip, H. S., Dusek, R. J., Bodenstein, B., Torchetti, M. K., DeBruyn, P., Mansfield, K. G., DeLiberto, T., and Sleeman, J. M. High Rates of Detection of Clade 2.3.4.4 Highly Pathogenic Avian Influenza H5 Viruses in Wild Birds in the Pacific Northwest During the Winter of 201415. *Avian Diseases 60*, 1s (2016), 354–358.

[192] Ip, H. S., Torchetti, M. K., Crespo, R., Kohrs, P., DeBruyn, P., Mansfield, K. G., Baszler, T., Badcoe, L., Bodenstein, B., Shearn-Bochsler, V., Killian, M. L., Pedersen, J. C., Hines, N., Gidlewski,

T., DeLiberto, T., and Sleeman, J. M. Novel Eurasian Highly Pathogenic Influenza A H5 Viruses in Wild Birds, Washington, USA, 2014. *Emerging Infectious Diseases 21*, 5 (2015).

[193] Jeong, J., Kang, H.-M., Lee, E.-K., Song, B.-M., Kwon, Y.-K., Kim, H.-R., Choi, K.-S., Kim, J.-Y., Lee, H.-J., Moon, O.-K., Jeong, W., Choi, J., Baek, J.-H., Joo, Y.-S., Park, Y. H., Lee, H.-S., and Lee, Y.-J. Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its relationship with migratory birds in South Korea during 2014. *Veterinary Microbiology 173*, 34 (2014), 249–257.

[194] Jhung, M. A., and Nelson, D. I. Outbreaks of Avian Influenza A (H5N2), (H5N8), and (H5N1) Among Birds - United States, December 2014-January 2015. *MMWR. Morbidity and mortality weekly report 64*, 4 (2015), 111.

[195] Johne, R., Buck, C. B., Allander, T., Atwood, W. J., Garcea, R. L., Imperiale, M. J., Major, E. O., Ramqvist, T., and Norkin, L. C. Taxonomical developments in the family Polyomaviridae. *Archives of virology 156*, 9 (2011), 1627–1634.

[196] Johnson, C. K., Hitchens, P. L., Evans, T. S., Goldstein, T., Thomas, K., Clements, A., Joly, D. O., Wolfe, N. D., Daszak, P., Karesh, W. B., and Mazet, J. K. Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Scientific Reports 5* (2015), srep14830.

[197] Jones, B. A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M. Y., McKeever, D., Mutua, F., Young, J., McDermott, J., and Pfeiffer, D. U. Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences 110*, 21 (2013), 8399–8404.

[198] Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. Global trends in emerging infectious diseases. *Nature 451*, 7181 (2008), 990–993.

[199] Jourdain, E., Gunnarsson, G., Wahlgren, J., Latorre-Margalef, N., Bröjer, C., Sahlin, S., Svensson, L., Waldenström, J., Lundkvist, ., and Olsen, B. Influenza Virus in a Natural Host, the Mallard: Experimental Infection Data. *PLoS ONE 5*, 1 (2010), e8935.

[200] Kalthoff, D., Breithaupt, A., Teifke, J. P., Globig, A., Harder, T., Mettenleiter, T. C., and Beer, M. Highly pathogenic avian influenza virus (H5N1) in experimentally infected adult mute swans. *Emerging Infectious Diseases 14*, 8 (2008), 1267–1270.

[201] Kalthoff, D., Breithaupt, A., Teifke, J. P., Globig, A., Harder, T., Mettenleiter, T. C., and Beer, M. Pathogenicity of Highly Pathogenic Avian Influenza Virus (H5N1) in Adult Mute Swans. *Emerging Infectious Diseases 14*, 8 (2008), 1267–1270.

[202] Kanehira, K., Uchida, Y., Takemae, N., Hikono, H., Tsunekuni, R., and Saito, T. Characterization of an H5N8 influenza A virus isolated from chickens during an outbreak of severe avian influenza in Japan in April 2014. *Archives of Virology* (2015).

[203] Kang, H.-M., Lee, E.-K., Song, B.-M., Jeong, J., Choi, J.-G., Jeong, J., Moon, O.-K., Yoon, H., Cho, Y., Kang, Y.-M., Lee, H.-S., and Lee, Y.-J. Novel Reassortant Influenza A(H5N8) Viruses among Inoculated Domestic and Wild Ducks, South Korea, 2014. *Emerging Infectious Diseases 21*, 2 (2015), 298–304.

[204] Kapczynski, D. R., Liljebjelke, K., Kulkarni, G., Hunt, H., Jiang, H. J., and Petkov, D. Cross reactive cellular immune responses in chickens previously exposed to low pathogenic avian influenza. *BMC Proceedings 5*, Suppl 4 (2011), S13.

[205] Kapoor, A., Simmonds, P., Lipkin, W. I., Zaidi, S., and Delwart, E. Use of Nucleotide Composition Analysis To Infer Hosts for Three Novel Picorna-Like Viruses. *Journal of Virology 84*, 19 (2010), 10322–10328.

[206] Kapoor, A., Simmonds, P., Scheel, T. K. H., Hjelle, B., Cullen, J. M., Burbelo, P. D., Chauhan, L. V., Duraisamy, R., Sanchez Leon, M., Jain, K., Vandegrift, K. J., Calisher, C. H., Rice, C. M., and Lipkin, W. I. Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses. *mBio 4*, 2 (2013).

[207] Kapusinszky, B., Mulvaney, U., Jasinska, A. J., Deng, X., Freimer, N., and Delwart, E. Local Virus Extinctions following a Host Population Bottleneck. *Journal of Virology 89*, 16 (2015), 8152–8161.

[208] Kawaoka, Y., Chambers, T. M., Sladen, W. L., and Webster, R. G. Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? *Virology 163*, 1 (1988), 247–250.

[209] Keawcharoen, J., van Riel, D., van Amerongen, G., Bestebroer, T., Beyer, W. E., van Lavieren, R., Osterhaus, A. D., Fouchier, R. A., and Kuiken, T. Wild Ducks as Long-Distance Vectors of Highly Pathogenic Avian Influenza Virus (H5N1). *Emerging Infectious Diseases 14*, 4 (2008), 600–607.

[210] Khalenkov, A., Perk, S., Panshin, A., Golender, N., and Webster, R. G. Modulation of the severity of highly pathogenic H5N1 influenza in chickens previously inoculated with Israeli H9n2 influenza viruses. *Virology 383*, 1 (2009), 32–38.

[211] Kim, H.-J., Kim, D.-W., Kwon, I.-K., Hwang, J.-W., and Han, S.-H. 2014 Winter Waterbird Census of Korea. Tech. rep., National Institute of Biological Resources, 2014.

[212] Kim, H.-R., Kim, B.-S., Bae, Y.-C., Moon, O.-K., Oem, J.-K., Kang, H.-M., Choi, J.-G., Lee, O. S., and Lee, Y.-J. H5n1 subtype highly pathogenic avian influenza virus isolated from healthy mallard captured in South Korea. *Veterinary Microbiology 151*, 34 (2011), 386–389.

[213] Kim, H.-R., Kwon, Y.-K., Jang, I., Lee, Y.-J., Kang, H.-M., Lee, E.-K., Song, B.-M., Lee, H.-S., Joo, Y.-S., Lee, K.-H., Lee, H.-K., Baek, K.-H., and Bae, Y.-C. Pathologic Changes in Wild Birds Infected with Highly Pathogenic Avian Influenza A(H5N8) Viruses, South Korea, 2014. *Emerging Infectious Diseases 21*, 5 (2015), 775–780.

[214] Kim, H.-R., Park, C.-K., Lee, Y.-J., Woo, G.-H., Lee, K.-K., Oem, J.-K., Kim, S.-H., Jean, Y.-H., Bae, Y.-C., Yoon, S.-S., Roh, I.-S., Jeong, O.-M., Kim, H.-Y., Choi, J.-S., Byun, J.-W., Song, Y.-K., Kwon, J.-H., and Joo, Y.-S. An outbreak of highly pathogenic H5N1 avian influenza in Korea, 2008. *Veterinary Microbiology 141*, 34 (2010), 362–366.

[215] Kim, H.-R., Yoon, S.-J., Lee, H.-S., and Kwon, Y.-K. Identification of a picornavirus from chickens with transmissible viral proventriculitis using metagenomic analysis. *Archives of Virology 160*, 3 (2015), 701–709.

[216] Kim, Y.-I., Pascua, P. N. Q., Kwon, H.-I., Lim, G.-J., Kim, E.-H., Yoon, S.-W., Park, S.-J., Kim, S. M., Choi, E.-J., Si, Y.-J., Lee, O.-J., Shim, W.-S., Kim, S.-W., Mo, I.-P., Bae, Y., Lim, Y. T., Sung, M. H., Kim, C.-J., Webby, R. J., Webster, R. G., and Choi, Y. K. Pathobiological features of a novel, highly pathogenic avian influenza A(H5N8) virus. *Emerging Microbes & Infections 3*, 10 (2014), e75.

[217] King, A. M. Q., Adams, M. J., Lefkowitz, E. J., and Carstens, E. B. *Virus Taxonomy: Classification and Nomenclature of Viruses : Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier, 2012.

[218] King, D. R., Mutukwa, N., Lesellier, S., Cheeseman, C., Chambers, M. A., and Banks, M. Detection of Mustelid herpesvirus-1 infected European badgers (Meles meles) in the British Isles. *Journal of Wildlife Diseases 40*, 1 (2004), 99–102.

[219] Kingman, J. The coalescent. *Stochastic Processes and their Applications* (1982).

[220] Kirby, J. S., Stattersfield, A. J., Butchart, S. H. M., Evans, M. I., Grimmett, R. F. A., Jones, V. R., O'Sullivan, J., Tucker, G. M., and Newton, I. Key conservation issues for migratory land- and waterbird species on the world's major flyways. *Bird Conservation International 18*, S1 (2008).

[221] Kistler, A. L., Smith, J. M., Greninger, A. L., DeRisi, J. L., and Ganem, D. Analysis of Naturally Occurring Avian Bornavirus Infection and Transmission during an Outbreak of Proventricular Dilatation Disease among Captive Psittacine Birds. *Journal of Virology 84*, 4 (2010), 2176–2179.

[222] Kistler, W. M., Stallknecht, D. E., Deliberto, T. J., Swafford, S., Pedersen, K., Van Why, K., Wolf, P. C., Hill, J. A., Bruning, D. L., Cumbee, J. C., Mickley, R. M., Betsill, C. W., Randall, A. R., Berghaus, R. D., and Yabsley, M. J. Antibodies to avian influenza viruses in Canada geese (Branta canadensis): a potential surveillance tool? *Journal of Wildlife Diseases 48*, 4 (2012), 1097–1101.

[223] Kitamura, T., Sugimoto, C., Kato, A., Ebihara, H., Suzuki, M., Taguchi, F., Kawabe, K., and Yogo, Y. Persistent JC virus (JCV) infection is demonstrated by continuous shedding of the same JCV strains. *Journal of Clinical Microbiology 35*, 5 (1997), 1255–1257.

[224] Kleijn, D., Munster, V. J., Ebbinge, B. S., Jonkers, D. A., Müskens, G. J. D. M., Van Randen, Y., and Fouchier, R. a. M. Dynamics and ecological consequences of avian influenza virus infection in greater white-fronted geese in their winter staging areas. *Proceedings of the Royal Society B: Biological Sciences 277*, 1690 (2010), 2041–2048.

[225] Klein, S. L., and Flanagan, K. L. Sex differences in immune responses. *Nature Reviews Immunology 16*, 10 (2016), 626–638.

[226] Koelle, K., Cobey, S., Grenfell, B., and Pascual, M. Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. *Science 314*, 5807 (2006), 1898–1903.

[227] Kofstad, T., and Jonassen, C. M. Screening of Feral and Wood Pigeons for Viruses Harbouring a Conserved Mobile Viral Element: Characterization of Novel Astroviruses and Picornaviruses. *PLoS ONE 6*, 10 (2011).

[228] Krauss, S., Walker, D., Pryor, S. P., Niles, L., Chenghong, L., Hinshaw, V. S., and Webster, R. G. Influenza A Viruses of Migrating Wild Aquatic Birds in North America. *Vector-Borne and Zoonotic Diseases 4*, 3 (2004), 177–189.

[229] Krishnamurthy, S. R., and Wang, D. Origins and challenges of viral dark matter. *Virus Research* (2017).

[230] Krumbholz, A., Bininda-Emonds, O. R. P., Wutzler, P., and Zell, R. Phylogenetics, evolution, and medical importance of polyomaviruses. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases 9*, 5 (2009), 784–799.

[231] Kuiken, T. Is low pathogenic avian influenza virus virulent for wild waterbirds? *Proceedings of the Royal Society B: Biological Sciences 280*, 1763 (2013), 20130990.

[232] Kumar, M., Chu, H.-J., Rodenberg, J., Krauss, S., and Webster, R. G. Association of serologic and protective responses of avian influenza vaccines in chickens. *Avian Diseases 51*, 1 Suppl (2007), 481–483.

[233] Kwon, J.-H., Noh, Y. K., Lee, D.-H., Yuk, S.-S., Erdene-Ochir, T.-O., Noh, J.-Y., Hong, W.-T., Jeong, J.-H., Jeong, S., Gwon, G.-B., Song, C.-S., and Nahm, S.-S. Experimental infection with highly pathogenic H5N8 avian influenza viruses in the Mandarin duck (Aix galericulata) and domestic pigeon (Columba livia domestica). *Veterinary Microbiology 203* (2017), 95–102.

[234] Lam, T. T.-Y., Hon, C.-C., and Tang, J. W. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Critical Reviews in Clinical Laboratory Sciences 47*, 1 (2010), 5–49.

[235] Lam, T. T.-Y., Ip, H. S., Ghedin, E., Wentworth, D. E., Halpin, R. A., Stockwell, T. B., Spiro, D. J., Dusek, R. J., Bortner, J. B., Hoskins, J., Bales, B. D., Yparraguirre, D. R., and Holmes, E. C. Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among North American birds. *Ecology Letters 15*, 1 (2012), 24–33.

[236] Lambrecht, B., Marché, S., Houdart, P., van den berg, T., and Vangeluwe, D. Impact of age, season and flowing versus stagnant water habitat on Avian Influenza prevalence in Belgian Mute Swan. *Avian Diseases* (2015).

[237] LaMere, M. W., Lam, H.-T., Moquin, A., Haynes, L., Lund, F. E., Randall, T. D., and Kaminski, D. A. Contributions of Antinucleoprotein IgG to Heterosubtypic Immunity against Influenza Virus. *The Journal of Immunology 186*, 7 (2011), 4331–4339.

[238] Latorre-Margalef, N., Brown, J. D., Fojtik, A., Poulson, R. L., Carter, D., Franca, M., and Stallknecht, D. E. Competition between influenza A virus subtypes through heterosubtypic immunity modulates re-infection and antibody dynamics in the mallard duck. *PLoS Pathogens 13*, 6 (2017), e1006419.

[239] Latorre-Margalef, N., Gunnarsson, G., Munster, V. J., Fouchier, R. A. M., Osterhaus, A. D. M. E., Elmberg, J., Olsen, B., Wallensten, A., Haemig, P. D., Fransson, T., Brudin, L., and Waldenström, J. Effects of influenza A virus infection on migrating mallard ducks. *Proceedings of the Royal Society B: Biological Sciences 276*, 1659 (2009), 1029–1036.

[240] Latorre-Margalef, N., Tolf, C., Grosbois, V., Avril, A., Bengtsson, D., Wille, M., Osterhaus, A. D. M. E., Fouchier, R. A. M., Olsen, B., and Waldenström, J. Long-term variation in influenza A virus prevalence and subtype diversity in migratory mallards in northern Europe. *Proceedings of the Royal Society B: Biological Sciences 281*, 1781 (2014), 20140098.

[241] Laurenson, K., Sillero-Zubiri, C., Thompson, H., Shiferaw, F., Thirgood, S., and Malcolm, J. Disease as a threat to endangered species:

Ethiopian wolves, domestic dogs and canine pathogens. *Animal Conservation 1*, 4 (1998), 273–280.

[242] Lavoie, E. T. Avian immunosenescence. *Age 27*, 4 (2005), 281–285.

[243] Lecoq, H. Découverte du premier virus, le virus de la mosaïque du tabac: 1892ou1898? *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 324*, 10 (2001), 929–933.

[244] Lee, C.-W., Jang, J.-D., Jeong, K.-S., Kim, D.-K., and Joo, G.-J. Patterning habitat preference of avifaunal assemblage on the Nakdong River estuary (South Korea) using self-organizing map. *Ecological Informatics 5*, 2 (2010), 89–96.

[245] Lee, D.-H., Bahl, J., Torchetti, M. K., Killian, M. L., Ip, H. S., DeLiberto, T. J., and Swayne, D. E. Highly Pathogenic Avian Influenza Viruses and Generation of Novel Reassortants, United States, 2014-2015. *Emerging Infectious Diseases 22*, 7 (2016), 1283–1285.

[246] Lee, D.-H., Kwon, J.-H., Noh, J.-Y., Park, J.-K., Yuk, S.-S., Erdene-Ochir, T.-O., Lee, J.-B., Park, S.-Y., Choi, I.-S., Lee, S.-W., and et al. Pathogenicity of the Korean H5N8 highly pathogenic avian influenza virus in commercial domestic poultry species. *Avian Pathology 45*, 2 (Mar 2016), 208211.

[247] Lee, D.-H., Torchetti, M. K., Winker, K., Ip, H. S., Song, C.-S., and Swayne, D. E. Intercontinental Spread of Asian-origin H5N8 to North America through Beringia by Migratory Birds. *Journal of Virology* (2015), JVI.00728–15.

[248] Lee, Y.-J., Choi, Y.-K., Kim, Y.-J., Song, M.-S., Jeong, O.-M., Lee, E.-K., Jeon, W.-J., Jeong, W., Joh, S.-J., Choi, K.-s., Her, M., Kim, M.-C., Kim, A., Kim, M.-J., ho Lee, E., Oh, T.-G., Moon, H.-J., Yoo, D.-W., Kim, J.-H., Sung, M.-H., Poo, H., Kwon, J.-H., and Kim, C.-J. Highly Pathogenic Avian Influenza Virus (H5N1) in Domestic Poultry and Relationship with Migratory Birds, South Korea. *Emerging Infectious Diseases 14*, 3 (2008), 487–490.

[249] Lee, Y.-J., Kang, H.-M., Lee, E.-K., Song, B.-M., Jeong, J., Kwon, Y.-K., Kim, H.-R., Lee, K.-J., Hong, M.-S., Jang, I., Choi, K.-S., Kim, J.-Y., Lee, H.-J., Kang, M.-S., Jeong, O.-M., Baek, J.-H., Joo, Y.-S., Park, Y. H., and Lee, H.-S. Novel Reassortant Influenza A(H5N8) Viruses, South Korea, 2014. *Emerging Infectious Diseases 20*, 6 (2014), 1087–1089.

[250] Leibovitz, L. Natural occurrence and experimental study of pox and Haemoproteus infections in a mute swan (Cygnus olor). *Wildlife Disease 5*, 3 (1969), 130–136.

[251] Leland, D. S., and Ginocchio, C. C. Role of Cell Culture for Virus Detection in the Age of Technology. *Clinical Microbiology Reviews 20*, 1 (2007), 49–78.

[252] Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., and Suchard, M. A. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens 10*, 2 (2014), e1003932.

[253] Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. Bayesian Phylogeography Finds Its Roots. *PLoS Computational Biology 5*, 9 (2009), e1000520.

[254] Lemey, P., Rambaut, A., Welch, J. J., and Suchard, M. A. Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. *Molecular Biology and Evolution 27*, 8 (2010), 1877–1885.

[255] Lemey, P., Vandamme, A.-M., and Salemi, M., Eds. *The Phylogenetic Handbook: A practical approach to phylognetic analysis and hypothesis testing.*, 2 ed. Cambridge University Press, 2009.

[256] Lemmon, A. R., Lemmon, E. M., and Jockusch, E. A Likelihood Framework for Estimating Phylogeographic History on a Continuous Landscape. *Systematic Biology 57*, 4 (2008), 544–561.

[257] Lewis, N. S., Verhagen, J. H., Javakhishvili, Z., Russell, C. A., Lexmond, P., Westgeest, K. B., Bestebroer, T. M., Halpin, R. A., Lin, X., Ransier, A., Fedorova, N. B., Stockwell, T. B., Latorre-Margalef, N., Olsen, B., Smith, G., Bahl, J., Wentworth, D. E., Waldenström, J., Fouchier, R. A. M., and Graaf, M. d. Influenza A virus evolution and spatio-temporal dynamics in Eurasian Wild Birds: A phylogenetic and phylogeographic study of whole-genome sequence data. *Journal of General Virology* (2015), vir.0.000155.

[258] Li, H. seqtk. *https://github.com/lh3/seqtk* (Accessed 2017).

[259] Li, H., and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics 25*, 14 (2009), 1754–1760.

[260] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics 25*, 16 (2009), 2078–2079.

[261] Li, J., Quinque, D., Horz, H.-P., Li, M., Rzhetskaya, M., Raff, J. A., Hayes, M. G., and Stoneking, M. Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC microbiology 14* (2014), 316.

[262] Li, K. S., Guan, Y., Wang, J., Smith, G. J. D., Xu, K. M., Duan, L., Rahardjo, A. P., Puthavathana, P., Buranathai, C., Nguyen, T. D., Estoepangestie, A. T. S., Chaisingh, A., Auewarakul, P., Long, H. T., Hanh, N. T. H., Webby, R. J., Poon, L. L. M., Chen, H., Shortridge, K. F., Yuen, K. Y., Webster, R. G., and Peiris, J. S. M. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature 430*, 6996 (2004), 209–213.

[263] Li, L., Pesavento, P. A., Gaynor, A. M., Duerr, R. S., Phan, T. G., Zhang, W., Deng, X., and Delwart, E. A gyrovirus infecting a sea bird. *Archives of virology 160*, 8 (2015), 2105–2109.

[264] Li, Z., Hou, X., and Cao, G. Is mother-to-infant transmission the most important factor for persistent HBV infection? *Emerging Microbes & Infections 4*, 5 (2015).

[265] Liais, E., Croville, G., Mariette, J., Delverdier, M., Lucas, M.-N., Klopp, C., Lluch, J., Donnadieu, C., Guy, J. S., Corrand, L., Ducatez, M. F., and Guérin, J.-L. Novel Avian Coronavirus and Fulminating Disease in Guinea Fowl, France. *Emerging Infectious Diseases 20*, 1 (2014), 105–108.

[266] Lim, E. S., Zhou, Y., Zhao, G., Bauer, I. K., Droit, L., Ndao, I. M., Warner, B. B., Tarr, P. I., Wang, D., and Holtz, L. R. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature Medicine 21*, 10 (2015), 1228–1234.

[267] Lima, D. A., Cibulski, S. P., Finkler, F., Teixeira, T. F., Varela, A. P. M., Cerva, C., Loiko, M. R., Scheffer, C. M., Dos Santos, H. F., Mayer, F. Q., and Roehe, P. M. Faecal virome of healthy chickens reveals a large diversity of the eukaryote viral community, including novel circular ssDNA viruses. *Journal of General Virology 98*, 4 (2017), 690–703.

[268] Liu, C.-G., Liu, M., Liu, F., Lv, R., Liu, D.-F., Qu, L.-D., and Zhang, Y. Emerging multiple reassortant H5N5 avian influenza viruses in ducks, China, 2008. *Veterinary Microbiology 167*, 3-4 (2013), 296–306.

[269] Loman, N. J., Quick, J., and Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods 12*, 8 (2015), 733–735.

[270] Löndt, B. Z., Núñez, A., Banks, J., Alexander, D. J., Russell, C., Richard-Löndt, A. C., and Brown, I. H. The effect of age on the pathogenesis of a highly pathogenic avian influenza (HPAI) H5N1 virus in Pekin ducks (Anas platyrhynchos) infected experimentally. *Influenza and Other Respiratory Viruses 4*, 1 (2010), 17–25.

[271] Lowen, A. C., and Steel, J. Roles of Humidity and Temperature in Shaping Influenza Seasonality. *Journal of Virology 88*, 14 (2014), 7692–7695.

[272] Lu, L., Leigh Brown, A. J., and Lycett, S. J. Quantifying predictors for the spatial diffusion of avian influenza virus in China. *BMC Evolutionary Biology 17* (2017), 16.

[273] Lupiani, B., and Reddy, S. M. The history of avian influenza. *Comparative Immunology, Microbiology and Infectious Diseases 32*, 4 (2009), 311–323.

[274] Ma, D., Jasinska, A., Kristoff, J., Grobler, J. P., Turner, T., Jung, Y., Schmitt, C., Raehtz, K., Feyertag, F., Sosa, N. M., Wijewardana, V., Burke, D. S., Robertson, D. L., Tracy, R., Pandrea, I., Freimer, N., and Apetrei, C. SIVagm Infection in Wild African Green Monkeys from South Africa: Epidemiology, Natural History, and Evolutionary Considerations. *PLoS Pathogens 9*, 1 (2013), e1003011.

[275] Magee, D., Beard, R., Suchard, M. A., Lemey, P., and Scotch, M. Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion. *Archives of Virology 160*, 1 (2014), 215–224.

[276] Maio, N. D., Wu, C.-H., OReilly, K. M., and Wilson, D. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. *PLoS Genetics 11*, 8 (2015), e1005421.

[277] Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J.-Y., Dijoux, B., Vincent, C., Reynaud, B., and Lett, J.-M. Complex Recombination Patterns Arising during Geminivirus Coinfections Preserve and Demarcate Biologically Important Intra-Genome Interaction Networks. *PLoS Pathogens 7*, 9 (2011), e1002203.

[278] Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal 17*, 1 (2011), pp. 10–12.

[279] Martin, V., Pfeiffer, D. U., Zhou, X., Xiao, X., Prosser, D. J., Guo, F., and Gilbert, M. Spatial Distribution and Risk Factors of Highly Pathogenic Avian Influenza (HPAI) H5N1 in China. *PLoS Pathogens 7*, 3 (2011).

[280] Martinez, P. P., King, A. A., Yunus, M., Faruque, A. S. G., and Pascual, M. Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world. *Proceedings of the National Academy of Sciences 113*, 15 (2016), 4092–4097.

[281] Matrosovich, M., Zhou, N., Kawaoka, Y., and Webster, R. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *Journal of Virology 73*, 2 (1999), 1146–1155.

[282] Maydt, J., and Lengauer, T. Recco: recombination analysis using cost optimization. *Bioinformatics 22*, 9 (2006), 1064–1071.

[283] McCleery, R., Perrins, C., Sheldon, B., and Charmantier, A. Age-specific reproduction in a long-lived species: the combined effects of senescence and individual quality. *Proceedings of the Royal Society B: Biological Sciences 275*, 1637 (2008), 963–970.

[284] McCleery, R. H., Perrins, C., Wheeler, D., and Groves, S. Population structure, survival rates and productivity of mute swans breeding in a colony at Abbotsbury, Dorset, England. *Waterbirds 25* (2002), 192–201. WOS:000180160100027.

[285] McFerran, J., and Smyth, J. Avian adenoviruses. *Scientific and Technical Review of the Office International des Epizooties 19*, 2 (2000), 589–601.

[286] McHardy, A. C., Martín, H. G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods 4*, 1 (2007), 63–72.

[287] Minin, V. N., and Suchard, M. A. Counting labeled transitions in continuous-time Markov models of evolution. *Journal of Mathematical Biology 56*, 3 (2008), 391–412.

[288] Misra, V., Dumonceaux, T., Dubois, J., Willis, C., Nadin-Davis, S., Severini, A., Wandeler, A., Lindsay, R., and Artsob, H. Detection of polyoma and corona viruses in bats of Canada. *Journal of General Virology 90*, 8 (2009), 2015–2022.

[289] Mohiuddin, M., and Schellhorn, H. E. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. *Frontiers in Microbiology 6* (2015), 960.

[290] Mokili, J. L., Rohwer, F., and Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology 2*, 1 (2012), 63–77.

[291] Mollentze, N., Nel, L. H., Townsend, S., Roux, K. l., Hampson, K., Haydon, D. T., and Soubeyrand, S. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proceedings of the Royal Society B: Biological Sciences 281*, 1782 (2014), 20133251.

[292] Møller, A. P., and Haussy, C. Fitness consequences of variation in natural antibodies and complement in the Barn Swallow Hirundo rustica. *Functional Ecology 21*, 2 (2007), 363–371.

[293] Montgomery, R. D., Chowdhury, K. A., and Reese, J. G. Avian pox in a whistling swan. *Journal of the American Veterinary Medical Association 177*, 9 (1980), 930–931.

[294] Morelli, M. J., Thébaud, G., Chadœuf, J., King, D. P., Haydon, D. T., and Soubeyrand, S. A Bayesian Inference Framework to Reconstruct Transmission Trees Using Epidemiological and Genetic Data. *PLoS Computational Biology 8*, 11 (2012), e1002768.

[295] Moustafa, A., Xie, C., Kirkness, E., Biggs, W., Wong, E., Turpaz, Y., Bloom, K., Delwart, E., Nelson, K. E., Venter, J. C., and Telenti, A. The blood DNA virome in 8,000 humans. *PLoS Pathogens 13*, 3 (2017).

[296] Munster, V. J., Baas, C., Lexmond, P., Waldenstrom, J., Wallensten, A., Fransson, T., Rimmelzwaan, G. F., Beyer, W. E. P., Schutten, M., Olsen, B., Osterhaus, A. D. M. E., and Fouchier, R. A. M. Spatial, Temporal, and Species Variation in Prevalence of Influenza A Viruses in Wild Migratory Birds. *PLoS Pathogens 3*, 5 (2007).

[297] Munster, V. J., and Fouchier, R. A. M. Avian influenza virus: Of virus and bird ecology. *Vaccine 27*, 45 (2009), 6340–6344.

[298] Muradrasoli, S., Bálint, Á., Wahlgren, J., Waldenström, J., Belák, S., Blomberg, J., and Olsen, B. Prevalence and Phylogeny of Coronaviruses in Wild Birds from the Bering Strait Area (Beringia). *PLoS ONE 5*, 10 (2010), e13640.

[299] Murcia, P. R., Hughes, J., Battista, P., Lloyd, L., Baillie, G. J., Ramirez-Gonzalez, R. H., Ormond, D., Oliver, K., Elton, D., Mumford, J. A., Caccamo, M., Kellam, P., Grenfell, B. T., Holmes, E. C., and Wood, J. L. N. Evolution of an Eurasian avian-like influenza virus in naïve and vaccinated pigs. *PLoS Pathogens 8*, 5 (2012), e1002730.

[300] Nagler, F. P. O., and Rake, G. The Use of the Electron Microscope in Diagnosis of Variola, Vaccinia, and Varicella. *Journal of Bacteriology 55*, 1 (1948), 45–51.

[301] Nagy, A., Vostinakova, V., Pirchanova, Z., Cernikova, L., Dirbakova, Z., Mojzis, M., Jirincova, H., Havlickova, M., Dan, A., Ursu, K., Vilcek, S., and Hornickova, J. Development and evaluation of a one-step real-time RT-PCR assay for universal detection of influenza A viruses from avian and mammal species. *Archives of Virology 155*, 5 (2010), 665–673.

[302] Nallar, R., Papp, Z., Epp, T., Leighton, F. A., Swafford, S. R., DeLiberto, T. J., Dusek, R. J., Ip, H. S., Hall, J., Berhane, Y., and others. Demographic and Spatiotemporal Patterns of Avian Influenza Infection at the Continental Scale, and in Relation to Annual Life Cycle of a Migratory Host. *PLoS ONE 10*, 6 (2015), e0130662.

[303] Nam, H.-K., Choi, S.-H., Choi, Y.-S., and Yoo, J.-C. Patterns of Waterbirds Abundance and Habitat Use in Rice Fields. *Korean Journal of Environmental Agriculture 31*, 4 (2012), 359–367.

[304] Nasir, A., Forterre, P., Kim, K. M., and Caetano-Anollés, G. The distribution and impact of viral lineages in domains of life. *Frontiers in Microbiology 5* (2014).

[305] Nazir, J., Haumacher, R., Ike, A. C., and Marschang, R. E. Persistence of Avian Influenza Viruses in Lake Sediment, Duck Feces, and Duck Meat. *Applied and Environmental Microbiology 77*, 14 (2011), 4981–4985.

[306] Neu, U., Stehle, T., and Atwood, W. The Polyomaviridae: Contributions of virus structure to our understanding of virus receptors and infectious entry. *Virology 384*, 2 (2009), 389–399.

[307] Neumann, W., Martinuzzi, S., Estes, A. B., Pidgeon, A. M., Dettki, H., Ericsson, G., and Radeloff, V. C. Opportunities for the application of advanced remotely-sensed data in ecological studies of terrestrial animal movement. *Movement Ecology 3*, 1 (2015), 8.

[308] Noreen, E., Bourgeon, S., and Bech, C. Growing old with the immune system: a study of immunosenescence in the zebra finch (Taeniopygia guttata). *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology 181*, 5 (2011), 649–656.

[309] Nunes, M. R. T., Palacios, G., Faria, N. R., Jr, E. C. S., Pantoja, J. A., Rodrigues, S. G., Carvalho, V. L., Medeiros, D. B. A., Savji, N., Baele, G., Suchard, M. A., Lemey, P., Vasconcelos, P. F. C., and Lipkin, W. I. Air Travel Is Associated with Intracontinental Spread of Dengue Virus Serotypes 13 in Brazil. *PLoS Neglected Tropical Diseases 8*, 4 (2014), e2769.

[310] Obadia, T., Haneef, R., and Boëlle, P.-Y. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Medical Informatics and Decision Making 12*, 1 (2012).

[311] O'Brien, J. D., Minin, V. N., and Suchard, M. A. Learning to Count: Robust Estimates for Labeled Distances between Molecular Sequences. *Molecular Biology and Evolution 26*, 4 (2009), 801–814.

[312] OIE. Chapter 2.3.4, Avian influenza. In *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*. 2015.

[313] OIE. Chapter 2.3.8: Duck Virus Hepatitis. In *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2017*. 2017.

[314] Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D. M. E., and Fouchier, R. A. M. Global Patterns of Influenza A Virus in Wild Birds. *Science 312*, 5772 (2006), 384–388.

[315] Olson, S. H., Parmley, J., Soos, C., Gilbert, M., Latorre-Margalef, N., Hall, J. S., Hansbro, P. M., Leighton, F., Munster, V., and Joly, D. Sampling Strategies and Biodiversity of Influenza A Subtypes in Wild Birds. *PLoS ONE 9*, 3 (2014), e90826.

[316] Ozawa, M., Matsuu, A., Tokorozaki, K., Horie, M., Masatani, T., Nakagawa, H., Okuya, K., Kawabata, T., and Toda, S. Genetic diversity of highly pathogenic H5N8 avian influenza viruses at a single overwintering site of migratory birds in Japan, 2014/15. *Eurosurveillance 20*, 20 (2015), 21132.

[317] Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N. N., and Kyrpides, N. C. Uncovering Earth's virome. *Nature 536*, 7617 (2016).

[318] Palacios, M. G., Cunnick, J. E., Winkler, D. W., and Vleck, C. M. Immunosenescence in some but not all immune components in a free-living vertebrate, the tree swallow. *Proceedings of the Royal Society B: Biological Sciences 274*, 1612 (2007), 951–957.

[319] Palya, V., Glávits, R., Dobos-Kovács, M., Ivanics, É., Nagy, E., Bányai, K., Szücs, G., Dá, Á., and Benkö, M. Reovirus identified as cause of disease in young geese. *Avian Pathology 32*, 2 (2003), 129–138.

[320] Pankovics, P., Boros, Á., Kiss, T., Delwart, E., and Reuter, G. Detection of a mammalian-like astrovirus in bird, European roller (Coracias garrulus). *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 34* (2015), 114–121.

[321] Pantin-Jackwood, M. J., Costa-Hurtado, M., Bertran, K., DeJesus, E., Smith, D., and Swayne, D. E. Infectivity, transmission and pathogenicity of H5 highly pathogenic avian influenza clade 2.3.4.4 (H5N8 and H5N2) United States index viruses in Pekin ducks and Chinese geese. *Veterinary Research 48* (2017).

[322] Pantin-Jackwood, M. J., Costa-Hurtado, M., Shepherd, E., DeJesus, E., Smith, D., Spackman, E., Kapczynski, D. R., Suarez, D. L., Stallknecht, D., and Swayne, D. E. Pathogenicity and transmission of H5 and H7 highly pathogenic avian influenza viruses in mallards. *Journal of Virology* (2016), JVI.01165–16.

[323] Pantin-Jackwood, M. J., Suarez, D. L., Spackman, E., and Swayne, D. E. Age at infection affects the pathogenicity of Asian highly pathogenic avian influenza H5N1 viruses in ducks. *Virus Research 130*, 1 (2007), 151–161.

[324] Pantin-Jackwood, M. J., and Swayne, D. E. Pathobiology of Asian Highly Pathogenic Avian Influenza H5N1 Virus Infections in Ducks. *Avian Diseases 51*, s1 (2007), 250–259.

[325] Papp, Z., Clark, R. G., Parmley, E. J., Leighton, F. A., Waldner, C., and Soos, C. The ecology of avian influenza viruses in wild dabbling ducks (Anas spp.) in Canada. *PLoS ONE 12*, 5 (2017), e0176297.

[326] Parker, J., Rambaut, A., and Pybus, O. G. Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution 8*, 3 (2008), 239–246.

[327] Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E.-C., Burke, D. S., Calisher, C. H., Laughlin, C. A., Saif, L. J., and Daszak, P. Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. *Microbiology and Molecular Biology Reviews 72*, 3 (2008), 457–470.

[328] Parry, J. V., Lucas, M. H., Richmond, J. E., and Gardner, S. D. Evidence for a bovine origin of the polyomavirus detected in foetal rhesus monkey kidney cells, FRhK-4 and -6. *Archives of Virology 78*, 3-4 (1983), 151–165.

[329] Pasick, J., Berhane, Y., Embury-Hyatt, C., Copps, J., Kehler, H., Handel, K., Babiuk, S., Hooper-McGrevy, K., Li, Y., Le, Q. M., and Song, L. P. Susceptibility of Canada geese (Branta canadensis) to highly pathogenic avian influenza virus (H5N1). *Emerging infectious diseases 13*, 12 (2007), 1821.

[330] Pasick, J., Berhane, Y., Joseph, T., Bowes, V., Hisanaga, T., Handel, K., and Alexandersen, S. Reassortant Highly Pathogenic Influenza A H5N2 Virus Containing Gene Segments Related to Eurasian H5N8 in British Columbia, Canada, 2014. *Scientific Reports 5* (2015).

[331] Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods 10*, 12 (2013), 1200–1202.

[332] Pearce, J. M., Ramey, A. M., Ip, H. S., and Gill, R. E. Limited evidence of trans-hemispheric movement of avian influenza viruses among contemporary North American shorebird isolates. *Virus Research 148*, 1-2 (2010), 44–50.

[333] Pedersen, K., Marks, D. R., Arsnoe, D. M., Afonso, C. L., Bevins, S. N., Miller, P. J., Randall, A. R., and DeLiberto, T. J. Avian Paramyxovirus Serotype 1 (Newcastle Disease Virus), Avian Influenza Virus, and Salmonella spp. in Mute Swans (Cygnus olor) in the Great Lakes Region and Atlantic Coast of the United States. *Avian Diseases 58*, 1 (2013), 129–136.

[334] Peretti, A., FitzGerald, P. C., Bliskovsky, V., Buck, C. B., and Pastrana, D. V. Hamburger polyomaviruses. *Journal of General Virology* (2015), vir.0.000033.

[335] Peretti, A., FitzGerald, P. C., Bliskovsky, V., Pastrana, D. V., and Buck, C. B. Genome Sequence of a Fish-Associated Polyomavirus, Black Sea Bass (Centropristis striata) Polyomavirus 1. *Genome Announcements 3*, 1 (2015).

[336] Pérez-Losada, M., Christensen, R. G., McClellan, D. A., Adams, B. J., Viscidi, R. P., Demma, J. C., and Crandall, K. A. Comparing phylogenetic codivergence between polyomaviruses and their hosts. *Journal of Virology 80*, 12 (2006), 5663–5669.

[337] Perrins, C. M. Survival rates of young Mute Swans Cygnus olor. *Wildfowl 0*, 0 (2013), 95–102.

[338] Perrins, C. M., McCleery, R. H., and Ogilvie, M. A. A study of the breeding Mute Swans Cygnus olor at Abbotsbury. *Wildfowl 45*, 45 (1994), 1–14.

[339] Perrins, C. M., and Ogilvie, M. A. A study of the Abbotsbury Mute Swans. *Wildfowl 32*, 32 (1981), 35–47.

[340] Peterhans, E., Jungi, T. W., and Schweizer, M. BVDV and innate immunity. *Biologicals 31*, 2 (2003), 107112.

[341] Peters, A., Patterson, E. I., Baker, B. G. B., Holdsworth, M., Sarker, S., Ghorashi, S. A., and Raidal, S. R. Evidence of psittacine beak and feather disease virus spillover into wild critically endangered Orange-bellied Parrots (Neophema chrysogaster). *Journal of Wildlife Diseases 50*, 2 (2014), 288–296.

[342] Phan, T. G., Vo, N. P., Boros, Á., Pankovics, P., Reuter, G., Li, O. T. W., Wang, C., Deng, X., Poon, L. L. M., and Delwart, E. The Viruses of Wild Pigeon Droppings. *PLoS ONE 8*, 9 (2013).

[343] Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.-M., and Abergel, C. Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science 341*, 6143 (2013), 281–286.

[344] Pipas, J. M. Common and unique features of T antigens encoded by the polyomavirus group. *Journal of Virology 66*, 7 (1992), 3979–3985.

[345] Plotkin, J. B., Dushoff, J., and Levin, S. A. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proceedings of the National Academy of Sciences 99*, 9 (2002), 6263–6268.

[346] Polyomaviridae Study Group of the International Committee on Taxonomy of Viruses. A taxonomy update for the family Polyomaviridae. *Archives of Virology 161*, 6 (2016), 1739–1750.

[347] Pond, S. L. K., Frost, S. D. W., and Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics 21*, 5 (2005), 676–679.

[348] Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Molecular Biology and Evolution 23*, 10 (2006), 1891–1901.

[349] Posada, D. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution 25*, 7 (2008), 1253–1256.

[350] Potter, C. A history of influenza. *Journal of Applied Microbiology 91*, 4 (2001), 572–579.

[351] Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C., and Harvey, P. H. The Epidemic Behavior of the Hepatitis C Virus. *Science 292*, 5525 (2001), 2323–2325.

[352] Pybus, O. G., Perrins, C. M., Choudhury, B., Manvell, R. J., Nunez, A., Schulenburg, B., Sheldon, B. C., and Brown, I. H. The ecology and age structure of a highly pathogenic avian influenza virus outbreak in wild mute swans. *Parasitology 139*, Special Issue 14 (2012), 1914–1923.

[353] Pybus, O. G., and Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics 10*, 8 (2009), 540–550.

[354] 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., Jr, L. C. A., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., and Loman, N. J. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols 12*, 6 (2017), 1261–1276.

[355] Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J.-P., Cabeza-Cabrerizo, M., Camino-Sanchez, A., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G. G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallash, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Trina, R., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., O'Shea, M. K., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Woelfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keita, S., Rambaut, A., Formenty, P., Gunther, S., and Carroll, M. W. Real-time, portable genome sequencing for Ebola surveillance. *Nature 530*, 7589 (2016), 228–232.

[356] R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2014).

[357] Raghwani, J., Rose, R., Sheridan, I., Lemey, P., Suchard, M. A., Santantonio, T., Farci, P., Klenerman, P., and Pybus, O. G. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection. *PLoS Pathogens 12*, 9 (2016), e1005894.

[358] Rambaut, A. Path-O-Gen. *http://tree.bio.ed.ac.uk/software/pathogen/* (2009).

[359] Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution 2*, 1 (2016).

[360] Rambaut, A., Suchard, M., and Drummond, A. Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). *http://tree.bio.ed.ac.uk/software/tracer/* (2013).

[361] Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., Scola, B. L., Suzan, M., and Claverie, J.-M. The 1.2-Megabase Genome Sequence of Mimivirus. *Science 306*, 5700 (2004), 1344–1350.

[362] Recker, M., Pybus, O. G., Nee, S., and Gupta, S. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proceedings of the National Academy of Sciences 104*, 18 (2007), 7711–7716.

[363] Reyes, G. R., and Kim, J. P. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molecular and Cellular Probes 5*, 6 (1991), 473–481.

[364] Rice, P., Longden, I., and Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics 16*, 6 (2000), 276–277.

[365] Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. Integrative Genomics Viewer. *Nature Biotechnology 29*, 1 (2011), 24–26.

[366] Robinson, T. P., Wint, G. R. W., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S. I., and Gilbert, M. Mapping the Global Distribution of Livestock. *PLoS ONE 9*, 5 (2014), e96084.

[367] Roche, B., Drake, J. M., Brown, J., Stallknecht, D. E., Bedford, T., and Rohani, P. Adaptive Evolution and Environmental Durability Jointly Structure Phylodynamic Patterns in Avian Influenza Viruses. *PLoS Biology 12*, 8 (2014), e1001931.

[368] Ronquist, F., and Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics 19*, 12 (2003), 1572–1574.

[369] Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L., and Prosperi, M. Challenges in the analysis of viral metagenomes. *Virus Evolution 2*, 2 (2016).

[370] Rott, R., and Siddell, S. One hundred years of animal virology. *Journal of General Virology 79*, 11 (1998), 2871–2874.

[371] RStudio Team. *RStudio: Integrated Development Environment for R.* RStudio, Inc., Boston, MA, 2015.

[372] Runckel, C., Flenniken, M. L., Engel, J. C., Ruby, J. G., Ganem, D., Andino, R., and DeRisi, J. L. Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia. *PLoS ONE 6*, 6 (2011).

[373] Russell, C. A. Sick birds don't fly...or do they? *Science 354*, 6309 (2016), 174–175.

[374] Rutz, C., and Hays, G. C. New frontiers in biologging science. *Biology Letters 5*, 3 (2009), 289–292.

[375] Sachsenröder, J., Twardziok, S. O., Scheuch, M., and Johne, R. The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium. *PLoS ONE 9*, 2 (2014), e88888.

[376] Sadeghi, M., Riipinen, A., Väisänen, E., Chen, T., Kantola, K., Surcel, H.-M., Karikoski, R., Taskinen, H., Söderlund-Venermo, M., and Hedman, K. Newly discovered KI, WU, and Merkel cell polyomaviruses: No evidence of mother-to-fetus transmission. *Virology Journal 7*, 1 (2010), 251.

[377] Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences 74*, 12 (1977), 5463–5467.

[378] Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., and Weinstock, G. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Archives of Pathology and Laboratory Medicine 141* (2017), 776–786.

[379] Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., Khan, A. M., Woodford, N., Saunders, N. J., Wain, J., O'Grady, J., and Livermore, D. M. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *Journal of Antimicrobial Chemotherapy 72*, 1 (2017), 104–114.

[380] Schuurman, R., van Steenis, B., van Strien, A., van der Noordaa, J., and Sol, C. Frequent detection of bovine polyomavirus in commercial batches of calf serum by using the polymerase chain reaction. *Journal of General Virology 72 (Pt 11)* (1991), 2739–2745.

[381] Scotch, M., Lam, T. T.-Y., Pabilonia, K. L., Anderson, T., Baroch, J., Kohler, D., and DeLiberto, T. J. Diffusion of influenza viruses among migratory birds with a focus on the Southwest United States. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 26* (2014), 185–193.

[382] Scuda, N., Madinda, N. F., Akoua-Koffi, C., Adjogoua, E. V., Wevers, D., Hofmann, J., Cameron, K. N., Leendertz, S. A. J., Couacy-Hymann, E., Robbins, M., Boesch, C., Jarvis, M. A., Moens, U., Mugisha, L., Calvignac-Spencer, S., Leendertz, F. H., and Ehlers, B. Novel Polyomaviruses of Nonhuman Primates: Genetic and Serological Predictors for the Existence of Multiple Unknown Polyomaviruses within the Human Population. *PLoS Pathogens 9*, 6 (2013), e1003429.

[383] Seo, S. H., and Webster, R. G. Cross-Reactive, Cell-Mediated Immunity and Protection of Chickens from Lethal H5N1 Influenza Virus Infection in Hong Kong Poultry Markets. *Journal of Virology 75*, 6 (2001), 2516–2525.

[384] Shadan, F. F., and Villarreal, L. P. Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. *Proceedings of the National Academy of Sciences 90*, 9 (1993), 4117–4121.

[385] Shah, J. D., Baller, J., Zhang, Y., Silverstein, K., Xing, Z., and Cardona, C. J. Comparison of tissue sample processing methods for harvesting the viral metagenome and a snapshot of the RNA viral community in a turkey gut. *Journal of Virological Methods 209* (2014), 15–24.

[386] Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A., and Delwart, E. The fecal virome of pigs on a high-density farm. *Journal of Virology 85*, 22 (2011), 11697–11708.

[387] Shao, H., Lv, Y., Ye, J., Qian, K., Jin, W., and Qin, A. Isolation of a goose parvovirus from swan and its molecular characteristics. *Acta Virologica 58*, 2 (2014), 194–198.

[388] Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., and Zhang, Y.-Z. Redefining the invertebrate RNA virosphere. *Nature* (2016).

[389] Shin, Y.-U., Cho, H.-J., Lee, S.-W., Lee, H.-S., and Lee, D.-P. A study on the community of waterbirds and protective value at Gangjin bay in Joellananamdo, Korea. *Journal of Korean Nature 4*, 1 (2011), 55–59.

[390] Shoham, D., Jahangir, A., Ruenphet, S., and Takehara, K. Persistence of Avian Influenza Viruses in Various Artificially Frozen Environmental Water Types. *Influenza Research and Treatment 2012* (2012).

[391] Shortridge, K. F., Zhou, N. N., Guan, Y., Gao, P., Ito, T., Kawaoka, Y., Kodihalli, S., Krauss, S., Markwell, D., Murti, K. G., Norwood, M., Senne, D., Sims, L., Takada, A., and Webster, R. G. Characterization of avian H5N1 influenza viruses from poultry in Hong Kong. *Virology 252*, 2 (1998), 331–342.

[392] Siengsanan, J., Chaichoune, K., Phonaknguen, R., Sariya, L., Prompiram, P., Kocharin, W., Tangsudjai, S., Suwanpukdee, S.,

Wiriyarat, W., Pattanarangsan, R., Robertson, I., Blacksell, S. D., and Ratanakorn, P. Comparison of outbreaks of H5N1 highly pathogenic avian influenza in wild birds and poultry in thailand. *Journal of Wildlife Diseases 45*, 3 (2009), 740–747.

[393] Singh, S., Briles, W. E., Lupiani, B., and Collisson, E. W. Avian influenza viral nucleocapsid and hemagglutinin proteins induce chicken CD8+ memory T lymphocytes. *Virology 399*, 2 (2010), 231–238.

[394] Sinnecker, R., Sinnecker, H., Zilske, E., and Köhler, D. Surveillance of pelagic birds for influenza A viruses. *Acta Virologica 27*, 1 (1983), 75–79.

[395] Skowronski, D. M., Hottes, T. S., McElhaney, J. E., Janjua, N. Z., Sabaiduc, S., Chan, T., Gentleman, B., Purych, D., Gardy, J., Patrick, D. M., Brunham, R. C., Serres, G. D., and Petric, M. Immuno-epidemiologic Correlates of Pandemic H1N1 Surveillance Observations: Higher Antibody and Lower Cell-Mediated Immune Responses with Advanced Age. *Journal of Infectious Diseases 203*, 2 (2011), 158–167.

[396] Slatkin, M., and Maddison, W. P. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics 123*, 3 (1989), 603–613.

[397] Slomka, M. J., Pavlidis, T., Banks, J., Shell, W., McNally, A., Essen, S., and Brown, I. H. Validated H5 Eurasian real-time reverse transcriptase-polymerase chain reaction and its application in H5N1 outbreaks in 2005-2006. *Avian Diseases 51*, 1 Suppl (2007), 373–377.

[398] Slomka, M. J., Pavlidis, T., Coward, V. J., Voermans, J., Koch, G., Hanna, A., Banks, J., and Brown, I. H. Validated RealTime reverse transcriptase PCR methods for the diagnosis and pathotyping of Eurasian H7 avian influenza viruses. *Influenza and Other Respiratory Viruses 3*, 4 (2009), 151–164.

[399] Śmietanka, K., Minta, Z., Włodarczyk, R., Wyrostek, K., Jóźwiak, M., Olszewska, M., Minias, P., Kaczmarek, K., Janiszewski, T., and Kleszcz, A. Avian influenza viruses in wild birds at the Jeziorsko reservoir in Poland in 2008-2010. *Polish Journal of Veterinary Sciences 15*, 2 (2012).

[400] Smith, D. J. Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science 305*, 5682 (2004), 371–376.

[401] Smith, J., Sadeyen, J.-R., Butter, C., Kaiser, P., and Burt, D. W. Analysis of the Early Immune Response to Infection by Infectious Bursal Disease Virus in Chickens Differing in Their Resistance to the Disease. *Journal of Virology 89*, 5 (2015), 2469–2482.

[402] Smits, S. L., Bodewes, R., Ruiz-Gonzalez, A., Baumgärtner, W., Koopmans, M. P., Osterhaus, A. D. M. E., and Schürch, A. C. Assembly of viral genomes from metagenomes. *Frontiers in Microbiology 5* (2014).

[403] Smits, S. L., Raj, V. S., Oduber, M. D., Schapendonk, C. M. E., Bodewes, R., Provacia, L., Stittelaar, K. J., Osterhaus, A. D. M. E., and Haagmans, B. L. Metagenomic Analysis of the Ferret Fecal Viral Flora. *PLoS ONE 8*, 8 (2013), e71595.

[404] Somveille, M., Manica, A., Butchart, S. H. M., and Rodrigues, A. S. L. Mapping Global Diversity Patterns for Migratory Birds. *PLoS ONE 8*, 8 (2013), e70907.

[405] Song, B.-M., Kang, H.-M., Lee, E.-K., Jeong, J., Kang, Y., Lee, H.-S., and Lee, Y.-J. Pathogenicity of H5N8 Highly Pathogenic Avian Influenza Virus in Chickens, South Korea, 2014. *Journal of Veterinary Science* (2014).

[406] Soper, H. E. The Interpretation of Periodicity in Disease Prevalence. *Journal of the Royal Statistical Society 92*, 1 (1929), 34–73.

[407] Spackman, E., Prosser, D. J., Pantin-Jackwood, M. J., Berlin, A. M., and Stephens, C. B. The Pathogenesis of Clade 2.3.4.4 H5 Highly Pathogenic Avian Influenza Viruses in Ruddy Duck (Oxyura jamaicensis) and Lesser Scaup (Aythya affinis). *Journal of Wildlife Diseases* (2017).

[408] Spoor, L. E., McAdam, P. R., Weinert, L. A., Rambaut, A., Hasman, H., Aarestrup, F. M., Kearns, A. M., Larsen, A. R., Skov, R. L., and Fitzgerald, J. R. Livestock Origin for a Human Pandemic Clone of Community-Associated Methicillin-Resistant Staphylococcus aureus. *mBio 4*, 4 (2013), e00356–13.

[409] Stack, J. C., Murcia, P. R., Grenfell, B. T., Wood, J. L. N., and Holmes, E. C. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. *Proceedings of the Royal Society B: Biological Sciences 280*, 1750 (2013), 20122173.

[410] Stadler, T., Kühnert, D., Rasmussen, D. A., and Plessis, L. d. Insights into the Early Epidemic Spread of Ebola in Sierra Leone Provided by Viral Sequence Data. *PLoS Currents Outbreaks* (2014).

[411] Stallknecht, D. E., and Shane, S. M. Host range of avian influenza virus in free-living birds. *Veterinary Research Communications 12*, 2-3 (1988), 125–141.

[412] Stallknecht, D. E., Shane, S. M., Zwank, P. J., Senne, D. A., and Kearney, M. T. Avian influenza viruses from migratory and resident ducks of coastal Louisiana. *Avian Diseases 34*, 2 (1990), 398–405.

[413] Staples, J. E., and Monath, T. P. Yellow Fever: 100 Years of Discovery. *Journal of the American Medical Association 300*, 8 (2008), 960–962.

[414] Stein, L. D. The case for cloud computing in genome informatics. *Genome biology 11*, 5 (2010), 207.

[415] Steinhauer, D. A. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology 258*, 1 (1999), 1–20.

[416] Stevens, H., Bertelsen, M. F., Sijmons, S., Van Ranst, M., and Maes, P. Characterization of a Novel Polyomavirus Isolated from a Fibroma on the Trunk of an African Elephant (Loxodonta africana). *PLoS ONE 8*, 10 (2013), e77884.

[417] Stieneke-Gröber, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H. D., and Garten, W. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *The EMBO journal 11*, 7 (1992), 2407–2414.

[418] Stoehr, A. M., and Kokko, H. Sexual dimorphism in immunocompetence: what does life-history theory predict? *Behavioral Ecology 17*, 5 (2006), 751–756.

[419] Streicker, D. G., Turmelle, A. S., Vonhof, M. J., Kuzmin, I. V., McCracken, G. F., and Rupprecht, C. E. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science 329*, 5992 (2010), 676–679.

[420] Stroud, D., Chambers, D., Cook, S., Buxton, N., Fraser, B., Clement, P., Lewis, P., McLean, I., Baker, H., and Whitehead, S., Eds. *The UK SPA network: its scope and content*. JNCC, Peterborough, 2001.

[421] Suarez, D., Perdue, M., Cox, N., Rowe, T., Bender, C., Huang, J., and Swayne, D. Comparisons of highly virulent H5N1 influenza A viruses isolated from humans and chickens from Hong Kong. *Journal of Virology 72*, 8 (1998), 6678–6688.

[422] Suarez, D. L., and Schultz-Cherry, S. Immunology of avian influenza virus: a review. *Developmental & Comparative Immunology 24*, 23 (2000), 269–283.

[423] Sukumaran, J., and Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics 26*, 12 (2010), 1569–1571.

[424] Sun, H., Liu, L., Xu, G., Gao, G., Lu, X., and Lu, J. Characterization of clade 2.3.4.4 highly pathogenic H5 avian influenza viruses in ducks and chickens. *Veterinary Microbiology 182* (Jan 2016), 116122.

[425] Süss, J., Schäfer, J., Sinnecker, H., and Webster, R. G. Influenza virus subtypes in aquatic birds of eastern Germany. *Archives of Virology 135*, 1-2 (1994), 101–114.

[426] Suttle, C. A. Viruses in the sea. *Nature 437*, 7057 (2005), 356–361.

[427] Suttle, C. A. Viruses: unlocking the greatest biodiversity on Earth. *Genome 56*, 10 (2013), 542–544.

[428] Swayne, D. E. Impact of Vaccines and Vaccination on Global Control of Avian Influenza. *Avian Diseases 56*, 4s1 (2012), 818–828.

[429] Swayne, D. E., Beck, J. R., Garcia, M., and Stone, H. D. Influence of virus strain and antigen mass on efficacy of H5 avian influenza inactivated vaccines. *Avian Pathology 28*, 3 (1999), 245–255.

[430] Swayne, D. E., Lee, C.-W., and Spackman, E. Inactivated North American and European H5N2 avian influenza virus vaccines protect chickens from Asian H5N1 high pathogenicity avian influenza virus. *Avian Pathology: Journal of the W.V.P.A 35*, 2 (2006), 141–146.

[431] Swayne, D. E., Suarez, D. L., Spackman, E., Jadhao, S., Dauphin, G., Kim-Torchetti, M., McGrane, J., Weaver, J., Daniels, P., Wong, F., Selleck, P., Wiyono, A., Indriani, R., Yupiana, Y., Sawitri Siregar, E., Prajitno, T., Smith, D., and Fouchier, R. Antibody Titer Has Positive Predictive Value for Vaccine Protection against Challenge with Natural Antigenic-Drift Variants of H5N1 High-Pathogenicity Avian Influenza Viruses from Indonesia. *Journal of Virology 89*, 7 (2015), 3746–3762.

[432] Takekawa, J. Y., Newman, S. H., Xiao, X., Prosser, D. J., Spragens, K. A., Palm, E. C., Yan, B., Li, T., Lei, F., Zhao, D., Douglas, D. C., Muzaffar, S. B., and Ji, W. Migration of Waterfowl in the East Asian Flyway and Spatial Relationship to HPAI H5N1 Outbreaks. *Avian Diseases 54*, 1 (2010), 466–476.

[433] Talbi, C., Lemey, P., Suchard, M. A., Abdelatif, E., Elharrak, M., Jalal, N., Faouzi, A., Echevarría, J. E., Vazquez Morón, S., Rambaut, A., Campiz, N., Tatem, A. J., Holmes, E. C., and Bourhy, H. Phylodynamics and Human-Mediated Dispersal of a Zoonotic Virus. *PLoS Pathogens 6*, 10 (2010), e1001166.

[434] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution 28*, 10 (2011), 2731–2739.

[435] Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution 30*, 12 (2013), 2725–2729.

[436] Tao, Y., Shi, M., Conrardy, C., Kuzmin, I. V., Recuenco, S., Agwanda, B., Alvarez, D. A., Ellison, J. A., Gilbert, A. T., Moran, D., Niezgoda, M., Lindblade, K. A., Holmes, E. C., Breiman, R. F., Rupprecht, C. E., and Tong, S. Discovery of diverse polyomaviruses in bats and the evolutionary history of the Polyomaviridae. *Journal of General Virology 94*, Pt 4 (2013), 738–748.

[437] Teifke, J. P., Klopfleisch, R., Globig, A., Starick, E., Hoffmann, B., Wolf, P. U., Beer, M., Mettenleiter, T. C., and Harder, T. C. Pathology of Natural Infections by H5N1 Highly Pathogenic Avian Influenza Virus in Mute (Cygnus olor) and Whooper (Cygnus cygnus) Swans. *Veterinary Pathology 44*, 2 (2007), 137–143.

[438] Teske, L., Rubbenstroth, D., Meixner, M., Liere, K., Bartels, H., and Rautenschlein, S. Identification of a novel aviadenovirus, designated pigeon adenovirus 2 in domestic pigeons (Columba livia). *Virus Research 227* (2017), 15–22.

[439] The Global Consortium for H5N8 and Related Influenza Viruses. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science 354*, 6309 (2016), 213–217.

[440] Thrall, P. H., Antonovics, J., and Dobson, A. P. Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. *Proceedings of the Royal Society B: Biological Sciences 267*, 1452 (2000), 1555–1563.

[441] Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nature Protocols 4*, 4 (2009), 470–483.

[442] Tian, H., Zhou, S., Dong, L., Van Boeckel, T. P., Cui, Y., Wu, Y., Cazelles, B., Huang, S., Yang, R., Grenfell, B. T., and Xu, B. Avian influenza H5N1 viral and bird migration networks in Asia. *Proceedings of the National Academy of Sciences 112*, 1 (2015), 172–177.

[443] Tolf, C., Latorre-Margalef, N., Wille, M., Bengtsson, D., Gunnarsson, G., Grosbois, V., Hasselquist, D., Olsen, B., Elmberg, J., and Waldenström, J. Individual Variation in Influenza A Virus Infection Histories and Long-Term Immune Responses in Mallards. *PLoS ONE 8*, 4 (2013), e61201.

[444] Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., and Bates, K. K. Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 365*, 1550 (2010), 2163–2176.

[445] Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo, D. A. A., Chen, L.-M., Recuenco, S., Ellison, J. A., Davis, C. T., York, I. A., Turmelle, A. S., Moran, D., Rogers, S., Shi, M., Tao, Y., Weil, M. R., Tang, K., Rowe, L. A., Sammons, S., Xu, X., Frace, M., Lindblade, K. A., Cox, N. J., Anderson, L. J., Rupprecht, C. E., and Donis, R. O. A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences* (2012), 201116200.

[446] Tong, S., Zhu, X., Li, Y., Shi, M., Zhang, J., Bourgeois, M., Yang, H., Chen, X., Recuenco, S., Gomez, J., Chen, L.-M., Johnson, A., Tao, Y., Dreyfus, C., Yu, W., McBride, R., Carney, P. J., Gilbert, A. T., Chang, J., Guo, Z., Davis, C. T., Paulson, J. C., Stevens, J., Rupprecht, C. E., Holmes, E. C., Wilson, I. A., and Donis, R. O. New World Bats Harbor Diverse Influenza A Viruses. *PLoS Pathogens 9*, 10 (2013), e1003657.

[447] Tsai, B., Gilbert, J. M., Stehle, T., Lencer, W., Benjamin, T. L., and Rapoport, T. A. Gangliosides are receptors for murine polyoma virus and SV40. *The EMBO Journal 22*, 17 (Sep 2003), 4346–4355.

[448] Úbeda, F., and Jansen, V. A. A. The evolution of sex-specific virulence in infectious diseases. *Nature Communications 7* (2016), ncomms13849.

[449] United States Department of Agriculture. Update on Avian Influenza Findings in the Pacific Flyway. *USDA APHIS Technical Report* (2015).

[450] Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., and Neteler, M. Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*, 1550 (2010), 2177–2185.

[451] van den Brand, J. M. A. v. d., van Leeuwen, M., Schapendonk, C. M., Simon, J. H., Haagmans, B. L., Osterhaus, A. D. M. E., and Smits, S. L. Metagenomic Analysis of the Viral Flora of Pine Marten and European Badger Feces. *Journal of Virology 86*, 4 (2012), 2360–2365.

[452] van Dijk, J. G. B., Hoye, B. J., Verhagen, J. H., Nolet, B. A., Fouchier, R. A. M., and Klaassen, M. Juveniles and migrants as drivers for seasonal epizootics of avian influenza virus. *Journal of Animal Ecology 83*, 1 (2014), 266–275.

[453] Van Ghelue, M., Khan, M. T. H., Ehlers, B., and Moens, U. Genome analysis of the new human polyomaviruses. *Reviews in Medical Virology 22*, 6 (2012), 354–377.

[454] van Gils, J. A., Munster, V. J., Radersma, R., Liefhebber, D., Fouchier, R. A., and Klaassen, M. Hampered Foraging and Migratory Performance in Swans Infected with Low-Pathogenic Avian Influenza A Virus. *PLoS ONE 2*, 1 (2007), e184.

[455] Vatti, A., Monsalve, D. M., Pacheco, Y., Chang, C., Anaya, J.-M., and Gershwin, M. E. Original antigenic sin: A comprehensive review. *Journal of Autoimmunity 83*, Supplement C (2017), 12–21.

[456] Verhagen, J. H., Herfst, S., and Fouchier, R. A. M. How a virus travels the world. *Science 347*, 6222 (2015), 616–617.

[457] Verhagen, J. H., Höfle, U., Amerongen, G. v., Bildt, M. v. d., Majoor, F., Fouchier, R. A. M., and Kuiken, T. Long-Term Effect of Serial Infections with H13 and H16 Low-Pathogenic Avian Influenza Viruses in Black-Headed Gulls. *Journal of Virology 89*, 22 (2015), 11507–11522.

[458] Verhagen, J. H., van der Jeugd, H. P., Nolet, B. A., Slaterus, R., Kharitonov, S. P., de Vries, P. P., Vuong, O., Majoor, F., Kuiken, T., and Fouchier, R. A. Wild bird surveillance around outbreaks of highly pathogenic avian influenza A(H5N8) virus in the Netherlands, 2014, within the context of global flyways. *Euro Surveillance 20*, 12 (2015), 21–32.

[459] Viscidi, R. P., Rollison, D. E., Sondak, V. K., Silver, B., Messina, J. L., Giuliano, A. R., Fulp, W., Ajidahun, A., and Rivanera, D. Age-Specific Seroprevalence of Merkel Cell Polyomavirus, BK Virus, and JC Virus. *Clinical and Vaccine Immunology 18*, 10 (2011), 17371743.

[460] Volz, E. M., Koelle, K., and Bedford, T. Viral Phylodynamics. *PLoS Computational Biology 9*, 3 (2013).

[461] Walker, J. A., Sakaguchi, T., Matsuda, Y., Yoshida, T., and Kawaoka, Y. Location and character of the cellular enzyme that cleaves the hemagglutinin of a virulent avian influenza virus. *Virology 190*, 1 (1992), 278–287.

[462] Wallensten, A., Munster, V. J., Latorre-Margalef, N., Brytting, M., Elmberg, J., Fouchier, R. A., Fransson, T., Haemig, P. D., Karlsson, M., Lundkvist, ., Osterhaus, A. D., Stervander, M., Waldenström, J., and Olsen, B. Surveillance of Influenza Virus A in Migratory Waterfowl in Northern Europe. *Emerging Infectious Diseases 13*, 3 (2007), 404–411.

[463] Wallinga, J., and Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences 274*, 1609 (2007), 599–604.

[464] Wang, D. Fruits of Virus Discovery: New Pathogens and New Experimental Models. *Journal of Virology 89*, 3 (2015), 1486–1488.

[465] Wang, F., Wang, M., Dong, Y., Zhang, B., and Zhang, D. Genetic characterization of a novel calicivirus from a goose. *Archives of Virology 162*, 7 (2017), 2115–2118.

[466] Wang, J., Horner, G. W., and O'Keefe, J. S. Genetic characterisation of bovine herpesvirus 1 in New Zealand. *New Zealand Veterinary Journal 54*, 2 (2006), 61–66.

[467] Wang, J., Moore, N. E., Deng, Y.-M., Eccles, D. A., and Hall, R. J. MinION nanopore sequencing of an influenza genome. *Frontiers in Microbiology 6* (2015).

[468] Wang, X., Liu, N., Wang, F., Ning, K., Li, Y., and Zhang, D. Genetic characterization of a novel duck-origin picornavirus with six 2a proteins. *Journal of General Virology 95*, Pt 6 (2014), 1289–1296.

[469] Warden, C. D., and Lacey, S. F. Updated phylogenetic analysis of polyomavirus-host co-evolution. *Journal of Bioinformatics and Research 1*, 4 (2012), 46–49.

[470] Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., and Kellam, P. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 368*, 1614 (2013), 20120205.

[471] Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiological Reviews 56*, 1 (1992), 152–179.

[472] Webster, R. G., Yakhno, M., Hinshaw, V. S., Bean, W. J., and Copal Murti, K. Intestinal influenza: Replication and characterization of influenza viruses in ducks. *Virology 84*, 2 (1978), 268–278.

[473] Wei, C.-J., Boyington, J. C., Dai, K., Houser, K. V., Pearce, M. B., Kong, W.-P., Yang, Z.-y., Tumpey, T. M., and Nabel, G. J. Cross-Neutralization of 1918 and 2009 Influenza Viruses: Role of Glycans in Viral Evolution and Vaccine Design. *Science Translational Medicine 2*, 24 (2010), 24ra21.

[474] Wellehan, J. F., Rivera, R., Archer, L. L., Benham, C., Muller, J. K., Colegrove, K. M., Gulland, F. M., St. Leger, J. A., Venn-Watson, S. K., and Nollens, H. H. Characterization of California sea lion polyomavirus 1: Expansion of the known host range of the Polyomaviridae to Carnivora. *Infection, Genetics and Evolution 11*, 5 (2011), 987–996.

[475] Westover, K. M., and Hughes, A. L. Molecular Evolution of Viral Fusion and Matrix Protein Genes and Phylogenetic Relationships among the Paramyxoviridae. *Molecular Phylogenetics and Evolution 21*, 1 (2001), 128–134.

[476] White, D. J., Hall, R. J., Wang, J., Moore, N. E., Park, D., McInnes, K., Gartrell, B. D., and Tompkins, D. M. Discovery and complete genome sequence of a novel circovirus-like virus in the endangered rowi kiwi, Apteryx rowi. *Virus Genes 52*, 5 (2016), 727–731.

[477] White, L. F., and Pagano, M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Statistics in Medicine 27*, 16 (2008), 2999–3016.

[478] White, P. A. Evolution of norovirus. *Clinical Microbiology and Infection 20*, 8 (2014), 741–745.

[479] WHO. Antigenic and genetic characteristics of zoonotic influenza viruses and development of candidate vaccine viruses for pandemic preparedness (September 2014, http://www.who.int/influenza/vaccines/virus/201409_zoonotic_vaccinevirusupdate.pdf).

[480] Whon, T. W., Kim, M.-S., Roh, S. W., Shin, N.-R., Lee, H.-W., and Bae, J.-W. Metagenomic Characterization of Airborne Viral DNA Diversity in the Near-Surface Atmosphere. *Journal of Virology 86*, 15 (2012), 8221–8231.

[481] Wickham, H. *ggplot2: elegant graphics for data analysis*. Springer, New York, NY, 2009.

[482] Wigington, C. H., Sonderegger, D., Brussaard, C. P. D., Buchan, A., Finke, J. F., Fuhrman, J. A., Lennon, J. T., Middelboe, M.,

Suttle, C. A., Stock, C., Wilson, W. H., Wommack, K. E., Wilhelm, S. W., and Weitz, J. S. Re-examination of the relationship between marine virus and microbial cell abundances. *Nature Microbiology 1*, 3 (2016), nmicrobiol201524.

[483] Wikramaratna, P. S., Pybus, O. G., and Gupta, S. Contact between bird species of different lifespans can promote the emergence of highly pathogenic avian influenza strains. *Proceedings of the National Academy of Sciences* (2014), 201401849.

[484] Wikramaratna, P. S., Sandeman, M., Recker, M., and Gupta, S. The antigenic evolution of influenza: drift or thrift? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 368*, 1614 (2013), 20120200.

[485] Wille, M., Lindqvist, K., Muradrasoli, S., Olsen, B., and Järhult, J. D. Urbanization and the dynamics of RNA viruses in Mallards (Anas platyrhynchos). *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 51* (2017), 89–97.

[486] Wille, M., Muradrasoli, S., Nilsson, A., and Järhult, J. D. High Prevalence and Putative Lineage Maintenance of Avian Coronaviruses in Scandinavian Waterfowl. *PLoS ONE 11*, 3 (2016), e0150198.

[487] Wilson, H. M., Hall, J. S., Flint, P. L., Franson, J. C., Ely, C. R., Schmutz, J. A., and Samuel, M. D. High Seroprevalence of Antibodies to Avian Influenza Viruses among Wild Waterfowl in Alaska: Implications for Surveillance. *PLoS ONE 8*, 3 (2013), e58308.

[488] Wint, G., and Robinson, T. Gridded livestock of the world 2007. *FAO and Environmental Research Group Oxford* (2007).

[489] Wong, F. Y., Phommachanh, P., Kalpravidh, W., Chanthavisouk, C., Gilbert, J., Bingham, J., Davies, K. R., Cooke, J., Eagles, D., Phiphakhavong, S., Shan, S., Stevens, V., Williams, D. T., Bounma, P., Khambounheuang, B., Morrissy, C., Douangngeun, B., and Morzaria, S. Reassortant Highly Pathogenic Influenza A(H5n6) Virus in Laos. *Emerging Infectious Diseases 21*, 3 (2015).

[490] Wood, D. E., and Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology 15* (2014), R46.

[491] Woolhouse, M. E., and Gowtage-Sequeria, S. Host Range and Emerging and Reemerging Pathogens. *Emerging Infectious Diseases 11*, 12 (2005), 1842–1847.

[492] Worobey, M., Rambaut, A., Pybus, O. G., and Robertson, D. L. Questioning the Evidence for Genetic Recombination in the 1918 "Spanish Flu" Virus. *Science 296*, 5566 (2002), 211–211.

[493] Woźniakowski, G., and Samorek-Salamonowicz, E. First survey of the occurrence of duck enteritis virus (DEV) in free-ranging Polish water birds. *Archives of Virology 159*, 6 (2014), 1439–1444.

[494] Wright, P. F., G., N., and Y., K. Oxthomyxoviruses. In *Fields Virology*, D. Knipe, P. Howley, D. Griffen, R. Lamb, R. Martin, B. Roizman, and S. Straus, Eds., 6th edition ed. Lippincott Williams and Wilkins, Philadelphia, 2013, pp. 1186–1242.

[495] Wu, H., Peng, X., Xu, L., Jin, C., Cheng, L., Lu, X., Xie, T., Yao, H., and Wu, N. Novel Reassortant Influenza A(H5N8) Viruses in Domestic Ducks, Eastern China. *Emerging Infectious Diseases 20*, 8 (2014), 1315–1318.

[496] Wu, Y., Cui, L., Zhu, E., Zhou, W., Wang, Q., Wu, X., Wu, B., Huang, Y., and Liu, H.-J. Muscovy duck reovirus sigmaNS protein triggers autophagy enhancing virus replication. *Virology Journal 14* (2017).

[497] 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., and Jin, Q. 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 (2016), 609–620.

[498] Xia, X., and Lemey, P. Assessing substitution saturation with DAMBE. In *The Phylogenetic Handbook.*, Second edition. ed. Cambridge University Press, 2009, pp. 613–630.

[499] Xia, X., and Xie, Z. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. *Journal of Heredity 92*, 4 (2001), 371–373.

[500] Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution 26*, 1 (2003), 1–7.

[501] Xu, R., Ekiert, D. C., Krause, J. C., Hai, R., Crowe, J. E., and Wilson, I. A. Structural Basis of Preexisting Immunity to the 2009 H1N1 Pandemic Influenza Virus. *Science 328*, 5976 (2010), 357–360.

[502] Xu, X., Subbarao, n., Cox, N. J., and Guo, Y. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology 261*, 1 (1999), 15–19.

[503] Xue, J., Han, T., Xu, M., Zhao, J., and Zhang, G. The first serological investigation of Chicken astrovirus infection in China. *Biologicals: Journal of the International Association of Biological Standardization 47* (2017), 22–24.

[504] Yoon, H., Moon, O.-K., Jeong, W., Choi, J., Kang, Y.-M., Ahn, H.-Y., Kim, J.-H., Yoo, D.-S., Kwon, Y.-J., Chang, W.-S., Kim, M.-S., Kim, D.-S., Kim, Y.-S., and Joo, Y.-S. H5n8 Highly Pathogenic Avian Influenza in the Republic of Korea: Epidemiology During the First Wave, from

January Through July 2014. *Osong Public Health and Research Perspectives 6*, 2 (2015), 106–111.

[505] Yoshikawa, T. T. Epidemiology and Unique Aspects of Aging and Infectious Diseases. *Clinical Infectious Diseases 30*, 6 (2000), 931–933.

[506] Ypma, R. J. F., Ballegooijen, W. M. v., and Wallinga, J. Relating Phylogenetic Trees to Transmission Trees of Infectious Disease Outbreaks. *Genetics* (2013), genetics.113.154856.

[507] Ypma, R. J. F., Bataille, A. M. A., Stegeman, A., Koch, G., Wallinga, J., and Ballegooijen, W. M. v. Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proceedings of the Royal Society B: Biological Sciences 279*, 1728 (2012), 444–450.

[508] Yu, M., Ismail, M. M., Qureshi, M. A., Dearth, R. N., Barnes, H. J., and Saif, Y. M. Viral Agents Associated with Poult Enteritis and Mortality Syndrome: The Role of a Small Round Virus and a Turkey Coronavirus. *Avian Diseases 44*, 2 (2000), 297–304.

[509] Yu, X., Tsibane, T., McGraw, P. A., House, F. S., Keefer, C. J., Hicar, M. D., Tumpey, T. M., Pappas, C., Perrone, L. A., Martinez, O., Stevens, J., Wilson, I. A., Aguilar, P. V., Altschuler, E. L., Basler, C. F., and Jr, J. E. C. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature 455*, 7212 (2008), 532–536.

[510] Zaitlin, M. The Discovery of the Causal Agent of the Tobacco Mosaic Disease. In *Discoveries in Plant Biology*. World Publishing Co., Hong Kong, 1998.

[511] Zerbino, D. R. Using the Velvet de novo assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics 11* (2010), 11.11.5.

[512] Zerbino, D. R., and Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research 18*, 5 (2008), 821–829.

[513] Zhang, W., Li, L., Deng, X., Kapusinszky, B., and Delwart, E. What is for dinner? Viral metagenomics of US store bought beef, pork, and chicken. *Virology 468470* (2014), 303–310.

[514] Zhao, G., Gu, X., Lu, X., Pan, J., Duan, Z., Zhao, K., Gu, M., Liu, Q., He, L., Chen, J., Ge, S., Wang, Y., Chen, S., Wang, X., Peng, D., Wan, H., and Liu, X. Novel Reassortant Highly Pathogenic H5N2 Avian Influenza Viruses in Poultry in China. *PLoS ONE 7*, 9 (2012), e46183.

[515] Zhao, K., Gu, M., Zhong, L., Duan, Z., Zhang, Y., Zhu, Y., Zhao, G., Zhao, M., Chen, Z., Hu, S., Liu, W., Liu, X., Peng, D., and Liu, X. Characterization of three H5N5 and one H5N8 highly pathogenic avian influenza viruses in China. *Veterinary Microbiology 163*, 34 (2013), 351–357.

[516] Zhou, W., Ullman, K., Chowdry, V., Reining, M., Benyeda, Z., Baule, C., Juremalm, M., Wallgren, P., Schwarz, L., Zhou, E., Pedrero, S. P., Hennig-Pauka, I., Segales, J., and Liu, L. Molecular investigations on the prevalence and viral load of enteric viruses in pigs from five European countries. *Veterinary Microbiology 182* (2016), 75–81.

[517] Zhu, Y., Yongky, A., and Yin, J. Growth of an RNA virus in single cells reveals a broad fitness distribution. *Virology 385*, 1 (2009), 39–46.

[518] Zsak, L., Strother, K. O., and Kisary, J. Partial genome sequence analysis of parvoviruses associated with enteric disease in poultry. *Avian Pathology: Journal of the W.V.P.A 37*, 4 (2008), 435–441.

[519] Zwickl, J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion.* Thesis, University of Texas at Austin, 2006.

[520] Zylberberg, M., Van Hemert, C., Dumbacher, J. P., Handel, C. M., Tihan, T., and DeRisi, J. L. Novel Picornavirus Associated with Avian Keratin Disorder in Alaskan Birds. *mBio 7*, 4 (2016).