THE UNIVERSITY *of* EDINBURGH

Transmission Dynamics of Avian Influenza A Virus

Lu Lu

Submitted for the degree of Doctor of Philosophy
University of Edinburgh

2015

Declaration

This thesis has been composed by my own. The work is my own work except where otherwise indicated in the text. The work has not been submitted for any other degree or professional qualification except as specified.

Lu Lu

2014.12.18

Acknowledgements

I appreciate my funding from the China Scholarship Council/University of Edinburgh Joint Scholarship. This work could not have been completed without it. I would like to thank many people who provided great help to not only my PhD study but the life in UK during the past three years. They have made it possible for me to complete this thesis.

First of all, I am eternally grateful to my first and second supervisor: Prof. Andy J. Leigh Brown and Dr. Samantha J. Lycett, for providing me with the opportunity to undertake a PhD and for guidance during my PhD. They have put a considerable amount of time into meetings and discussions which were invaluable for the progress of my PhD. They have been incredibly patient and provided great encouragement when I met difficulties. They are the critical people to help me become confident and independent in my research. I also appreciate Prof. Andrew Rambaut, Prof. Ian White, Dr. Darren J Obbard for very helpful technical suggestions.

It is a great pleasure to work in this friendly environment in the Ashworth Laboratories and I'd like to thank all my friends here. Especially, I appreciate all the numbers in the Leigh Brown group: Dr. Melissa Ward, Emma Hodcroft, Manon Ragonnet and Mojca Zelnikar. Their sincere friendship and selfless help make me feel extremely touched and warm. I also thank the numbers in virus group: Gytis Dudas, Matthew Hall and Dr. Jessica Hedge, Dr. Trevor Bedford, they always give me suggestions with patience. Finally, I thank my family for their great love, encouragement, and for the confidence they have placed in me from start to finish.

Publications

The following papers have arisen from this thesis:

Lu Lu, Samantha J. Lycett, and Andrew J. Leigh Brown. Determining the genetic and phylogeographic origin of highly pathogenic avian influenza H7N3 in Mexico.*PLos one.* 2014, 9 doi: 10.1371/journal.pone.0107330

Lu Lu, Samantha J. Lycett, and Andrew J. Leigh Brown. Reassortment patterns of avian influenza viruses among different subtypes in internal segments. *BMC Evolutionary Biology.* 2014, **14**:16 doi:10.1186/1471-2148-14-16

List of Abbreviations

AI	Association Index
AIC	Akaike's information criterion
AIV	Avian influenza A virus
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BER	Bohai Economic Rim
BF	Bayes Factor
Cls	Credible intervals
CTMC	Continuous-time Markov chain
ED	Economic divided zones
ESS	Effective sample size
GLM	Generalized linear model
GTR	General reversible model
HKY	Hasegawa-Kishino-Yano
HME	Harmonic mean estimator
HPAI	Highly Pathogenic Avian Influenza
HPD	Highest posterior density
IP	Inclusion probabilities
LPAI	Low Pathogenic Avian Influenza
MC	Maximum monophyletic clade size
MCC	Maximum Clade Credibility
MCMC	Markov chain Monte Carlo
ML	Maximum likelihood
NCBI	the National Centre for Biotechnology Information
NIAID	the National Institute of Allergy and Infectious Diseases
NJ	Neighbour-joining
OIE	World Organisation for Animal Health
PRD	Pan-Pearl River Delta
PS	Parsimony Score
PS	Path sampling
RRW	Relaxed random walk
SA	Sialic acid
SIV	Swine influenza A viruses
SLAC	Single Likelihood Ancestor Counting
S-OIV	H1N1 swine-originated H1N1 pandemic strain
SS	Stepping stone
TMRCA	the most recent common ancestor
TR	Traditional regions
WHO	World Health Organization
YRD	Yangtze River Delta

Contents

List of Figures

List of Tables

Abstract

Influenza A virus (AIV) has an extremely high rate of mutation. Frequent exchanges of gene segments between different AIV (reassortment) have been responsible for major pandemics in recent human history. The presence of a wild bird reservoir maintains the threat of incursion of AIV into domestic birds, humans and other animals. In this thesis, I addressed unanswered questions of how diverse AIV subtypes (classified according to antigenicity of the two surface proteins, haemagglutinin and neuraminidase) evolve and interact among different bird populations in different parts of the world, using Bayesian phylogenetic methods with large datasets of full genome sequences.

Firstly, I explored the reassortment patterns of AIV internal segments among different subtypes by quantifying evolutionary parameters including reassortment rate, evolutionary rate and selective constraint in time-resolved Bayesian tree phylogenies. A major conclusion was that reassortment rate is negatively associated with selective constraint and that infection of wild rather than domestic birds was associated with a higher reassortment rate. Secondly, I described the spatial transmission pattern of AIV in China. Clustering of related viruses in particular geographic areas and economic zones was identified from the viral phylogeographic diffusion networks. The results indicated that Central China and the Pearl River Delta are two main sources of viral out flow; while the East Coast, especially the Yangtze River delta, is the major recipient area. Simultaneously, by applying a general linear model, the predictors that have the strongest impact on viral spatial diffusion were identified, including economic (agricultural) activity, climate, and ecology. Thirdly, I determined the genetic and phylogeographic origin of a recent H7N3 highly pathogenic avian influenza outbreak in Mexico. Location, subtype, avian host species and pathogenicity were modelled as discrete traits and jointly analysed using all eight viral gene segments. The results indicated that the outbreak AIV is a novel reassortant carried by wild waterfowl from different migration flyways in North America during the time period studied. Importantly, I concluded that Mexico, and Central America in general, might be a potential hotspot for AIV

reassortment events, a possibility which to date has not attracted widespread attention.

Overall, the work carried out in this thesis described the evolutionary dynamics of AIV from which important conclusions regarding its epidemiological impact in both Eurasia and North America can be drawn.

Chapter 1

Introduction

Chapter abstract

In this chapter, the general background of influenza viruses is provided, including viral structure, life cycle and major mechanisms of evolution. A detailed overview is given on avian influenza viruses (AIV), including the pathogenicity, the wild and domestic bird host populations, the inter/cross species transmission, the pandemic and outbreak threat to humans and animals, and a brief introduction to the phylogenetic analysis of viral sequence data.

1.1 Influenza viruses

Influenza viruses are negative-sense single-stranded RNA viruses with segmented genomes. They belong to the family Orthomyxoviridae (Smith *et al.*, 1933). There are three types of influenza viruses which differ vastly in host range and pathogenicity: influenza type A and type B viruses, and influenza C virus. Influenza A viruses are able to infect a wide variety of host species including wild ducks, chickens, ferrets, turkeys, pigs, horses, minks, seals and humans. All currently known avian influenza viruses (AIV) are influenza A viruses. In contrast, influenza B and C viruses have a much limited host range. They primarily infect humans and occasionally pigs and seals. Avian influenza A virus is the focus of this thesis.

Influenza A viruses are further classified into distinct subtypes based on the genetic and antigenic characteristics of two viral surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). Sixteen hemagglutinin (H1–16) and nine neuraminidase subtypes (N1–9) are known to be circulating in the reservoir of bird populations (McHardy & Adams, 2009; Webster *et al.*, 1992). However, in recent years, novel subtypes of influenza A viruses (H18N11 and H17N10) have been discovered in fruit bats in South and Central America, with HA and NA genes highly divergent from previously known influenza viruses (Tong *et al.*, 2012; Tong *et al.*, 2013). These findings indicate that our knowledge of the host range and genetic diversity of influenza A viruses is still incomplete. Influenza viruses can also be differentiated by serological tests, based on the antigenic differences in the conserved internal proteins, e.g., the nucleoprotein (NP) and matrix genes (M) (Plague, 2010).

1.1.1 Virion structure and life cycle

The genome of influenza A virus is composed of eight segments and has a total length of approximately 13.6 kb. They encode at least 13 proteins. In order of genomic length: segment 1 encodes the polymerase basic protein 2 (PB2); segment 2 encodes the polymerase basic protein 1 (PB1) and two other discovered proteins PB1-F2 and N40; segment 3 encodes the polymerase acidic protein (PA); segment 4 encodes the HA; segment 5 encodes the nucleoprotein (NP); The PA, PB1, PB2 and NP proteins consist of the viral polymerase complex, which packs the viral negative stranded RNAs inside. This ribonucleoproteins (vRNPs) structure comprises the virion core. Segment 6 encoding the NA; segment 7 encodes M1 and M2; and segment 8 encodes non-structural proteins (NS1) and the nuclear export protein (NEP, formally referred to asNS2). The virion is covered with the projection of three viral proteins: HA and NA, and the M2 transmembrane proton channel. The projections anchored a lipid bilayer which forms the viral envelope. The large, external domains of HA and NA are the major targets for neutralizing antibodies of the host immune response. The M1 matrix protein is located below the membrane. M1 forms a matrix holding the viral vRNPs (McHardy & Adams, 2009) (Figure1.1).

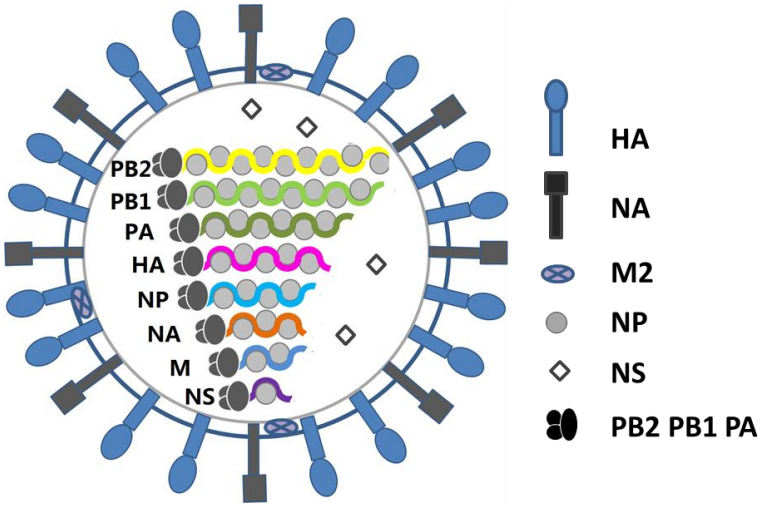


Figure 1.1 Schematic representation of influenza A virion

Like all viruses, influenza viruses replicate in living host cells. Early in the infection cycle, HA is responsible for binding to the receptor on the cell surface. The single polypeptide precursor (HA0) is first cleaved into two subunits: HA1 and HA2, by host cellular enzymes (Chen *et al.*, 1998). HA2 forms a highly conserved

stem-like structure, which anchors the globular domain to the viral membrane and contains the viral fusion peptide (Chen *et al.*, 1995); The mature HA1 protein mediates the attachment of the virus to sialic acid residues on the host cell surface and fuses the virus membrane envelope with the host cell membrane, thus delivering the viral genome into the cell in a process called receptor-mediated endocytosis (McHardy & Adams, 2009). The vRNP complexes are released into the cytoplasm and subsequently transported to the nucleus of the host cell, where the replication of new viral RNA and the transcription into proteins take place. Polymerases PB2, PB1 and PA are essential for viral transcription and replication: PB2 helps to generate primers for viral gene transcription (Engelhardt & Fodor, 2006; Guilligay *et al.*, 2008); PB1 is directly involved in viral RNA synthesis (Biswas & Nayak, 1994); PB1-F2 and N40 are two relatively discovered PB1 proteins which are encoded within the PB1 gene, both of which may be associated with the virus pathogenicity. PB1-F2 is associated with the severe symptoms (Kash *et al.*, 2006); N40 is found to have impact on the viral replication under certain conditions (Wise *et al.*, 2009). PA mainly acts as the endonuclease, and has promoter binding and cap binding function (Fodor *et al.*, 2002). The three polymerase proteins also have impacts on host-range restriction and genetic reassortment (Labadie *et al.*, 2007; Naffakh *et al.*, 2000; Naffakh *et al.*, 2008). Late in the infection cycle, M1 and NS2 facilitate the export of newly synthesized vRNPs the nuclear membrane. At the plasma membrane, the viral components are assembled into new viral particles. Finally, NA cleaves the terminal sialic acid residues from glycoproteins and glycolipids on the host cell surface, thus releasing budding viral particles from the infected host cell (Neumann *et al.*, 2009) .

1.1.2 Mechanisms of Evolution: antigenic shift and drift

The genetic diversity of influenza viruses is achieved through both mutation and reassortment (Figure 1.2). The viral RNA-dependent RNA polymerase is responsible for catalysing the replication of the RNA from the RNA template. The replication process is error-prone and there is no proofreading during replication, so the virus exhibits a high rate of nucleotide substitution. Variation in viruses that involves the accumulation of mutations within the genes that code for

antibody-binding sites,, this is called antigenic drift, resulting from natural selection of variants circulating among an immune or partially immune population (Scholtissek, 1995; Wolf *et al.*, 2006). The exchange of entire genetic segments between two co-infecting parental viruses during replication is called reassortment (Webster, 1998). When influenza viruses reassort, especially occurring among HA or NA gene segments, they may acquire completely new antigens (also called antigenic shift). The novel variant may not be recognized by the immune system in the host population and can be transmitted efficiently, leading to an epidemic or pandemic (Parrish & Kawaoka, 2005).

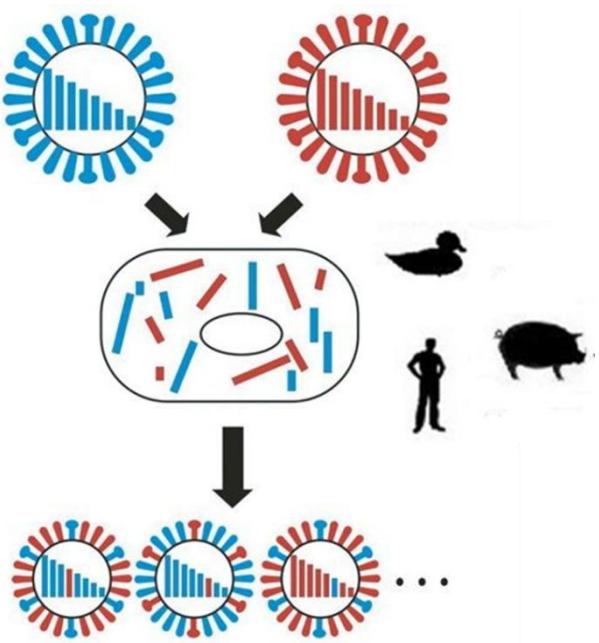


Figure 1.2 Genome reassortment of influenza A virus
The structure of the influenza virus genome allows for exchange of the eight RNA segments between two different viruses (blue and red) co-infecting a host cell.

1.2 Avian influenza viruses (AIV)

Avian influenza is a disease of great importance to both animal and human health. It's ecology and epidemiology have undergone substantial changes over time (Kalthoff *et al.*, 2010).

1.2.1 Pathogenicity of AIV

AIV can be classified into two groups according to their pathogenicity: highly pathogenic avian influenza viruses (HPAIV) and low pathogenic avian influenza

viruses (LPAIV). LPAIV are common in wild birds and poultry. The cleavage site between HA1 and HA2 proteins of LPAIV is composed of only one to two basic amino acids at distinct positions. Trypsin-like enzymes preferentially expressed at the surface of respiratory and gastrointestinal epithelia of host species recognise this monobasic cleavage motif (Wood *et al.*, 1993). As a result, efficient replication of AIV with a monobasic HA cleavage site is restricted to these tissues - leading to only mild, primarily respiratory disease in bird populations, unless exacerbated by other infections or environmental conditions.

In contrast, the main factor contributing to virulence in HPAIV is the insertion of a polybasic motif in its cleavage site, which facilitates systemic virus replication (Kalthoff *et al.*, 2010). HPAIV can cause significant mortality (from 75 to 100%) in unvaccinated flocks. These viruses have mostly been restricted to subtypes H5 and H7. However, other subtypes of LPAIV like H9N2 may become lethal if subjected to artificial reassortment with high pathogenic virus gene segments in the laboratory (Sun *et al.*, 2011). Though the trigger for LPAIV to mutate into HPAIV is still unknown, HPAI outbreaks in poultry seem to take place rapidly after LPAI introduction from wild birds (Kalthoff *et al.*, 2010; Perez *et al.*, 2008; Webster, 1998), and the wider the circulation of LPAI in poultry, the higher the probability of mutation to HPAI (Alexander, 2007). Phylogenetic analysis of AIV of the H7 subtype also showed that the HPAI viruses did not evolve into a separate phylogenetic lineage, but fell in the same lineages with non-pathogenic and low-pathogenic strains (Banks *et al.*, 2000).

In addition, HPAI viruses were rarely isolated from wild birds in the past, compared to the extremely high isolation rates of LPAI viruses (particularly in waterfowl) recorded in surveillance studies (Capua & Alexander, 2007). However, varieties of sub-strains of HPAI H5N1 have appeared in recent years and have probably become endemic in both wild birds and poultry, with varied virulence in different host species (Lycett *et al.*, 2009).

1.2.2 AIV in wild birds

1.2.2.1 Wild birds as reservoirs of AIV

Before the recent discovery of influenza virus in bats (Sun *et al.*, 2011), wild bird populations had long been considered as the ancestral hosts of influenza A viruses, or the genetic donor of the pandemic influenza strains. There are 16 subtypes of influenza virus HA and 9 subtypes of NA known to infect birds. More than 100 of the theoretical total 144 HA-NA combinations of influenza A virus have been isolated from wild birds, indicating an extensive reservoir of influenza viruses is circulating in wild bird populations (Bahl *et al.*, 2009; Perez *et al.*, 2008).

At least 105 out of a total of 9000 wild bird species are identified as the hosts of influenza A virus (Huang *et al.*, 2012). Two major wild bird orders are Anseriformes (e.g., ducks, geese and swan) and Charadriiformes (e.g., gulls, terns, surfbirds and sandpipers), and most surveillance efforts have focused on these species (Schnebel *et al.*, 2007). The distribution of HA and NA subtypes is not the same among wild bird AIV isolates, with certain HA and NA subtypes more common than others. To be specific, subtypes H3, H4 and H6 of HA and N2, N6, N8 of NA are detected most often in North American and northern European wild ducks (Swayne, 2008), while subtypes of H4, H9, H11, H13 and N6, N9 predominate in shorebirds and gulls. Isolates from cage birds usually contain H3 or H4; however, infections with high pathogenicity subtypes containing H7 or H5 can also occur. The observation that AIV with H13 and H16 HA subtypes are isolated predominantly from gulls suggests that the two subtypes are more promiscuous in host range than other subtypes (Taubenberger & Kash, 2010).

1.2.2.2 Global flyway of wild birds

Three major global flyways of wild migratory birds have been identified: the Americas Flyway, the African-Eurasian Flyway and the East Asian-Australasian Flyway (Figure 1.3). The Americas Flyway connects North American breeding grounds with wintering grounds in the Caribbean and Central and South America. The Africa-Eurasia Flyway connects the breeding grounds of Europe and northern Asia with the wintering grounds in Africa, and including vital stop-over sites in the

Middle East and Mediterranean. The East Asia-Australasia Flyway connects north-east Asian breeding grounds with the wintering grounds in south-east Asia and Australia, with major stop-over sites in China and the Korean Peninsula. From the overlap of flyways there is a high chance of AIV exchange between Europe and Asia, whereas the AIV gene exchange between the Americas and Eurasia is more restricted (Olsen *et al.*, 2006; Perez *et al.*, 2008). Studies of the HPAI H5N1 outbreaks at Qinghai Lake in China suggest that the influenza viruses were circulated by wild birds through the Central Asian flyway, and are still continuously undergoing evolution (Chen *et al.*, 2005; Hu *et al.*, 2011; Kilpatrick *et al.*, 2006).

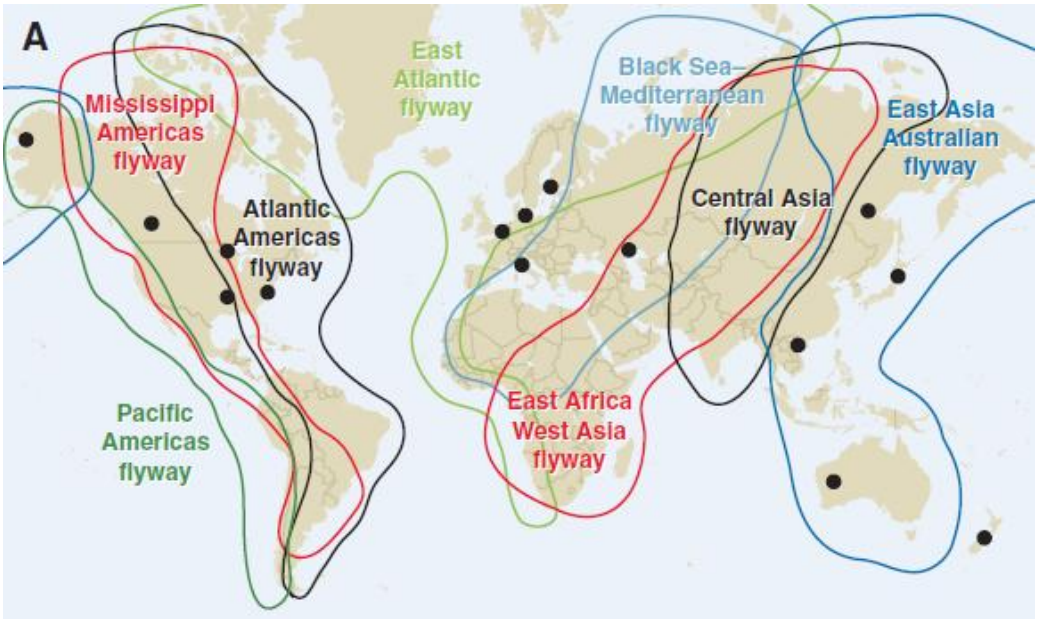


Figure 1.3 Migratory flyways of wild bird populations
(Adapted from Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus AD, Fouchier RA. Global patterns of influenza A virus in wild birds. *Science* 2006, 312:384-388) Black dots indicate the locations of historical and current influenza virus surveillance sites from which data have been used in this paper.

Transmission in feeding pools provides a route of virus transmission between wild birds, which is of great ecological concern (Webster, 2002). Studies have been conducted to determine how migratory birds might contribute to disease in wild birds and poultry. The susceptibility of migratory birds to infection has been analyzed using a model to trace their path along flyways (Bourouiba, 2011). Geographical separation of their hosts has shaped the gene pool of AIV into two independently

evolving Eurasian and American lineages. However, phylogenetic evidence indicates that these lineages mix on occasion (Olsen *et al.*, 2006; Perez *et al.*, 2008). Through characterizing changes in rates of nucleotide substitution and selection pressures, previous study has shown viral gene flow from Eurasia led to the replacement of endemic AIV in North America, likely through competition for susceptible hosts (Bahl *et al.*, 2009). As a result, valid concern has been raised over the importation of the virus into North America via migratory birds.

In addition, mixed infection and reassortment have been shown to be extremely common in AIV in wild birds (Wang *et al.*, 2008). The large number of different HA-NA subtype combinations recovered also highlights the frequency of reassortment in AIV in wild birds, but provides little evidence for elevated fitness of specific HA-NA combinations (Taubenberger & Kash, 2010). Studies of the transmission of AIV in wild birds at the Asian–North American interface demonstrated hemispheric mixing with reassortment (Perez *et al.*, 2008). However, most reassortants possess only one gene segment derived from the other hemisphere, indicating that there is little or no survival advantage for such hemispheric crossovers in the new gene pool (Perez *et al.*, 2008). The reassortment of AIV along the wild bird North America flyways will be analysed in Chapter 5.

1.2.3 AIV in domestic birds

Influenza A viruses have been isolated from domestic birds such as chickens, turkeys, domestic ducks, quails and pheasants. Surveillance studies have found that influenza viruses are endemic in poultry populations globally (Liu *et al.*, 2003; Senne *et al.*, 2003; Suarez *et al.*, 2003). Multiple co-circulating lineages of H5, H6, H7 and H9 subtypes exist in poultry (Chen *et al.*, 2006b; Xu *et al.*, 2007), while other subtypes harboured by these hosts are unlikely to become endemics. Domestic birds are susceptible to infection with wild-bird-derived AIV (Taubenberger & Kash, 2010). Phylogenetic analyses have clearly demonstrated the stable adaptation of AIV to poultry hosts, in contrast to the apparent transient nature of viral evolution within wild bird populations (Dugan *et al.*, 2008). The molecular features of host adaptation to domestic poultry include positive selection for mutations in HA, NA, and viral

RNP proteins (Wasilenko *et al.*, 2008). The characteristic feature of poultry-adapted AIV is an in-frame deletion of approximately 20 amino acids from the stalk region of the NA (Blok & Air, 1982). In contrast to mild or asymptomatic AIV infection in wild birds, AIV infection in domestic birds often develop respiratory illnesses with mortality rates of up to 100% and may result in economic loss and disruption to the poultry industry.

Lack of hygiene, overstocking and mixing of different domestic animals greatly increases the risk of spreading the infection. Live bird markets, farms and slaughterhouses are the major disseminating points of AIV in poultry flocks because of poor biosecurity measures. In particular the live bird markets are an ideal environment for the emergence of virulent AIV, as multiple influenza subtypes circulate and where the virus can persist and replicate in poultry for extended periods of time (Cardona *et al.*, 2009; Wan *et al.*, 2011; Webster, 2004). More importantly, the movement of humans as well as the trade of birds can be important factors in the spread of the virus. With globalization, the extensive and intensive movements of humans and poultry around the world at an unprecedented pace provides more opportunities for the spreading of influenza virus (FAO, 2007). With the attempts to eradicate the virus, culling and vaccination of flocks, as well as the implementation of local and international trade restrictions are often introduced (Davison *et al.*, 1999; Pearson, 2003). The human impact on AIV diffusion in China will be explored in Chapter 4.

1.2.4 Inter-species and cross-species transmission

Genetic exchanges between viruses circulating in wild and domestic birds have been documented, especially when LPAIV from wild birds infects poultry (Campitelli *et al.*, 2004; Duan *et al.*, 2007). AIV have been isolated from faecal material and from lake water, where they can be present at extremely high concentrations (Webster, 2002). Infected waterfowl may spread AI viruses to poultry farms by a variety of mechanisms. Surface water contaminated with influenza viruses used as drinking water may be a source of infection.

Wild birds and domestic poultry can exchange viruses when they share the same environment. The spillover of Asian HP H5N1 viruses from domestic to wild birds revealed the potential for reverse flow (Chen *et al.*, 2005; Feare, 2010). A typical example is that among the outbreaks reported in poultry or wild birds on the Qinghai-Tibet Plateau, the initial outbreaks occurred in poultry, followed by outbreaks in wild birds, and the outbreaks occurred first in Lasa (in Xinjiang) then transmitted to Tibet (Prosser *et al.*, 2011a). In addition, human activities play a very important role in the secondary spread of AI among domestic poultry. Caretakers, farmers, workers, trucks and drivers visiting farms, moving birds or delivering food have caused the spread of AI virus both between and within farms (Bourouiba, 2011). The most significant example is the persistence and recurrence of H5N1 avian influenza, which can largely be blamed on the movement and infection by migratory birds. Trade in poultry, poultry products and caged birds also account for H5N1 prevalence in the epidemic regions (Bourouiba, 2011).

Influenza A viruses have also been isolated from animals including pigs, horses, ferrets, cats (Ito *et al.*, 1997) and marine mammals such as whales and seals (Hinshaw *et al.*, 1986; Hinshaw *et al.*, 1984). The different binding preference of the viruses is believed to be one of the major factors impeding cross-species transmission. AIV bind preferentially to sialic acid (SA)-a2, 3-Gal terminated polysaccharides which are prominent in avian cells, whereas human influenza viruses bind preferentially to SA-a2, 6-Gal terminated polysaccharides which are present in human epithelial cells. However, AI viruses do occasionally infect people which indicates that this barrier is not insurmountable (Capua & Alexander, 2008). AIV from wild and domestic bird populations have been associated with stable host-switching events to novel hosts including horses and swine, which led to the emergence of influenza A viruses transmissible in the new host. Importantly, AIV can be transmitted to humans and cause zoonotic outbreaks (Capua & Alexander, 2008).

1.2.5 Epidemics and Reassortment of AIV

1.2.5.1 Human Pandemics

Reassortments of influenza A viruses are responsible for some of the major genetic shifts in the history of the influenza virus, including those of avian origin which were responsible for significant human mortality. Two of the three influenza pandemics in humans in the 20th century (the H2N2 Asian flu in 1957 and the H3N2 Hong Kong flu in 1968) were the result of reassortment between human and avian strains. In addition, it has long been suggested that the Spanish H1N1 flu in 1918 were directly derived from an avian source (Kalthoff *et al.*, 2010; Reid *et al.*, 2002; Reid *et al.*, 2000). However, more recent analyses suggested that this pandemic strain originated upon the acquisition of N1 neuraminidase and the internal protein genes from avian flu viruses by an circulating human H1 virus (Worobey *et al.*, 2014).

The first flu pandemic of the 21st century occurred in April 2009 at the border between the United States and Mexico, which is caused by the swine-originated H1N1 pandemic strain (S-OIV H1N1). Phylogenetic analyses of the viral genomes indicated that the virus possessed a complex reassortment history with segments that were originally derived from avian-to-mammalian cross-species transmissions (Zimmer & Burke, 2009). In this virus, the NA and M segments were originated from the avian-like Eurasian swine influenza A viruses (SIV), which was first observed in Eurasian swine in 1979, and from a triple reassortant virus. The other segments of S-OIV were derived from a triple reassortant identified in North American swine after 1998, which has PB1 from human H3N2 influenza; PB2 and PA from avian influenza A virus; and HA, NP and NS segments from classical North American swine influenza A viruses which have a common ancestry with the 1918 H1N1 virus (McHardy & Adams, 2009).

1.2.5.2 Human outbreaks

A novel H7N9 influenza A virus has caused 375 human infections in China, resulting in 115 deaths which raised concerns of a new pandemic, The first wave occurred in southeast China from February to May 2013 and the second wave occurred since October 2013, mainly in Guangdong province (WHO, 2014). Infections by this novel H7N9 virus are asymptomatic in bird populations, therefore the virus is likely to spread silently in birds. Human infection is associated with exposure to infected live poultry, or contaminated environments which include the markets where live poultry

are sold (Fang *et al.*, 2013; Lam *et al.*, 2013). Phylogenetic studies indicated that the H7N9 outbreak virus is a reassortant with the HA gene and NA gene derived from AIV carried by migratory wild birds, which subsequently reassorted with enzootic H9N2 viruses in poultry. An outbreak H7N9 lineage has been found which was related to a previously unrecognized H7N7 lineage. This lineage was introduced to chickens and obtained internal genes from co-circulating H9N2 viruses through reassortment events (Lam *et al.*, 2013).

1.2.5.3 HPAI outbreaks

Avian influenza viruses of HA subtypes H5 and H7 have been associated with all recorded outbreaks of HPAI. The outbreaks have mainly occurred in chickens and have been located all over the world. The worldwide HPAI outbreaks recorded since 1959 are summarized in Table 1.1, which shows the time (year), subtype and location (country, state) of each outbreak. From this list we can see that the diversity of subtypes (H5 and H7 with different NA) and the range of geographic regions have increased dramatically in the last ten years. The distribution of outbreaks on the global map provided by World Organisation for Animal Health (OIE) is shown in Figure 1.4. The continuous HPAI H7N3 outbreaks originated in Mexico from June 2012 to May 2014 are analysed in Chapter 5.

Year	Subtype	Country
1959	H5N1	Scotland
1961	H5N3	South Africa
1963	H7N3	England
1966	H5N9	Canada
1976	H7N7	Victoria
1979	H7N7	Germany, England
1983	H5N2	USA
	H5N8	Ireland
1985	H7N7	Australia (Victoria)
1991	H5N1	England
1992	H7N3	Australia (Victoria)
1994	H5N2	Mexico
1995	H5N2	Puebla, Mexico
	H7N3	Australia (Queensland), Pakistan
1997	H5N1	Hong Kong
	H7N4	New South Wales

	H5N2	Italy
1999	H7N1	Italy
2002	H7N3	Chile
2002	H5N1	Hong Kong
2003	H7N7	Netherlands
2004	H5N1	Cambodia, China, Hong Kong, Indonesia, Croatia, Kazakhstan, Malaysia, Mongolia, Philippines, Romania, Russia, Thailand, Turkey, Ukraine, Vietnam
	H5N2	Zimbabwe, Chinese Taipei
	H7N7	Korea (Dem. People's Rep.)
	H7N3	Canada
2005	H5N1	Cambodia, China, Hong Kong, Indonesia, Japan, Laos, Thailand, Vietnam
	H5N2	Zimbabwe
	H7N7	Korea (Dem. People's Rep.)
2006	H5N1	Afghanistan, Albania, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Burkina Faso, Cambodia, Cameroon, China, Croatia, Czech Republic, Denmark, Djibouti, Egypt, France, Georgia, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Iran, Italy, Israel, Italy, Jordan, Kazakhstan, Korea, Laos, Malaysia, Mongolia, Myanmar, Niger, Pakistan, Palestinian Auton, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sudan, Sweden, Switzerland, Thailand, Turkey, Ukraine, United Kingdom, Vietnam
	H5N2	South Africa
2007	H5N1	Afghanistan, Bangladesh, Benin, Cambodia, China, Czech Republic, Djibouti, Egypt, France, Germany, Ghana, Hong Kong, Hungary, India, Japan, Korea, Kuwait, Laos, Malaysia, Myanmar, Pakistan, Poland, Romania, Russia, Saudi Arabia, Slovenia, Sudan, Thailand, Togo, Turkey, Ukraine, United Kingdom, Vietnam
	H7N3	Canada
2008	H5N1	Bangladesh, Benin, Cambodia, China, Egypt, Germany, Hong Kong, India, Iran, Israel, Japan, Korea, Laos, Myanmar, Nigeria, Pakistan, Poland, Romania, Russia, Saudi Arabia, Switzerland, Thailand, Togo, Turkey, Ukraine, United Kingdom, Vietnam.
	H7N3	Canada
	H7N7	United Kingdom
2009	H5N1	Afghanistan, Bangladesh, Benin, Cambodia, China, Germany, Hong Kong, India, Japan, Laos, Mongolia, Nepal, Nigeria, Russia, Thailand, Togo, Vietnam
	H7N7	Spain
2010	H5N1	Bangladesh, Bhutan, Bulgaria, Cambodia, China, India, Israel, Japan , Korea, Laos, Mongolia, Myanmar, Nepal, Romania, Russia, Vietnam
	H7N7	Spain
2011	H5N1	Bangladesh, Cambodia, China, Hong Kong, India, Indonesia, Iran, Israel, Japan, Korea, Mongolia, Nepal, Palestinian Autonomous Territories, Vietnam
	H5N2	South Africa
2012	H5N1	Bangladesh, Bhutan, Cambodia, China, Chinese Taipei, Hong Kong, India, Israel, Myanmar, Nepal, Vietnam
	H5N2	Chinese Taipei, South Africa
	H7N7	Australia
	H7N3	Mexico
2013	H5N1	Bangladesh, Bhutan, Cambodia, China, Hong Kong, India, Korea, Nepal, Vietnam
	H5N2	China, Chinese Taipei, South Africa

	H7N2	Australia
	H7N3	Mexico
	H7N7	Australia, Italy
2014	H5N1	Cambodia, China, India, Korea, Libya, Nepal, Vietnam
	H5N2	China, Chinese Taipei
	H5N6	Laos
	H5N8	Japan, Korea, German. Holland, United Kingdom
	H7N2	Australia
	H7N3	Mexico

Table 1.1 Worldwide outbreaks of highly pathogenic avian influenza in animals
The data sources are all immediate notifications and follow-ups reports notified by WHO (data before 2005) and OIE (data after 2005).

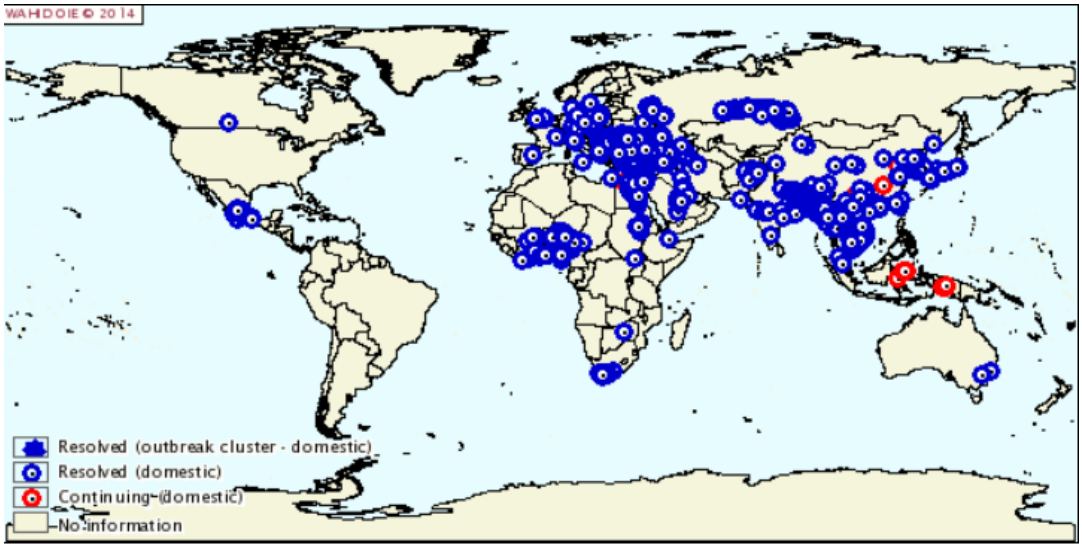


Figure 1.4 The HPAI outbreaks globally in bird populations
The map shows the HPAI outbreaks globally in bird populations from 1st January 2005 to 24th July 2014Map adapted adopted from World Animal Health Information Database (WAHID). http://www.oie.int/animal-health-in-the-world/update-on-avian-influenza/2014/

The pandemic of highly pathogenic H5N1 appears to be unique in the era of modern influenza virology (Watanabe *et al.*, 2012). The first HPAI H5N1 AIV was isolated from geese in south China in 1996 (Xu *et al.*, 1999). The virus rapidly spread and donated the HA gene segment to the Hong Kong/156/97–like virus, which caused an outbreak of "Hong Kong flu" in 1997 (Claas *et al.*, 1998). From 1998 to 2003, a series of HPAI H5N1 outbreaks occurred in southern China. Since 2003, the viruses have spread from south China into Southeast Asia resulting in the death of millions of poultry in 64 countries on three continents either from infection or culling, causing tremendous damage to the poultry industry. There are significant

zoonotic implications of the HPAI H5N1 outbreaks (Taubenberger & Kash, 2010) discussed below.

HPAI H5N1 viruses are continuously isolated from birds, humans, pigs, and other mammals worldwide. As of 7th June 2012, a total of 606 confirmed cases and 357 fatal cases of human infection in 15 countries have been reported. Phylogenetic analyses indicated that human infection with H5N1 viruses is most likely to be associated with direct or indirect contact with infected birds or wildfowl (Guan *et al.*, 2004; Nidom *et al.*, 2010). Genetic analysis of HPAI H5N1 isolated from pigs in China and Indonesia showed that the swine isolates all had an avian origin (Nidom *et al.*, 2010; Shi *et al.*, 2008; Zhu *et al.*, 2008).

Both inter-subtype reassortment and inter-host transmission play important roles in the evolution and variation of HP H5N1 AIV (Lei & Shi, 2011). H5N1 viruses are frequently replaced by other AIV through genetic reassortment, establishing multiple viral genotypes (Chen *et al.*, 2006b; Guan *et al.*, 2002; Li *et al.*, 2004). Phylogenetic analyses have provided evidence for frequent inter-subtype reassortment in the internal gene segments of HP H5N1 AIV (Guan *et al.*, 2004; Lei & Shi, 2011; Vijaykrishna *et al.*, 2008). A study of a swine influenza strain isolated in China revealed that it was a multiple reassortant, with its gene segments coming from avian H5N1, H9N2 and influenza viruses of other unknown subtype (Shi *et al.*, 2008). A recent phylogenetic study on full genome HPAI H5N1 viruses isolated from humans and poultry in Bangladesh indicated that various types of reassortment events have occurred with other sub-lineages in the same clade, those between different clades and with LP AIV subtypes (Gerloff *et al.*, 2014).

Using reverse genetics technology, many studies are being conducted to evaluate the pathogenic potential of reassortant viruses between different subtypes. Reassortment studies on currently co-circulating avian H5N1 and human H3N2 influenza viruses resulted in reassortants with high pathogenicity demonstrating that virulence may increase when a human virus PB2 segment functions in the background of an avian H5N1 virus (Jackson *et al.*, 2009). Reverse genetic analysis

also suggested that reassortment between co-circulating pandemic swine-origin influenza virus (S-OIV) and avian H5N1 influenza strains produced a reassortant with the potential for increased pathogenicity in mammals (Cline *et al.*, 2011). In addition, co-infection of S-OIV and a contemporary H5N1 virus in cultured human cells indicated that these two viruses have high genetic compatibility (Octaviani *et al.*, 2010).

Influenza virus H5N2 is usually present with low pathogenicity in wild birds. However, this virus may mutate into a HPAI strain after introduction into domestic poultry. Highly pathogenic strains of H5N2 have caused a series of outbreaks and have been devastating to the commercial poultry industry since the first report in the USA in 1983 (Kawaoka *et al.*, 1984), and then in 1994 in Mexico (Horimoto *et al.*, 1995a). The HPAI have infected poultry farms in South Africa and Asian countries including Korea, Japan, and China since 2004 (Table 1.1). Furthermore, poultry farm workers in Japan in 2005 may have been exposed to H5N2 which was not previously known to infect humans (Ogata *et al.*, 2008). Several reassortment events of HPAI H5N2 identified from domestic ducks and chickens in live bird markets in Eastern China as well as in Korea (Gu *et al.*, 2012; Kim *et al.*, 2010; Zhao *et al.*, 2012). In addition, a cross-hemispheric HPAI H5N2 sequence isolated from a parrot in Guangdong in southern China in 2004 was closely related with the North American H5N2 viruses (Jiao *et al.*, 2012). HPAI H5N2 sequences were detected from healthy wild waterfowl species sampled in Nigeria also indicating that they can be spread by wild bird migration (Gaidet *et al.*, 2008).

HPAIV H5N8 has been detected among wild birds in Asia and has caused several outbreaks in commercial poultry farms in South Korea, Japan and China (Lee *et al.*, 2014; Zhao *et al.*, 2013). A novel HPAI H5N8 outbreak occurred in turkey farms in Germany on 6 November 2014; Outbreaks caused by the same virus occurred in laying hens in north east of Rotterdam and breeding ducks in England, reported by OIE on 16 Novermber. This is the first time this virus has been detected in Europe. It remains unclear how this virus was introduced simultaneously into closed indoor holdings in European regions far from one another and different

poultry production sectors. Initial phylogenetic studies indicated that the virus is possibly being spread via wild bird migration.

H7 viruses have been associated with both low pathogenicity (e.g., H7N2 and H7N7) and high pathogenicity (e.g., H7N3 and H7N7), and have caused mild to fatal illness in humans. By 2014, nine NA subtypes matched with H7 have been reported. Three H7N3 HPAI events in poultry have occurred in the North Americas since 2000, and, in one case, it was reported that the outbreak H7N3 AIV was transmitted from poultry to humans (Tweed *et al.*, 2004). Phylogenetic analyses indicated that each of these H7N3 HPAI strains had a close relationship with LPAI isolated from wild birds sampled in neighbouring provinces (Berhane *et al.*, 2009; Hirst *et al.*, 2004; Suarez *et al.*, 2004). Since 2002, H7 subtype viruses have caused more than 100 cases of human infection in Europe and North America, resulting in both ocular and respiratory illness (Belser *et al.*, 2009). H7 AIV is one of the most important causes of influenza outbreaks in poultry in Europe and North America. For example, HPAI of H7 subtype spread to and caused outbreak in numerous farms, resulting in great economic losses in 1999 in Italy (caused by H7N1), and in Netherlands in 2003(caused by H7N7) (Table1.1).

1.2.5.4 AIV Endemism

Studies on the transmission and distribution of avian influenza virus are mainly focused on the highly pathogenic outbreaks, such as H5 and H7 (Li *et al.*, 2009; Martin *et al.*, 2011). However, phylogenetic studies have shown that other low pathogenic or asymptomatic subtypes of avian influenza could transmit their gene segments (especially internal segments) and contribute to the creation of new highly pathogenic viruses. Their transmission among birds has the potential to cause zoonotic outbreaks in humans and mammals. Surveillance studies have revealed that there has been significant activity with AIV endemics of H6 and H9 viruses in poultry populations, especially in Asia (Brown, 2010). Mixed infections of HPAI H5N1 with the two subtypes are often reported (Huang *et al.*, 2012; Liu *et al.*, 2014; Marinova-Petkova A., 2014; Perk *et al.*, 2007; Seifi *et al.*, 2010).

As a specific example, H9N2 influenza A viruses have become endemic in different types of terrestrial poultry and wild birds in Asia, and are occasionally transmitted to mammals. Human cases of H9N2 virus infection have also been reported in south China since the late 1990s (Peiris *et al.*, 1999), pointing to the need for further surveillance of H9N2 viruses in poultry, especially chickens. In contrast, there is no evidence of fixed lineages of the H9N2 avian influenza virus in land-based poultry in North America, even though it is found in wild ducks and shore birds (Hossain *et al.*, 2008). Phylogenetic studies of H9N2 viruses circulating in Asian countries indicated that the H9N2 AIV has undergone widespread inter- and intra-subtype reassortments. The generated viruses have unknown biological properties (Ge *et al.*, 2009). Genetic analysis also showed that multiple reassortments occur between co-circulating H9N2-like and H5N1-like viruses (Dong *et al.*, 2011). Avian influenza viruses of H9N2 subtype acted as the donor of the internal segments of two important pandemic viruses: the H5N1 Hong Kong outbreak in 1997 (Guan *et al.*, 1999) and the H7N9 human outbreaks in 2013 (Lam *et al.*, 2013).

Influenza viruses of the H6 subtype have been isolated from both wild and domestic avian species throughout the world since their first detection in a turkey in Massachusetts in 1965. A comprehensive phylogenetic analysis of the extent and timing of cross-hemispheric movements showed that the Eurasian H6 subtype has invaded North America several times and replace the North American lineage (Dohna *et al.*, 2009). Although no human infections have been reported with AIV of H6 subtype thus far, serological studies have detected low levels of H6 specific antibodies in humans exposed to birds (Kayali *et al.*, 2008). In addition, a recent study assessed H6 viruses isolated from live poultry markets in southern China, and identified several viruses among them which have the capability to replicate and transmit to mammals (Wang *et al.*, 2014).

1.2.6 Phylodynamics studies of AIV

High-throughput sequencing of complete AIV genomes sampled from both viral epidemics and infected hosts have provided an insight into virus evolution and inference of viral transmission dynamics from genetic data (Pybus & Rambaut,

2009). Recent advances of phylogenetic studies have provided novel methods which directly link patterns of genetic diversity to ecological processes, collectively described as "phylodynamics", with the aim of characterising the joint evolutionary and epidemiology behaviour of infectious diseases (Drummond & Rambaut, 2007; Lemey *et al.*, 2009; Lemey *et al.*, 2010; Pybus *et al.*, 2013; Pybus *et al.*, 2012). Phylodynamic processes can be recovered from genomic data, and allow quantitative studies to be conducted on how ecological features shape pathogen genetic diversity. Phylogenies reconstructed from spatial epidemics record the correlated histories of transmission among sampled infections as transitions along branches (Faria *et al.*, 2011; Pybus *et al.*, 2013). The phylodynamic methodologies applied in Chapter 3-5 will be further explained in Chapter 2.

1.3 Objective of this study

The three research chapters in this thesis aim to explore the evolution and dynamics of AIV by using a large dataset of complete AIV genomes compiled to date. The specific aims are (i) to describe the complete structure of genetic diversity of AIV, (ii) to quantify the frequency and pattern of segment reassortment, and (iii) to uncover the associations between phylogeny and different traits including subtypes, host range and temporal–spatial distribution.

The evolution of internal segments of AIV were systematically analysed with a comprehensive avian dataset in Chapter 3. With the aim of exploring the rate of genetic exchange among viral subtypes during evolution, a phylodynamic analysis was carried out upon the influenza A viruses circulating in avian populations (in both wild birds and domestic birds) from 1956 to 2011. In particular, I focused on investigating how avian influenza segments specific to the HA, NA and combined HA-NA subtypes evolve on the backbones of the six internal gene segments. I quantified the inter-subtype reassortment rates and tested possible ecological and evolutionary factors that are associated with such patterns of reassortment: the relative proportion of each host species, the inferred evolutionary relaxed clock rates and the selective constraint acting on the genes.

In Chapter 4, the spatial diffusion of AIV was explored. The determinants of spatial diffusion were tested using the six internal segments with an advanced discrete mapping methodology. This analysis was performed on AIV isolated in China, which plays a central role in maintaining a source population for global influenza diversity. This phenomenon has attracted worldwide concerns after the human infection with Avian Influenza A (H7N9) virus since April 2013. In this study, the phylogeographic diffusion network of AIV in China was reconstructed, and the genetic and ecology predictors of viral diffusion were integrated with phylogeographic inference. I explored the main sources and destinations of AIV spread in bird population in China while simultaneously testing and quantifying the risk factors on the viral transmission process using a homogenous generalized linear model.

Finally in Chapter 5, the origin of the novel H7N3 outbreak influenza virus in Mexico was explored using phylogenetic methodologies. The specific aims were to i) identify the origin of outbreaks of endemic and epidemic diseases, ii) characterise pathogen emergence at its earliest stages, and iii) determine viral and epidemiological characteristics that impact the evolution, emergence and persistence of pathogens in human and animal populations.

Chapter 2

Methods

Chapter Abstract

This thesis focuses on the evolutionary analysis of viral sequence data, in particular, to answer evolutionary questions quantitatively by analysis of avian influenza virussequence data using Bayesian approaches. Many of the methods employed are commonly used in all main chapters and are described in this chapter for reference: including the overview of evolutionary models, phylogenetic methods, the selection and subsampling procedure of the large sequence dataset and a description of the main software.

2.1 Models of molecular evolution

Genetic distance is a measure of how far apart two sequences are from each other. Genetic distance can be measured by counting changes between sequences and dividing by the number of sites. Simple measures of genetic distance only take the point nucleotides mutations into account. There are two different types of point mutation changes in 1) Transitions, which is a point mutation that changes a purine nucleotide to another purine A ↔ G or a pyrimidine nucleotide to another pyrimidine C ↔ T; 2) Transversions, which is the substitution of a purine for a pyrimidine or a pyrimidine for a purine (Collins & Jukes, 1994) (Figure 2.1) . When reconstructing evolutionary relationships among divergent sequences, some sites will undergo multiple substitutions at the same site (multiple hits) as genetic distance increases. Therefore the possibilities of reverting to the state they were originally in (reversion) need to be taken into account.

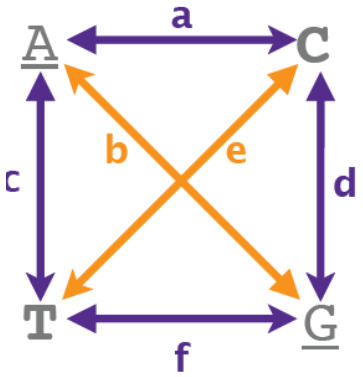


Figure 2.1 Diagram of nucleotide mutation changes among the four DNA nucleotides
Adenine (A), Thymine (T), Cytosine (C) and Guanine (G); a, b, c, d, e and f are relative mutation rate parameters. The edges of a, c, d, f represent transversions; the edges of b and e represent transitions.

Nucleotide substitution models differ in their treatment of different types of substitution and specification of base frequencies. The commonly used substitution models are listed from simple to complex as follows:

1) The Jukes-Cantor model which allows bases to change at equilibrium frequencies and all substitutions are equally likely (Jukes & Cantor, 1969);

2) The Kimura 2 parameter model which also allows bases to change at equal frequencies but transitions and transversions occur at different rates (Kimura, 1980);
3) Hasegawa-Kishino-Yano (HKY) model which allows bases to change at unequal frequencies and transitions and transversions occur at different rates (Drummond & Rambaut, 2007; Hasegawa *et al.*, 1985);
4) Tamura-nei 1993 model which allows unequal base frequencies, transversions occur at the same rate (parameter1) and transitions occur at different rates (parameter 2=A <->G; parameter 3=C <->T) (Tamura & Nei, 1993);
5) General Reversible Model (GTR) model which allows unequal base frequencies and all substitution types occur at different (reversible) rates (Figure 2.1) defined by a Markov process, which specifies the relative rates at which substitutions occur along a branch in the tree (Drummond & Rambaut, 2007; Tavaré, 1986a).

RNA viruses are often modelled with GTR or HKY nucleotide substitution models, depending on the amount of information within the sequence alignment from which to inform the model (Drummond & Rambaut, 2007). The most appropriate substitution model to use can be determined using the ModelTest program (e.g in HyPhy, or jModelTest). In addition to the nucleotide substitution models, the overall substitution rate can be allowed to vary from site to site, and typically site to site rate variation models assume that the overall rates have a Gamma distribution (Tamura & Nei, 1993), partitioned in to four rate categories.

The Goldman and Yang model is the basic model which often used to model codon evolution with protein-coding nucleotide sequence data. This model with different GTR substitution rates for each of the three positions of the codon which makes it more general (Goldman & Yang, 1994). In contrast, a common extension to nucleotide substitution models are models in which positions 1 and 2 are treated differently to position 3, or even all 3 positions are treated differently. For example, the SRD06 (Shapiro *et al.*, 2006). The SRD06 mode employs the HKY model of

nucleotide substitution which allows the base frequencies to be estimated from the data and a gamma-distributed rate across sites. This model gives the partitioning substitution rates: Position 1 and 2 approximate non-synonymous changes, and position 3 approximates synonymous changes, which reflect the increased tendency for nucleotide changes at the first and second positions relative to third position substitutions. The approach improves the fit of nucleotide models to protein-coding data and is more computationally efficient than using full-codon models. In addition, unlike GTR, HKY can be solved without matrix inversion (Felsenstein, 1981) and therefore models with HKY implementation are faster. Calculation of the phylogenetic likelihood can be extremely slow if the number of sequences and complexity of the evolutionary models are high. The usage of the BEAGLE library may drastically speed up computation time through parallelization of phylogenetic likelihood calculations, and can perform the core calculations at the heart of most Bayesian and Maximum Likelihood phylogenetics package (Ayres *et al.*, 2012).

2.2 Phylogenetic methods

2.2.1 Neighbour joining and maximum likelihood phylogenetics

Neighbour-joining (NJ) phylogenetic methods apply general data clustering techniques to sequence analysis, and use a distance matrix to specify the genetic distances (e.g., Tamura-Nei model；Jukes-Cantor model；Kimura two-parameter, see section 2.1) between each pair of taxa in a polynomial-time algorithm. NJ is fast as the algorithm adopts the hierarchical clustering strategy (Saitou & Nei, 1987). NJ phylogenetic methods have been shown to provide a good approximation to the evolution tree with the smallest least-squares measure of observed distances between sequences and the distances predicted by the tree (Hollich *et al.*, 2005). However, it can be difficult to arrive at reliable values for the distance matrix if the input sequences are distantly related or the input sequence data is of poor quality (St John *et al.*, 2003). Therefore, the NJ method is practical for analysing large datasets (including of RNA viruses of the same subtype or lineage) and serves as the starting point to select data for further computationally intensive search of the best phylogeny.

A common obstacle in NJ phylogenetic reconstruction methods is that they produce point estimates of the phylogenies by assuming a single value as the best estimation of an unknown parameter. They yield a single tree per run, and rely on bootstrapping to offer a measure of how strongly the data supports each of the relationships depicted in the tree. When an NJ phylogeny is reconstructed with bootstrapping, the original data matrix is randomly re-sampled with replacement to produce pseudo-replicate datasets. The tree-building algorithm is performed on each of these replicate datasets (Felsenstein, 1985). The resultant bootstrap proportions are conservative measures of support, with a 70% threshold indicating acceptable support for a group (Efron *et al.*, 1996).

The maximum likelihood (ML) methods infers probability distributions to assign probabilities to particular possible phylogenetic trees (Felsenstein, 1981) and require a substitution model (e.g., GTR and HKY see section 2.1) to assess the probability of particular mutations. The ML methods first calculate the probability of the topology, branch lengths and substitution model parameters, then select the phylogeny with the highest overall likelihood score (Fukami-Kobayashi & Tateno, 1991). If a tree requires fewer mutations in internal nodes to explain the observed phylogeny, a higher probability (likelihood) of the tree will be assessed. The ML algorithm need to search through a multidimensional space of parameters which could be a computational burden, especially when ML methods are used with complicated parameter-rich models (Holder & Lewis, 2003; Rogers & Swofford, 1998).

2.2.2 Bayesian phylogenetics

Bayesian phylogenetic methods provide a statistical framework in which to estimate the probabilities of various evolutionary and epidemiological models from sequence data. These methods also provide a measure of uncertainty regarding the

estimated genealogy and evolutionary parameters, which offer opportunities to explore a wide diversity of models, each of which corresponds to specific assumptions concerning the shape of the tree and the rate of substitution over time.

In Bayesian phylogenetic models, the probability of a specific tree is estimated as a function of the prior probability describing expectations of the model, and likelihood of the virus sequence alignment over the likelihood of observing the data. To be specific, for parameters θ, conditioned on a model M, with observed data D:

$$p(\theta|D, M) = \frac{p(D|\theta,M)p(\theta|M)}{p(D|M)} \quad (2.1)$$

p(θ | D, M) is the posterior probability distribution for the parameters given the data and model; p(D | θ, M) is the likelihood of the data given the parameters and p(θ|M) is the prior distribution representing existing beliefs about the parameters.

Multidimensional integrals are required to calculate the posterior distribution, the prior distribution and the likelihood probabilities of biological probability models. Such integrals also need to combine both continuous and discrete calculations. Markov chain Monte Carlo (MCMC) algorithms can be employed in the Bayesian framework to sample the distribution of likely model parameters that explain the data (Hastings, 1970; Metropolis *et al.*, 1953). MCMC takes a series of steps that form a conceptual chain. At each step, a new location in parameter space is proposed as the next link in the chain. The relative posterior-probability density at the new state is then calculated. If the new state has a higher probability than that of the previous state of the chain, the move is accepted. Then the proposed state becomes the next link in the chain and the cycle is repeated. If the probability is lower, it is accepted with a probability which calculated as a function of the difference between the original log likelihood and the new proposed log likelihood. And if the proposed state is rejected, the present state is retained as the next link in the chain. By repeating this procedure (millions of times), a long chain of positions in parameter space is created and the chain tends to converge onto sample regions of high posterior probability. In this way, this function allows the lower log likelihood to be accepted with a certain probability such that a representative sample is obtained.

Improved computational efficiency enables the analysis of large datasets on a practical timescale. A diagnostic plot of the chain trace is widely used to assess convergence, by showing the likelihood of each model against the sample generation. A burn-in period is required to obtain convergence to the stationary distribution in MCMC, after which the sampling from the MCMC is undertaken at periodic intervals as specified by the user (Meyn & Tweedie, 1993). The sampling interval should be sufficiently wide as to ensure that the samples are not auto-correlated. The effective sample size (ESS) of a post-burn-in sample, is given by the post-burn-in chain length divided by the average number of states in the chain by which two samples must be separated for them to be uncorrelated (Kass *et al.*, 1998). The ESS provides an estimate of the number of effectively independent samples from the posterior distribution to which the MCMC is equivalent. Comparing MCMC output from multiple independent runs can indicate whether chains are converging on the same distribution. In addition, the uncertainty in parameter estimates using MCMC can be reported using Bayesian credible intervals. A Bayesian credible interval for a posterior distribution for a certain parameter can be any interval in the domain of the distribution (Edwards *et al.*, 1963). The highest posterior density (HPD) interval, typically the 95% interval, is the shortest interval in parameter space which covers 95% area of the posterior probability density(Okada, 1974). However, HPD intervals is obsolete and problematic in the Bayesian framework as the prior needs to be taken into account. Therefore, the Bayes Factor (BF) estimation is preferred.

Phylogenetic analysis of sequence data from measurably evolving populations which have been sampled at sufficiently diverse time-points can provide an insight into the evolutionary dynamics and demographic history of a population (Drummond *et al.*, 2002). BEAST (Bayesian Evolutionary Analysis by Sampling Trees) is a software package that performs Bayesian MCMC phylogenetic analysis of sequence data and provides a coalescent-based framework for concomitant inference of phylogenetic and population genetic parameters (Drummond & Rambaut, 2007). The estimation of the dates associated with every internal node in a phylogenetic tree

is governed by the parameters in the Bayesian context. In a time-scaled phylogeny the height of the root node of the tree, relative to the age of the youngest sampled sequence, provides an estimate of the time to the most recent common ancestor (TMRCA) of all the sequences in the dataset.

The BEAST software contains a range of complementary evolutionary models for nucleotide and protein substitution, molecular clock variation, demographic history, spatial distribution and phenotypic trait evolution that can be combined into a full probabilistic model, and has therefore been used extensively in analyses of viral evolution (Pybus *et al.*, 2009; Rabaa *et al.*, 2010; Rambaut *et al.*, 2008). Phylogenetic analysis of avian influenza sequences in BEAST forms the core component of Chapters 3-5.

2.2.3 Molecular clock models

The sampling time of sequences can help to calibrate a phylogenetic tree by scaling internal branches so that the substitution rate can be estimated under a molecular clock assumption. BEAST incorporates the sampling date information of sequences when constructing trees, and models the rate of molecular evolution on each branch in the tree. The rate can be a uniform rate known in advance or can be estimated from calibration information over the entire tree. BEAST also incorporates relaxed molecular clock models that do not assume a constant molecular evolution rate across lineages (Drummond & Rambaut, 2007). The strict molecular clock model is the basic model for estimating substitution rates among branches, through specifying a constant rate (or the date) of a node or set of nodes, such that dates of divergence for particular clades can be estimated. However, strict molecular clock models are not necessarily the most appropriate for fast evolving viral sequence datasets over relatively long timescales or including different host species, since rate of substitution may differ among lineages (Ayala, 1997; Hasegawa & Kishino, 1989). If there is true rate variation between branches, but a strict clock is used, then both divergence date estimation and phylogenetic inference can become inaccurate (Drummond *et al.*, 2006).

In comparison, uncorrelated relaxed clock models relax the molecular clock assumption by allowing the rate of evolution to vary among the branches of the tree (Drummond *et al.*, 2006). In uncorrelated relaxed clock models, the rates are drawn independently from an underlying distribution (including exponential and lognormal distributions) and they have been shown to provide a better fit to viral sequence data than either a strict molecular clock or the absence of a molecular clock (Drummond *et al.*, 2006); The uncorrelated exponential relaxed clock model assumes branch lengths are distributed according to an exponential distribution of mean (Drummond *et al.*, 2006; Lepage *et al.*, 2007); In the uncorrelated lognormal relaxed clock model, the substitution rate varies according to a lognormal distribution with mean and variance(Kishino *et al.*, 2001). The lognormal distribution generally provides better precision than the exponential distribution, while maintaining the accuracy and fitting for data that are relatively clock-like (Drummond *et al.*, 2006).

2.3.4 Tree priors (the coalescent models)

Coalescent theory traces the lineages of a gene shared by all members of a population to the most recent common ancestor (MRCA). These coalescent lineages are represented in the form of a phylogenetic tree (Kingman, 1982). This theory is widely used to estimate population genetic parameters in natural populations. In BEAST, several coalescent models are implemented in inferring population dynamics (Drummond & Rambaut, 2007). Models with simple parametric functions include the assumed constant size model, exponential growth model and logistical growth model (Drummond *et al.*, 2002; Kuhner *et al.*, 1998). In a constant population size model, the number of individuals in a population is assumed to remain constant over time, and therefore suitable to describe the evolutionary history of a virus when it reaches endemic equilibrium; Exponential and logistical growth models are used for virus epidemics and pandemics which represent a rapid growth of the virus population upon emergence into the host population and with subsequent changes or increases in population size.

Other models in BEAST are highly parametric. The Bayesian Skyline plot (Drummond *et al.*, 2005) is based on a piecewise constant function and can only be

used when the data are strongly informative about population history; The Bayesian Skyride model (Minin *et al.*, 2008) exploits a Gaussian Markov random field (GMRF) which allows the estimated trajectory to change at coalescent times, so that it achieves temporal smoothing of the effective population size (Minin *et al.*, 2008); In addition, a recently developed method in BEAST is the Bayesian Skygrid model, which is able to estimate the effective population size trajectories using multi-locus sequence data (Gill *et al.*, 2013) (see Chapter 6, sectionsection 6.2).

2.3 Model selection

In a Bayesian analysis, the best model can be selected from a set of models by comparing marginal likelihoods. The marginal likelihood for a model M is the probability of the observed data D, given the model M, averaged over the model parameters θ:

$$p(D|M) = \int p(D|\theta, M)\, p(\theta \mid M) d\theta \qquad (2.2)$$

p (D|M) is the denominator in the Bayes theorem expression for parameters θ conditioned on a model M, with data D. The theoretical framework for model comparison in a Bayesian framework is the Bayes factor (BF) test. BF is the ratio of the posterior odds to the prior odds for two hypotheses of interest, which attempts to measure how strongly the data support or refute a hypothesis. In other words, the test gives an estimate of how much the data support one hypothesis over another. In order to calculate BF, the marginal likelihoods of the models under comparison are evaluated using Bayesian MCMC over both models (Green, 1995).

There are several ways to estimate the marginal likelihood of each model by processing the output of two BEAST analyses. The Bayes factor (BF) between them can be calculated. In Tracer (a graphical tool for visualization and diagnostics of MCMC output), the simplest method incorporated as default is the posterior harmonic mean estimator (HME), which employs importance sampling of the posterior distribution. Importance sampling is a statistical method for estimating properties of a particular distribution, while only having samples generated from a

different distribution rather than the distribution of interest. Importance sampling increases the density of points in interest regions in Monte Carlo integration, hence improve the overall efficiency (Newton & Raftery, 1994). In HME, the harmonic mean of the sampled likelihoods provides a reasonable estimate of the marginal likelihood. Since the HME does not require any additional calculations, it is relatively quick to perform and has gained widespread popularity for model selection in viral evolution analyses (Pybus *et al.*, 2009; Smith *et al.*, 2009a). In general, a logBF > 20 is strong support for the favoured model. The estimation can be processed in Tracer software version 1.4 and higher.

Akaike's information criterion (AIC) is a measure of the relative quality of a model with a given set of data. It is commonly used to compare models in a maximum likelihood context:

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln(\mathrm{L}) \qquad (2.3)$$

Where k is the number of parameters in the model, and L is the maximized value of the likelihood function for the model. A posterior simulation-based analogue of AIC (AICM) is approach to compare Bayesian models by measuring AIC in a Bayesian Monte Carlo context (Raftery *et al.*, 2007). AICM is estimated with the posterior sample mean and variance of the log likelihood. This measure can also be computed directly from MCMC runs produced by BEAST without resorting to additional analysis (Kitakado *et al.*, 2006). AICM performs better than HME and is a useful initial evaluation of model (Baele *et al.*, 2012). The preferred model would be the one with the lowest AICM score (available in Tracer V1.6).

Path sampling (PS) and stepping stone (SS) estimators use a Markov chain to sample a series of distributions along a path between the prior and posterior and have been used to select between molecular clock models. They have recently been shown to estimate the marginal likelihood with much greater accuracy than the HME (Baele *et al.*, 2012; Baele *et al.*, 2013). The method is most useful when HME estimates for alternate models are quite close to one another, i.e. the HME has problems discriminating between the different models in terms of performance, then the differences between the models are larger when using PS/SS (Baele *et al.*, 2012). PS

(or SS) methods of estimating marginal likelihoods are also relevant when fundamentally different models are being compared; e.g. two clock models or two population growth models, which contain different algorithms for the prior and likelihood calculations. PS and SS are performed after the main MCMC chain in BEAST and are specified in the XML via Beauti (Drummond *et al.*, 2012).

2.4 Ancestral state reconstruction under a Bayesian framework

Virus evolution occurs simultaneously with geographic dispersal. This interaction characterizes a spatial phylodynamic process that can be recovered from genomic data using phylogeographic analyses (Faria *et al.*, 2011). Therefore, the emergence of virus could be inferred to a certain extent by identifying key reservoir species and geographic areas from which new infections are likely to emerge and spread (Holmes & Grenfell, 2009). The impact of host movement or human impact on viral disease spread could also be similarly evaluated. Bayesian models of phylogenetic diffusion in discrete traits are integrated in the BEAST software package. The methods connect phylogenetic inference to a statistical description of trait evolution, which can be used to explore the molecular evolution and spatial dynamics of viruses at different evolutionary scales.

In a Bayesian statistical framework, ancestral reconstruction of discrete states (Rabaa *et al.*, 2010; Streicker *et al.*, 2012) is geared towards rooted, time-measured phylogenies. Discrete trait models compute the likelihood of the observed states at the tips of the tree and finite-time transition probabilities along each branch with continuous-time Markov chain (CTMC) substitution models, which is characterized by rate matrices that contain symmetrical/asymmetric instantaneous rates of discrete states exchange (Minin *et al.*, 2008). In CTMC, diffusion between n states is characterized by an $n \times n$ rate matrix, Λ, that contains $n \times (n - 1)$ off-diagonal, nonnegative rate parameters λ_{ij} for $i,j = 1,\ldots,n$. Symmetric diffusion assumes that $\lambda_{ij} = \lambda_{ji}$, in which Λ is a symmetric matrix. In asymmetric models, Λ is decomposed into $V \times D \times V^{-1}$, where matrix V is a set of real eigenvectors and matrix D is of block diagonal form with 1×1 sub-matrices of real eigenvalues and 2×2 sub-matrices of complex conjugate pairs (Figure 2.2 left). Asymmetric models allow for diffusion

rates to differ between locations depending on the directions, thereby providing a more realistic approximation of the geographic diffusion process through time (Edwards *et al.*, 2011).

In CTMC, the joint probabilities of the total number of labelled transitions (Nt) in time interval (0, t] are computed on a phylogenetic tree to characterize the evolutionary counting process: all possible transitions from state i to state j are counted according to the end state j within this time interval. Figure 2.2 (right) shows that in a rooted phylogeny, conditioning on the observed locations at the tips , CTMCs model the instantaneous locations along each branch of a tree to infer the ancestral states at the internal nodes (Minin & Suchard, 2008a).

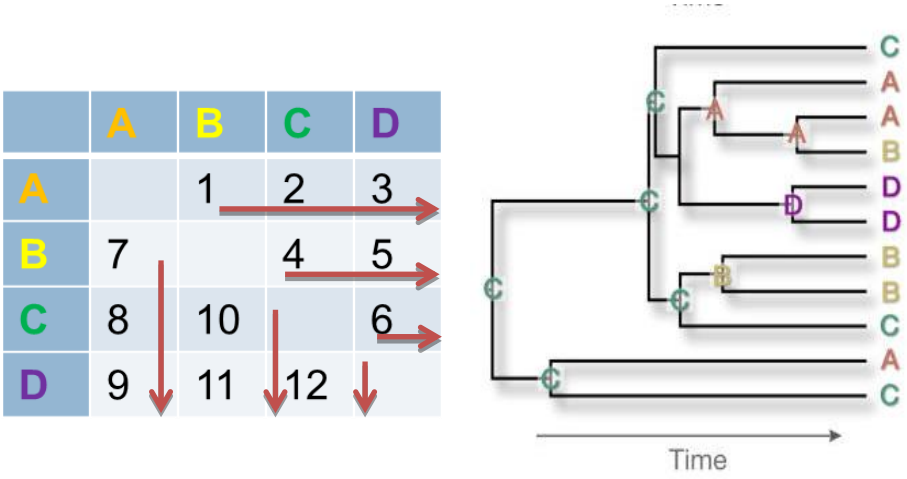


Figure 2.2 Diagram of discrete trait matrix and phylogeny
In a discrete model with a four-state CTMC path, CTMC consider diffusion between the 4 states (A, B, C and D) as a 4 × 4 rate matrix (left). Symmetric model take the right half matrix in to consideration; while asymmetric model include matrix of the two directions. In a rooted phylogeny (right), the observed locations (A, B, C and D) are labelled at the tips of tree, the inferred ancestral states are labelled at the internal nodes.

Standard Markov model inference is extended with a stochastic search variable selection procedure that identifies the parsimonious descriptions of the diffusion process in BEAST. Because all rates are generally not required to adequately explain the diffusion process, the estimation procedure would gain efficiency through focusing on a limited set of well-supported migration pathways (Faria *et al.*, 2011; Lemey *et al.*, 2009). The BSSVS procedure allows rates to shrink to zero with some probability based on prior specification. It extends the Bayesian implementation to a mixture model in which exchange rates in the Markov model.

The significant dispersal pathways than are identified by comparing the posterior to the prior odds that the transition rates are nonzero by a BF test (Lemey *et al.*, 2009; Suchard *et al.*, 2001).

In BSSVS, the relative rate is equivalent to the mean instantaneous substitution rate in nucleotide substitution models. The rates are the relative rate parameters describing how often diffusion between location A and B occurs during evolution with respect to the other locations transitions. The rate indicators are used to augment the standard rate matrix in order to perform the BSSVS procedure. Since BSSVS is aiming at finding a minimal set of rates to explain the diffusion process in the evolution history, a prior is set that prefers a small number of non-zero rates. Comparison of the prior probability (provided by the truncated Poisson prior) and the posterior expectation for each rate indicator informs us to what extent the data requires this rate to provide an adequate explanation of the diffusion process (Lemey *et al.*, 2009). The BF or a particular rate contributing to the diffusion process is essentially the posterior odds that rate is non-zero divided by the equivalent prior odds. The default Poisson priors on non-zero rates generated differ between the asymmetric and symmetric model. For the symmetric model prior is a Poisson with mean = 0.693 (log2) and the offset = n-1 (where n is the number of states). For the asymmetric model it is a simpler Poisson with mean of n-1. Asymmetrical models may provide a more realistic description of spatial dynamics in viral epidemics, as long as sufficient information is available to inform the more complex parameterization. Diffusion rates yielding a BF ≥3 are well supported.

The Markov jumps method is another implementation in BEAST which can be used to quantify the number of CTMC transitions between discrete traits and waiting times in given traits in the phylogeny. The counting process for the expected number of the discrete trait transitions (jumps) is conducted in this implementation, by which the CTMC transitions are enumerated along a branch of a phylogeny (Minin & Suchard, 2008a). The timings of the Markov jumps can be tracked so that the length of time spent in each state along a branch (the 'Markov rewards') can be recorded across the whole tree (Minin & Suchard, 2008b). These jumps and rewards

are possible on a branch-by-branch basis and can track specifically defined sets of transitions (O'Brien *et al.*, 2009).

2.5 AIV dataset

The influenza data analysed in this thesis (chapters 3 to 5) were obtained from the NCBI Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html), a database of influenza virus genome sequences collated by the National Centre for Biotechnology Information (NCBI) GenBank and from the National Institute of Allergy and Infectious Diseases (NIAID) Influenza Genome Sequencing project. In Chapter 3, I downloaded 3226 complete avian influenza genomes of all subtypes and sampling years on 20th November 2011 with all laboratory recombinants and highly-cultured sequences excluded. This dataset was updated with newly-released sequences on 1st March 2013 and 18th February 2014 for the analyses in Chapter 4 and 5, respectively.

Each sequence was assigned a unique ID, the name of each sequence was modified to ensure each name has a consistent naming pattern. In particular, I included the sampling location and the name of the host species. Sampling location is represented by region and then country. Host species of the viruses are recorded by their taxonomic order, for example Anseriformes (Ans), which represents water poultry like ducks and geese, or Galliformes (Gal), which represents land poultry such as chickens. Sequences from non-specific host species, such as the generic "avian" and "bird", were removed from the datasets. The full list can be found in Table 2.1. The scripts of R programs for host classification are provided by Dr. Samantha J. Lycett. The collection date of each sequence was converted into fractional years for the estimation of divergence time. For sequences which only have the year of sampling, the estimation of time was taken to be half way through the year; if only sampling month and year was available the date was taken to be the 15th of the month. A consistent nomenclature was used for each segment of all influenza genomes in the dataset.

Abbreviations	Order names	Species names
Acc	Accipitriformes	eagle, hawk, buzzard, hawk-eagle
Ans	Anseriformes	duck, shelduck, mallard, teal, goose, swan, wigeon, munia, goldeneye, pochard, goosander, pintail, gadwall, garganey, shoveler, widgeon, bufflehead, eider, scaup, scoter, redhead
Cha	Charadriiformes	shorebird, gull, tern, turnstone, dunlin, sandpiper, sanderling, redknot, knot, plover, stint, rednecked stint, murre
Cic	Ciconiiformes	stork, grebe
Gal	Galliformes	chicken, pheasant, quail, turkey, guineafowl, chukar, peafowl, partridge
Gru	Gruiformes	coot, moorhen, bustard
Fal	Falconiformes	falcon, kestrel
Pas	Passeriformes	sparrow, crow, magpie, pigeon, dove, blackbird, myna, starling, shrike, robin, softbill, Japanese white eye, common iora, bluebird, fairy bluebird, rook
Pel	Pelecaniforms	pelican, heron, egret
Pro	Procellariiformes	shearwater
Psi	Psittaciformes	parrot, conure, parakeet, psittacine, macaw
Rhe	Rheiformes	rhea
Str	Struthioniformes	ostrich, emu
Tin	Tinamiformes	tinamou
Sul	Suliformes	cormorant

Table 2.1 The abbreviations of host species and host order in this thesis

Discrete alignments of the eight gene segments were manually concatenated in the order of their length to generate a single alignment of complete genome sequences. Duplicate sequences having 100% nucleotide identity across the whole genome were removed. The multiple sequence alignments were created using MUSCLE (Edgar, 2004) and then adjusted manually using BioEdit version 7.0.9 (http://www.mbio.ncsu.edu/bioedit/) (Hall, 2013). To minimise sampling bias, random stratified sampling was conducted by keeping at least one sequence per influenza subtype per host species per location per year.

The dataset in this study was obtained from the NCBI database, where the number of sequences is unevenly distributed. To investigate a possible sample size effect, I conducted phylogenetic analysis using three subsampled datasets of the same number but comprised of different representative sequences. I further select representative sequences of different trait state in two other different ways: 1) by reducing the number of the most over-represented state to equal the number of the rare subtype in the original dataset; and 2) by reducing the number of sequences of all the state to be the same as the rare state. In this comparison, the reassortment rates among subtypes were similar in all the three datasets. The results of parameter estimates using different subsampled datasets were compared and little effect was found.

2.6 Phylogenetic analysis

Phylogenetic trees can be constructed using several methods such as distance-matrix, maximum likelihood and Bayesian inference (see section 2.2). In this sectionsection, a set of consistent steps were used to generate phylogenetic trees in chapters 3 to 5 are summarized.

Firstly, the preliminary phylogenetic tree of all AIV sequences obtained from the database was generated by NJ methods using MEGA 5 (Tamura *et al.*, 2011). The commonly chosen substation model is Tajima-Nei model and with 1000 bootstraps. Then candidate sequences with incorrect information were then detected by using Path-O-Gen version 1.2 (http://tree.bio.ed.ac.uk/software/pathogen/).

Path-O-Gen was used to investigate the temporal signal and 'clock-likeness' of molecular phylogenies. It tests the adherence of the data to a strict molecular clock by performing a linear regression analysis on the data. The genetic distance between each tip and the root is plotted against the sampling date (shown in Figure 2.3). If the virus is evolving at a constant rate, the plots will show a linear relationship between the genetic divergence and sampling time. Significant deviation of each sample from the trend line (residuals) indicates severe non-adherence to the molecular clock, with samples lying more than $3x10^{-3}$ units away likely to be misdated. These sequences were removed from the datasets.

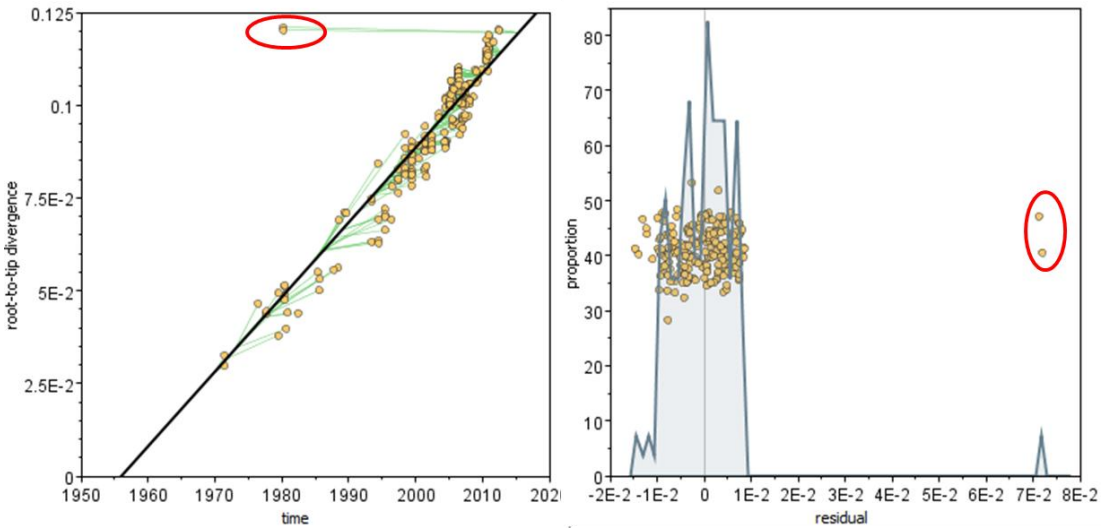


Figure 2.3. The regression line and residual plot in Path-O-Gen.
The regression line (left) and residual plot (right) of a neighbour-joining tree of the HA segment of a subsampled AIV dataset (223 sequences). The red circle indicates the strains with potentially wrongly-labelled sampling date.

Then the ML phylogenetic trees of AIV sequences were reconstructed with the refined dataset using Randomized Accelerated Maximum Likelihood (RAxML), which a program for sequential and parallel maximum likelihood-based inference of giant phylogenetic trees (Stamatakis, 2005). There are three different programs of RaxML: RAxML-HPC is the standard version of the program, meaning it does not take advantage of running more than one 'thread' at a time; RAxML-PTHREADS utilizes P-threads-parallelization to perform multiple calculations at the same time; RAxML-MPI utilizes MPI-parallelization to perform bootstraps and multiple inferences on the original alignment at the same time, which is best for doing bootstraps. In all my studies, the RAxML-MPI program was adopted with GTR substitution model and a bootstrapping of 500 replicates was applied to determine the support for the topology of the ML tree. The total running time for 2996 PB2 gene segment (2280nt) of AIV sequences is 63.15 hours on Haldane. These bootstrap values were compared against Bayesian posterior probabilities generated from the Bayesian-inferred tree to ensure the robustness of the tree topology using these two different tree-construction approaches.

Furthermore, time-scaled trees were generated for each internal protein-coding segments with a subsample of the total AIV dataset with known isolation dates using BEAST version 1.7.3 and version 1.8.0 (Drummond *et al.*, 2012). Results of model selection and settings are varied among different AIV datasets, which will be described individually in the following three chapters.

Chapter 3

Reassortment patterns of avian influenza virus internal segments among different subtypes

Published as Lu Lu, Samantha J. Lycett, and Andrew J. Leigh Brown. Reassortment patterns of avian influenza viruses among different subtypes in internal segments. *BMC Evolutionary Biology*. 2014, **14**:16 doi:10.1186/1471-2148-14-16

Chapter Abstract

The segmented RNA genome of avian Influenza viruses (AIV) allows genetic reassortment between co-infecting viruses, providing an evolutionary pathway to generate genetic innovation. The genetic diversity (16 haemagglutinin and 9 neuraminidase subtypes) of AIV indicates that an extensive reservoir of influenza viruses exists in bird populations, but how frequently subtypes reassort with each other is still unknown. Here I quantify the reassortment patterns among subtypes in the Eurasian avian viral pool by reconstructing the ancestral states of the subtypes as discrete states on time-scaled phylogenies with respect to the internal protein coding segments. I further analyzed how host species, the inferred evolutionary rates and the dN/dS ratio varied among segments and between discrete subtypes, and whether these factors may be associated with inter-subtype reassortment rate. The general patterns of reassortment are similar among five internal segments with the exception of segment 8, encoding the non-structural genes, which has a more divergent phylogeny. However, significant variation in rates between subtypes was observed. In particular, hemagglutinin-encoding segments of subtypes H5 to H9 reassort at a lower rate compared to those of H1 to H4, and Neuraminidase-encoding segments of subtypes N1 and N2 reassort less frequently than N3 to N9. Both host species and dN/dS ratio were significantly associated with reassortment rate, while evolutionary rate was not associated. The dN/dS ratio was negatively correlated with reassortment rate, as was the number of negatively selected sites for all segments. These results indicate that overall selective constraint and host species are both associated with reassortment rate. These results together identify the wild bird population as the major source of new reassortants, rather than domestic poultry. The lower reassortment rates observed for H5N1 and H9N2 may be explained by the large proportion of strains derived from domestic poultry populations. In contrast, the higher rates observed in the H1N1, H3N8 and H4N6 subtypes could be due to their primary origin as infections of wild birds with multiple low pathogenicity strains in the large avian reservoir.

3.1 Introduction

Influenza A viruses are highly adaptable, able to evade host immune responses and to transmit between different host species. The segmented genome is composed of eight negative sense RNA strands. Segments 4 and 6 encode haemagglutinin (HA) and neuraminidase (NA), respectively. The external domains of HA and NA are the major targets for host neutralizing antibodies and are used in the subtyping of influenza viruses. The other six are internal gene segments: the first three segments encode the viral polymerase complex: PB2, PB1 and PA, and segment 5 encodes the nucleoprotein (NP). These four proteins comprise the ribonucleoproteins (vRNPs), which are the core of the virion, with the viral negative stranded RNAs packaged inside. Segment 7 encodes both matrix protein (M1) and ion channel protein (M2). Two separate Non-Structural proteins NS1 and NS2 are encoded by segment 8 (McHardy & Adams, 2009; Webster *et al.*, 1992).

The structure of the virus genome allows for exchange among the eight RNA segments between viruses co-infecting a cell, a process termed reassortment (Webster, 1998). The viral reassortants derived from different virus subtypes can acquire completely new antigens (antigenic shift) thus avoiding recognition by previously infected hosts and allowing for efficient transmission, leading to a pandemic (Parrish & Kawaoka, 2005). Three human influenza pandemics in the 20th century had either complete or partial avian origins: the H1N1 Spanish flu in 1918 appears to have been derived from an earlier avian source and both the H2N2 Asian flu in 1957 and the H3N2 Hong Kong Flu in 1968 were reassortants between human and avian strains (Belshe, 2005; Kalthoff *et al.*, 2010; Lindstrom *et al.*, 2004). Genetic reassortment also played a prominent role in the origin of the swine-origin influenza A (H1N1) 2009 virus (Smith *et al.*, 2009b).

Wild avian species of wetlands and aquatic environments such as the Anseriformes (particularly ducks, geese, and swans) and Charadriiformes (particularly gulls, terns, and waders) have been recognized as the major natural reservoirs of influenza A viruses where the viruses are maintained predominantly by asymptomatic birds, indicating a long-standing history of endemic infection (Webster *et al.*, 1992).

Terrestrial poultry, such as chickens and turkeys are not traditionally seen as reservoir hosts of avian influenza virus (AIV), but are susceptible to infection with wild-bird-derived AIV (Taubenberger & Kash, 2010) through exposure to fecal material from wild birds when they migrate through an area (Webster, 2002). Previous studies indicated that domestic ducks may act as intermediaries between migratory ducks and terrestrial poultry in southern China (Cheung *et al.*, 2007), thus they could be thought as part of the larger reservoir within which different subtypes of influenza interact with each other (Huang *et al.*, 2010; Huang *et al.*, 2012).

Avian influenza A virus strains are classified as low pathogenicity (LP) or high pathogenicity (HP) due to their pathogenicity in chickens (Kalthoff *et al.*, 2010; Wood *et al.*, 1993). Highly pathogenic strains only arise in subtypes H5 and H7 and are distinguished by the appearance of a highly basic cleavage site in HA (Klenk & Garten, 1994; Stienekegrober *et al.*, 1992). Genetic exchanges between viruses circulating in wild and domestic birds have been documented, usually involving LPAI viruses in domestic poultry (Campitelli *et al.*, 2004; Duan *et al.*, 2007). More rarely, HPAI virus can be found to infect both wild and domestic birds, as seen in the case of HP-H5N1 avian influenza. Since 2003, HPAI H5N1 AIV spread from south China into Southeast Asia, causing sporadic transmissions to humans and tremendous damage to the poultry industry (Claas *et al.*, 1998; Lei & Shi, 2011). Repeated spill-over of Asian HPAI H5N1 viruses from domestic to wild birds has also been observed, which demonstrates the potential for reverse flow (Chen *et al.*, 2005; Feare, 2010), and supports a close interaction between influenza strains in wild and domestic birds (Capua & Alexander, 2008).

The 16 HA subtypes and 9 NA subtypes have allowed at least 103 of the possible 144 type A influenza A virus HA-NA combinations to arise (Horimoto & Kawaoka, 2005), which indicates a high frequency of reassortment among different HA and NA subtypes. However some subtypes are rarely detected, indicating that some restrictions on possible combinations may exist. Thus, two important questions, regarding the general pattern of genetic interaction between different influenza A virus subtypes remain: whether certain subtypes are more likely to reassort than the other subtypes, and why this might be so.

Though most studies of evolutionary dynamics of influenza viruses have focused on single segments, new methods have been developed to explore the interactions among them. The differences in phylogenetic history among segments for H3N2 human influenza viruses from New York state were described in a previous study (Rambaut *et al.*, 2008). They found the divergent position of HA was particularly related to NA. Similar phylogenetic histories among the HA and M1/2 segments, and among the NS and NP segments were also identified, but to a lesser extent among the PB2, PB1 and PA polymerase complex segments.

Unlike the dynamics of seasonal human influenza strains, characterized by a few co-circulating subtypes, little inter-subtype reassortment (comparing to multiple within-subtype resassortment) and broadly correlated evolution of lineages of the individual segments (Holmes *et al.*, 2005; Lycett *et al.*, 2012), avian influenza has many co-circulating subtypes which causes a complex distribution of internal segment lineages (Dong *et al.*, 2011; Li *et al.*, 2004). Conversely internal segment lineages have their own continuous histories, but with many different external segment partners (Perez *et al.*, 2008). Here I will consider the patterns of acquisition and loss of the external protein coding segments on the backbone of each internal protein coding segment, since this represents the mechanism for generating antigenic novelty in a circulating population of viruses from the perspective of the internal protein coding segments.

There are many statistical models to study the reassortment of influenza A viruses. Graph-incompatibility-based Reassortment Finder (GiRaF) is a computational method which robustly identifies reassortments by find incompatible gene segment phylogenies with a fast consensus-search algorithm. (Nagarajan & Kingsford, 2011). In addition, two-time test, which is analogous to phylogeny-based tests for recombination in non-segmented genomes. It acts on T1 and T2 genotypes and is performed in two steps. The modifier 'relative' makes explicit that the reassortant status of a T2 virus is evaluated relative to the T1 viral genotypes, but the time intervals can hardly be confirmed (Nagarajan & Kingsford, 2011). In addition, a recent study combined the experiments which examine reassortment patterns between the 2009 H1N1 pandemic and seasonal H1N1 strains with the mathematical studies, and further combined with virulence potentials to assess pandemic threats (Greenbaum *et al.*, 2012).

However, none of the studies illustrated to what extent the different subtypes of avian influenza viruses are likely to reassort among each other.

With the aim of exploring the rate of genetic exchange among viral subtypes during evolution, I carried out a phylodynamic analysis of the influenza A viruses circulating in avian populations (in both wild birds and domestic birds) from 1956 – 2011. In particular, I focused on investigating how avian influenza segments specific to the HA, NA and combined HA-NA subtypes evolve on the backbones of the six internal gene segments, by quantifying the inter-subtype reassortment rates and testing possible ecological and evolutionary factors that are associated with such pattern of reassortment: the relative proportion of each host species, the inferred evolutionary relaxed clock rates and the selective constraint acting on the genes.

3.2 Methods

3.2.1 Data preparation and Preliminary phylogenetic analysis

Initial analyses included 3226 full length complete genome avian influenza sequences of any subtype and sampling year from the NCBI Influenza Virus Resource (www.ncbi.nlm.nih.gov/genomes/FLU). Laboratory recombinants or highly cultured sequences were excluded.

The locations of samples were classified by countries and states/provinces separately. The host species were categorized by taxonomic bird Orders, and were further classified into wild birds and domestic birds: for instance, aquatic birds of the orders of Anseriformes (ducks, swans, etc.) and Charadriiformes (gulls, terns, etc.). Domestic birds can be further separated into domestic Anseriformes (ducks, geese etc.) and domestic Galliformes (chickens, turkeys etc.) (see Chapter 2 Table 2.1). Sequence alignments of coding regions were created using MUSCLE (Edgar, 2004) and then adjusted manually using BioEdit version 7.2.3. The gene segment alignments were manually concatenated to generate a single alignment of complete genome sequences, and duplicate genomes (having 100% nucleotide identity) were removed.

A preliminary phylogenetic tree of each segment was generated using the Neighbor Joining method in MEGA 5. These phylogenies were analyzed using

Path-O-Gen version 1.2 (Chapter 2 section 2.6). The plots of root-to-tip showed a good linear relationship between genetic divergence and sampling time of the AIV sequences, except strains A/pintail/ALB/179/1993 in PB2 and A/shorebird/DE/12/2004 and A/duck/Yokohama/aq10/2003 in HA, respectively, indicating that they might contain incorrect date information and were removed from the dataset. Thus, the final dataset comprised 2996 full genome sequences sampled over 55 years from 1956 to 2011. Phylogenetic trees for each segment were generated using RAxML (Stamatakis, 2005), employing maximum likelihood (ML) analyses with 500 bootstraps.

Phylogenetic analysis of the 2996 AIV sequences datasets confirms a clear separation of AIV sequences from east and west hemispheres, although a minority of sequences (0.87%) are found to cross the boundaries, which confirmed that geographical separation of host species has shaped the avian influenza A virus gene pool into almost completely independently evolving Eurasian and American lineages (Figure 3.1). I focused on the avian influenza sequences of the Eurasian lineage, which included 1140 strains and composed of 15 HA subtype (H1-16, without H14), 9 NA subtype and 60 combined HA-NA subtypes in total.

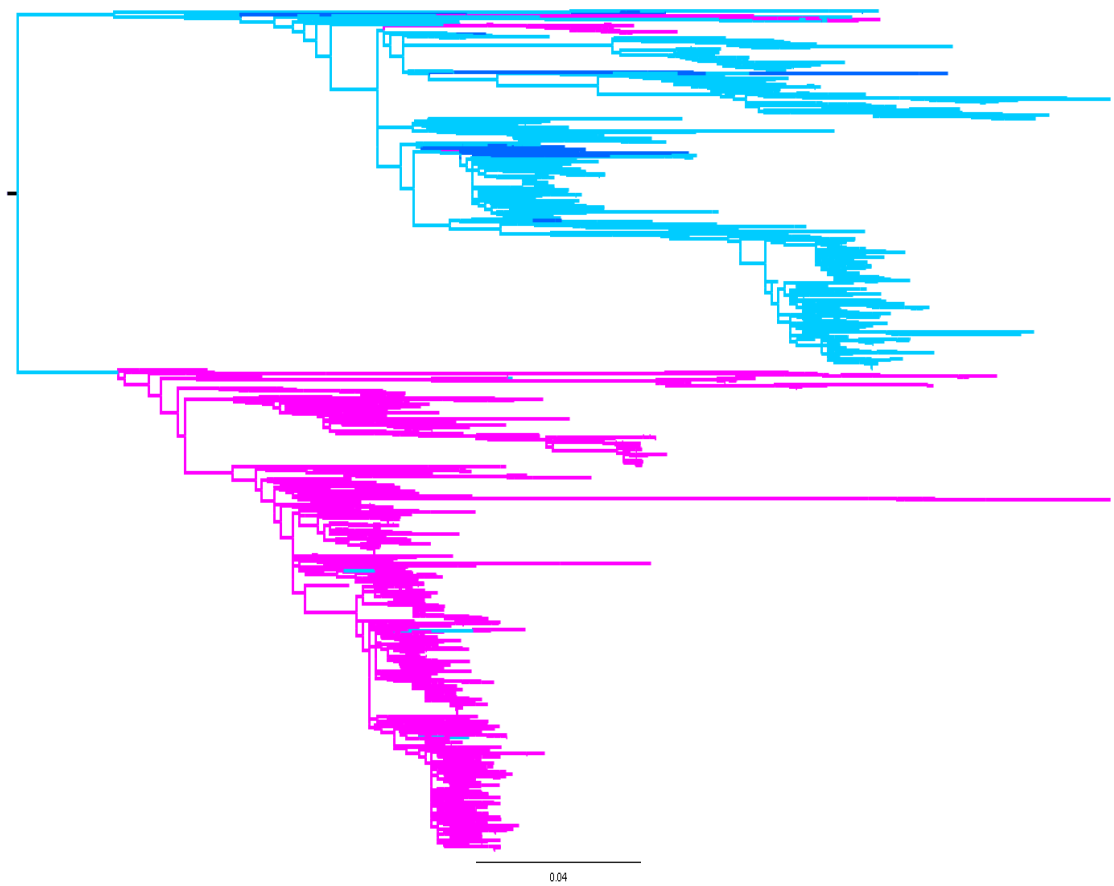


Figure 3.1 Maximum likelihood tree for the PB2 segment of 2996 AIV.
AIV from Eurasian region are coloured by blue (those from Oceania are coloured by dark blue), and those from the American regions are coloured by pink.

3.2.2 Time-scaled phylogenetic analysis

Time-scaled trees were generated for each internal protein-coding segment with a subsample of the 1140 sequence dataset with known isolation dates using BEAST (version 1.7.3). Different substitution models, clock models and tree models were evaluated for each segment by the calculation of the Bayes factor (BF),which is the ratio of the marginal likelihoods of the two models, and log BF > 3 was taken as indicating support for one model over another (Raftery, 1994) in Tracer v1.4 (http://tree.bio.ed.ac.uk/software/tracer/) (Drummond *et al.*, 2006). A general-time-reversal (GTR) model (Tavaré, 1986b) with gamma distributed rate heterogeneity of 4 rate categories (C4) was chosen for the M and NS segments for both of which have overlapping reading frames. For the PB2, PB1, PA and NP segments,

the SRD06 nucleotide substitution model (Shapiro *et al.*, 2006) was chosen as the default nucleotide substitution model (preferred over GTR model with log BF from 3242 to 3289), which allows one Hasegawa, Kishino and Yano (HKY) model (Hasegawa *et al.*, 1985) for codon positions 1 and 2 and a different HKY model for position 3, together with gamma-distributed rates across sites. A constant population size model (preferred to either an exponential model or a Bayesian skyride model with log BF from 7.23 to 7.52) and a relaxed uncorrelated log-normal molecular clock model (preferred to a strict clock model with log BF from 5.66 to 15.32) were chosen for all eight segments. The MCMC was run for 10^8 steps and sampled every 10^4 steps. Two independent runs were used in each segment to confirm the convergence and then combined. The Maximum Clade Credibility (MCC) trees were obtained and the mean ages of each node and the corresponding HPD were calculated by using Tree Annotator v1.7.3 in BEAST.

3.2.3 Discrete trait mapping and estimation of reassortment rates

To quantify the genetic association among subtypes (HA, NA and HA-NA) in AIV from Eurasia, I implemented the asymmetric discrete traits model with irreversible transitions (with the exponetial prior with mean = 1 for the trait.rate) to reconstruct the ancestral subtype states of the internal nodes from the posterior time-scaled tree distribution (Lycett *et al.*, 2012), using a relaxed uncorrelated log-normal molecular clock model, SRD06 model of nucleotide substitution and a constant-population coalescent process prior across the phylogeny of each segment. In addition, a BF test was constructed to identify the significant transition rates between discrete traits with the Bayesian stochastic search variable selection (BSSVS) extension of the discrete model. For the rates calculated from BSSVS, a BF test was adopted to identify significant non-zero transition rates (BF = 3) (Faria *et al.*, 2011).

For each of the six internal segments, I did 3 separate analyses on sequences labelled by HA subtype, NA subtype, and HA-NA combined subtype. The reassortment rate being analysed in this study was classified into three levels: a) The overall reassortment rate for an internal segment (e.g. PB2) reported for HA subtype and for NA and for combined HA-NA was the overall mutation rate of HA,NA or

combined HA-NA traits in substitutions per year; b) The rate of reassortment for an internal segment (e.g. PB2) starting from each different subtype (e.g. H1, N1, or H1N1) is the mean rate that averaged over the individual transition rates from certain subtype to all the other subtypes; c) The asymmetric rate of reassortment for an internal segment (e.g.PB2) starting from one subtype to another (e.g. H1 to H2, H2 to H1) was estimated from each relative transition rate multiplied by the corresponding clock rate per Monte Carlo Markov Chain (MCMC) sample.

For comparison, the reassortment rates were also estimated by MultiState in BayesTraits (version 1.0), using a maximum likelihood method with 1000 bootstrap replications. BayesTraits is a computer package for performing analyses of trait evolution among groups of species for which a phylogeny or sample of phylogenies is available, which can be applied to the analysis of traits that adopt a finite number of discrete states, or to the analysis of continuously varying traits. Hypotheses can be tested about models of evolution, about ancestral states and correlations among pairs of traits. MultiState implementation is used to reconstruct how traits that adopt a finite number of discrete states evolve on phylogenetic trees. (Pagel *et al.*, 2004). Only the rates of HA and NA subtypes (not HA-NA joint subtype) could be estimated using Multistate, as only a maximum of 10 distinct states were allowed. The mean rates estimated by two different reconstruction methods (between a single tree with maximum likelihood method in BayesTraits and a MCMC over a set of trees in BEAST) are similar, although there was a significant shrinkage in uncertainty using the former method (see Figure 3.2). This shows that the two implementations produce similar mean results, but reasortment rates estimated with the discrete trait model using the MCMC-based Bayesian framework were preferred because the results were integrated over a set of likely trees rather than a single best tree.

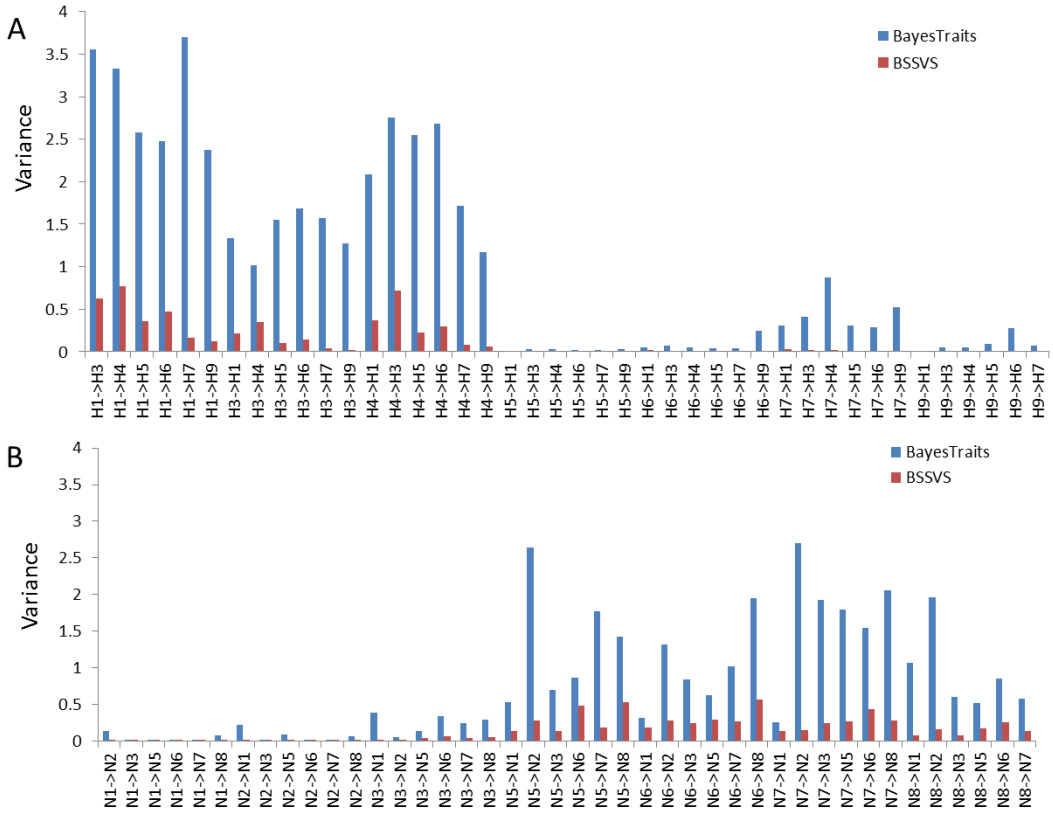


Figure 3.2 Distribution of variance of reassortment rate using BSSVS and Bayes Traits. PB2 segment of HA subtype (A) and NA subtype (B). Variance of estimated reassortment rate between pairs of subtypes using BayesTraits (blue); Variance of estimated reassortment rate between pairs of subtypes using BSSVS (red).

3.2.4 Robustness of random sampling

In order to generate datasets suitable for obtaining time-resolved trees, stratified subsampling was conducted using custom R scripts. Subtypes with more than 7 sequences were subjected to further analysis. In addition, to keep the pattern of the original phylogeny, sequences in isolated branches with extremely long branches were ignored. To retain a similar distribution of isolates for each subtype, random subsampling was conducted to keep one sequence per subtype per host species per location per year, and was repeated 3 times (Table 3.1, rows A1-A3). Among the three subgroups (A1, A2, and A3), the mean individual reassortment rate of each single transmission were similar to each other and with overlapping credible intervals (Cls) , which indicated that the estimated reassortment rates are robust by random sampling

(Figure 3.3). Thus, the reassortment rates shown in the results were the mean rate of the three subgroups.

Group	A1	A2	A3	B	C
H1N1	7	7	7	7	7
H3N8	19	18	19	18	10
H4N6	12	12	12	12	10
H5N1	100	129	126	10	10
H5N2	17	16	17	16	10
H5N3	12	12	12	12	10
H6N1	31	29	30	29	10
H6N2	36	24	23	24	10
H6N5	10	8	8	8	10
H7N1	10	10	10	10	10
H7N3	12	10	10	10	10
H7N7	9	8	8	8	9
H9N2	68	61	61	61	10
SUM	344	344	344	225	127

Table 3.1 Number of sequences by subtype in each subgroup (A1, A2, A3, B, C)

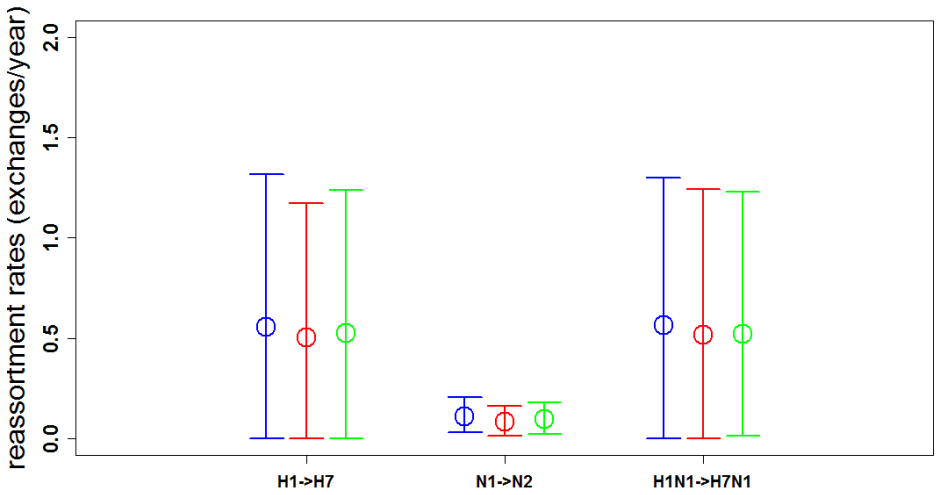


Figure 3.3 Robustness of reassortment rate (HA, NA and HA-NA combined subtype). As for clarity, only one reassortment rate per HA,NA and HA-NA combined subtype for PB2 segment was shown with mean reassortment rates and 95% Highest probability density (HPD) intervals among 3 random sampling subgroups, respectively. Blue: group1; Red: group 2; Green: group 3.The complete rate of each transition were shown in Table A 3.1.

3.2.5 Test of sample size effect

The dataset in this study was obtained from the NCBI database, where the number of sequences is unevenly distributed among different subtypes. HPAI H5N1 AIV in particular accounts for a very high proportion of sequences due to large-scale surveillance efforts. To investigate a possible sample size effect on the reassortment rate, I estimated reassortment rates using two other subsampled datasets which comprised different numbers of representative sequences from the original datasets of 344. To identify whether the sample size of each subtypes would affect the reassortment rate, I further selected representative sequences of different subtypes in two other different ways: 1) by reducing the number of the most over-represented subtype to equal the number of the rare subtype in the original dataset (Table 3.1, row B); and 2) by reducing the number of sequences of all the subtypes to be the same as the rare subtype (Table 3.1, row C). Representative isolates of the down-sampled datasets were selected according to the tree phylogeny of the original maximum likelihood tree generated with all the Eurasian AIV, using a software tool which searches for sequence clusters (Ragonnet-Cronin M, 2013)

(http://hiv.bio.ed.ac.uk/software.html), with 70% bootstrap support and a genetic distance lower than 0.045 specified. In this comparison, the reassortment rates among subtypes were similar in all the three datasets (Figure A 3.2, at the end of this Chapter). Thus, I can conclude that the sample size does not affect the estimation of the reassortment pattern of Eurasian AIV in this study.

3.2.6 Estimation of evolutionary rates

The time scaled Bayesian phylogenies were generated with an uncorrelated log-normal relaxed molecular clock, and each branch was annotated with its mean substitution rate (named as ucld.mean , which is the mean rate under the uncorrelated log-normal relaxed molecular clock). Additionally the nodes of the trees were annotated with the inferred HA, NA or HA-NA subtype states from the respective discrete trait models. The mean evolutionary rates per subtype were extracted from the trees, averaging over the rates on only those branches for which both the parent and child nodes were annotated with that subtype. The overall evolutionary rate of each segment was the ucld.mean rate of the MCMC tree phylogeny of each segment with 95% HPD intervals.

3.2.7 Estimation of the dN/dS Ratio

Selection within subtypes on the tree phylogeny was analysed with implementation of robust counting on the trees generated with Bayesian discrete models in Beast, which reconstructed the synonymous and non-synonymous change counts using a 3-partition codon model (O'Brien *et al.*, 2009). For segment 7 and 8, M1 and NS1 were analysed instead of full length segments due to overlapping reading frames (Ghedin *et al.*, 2005). A recent developed robust counting methodology was adopted for the dN/dS estimation (Lemey *et al.*, 2012). This method records the numbers of synonymous (S) substitutions in stochastic mapping realizations of the continuous-time Markov chains, with conditioning (C) on the codon data at the tips of the phylogeny and without conditioning (U) on the data. For either conditional (C) or unconditional (U) stochastic mapping realizations, the numbers of synonymous (S) substitutions C(S-C) , C(S-U) and nonsynonymous (N) substitutions C(N-C) , C(N-U) for each state on each branch were recorded, and then converted into dN/dS values (indicated by certain

state/subtype) by computing the ratio *C*(N-C)/*C*(S-C) / *C*(N-U)/*C*(S-U) (Lemey *et al.*, 2012). The ratio per subtype was estimated from the ratio per branch across all the MCMC trees. The overall selective pressure of each segment was the mean difference between the mean number of total N and total S of across all the MCMC trees of certain segment with 95%HPD intervals.

I also used Single Likelihood Ancestor Counting (SLAC) in HyPhy (http://www.datamonkey.org/) to estimate dN/dS per subtype as a comparison. Here I used the original datasets (1185 AIV in total) which have more sequences of each subtype than the subsampling dataset. The HKY85 nucleotide substitution model was used in SLAC-based analyses.

3.2.8 The estimation of partial correlation

I estimated the partial correlation coefficient between selective pressure and the reassortment rates aside of the effect of the host proportion in equations (Baba *et al.*, 2004):

$$\mathrm{R}(A,B|C) = \left(R(A,B) - R(A,C) * R(B,C)\right) \Big/ \sqrt{(1-R(A,C)^2)(1-R(B,C)^2)}$$

$$\mathrm{t} = \sqrt{\mathrm{df}}\sqrt{\mathrm{R}(A,B|C)/(1-R(A,B|C)^2} \qquad (3.1)$$

In which A is dN/dS ratio; B is inter-subtype reassortment rate; C is host proportion (domestic); R(A, B|C)is the partial correlation coefficient between reassortment rate and dN/dS independent of domestic proportion of 13 HA-NA subtypes; t is the t test P value of R(A, B|C) with degrees of freedom (df) =11.

3.3 Results

3.3.1 Time-scaled phylogenies of internal genes of Eurasian AIV

To explore how the HA and NA subtypes reassorted with the internal segments of Eurasian AIV, I used discrete trait models (in BEAST) upon empirical phylogenetic trees of the six internal gene segments. For reasons of computational tractability, I analysed subsamples of 344 sequences generated using a stratified approach which maintained the range of genetic and subtype diversity. These were composed of 7 HA

and 7 NA subtypes and 13 combined HA-NA subtypes respectively (see Methods and Figure A 3.1).

The reassortment history of Eurasian AIV was inferred by discrete trait models (HA, NA and combined HA-NA subtypes) on phylogenetic trees. Figure 3.4 shows the MCC tree of PB2 with branches coloured by the HA subtype at the child nodes (or tips), and the other trees can be found in Figure A 3.1. The two most over-represented subtypes can be easily seen to form single subtype clades; coloured green (H5N1) and orange (H9N2). However, it is apparent that some different subtypes were interspersed with each other, and the internal segments did not form monophyletic lineages with respect to antigenic subtype, indicating extensive reassortment between external and internal gene segments. A separation of two main clades can be seen in the PB2 phylogeny (Figure 3.4 A-C). The sequences in each clade cover a range of subtypes, host, sample location and time and do not correspond to the strains belonging to allele A and B in NS (fewer than 40% of the sequences in either lineage are the same) (Figure 3.4 D). In contrast, lower genetic diversity was observed in the remaining 4 gene segments (PB1, PA, NP and M) (Figure A 3.1).

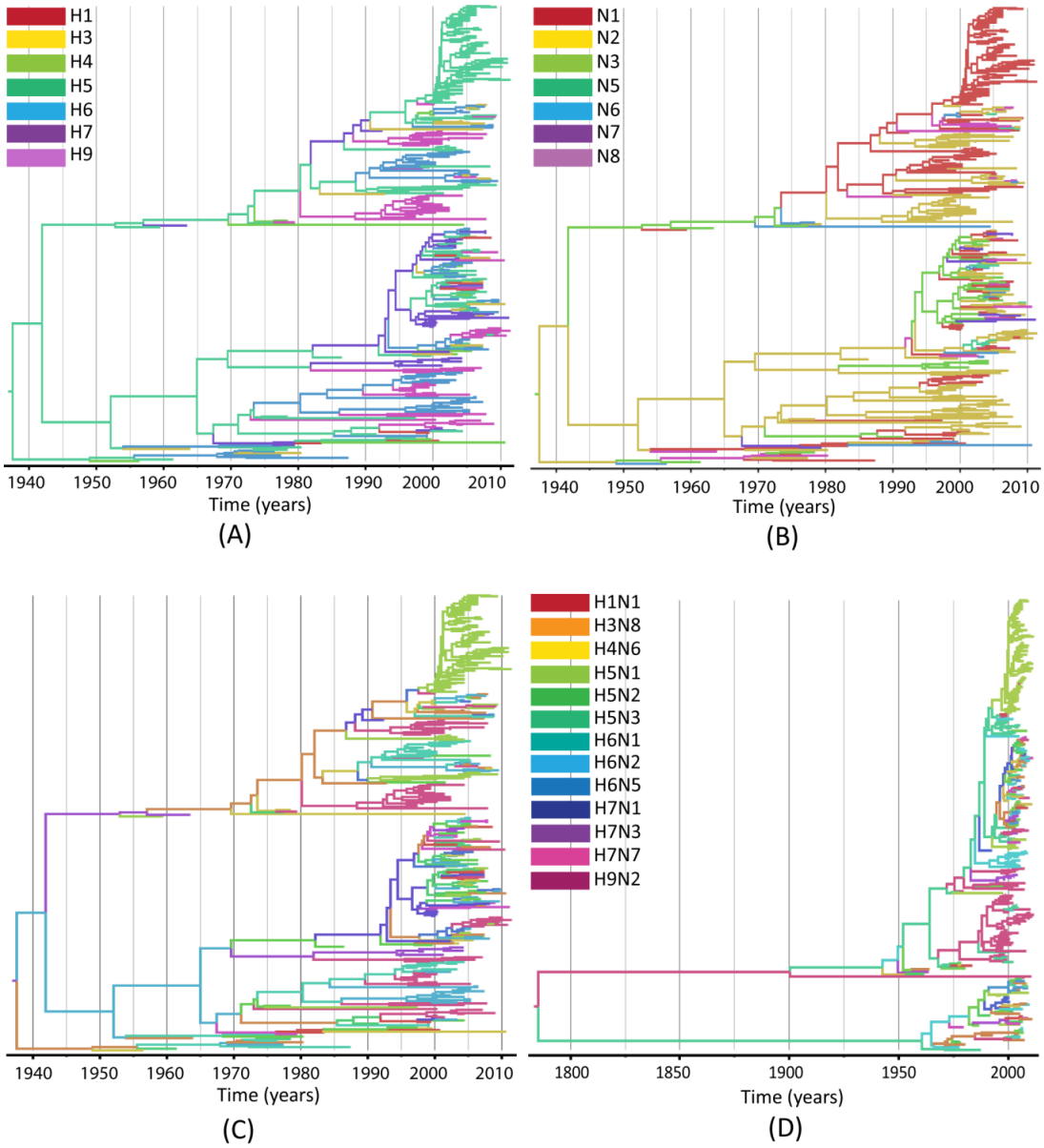


Figure 3.4 Bayesian MCC phylogenies for the internal segments encoding PB2 and NS. A: PB2 phylogeny coloured by HA subtype. B: PB2 phylogeny coloured by NA subtype. C: PB2 phylogeny coloured by combined HA-NA subtype. D: NS phylogeny coloured by combined HA-NA subtype. Branches are coloured according to the different subtypes of their descendent nodes (C and D share the same key).

3.3.2 Reassortment rates of discrete subtypes of Eurasian AIV

A range of inter-subtype reassortment rates was found with respect to the internal segments of Eurasian AIV. For both HA and NA subtypes, there were 7 distinct states and thus each discrete trait model contained 42 (asymmetric) transition pairs (e.g. H1 to H3, H3 to H1, N1 to N2, N2 to N1). For the combined HA-NA subtypes, there were 156 different possible transition pairs based on 13 distinct states (e.g. H1N1 to H3N8,

H3N8 to H1N1). I analysed the inter-subtype reassortment rate at three levels: rate per segment (Table 3.2), mean rate per subtype (Table 3.3) and absolute transition rate between pairs of subtypes (Table A 3.1).

In general, the overall reassortment rate at the segment level was slightly higher for the NA subtypes than for the HA and HA-NA combined subtypes (Table 3.1). In addition, the reassortment rates of the combined HA-NA subtypes (e.g. H5N1) were related to the rate of the corresponding independent HA (e.g. H5) and NA (e.g.N1) subtypes with P of correlation test < 0.01 in all six segments.

	PB2	**PB1**	**PA**	**NP**	**M**	**NS**
H	0.30 (0.13,0.39)	0.30 (0.13,0.45)	0.31 (0.11,0.44)	0.28 (0.13,0.38)	0.37 (0.17,0.46)	0.63 (0.28,0.70)
N	0.46 (0.22,0.57)	0.47 (0.18,0.63)	0.49 (0.19,0.64)	0.29 (0.14,0.35)	0.40 (0.29,0.49)	0.72 (0.43,0.97)
H-N	0.37 (0.21,0.45)	0.46 (0.26,0.52)	0.37 (0.21,0.45)	0.34 (0.21,0.40)	0.35 (0.20,0.43)	0.68 (0.41,0.76)

Table 3.2 Overall reassortment rates (substitutions per year) of six internal segments for H, N and HN combined subtype

When comparing the reassortment rates among the six internal segments, I found the rates for five (PB2, PB1, PA, NP and M) were similar. Among these NP had a slightly lower and PB1 slightly higher overall rate for HA, NA and the HA-NA combined subtypes. In contrast to these five segments, the NS segment possessed significantly higher overall reassortment rates than those of the other segments (Table 3.3).

	PB2	PB1	PA	NP	M	NS
H						
H1	0.47	0.54	0.39	0.45	0.54	1.15
H3	0.35	0.35	0.35	0.29	0.41	0.65
H4	0.44	0.46	0.53	0.41	0.53	0.78
H5	0.06	0.04	0.05	0.02	0.03	0.06
H6	0.09	0.10	0.11	0.09	0.12	0.17
H7	0.16	0.23	0.14	0.18	0.14	0.26
H9	0.02	0.05	0.04	0.04	0.03	0.04
N						
N1	0.06	0.07	0.07	0.04	0.06	0.05
N2	0.09	0.07	0.10	0.07	0.12	0.16
N3	0.29	0.28	0.29	0.20	0.35	0.59
N5	0.58	0.48	0.64	0.36	0.75	1.01
N6	0.72	0.58	0.69	0.37	0.76	1.06
N7	0.57	0.84	0.59	0.39	0.56	0.99
N8	0.51	0.42	0.49	0.28	0.60	0.86
H-N						
H1N1	0.48	0.60	0.41	0.45	0.46	0.90
H3N8	0.34	0.39	0.34	0.29	0.33	0.64
H4N6	0.45	0.54	0.47	0.38	0.43	0.77
H5N1	0.03	0.05	0.04	0.03	0.03	0.06
H5N2	0.35	0.46	0.38	0.31	0.32	0.66
H5N3	0.35	0.54	0.45	0.36	0.42	0.75
H6N1	0.10	0.16	0.15	0.09	0.12	0.16
H6N2	0.21	0.29	0.18	0.18	0.18	0.51
H6N5	0.46	0.44	0.40	0.29	0.42	0.84
H7N1	0.36	0.41	0.29	0.34	0.29	0.59
H7N3	0.22	0.16	0.17	0.21	0.16	0.25
H7N7	0.40	0.60	0.41	0.41	0.33	0.81
H9N2	0.02	0.05	0.04	0.03	0.03	0.05

Table 3.3 Inter-subtype reassortment rates (substitutions per year) per subtype for H, N and H-N combined subtypes of 6 internal segments

I compared the mean reassortment rates of different subtypes, and found a substantial range among each subtype in all six internal segments and all three subtype traits (Table 3.3). Specifically, of the HA subtypes, H5, H6, H7 and H9 showed lower reassortment rates relative to the others, and H5 and H9 had the lowest rates in the six internal segments. By contrast, the reassortment rates of the H1, H4, and H3 subtypes were high, especially for H1 with the NS segment, which had the highest rate (1.148 exchanges per year) compared to all the others. For NA subtypes, N1 and N2 presented lower reassortment rates compared to those of N3, N5, N6, N7 and N9. For the HA-NA joint subtypes, reassortment rates of H5N1 and H9N2 were considerably

lower than with the other subtypes in all segments, and H9N2 had the lowest rates in the PB2 segment (0.0220 exchanges per year) and NS segment (0.0487 exchanges per year), while H5N1 had the lowest rates in the other segments (PB1: 0.0449; PA: 0.0352; NP: 0.0249; and M: 0.0266 exchanges per year); other subtypes with low rates were H6N1, H6N2 and H7N3 subtypes. In contrast, H1N1, H3N8 and H4N6 and others exhibited high reassortment rates.

For clarity, the patterns of pairwise transition between individual subtypes are represented as a series of heat maps for the six internal segments in Figure 3.5 and Figure 3.6. In the two figures, reassortment rates are shown as a heat map. Both forward and reverse rates are shown, with subtypes labelled on the left side acting as donors and subtypes labelled at the bottom acting as recipients. The values of the mean reassortment rates, in exchanges per year, are represented as colours from yellow (low) to red (high). The scale of rates is on the left hand side of each plot. The inferred asymmetric reassortment transition rate matrixes are shown in Table A.3.1. The colour intensity in each column reflects the rate with which each subtype (labelled on the y-axis as donors) was found to reassort into a clade of other subtypes. The colour intensity in each row reflects the rate with which the reciprocal events were inferred (labelled on x-axis as recipients).

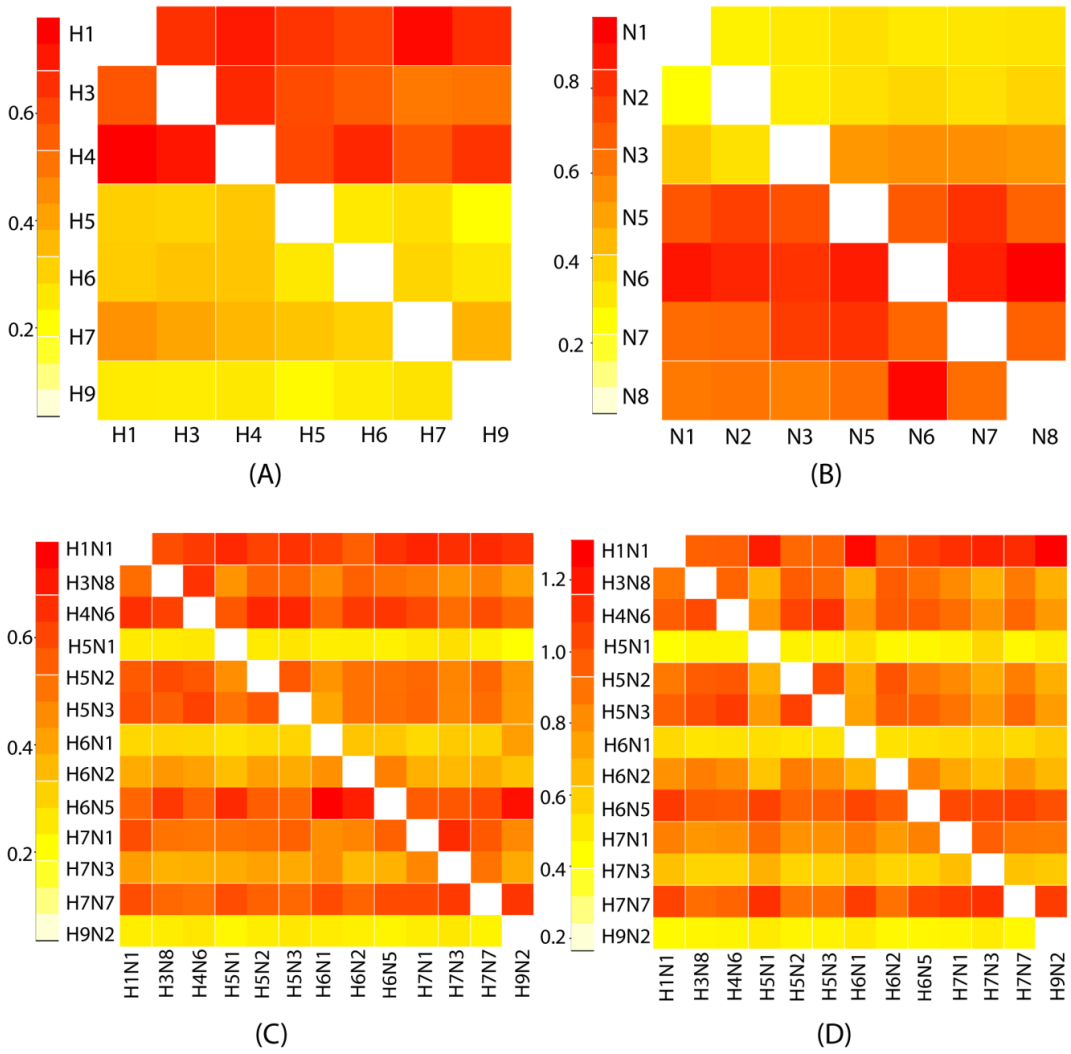


Figure 3.5 Reassortment rates between internal segments of different subtypes.
A: PB2 segments associated with HA subtypes shown, which means PB2 that are associated with H1, H3 and H4 subtypes rapidly reassort to become associated with any HA, while PB2 associated with H5, H6, H7 and H9 subtypes reassort more slowly to become associated with any HA. B: PB2 segments associated with NA subtypes shown; C: PB2 segments associated with combined HA- NA subtypes shown; D: NS segments associated with combined HA-NA subtypes shown. The reassortment rates from donors to other subtypes are shown in rows; the rates from all subtypes to recipients are shown in columns.

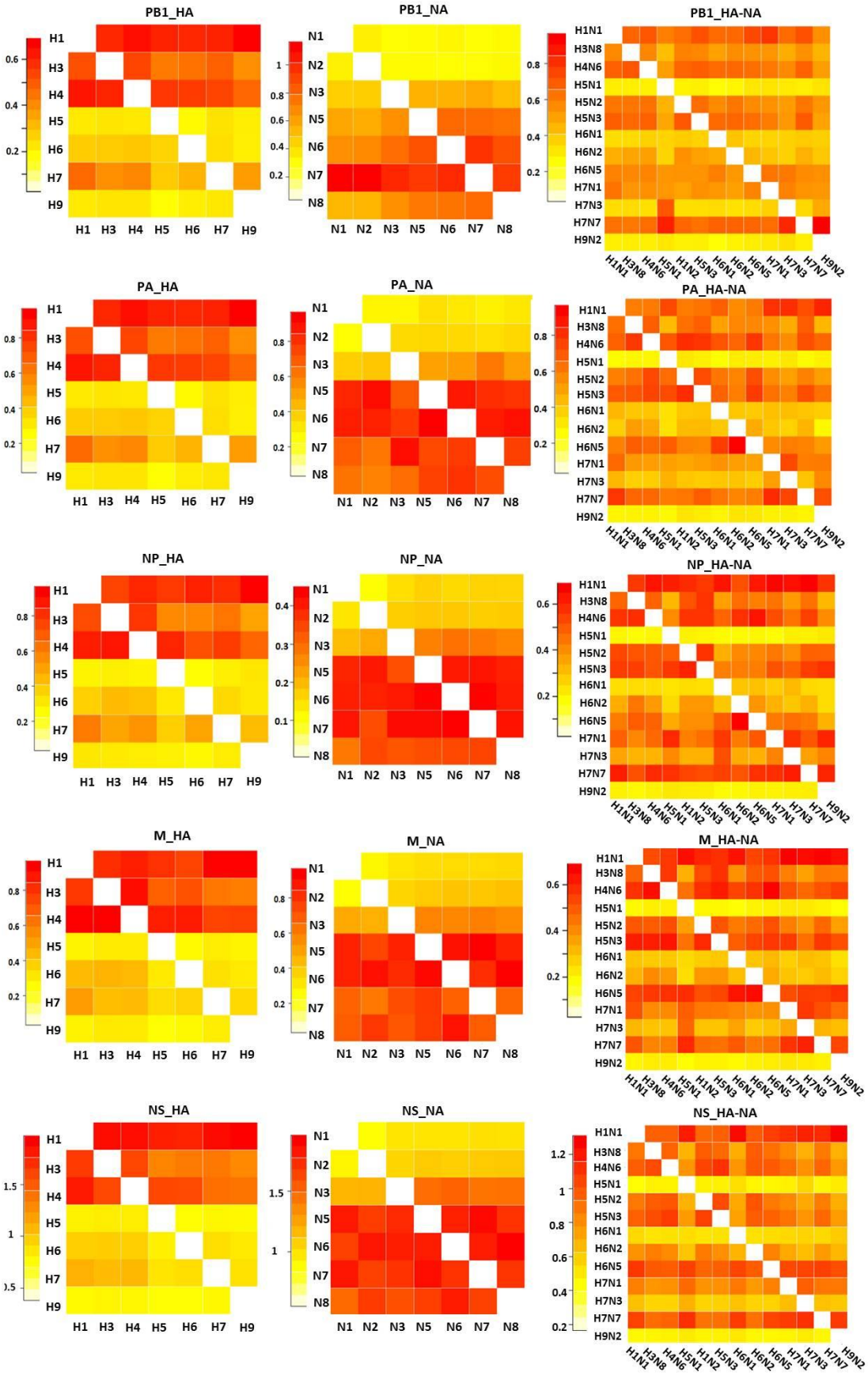

PB1_HA
H1
H3
H4
H5
H6
H7
H9
PB1_NA
N1
N2
N3
N5
N6
N7
N8
PB1_HA-NA
H1N1
H3N8
H4N6
H5N1
H5N2
H5N3
H6N1
H6N2
H6N5
H7N1
H7N3
H7N7
H9N2
PA_HA
PA_NA
PA_HA-NA
NP_HA
NP_NA
NP_HA-NA
M_HA
M_NA
M_HA-NA
NS_HA
NS_NA
NS_HA-NA

Figure 3.6 Heat map of reassortment rates of remain 5 internal segments (PB1, PA, NP, M, NS).
Subtypes (HA, NA and both HA and NA) labelled on left side are as donors, subtypes labelled in the bottom are as recipients. The values of rates from low to high are represented colours from yellow to red. The scale of rates is on the left hand side of each plot.

A range of reassortment rates were observed among subtypes as recipients, which indicates that each subtype possesses a different probability of giving the viral gene to viruses of other subtypes during co-infection. In comparison, the rates of subtypes as donors from the other subtypes were consistent which gives rise to the horizontal striping seen in the heat maps. The distinction between source and sink rates were further quantified by estimating the variance of rates in columns to the variance of rates in rows (Table A.3.1 and Figure 3.7), which showed the variances were significantly smaller in the source to sink direction than the reverse indicating the asymmetry of the reassortment rates. This pattern can be seen in the HA, NA subtypes independently and in the HA-NA combined subtypes, and is also similar in all six internal segments.

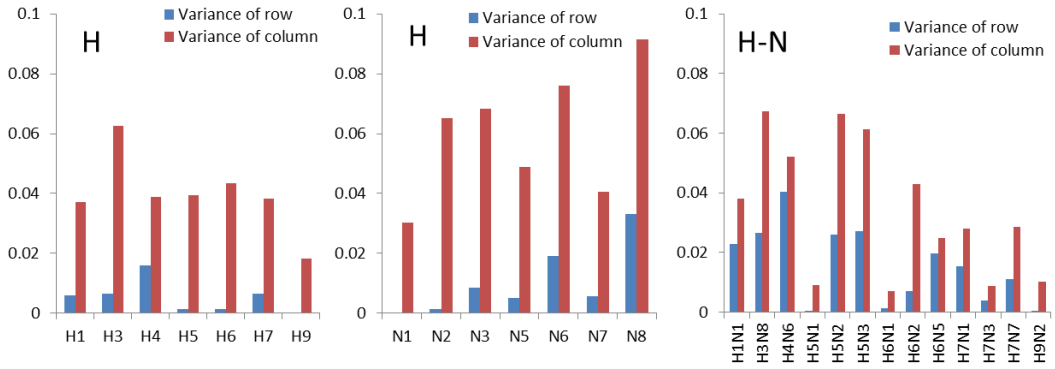


Figure 3.7 Distribution of inter-subtype reassortment rates as donors and recipients.
In the bar plot, the reassortment rates of H, N, and H-N subtypes as donors (rows) are shown in blue and those as recipients (columns) are shown in red, as described in Figure 3.5. The rates are from the PB2 tree.

3.3.3 Correlation of reassortment with genetic diversity

To further explore what factors are required to explain the pattern of reassortment rate in the segment level, I first tested the genetic diversity at the segment level which can be represented by the mean time to the most recent common ancestor (TMRCA) based on each phylogeny. The TMRCA of three internal gene segments were quite similar

and recent: PB2 (75.71 years, with 95% HPD (Highest Posterior Density): 64.70-87.21), PA (77.95 years with 95% HPD: 64.42-93.01) and the M segment (73.17 years, with 95% HPD of 61.56-87.87). The PB1 segment had an older mean TMRCA, of 105.12 years (95% HPD: 80.21-135.71) while NP was less divergent than the other segments (TMRCA 63.22 years; 95% HPD: 58.32-69.33). Compared to the other segments, the tree topology of the NS segment was characterized by a very deep divergence between the two sub-lineages formed by the A and B alleles which represent two distinct gene pools, and consequently has a much older TMRCA of 226 years (95% HPD: 112–273) (Figure 3.4 D).

The association between genetic diversity and the overall reassortment rate of six internal segments of Eurasian AIV was tested for HA, NA and HA-NA combined subtypes, respectively. The results (Figure 3.8) exhibited a strong positive correlation for the H subtype ($p = 4.7 \times 10^{-4}$, $R = 0.982$), the N subtype ($p = 0.01$, $R = 0.879$) and the H-N combined subtype ($p = 7.569 \times 10^{-5}$, $R = 0.99$).

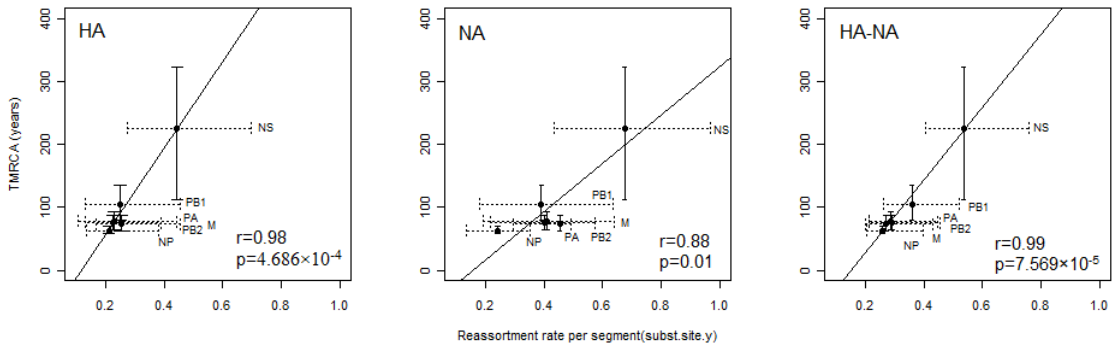


Figure 3.8 Correlation between TMRCA (years) and reassortment rate (exchanges per year).

For HA subtype, NA subtype and HA-NA combined subtype, respectively. Scatterplots of mean TMRCA (95% HPD intervals were shown in solid vertical lines) against the mean reassortment rate (95% HPD intervals shown in dotted horizontal lines) of each internal segment. The related segment names were labelled alongside.

3.3.4 Correlation of reassortment rates with host

The prevalence of AIV in different avian hosts could also influence the reassortment pattern. The host prevalence of different subtypes was calculated from the subsampled data sets used to make the phylogenetic trees, which was proportional to the original data. The primary host difference lies between wild birds and domestic birds which live in completely different ecosystems. Subtypes H4N6 had the highest prevalence in wild birds (Table 3.4). I tested the association between individual reassortment rates and the proportion of domestic birds for the HA-NA combined subtypes. I found that there was a significant positive correlation between the proportion of wild bird strains per subtype and the reassortment rates for those subtypes, and that the correlation held in each of the six segments (PB2: p = 0.009, R = 0.69; PB1: p = 0.004, R = 0.73; PA: P = 0.006, R = 0.71; NP: P = 0.02, R = 0.62; M: P = 0.06, R = 0.71; NS: p = 0.002,R = 0.76; Figure 3.8 and Figure 3.9).

Subtype	Wild[a]	Ans[b]	Ans-Domestic[c]	Gal-Domestic[d]
H1N1	0.50	1.00	0.50	0.00
H3N8	0.44	0.89	0.50	0.06
H4N6	0.58	0.92	0.42	0.00
H5N1	0.35	0.42	0.25	0.40
H5N2	0.38	0.81	0.38	0.19
H5N3	0.42	0.92	0.58	0.00
H6N1	0.34	0.45	0.24	0.41
H6N2	0.50	0.88	0.42	0.08
H6N5	0.50	1.00	0.50	0.00
H7N1	0.30	0.60	0.40	0.30
H7N3	0.20	0.40	0.20	0.60
H7N7	0.50	0.75	0.25	0.25
H9N2	0.15	0.26	0.18	0.67

Table 3.4 Proportion of domestic birds of each subtype analysed

[a]The overall proportion of AIV with a wild bird host

[b]The overall proportion of AIV with a host species belonging to Anseriformes

[c]The proportion of AIV with a domestic Anseriformes host

[d]The proportion of AIV with a domestic Galliformes host

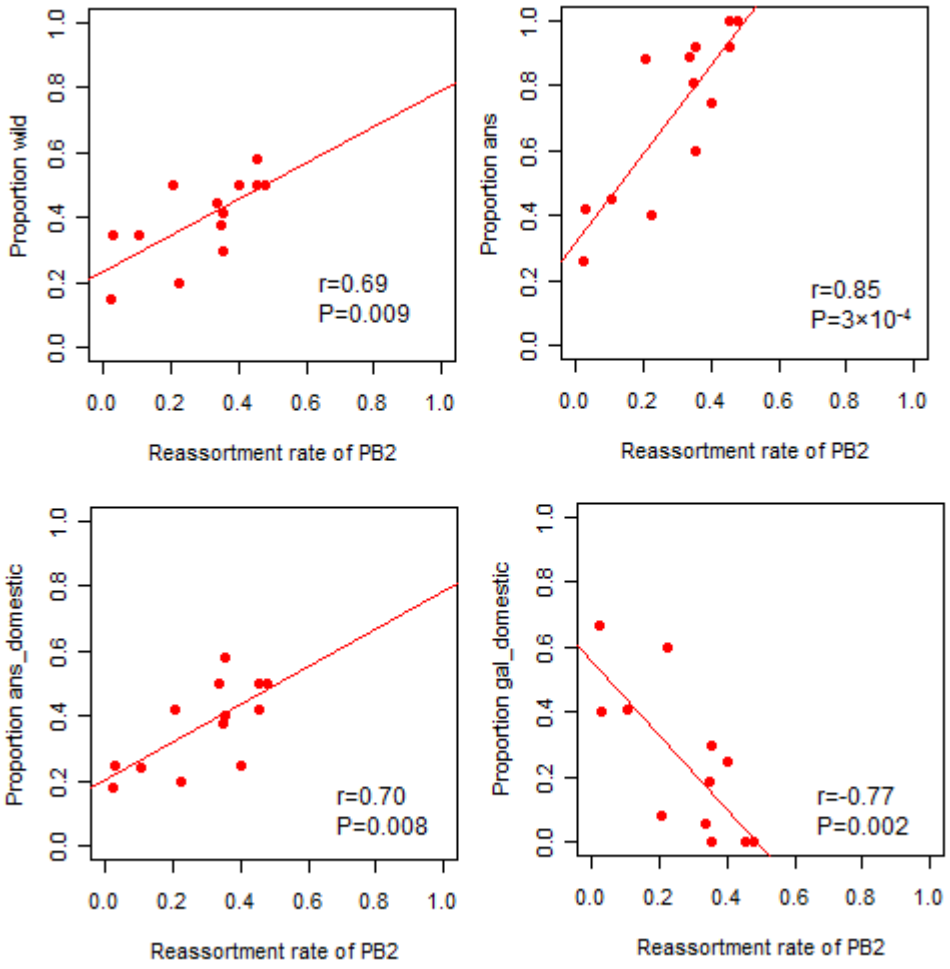


Figure 3.8 Correlation between the prevalence in host and reassortment rate per subtype in PB2 segment.
The scatterplot of the mean reassortment rates of 13H-N combined subtypes of each internal segment against the proportion of AIV 1) in wild birds, 2) in Anseriformes 3) in domestic birds of Anseriformes and 4) in domestic birds of Galliformes.

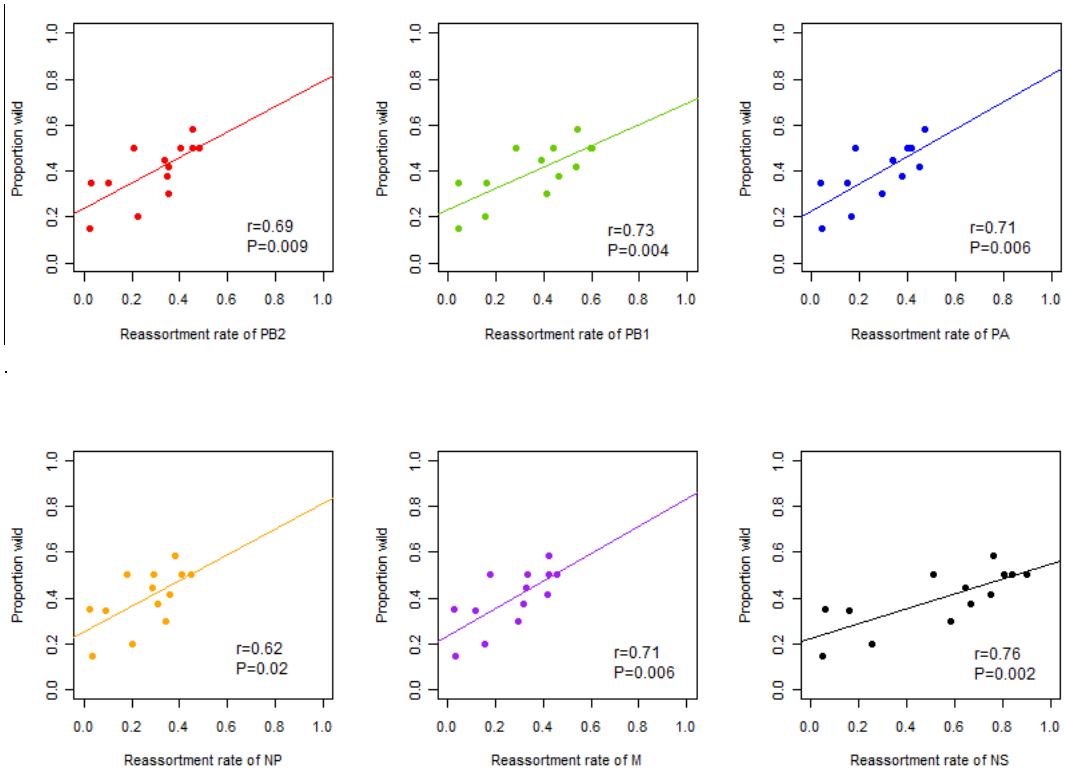


Figure 3.9 Correlation between the proportion of AIV in wild birds and reassortment rate per subtype in 6 segments.
The scatterplot of reassortment rate of HA-NA combined subtype against proportion of wild birds per subtype were represented for 6 internal segments by different colours: PB2 (red); PB1 (green); PA (blue); NP (orange); M1 (purple); NS1 (black).

The prevalence of AIV in different host species (by Order) was also analysed. In this dataset, subtypes with a high reassortment rate also have the highest prevalence in Anseriformes, which is the Order representing birds living in predominantly freshwater habitats. I found a very strong positive correlation between reassortment rates and the proportion of Anseriformes (PB2: $p = 3 \times 10^{-4}$, $R = 0.85$; PB1: $p = 2 \times 10^{-4}$, $R = 0.84$; PA: $p = 2 \times 10^{-4}$, $R = 0.84$; NP: $p = 0.002$, $R = 0.77$; M: $p = 4 \times 10^{-5}$, $R = 0.89$; NS: $p = 9 \times 10^{-6}$, $R = 0.90$) (Figure 3.8 and Figure 3.10), which may indicate the subtype with high reassortment rate could be better explained by their high proportion in waterfowl than wild birds.

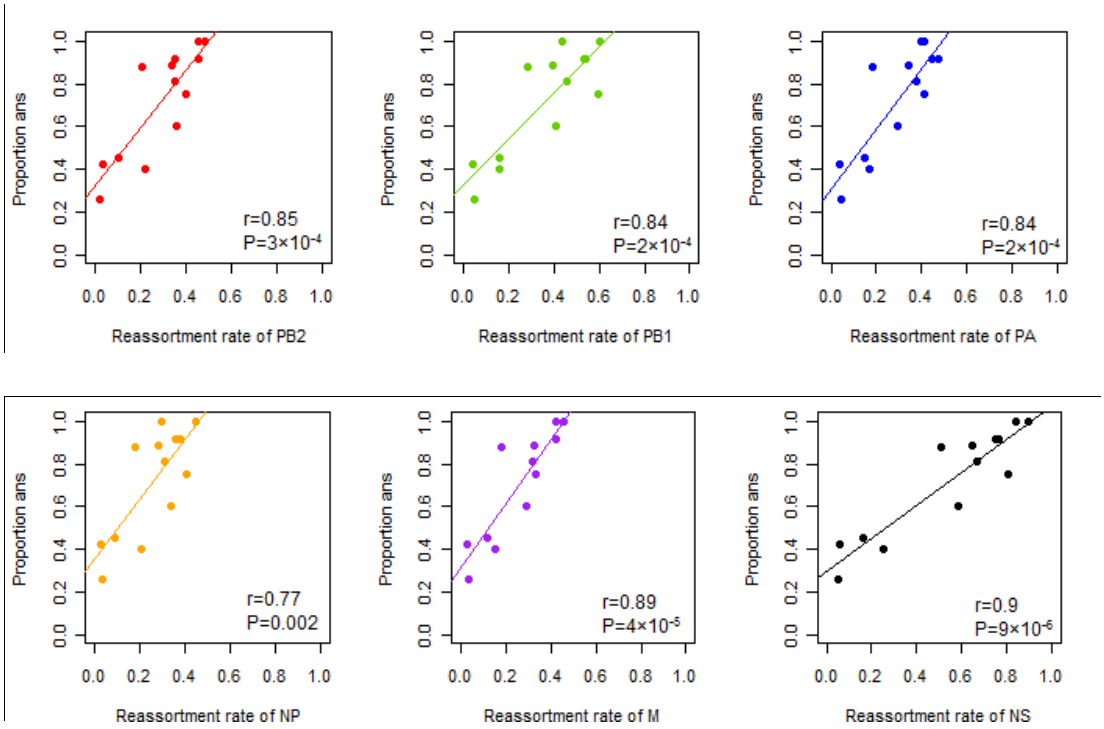


Figure 3.10 Correlation between the proportion of AIV in Anseriformes and reassortment rate per subtype in 6 segments.

We further tested wild (mainly mallard in this study) and domestic Anseriformes (mainly domestic duck in this study) separately (Table 3.4). Both of them show a strong positive correlation with reassortment rate per subtype (Figure 3.11 and Figure 3.12).

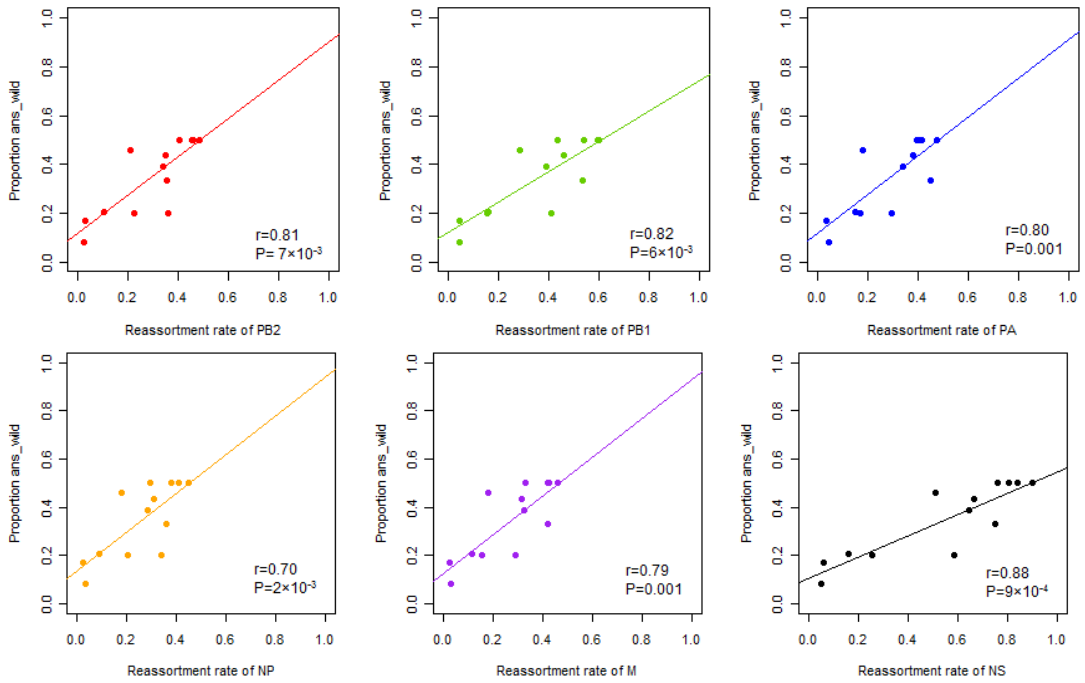


Figure 3.11 Correlation between the proportion of AIV in wild Anseriformes and the reassortment rate per subtype in 6 segments.

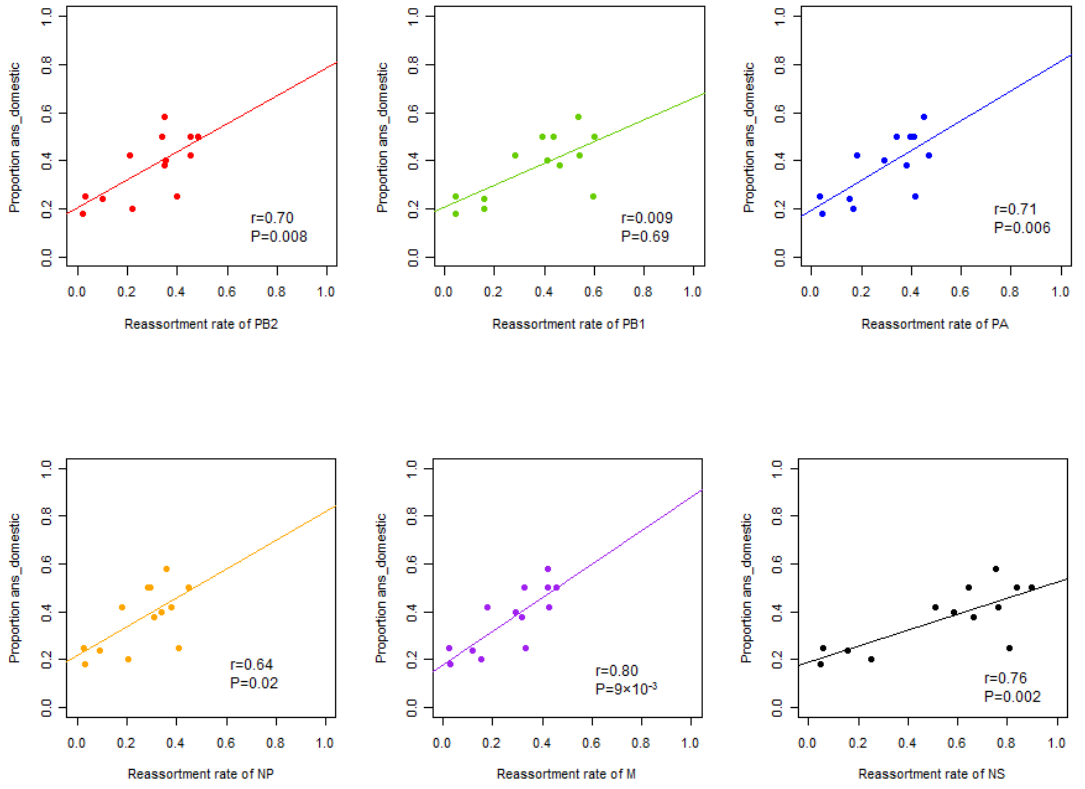


Figure 3.12 Correlation between the proportion of AIV in domestic Anseriformes and reassortment rate per subtype in 6 segments.

On the other hand, subtypes H9N2, H7N3, H7N1 and H5N1 had a high prevalence in domestic birds of the order Galliformes (chickens, turkeys, etc.). There was a strong negative correlation between proportion of AIV in domestic birds of the order Galliformes and rate of reassortment per subtype (PB2: $p = 0.002$, $R = -0.77$; PB1: $p = 0.001$, $R = -0.79$; PA: $p = 0.001$, $R = -0.70$; NP: $p = 0.008$, $R = -0.70$; M: $p = 2 \times 10^{-4}$, $R = -0.84$; NS: $p = 1 \times 10^{-4}$, $R = -0.85$) (Figure 3.8 and Figure 3.13).

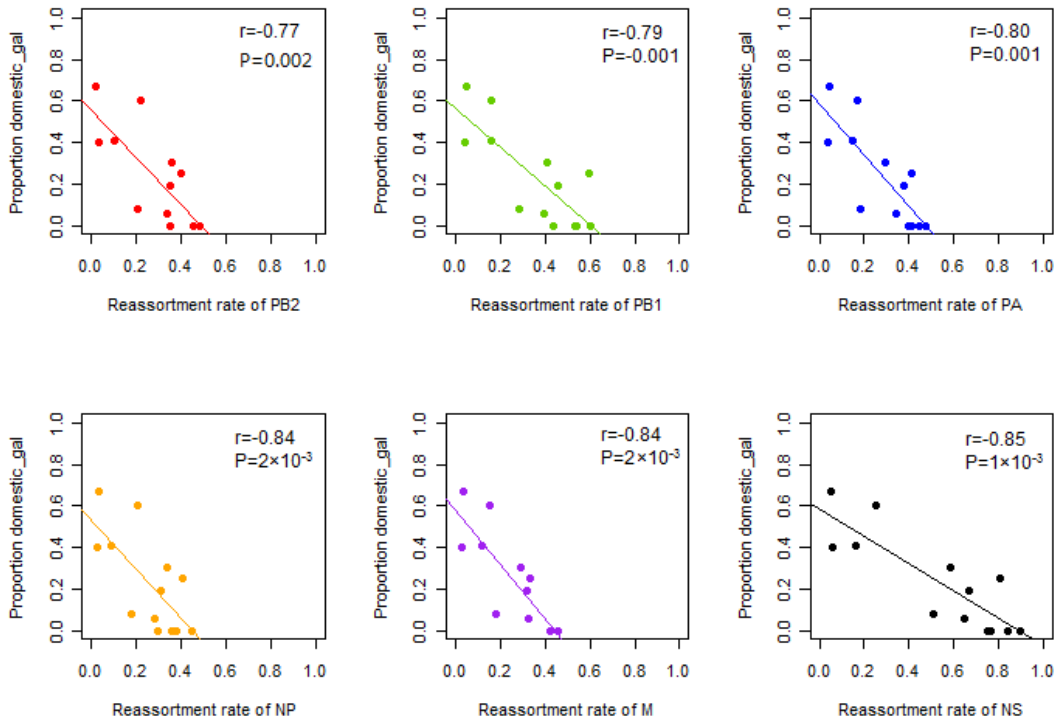


Figure 3.13 Correlation between the proportion of AIV in domestic Galliformes and reassortment rate per subtype in 6 segments.

To also address whether the domestic and wild birds should be analysed separately, I used the same dataset of 344 AIV to generate MCMC trees, using the same discrete trait models but labelling by host (wild or domestic). This showed that wild and domestic AIV do not fall into separate clades but are interspersed in the tree

(Figure 3.14). Therefore, it is not possible to analyse the wild and domestic hosts separately in these data.

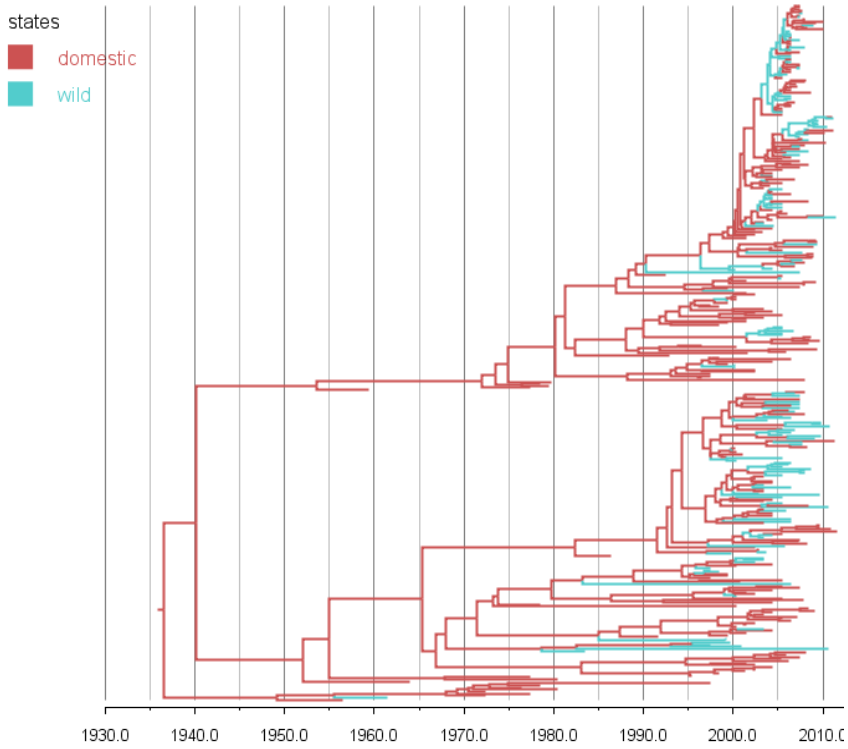


Figure 3.14 Bayesian MCC phylogenies coloured with host types traits.
In this phylogenetic tree for PB2 segment, branches are coloured according to the different host of their descendent nodes (domestic birds: red; wild bird: green).

3.3.5 Correlation of reassortment rate with evolutionary rate

We further tested whether the reassortment rates were correlated with evolutionary rates of AIV at both segment and subtype level. The overall evolutionary rates of Eurasian AIV in different segments were the relaxed clock rates per segment from the MCMC tree (Table 3.5) (Drummond & Rambaut, 2007). No significant difference in evolutionary relaxed clock rate was found among six internal genes of the analysed AIV sequences, with estimated mean rates from 2.27 to 2.85 $\times 10^{-3}$ substitutions per site per year (subs/site/year). In addition, no significant correlation was identified between overall reassortment per segment and evolutionary rate per segment (P = 0.7).

	PB2	PB1	PA	NP	M	NS
segment[a]	**2.85 (2.54,3.19)**	**2.27 (1.96,2.61)**	**2.62 (2.32,2.91)**	**2.65 (2.36,2.95)**	**2.41 (2.07,2.76)**	**2.56 (2.18,2.95)**
subtype[b]						
H1N1	2.63 (1.88,3.38)	2.36 (1.62,3.10)	2.52 (1.41,3.64)	2.70 (1.69,3.72)	2.62 (1.32,3.93)	2.28 (1.46,3.11)
H3N8	2.40 (1.87,2.93)	2.86 (2.12,3.60)	2.38 (1.62,3.13)	2.30 (1.61,2.99)	2.13 (1.33,2.92)	2.20 (1.60,2.81)
H4N6	2.80 (2.42,3.18)	2.59 (1.85,3.33)	2.33 (1.50,3.15)	2.62 (1.73,3.51)	2.28 (1.39,3.17)	2.56 (1.80,3.31)
H5N1	2.93 (2.61,3.25)	2.84 (2.10,3.58)	2.62 (1.85,3.38)	2.60 (1.94,3.26)	2.32 (1.60,3.04)	2.73 (2.30,3.16)
H5N2	2.91 (2.28,3.54)	2.78 (2.04,3.52)	2.39 (1.57,3.21)	2.74 (1.86,3.62)	2.50 (1.57,3.43)	2.45 (1.83,3.07)
H5N3	3.01 (2.29,3.72)	2.44 (1.70,3.18)	2.42 (1.54,3.30)	2.54 (1.66,3.43)	2.68 (1.51,3.85)	2.25 (1.61,2.88)
H6N1	3.20 (2.67,3.74)	2.78 (2.04,3.52)	2.54 (1.71,3.36)	2.53 (1.81,3.24)	2.16 (1.40,2.93)	2.61 (2.03,3.19)
H6N2	2.85 (2.32,3.38)	2.90 (2.16,3.64)	2.61 (1.78,3.45)	2.52 (1.79,3.25)	2.51 (1.59,3.43)	2.51 (1.88,3.14)
H6N5	2.72 (1.97,3.47)	2.99 (2.25,3.73)	2.45 (1.55,3.34)	2.54 (1.68,3.41)	2.18 (1.16,3.20)	2.30 (1.47,3.13)
H7N1	2.77 (2.08,3.45)	2.68 (1.94,3.42)	2.35 (1.47,3.24)	2.57 (1.71,3.44)	2.37 (1.32,3.41)	2.33 (1.57,3.09)
H7N3	2.54 (1.89,3.18)	2.42 (1.68,3.16)	2.08 (1.31,2.85)	2.47 (1.66,3.28)	2.17 (1.19,3.16)	1.92 (1.25,2.59)
H7N7	2.98 (2.15,3.80)	2.55 (1.81,3.29)	2.31 (1.39,3.22)	2.46 (1.59,3.32)	2.06 (1.07,3.05)	2.32 (1.35,3.28)
H9N2	3.11 (2.71,3.52)	2.87 (2.13,3.61)	2.74 (1.92,3.56)	2.72 (2.00,3.45)	2.36 (1.61,3.11)	2.62 (2.18,3.05)

Table 3.5 Evolutionary Rate of internal segment of Eurasian AIV

[a] Estimated overall evolutionary rate ($\times 10^3$ subst/site/year) per internal segment. Mean rate under the uncorrelated log-normal relaxed molecular clock with 95% HPD of all MCMC samples are shown in brackets.

b Estimated evolutionary rate ($\times 10^3$ subst/site/year) per H-N combined subtype in each internal segment. Mean clock rate of certain subtype on MCMC tree nodes with 95% confidential interval (CI) of all MCMC samples are shown in brackets.

The evolutionary rates in different subtypes were estimated by considering the relaxed clock rates corresponding to branches of particular subtypes. The rates per subtype were also similar between different subtypes in the same segment and not significantly different across different segments (Table 3.5). There was no significant correlation between evolutionary rate and reassortment rate per subtype in each internal segment ($P > 0.1$) (Figure 3.15), which is also confirmed by the distribution of correlation coefficients between the evolutionary rates (relaxed clock rates from 1000 random MCMC samples post-burnin) and reassortment rates (rates were taken from 1000 random MCMC samples post-burnin). In addition, there was no indication of any interaction between evolutionary rate and the proportion of domestic birds ($P > 0.5$).

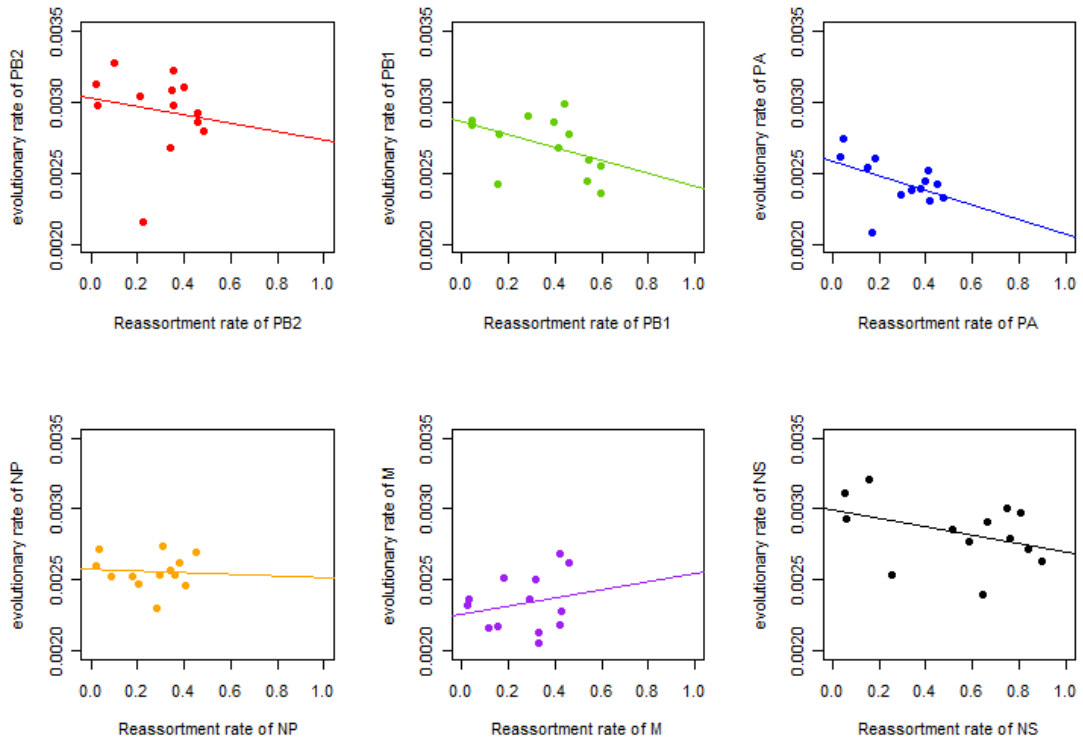


Figure 3.15 Correlation between evolutionary rate and reassortment rate per subtype in 6 segments.
The scatter plot of reassortment rate of HA-NA combined subtype against relaxed clock rate per subtype were represented for 6 internal segments by different colours: PB2 (red); PB1 (green); PA (blue); NP (orange); M1 (purple); NS1 (black). No significant correlation was identified in the six segments ($P > 0.1$).

3.3.6 Correlation of reassortment with selection

I applied the same Bayesian method used in estimating theper trait (subtype) reassortment rate to estimating the dN/dS ratio and the evolutionary rate. The dN/dS ratio of each branch across all the MCMC trees can be estimated using the robust counting methodology (Lemey *et al.*, 2012), and with the Bayesian sampling, the mean and 95% HPD interval per branch accordingly. In addition, by applying a discrete trait model (Lemey *et al.*, 2009) in the same tree phylogeny, the AIV subtypes can be labelled as trait on the tree nodes (and tips), then the branches with the tip and nodes are labelled as the same subtype are selected and counted together as the dN/dS of a certain subtype.

I estimated the total dN/dS ratio for each internal segment as well as the individual ratio per subtype in each segment (Table 3.6). Initial analysis of the coding regions in the five segments PB2, PB1, PA, NP and M1 showed the d_N/d_S ratios were similar to each other with mean values from 0.11 to 0.17, indicating, as expected, that at the amino acid level purifying selection predominates in these internal segments; however, the ratio for the NS segment (mean 0.59) was significantly higher than those of the other segments. This suggested a strong correlation between reassortment rate and d_N/d_S ratio at the segment level ($P = 0.006$, $R = 0.93$). However, this result is entirely dependent on the high value for NS1 and the correlation became negative and non-significant when NS1 was left out ($R = -0.23$, $p = 0.83$). Compared to the evolutionary rate, the dN/dS ratio for each subtype differs noticeably between different subtypes in the same segment and is very different across different segments (Table 3.6). For instance, in PB2, H4N6 had the lowest dN/dS ratio (0.22, 95% HPD: 0.13-0.30) and H6N1 had the highest (0.76, 95% HPD: 0.55-0.96). In NS1, H7N7 had the lowest ratio (0.351, with 95% HPD: 0–0.842) and H7N3 had the highest ratio which was equal to 1, but this difference was not significant given a wide range of statistical uncertainty for these subtypes in the NS1 segment. I found strong negative correlations between mean dN/dS and the reassortment rate per subtype in each of the six segments (PB2: $p = 0.01$, $R = -0.67$; PB1: $p = 0.009$, $R = -0.69$; PA: $P = 2 \times 10^{-4}$, $R = -0.86$; NP: $P = 0.01$, $R = -0.67$; M: $P = 0.002$, $R = -0.78$; NS: $p = 0.001$, $R = -0.80$) (Figure 3.16).

	PB2	PB1	PA	NP	M1	NS1
segment[a]	**0.14 (0.14,0.15)**	**0.12 (0.12,0.13)**	**0.17 (0.17,0.17)**	**0.13 (0.13,0.14)**	**0.11 (0.11,0.12)**	**0.59 (0.15,1.03)**
subtype[b]						
H1N1	0.31 (0.15,0.46)	0.23 (0.09,0.37)	0.30 (0.16,0.43)	0.64 (0.11,0.38)	0.02 (0.02,0.11)	0.49 (0.06,0.91)
H3N8	0.31 (0.19,0.43)	0.25 (0.16,0.34)	0.31 (0.21,0.42)	0.35 (0.14,0.24)	0.15 (0.03,0.28)	0.77 (0.39,1.15)
H4N6	0.22 (0.13,0.30)	0.41 (0.26,0.55)	0.31 (0.19,0.43)	0.47 (0.16,0.31)	0.34 (0.06,0.62)	0.46 (0.15,0.78)
H5N1	0.56 (0.48,0.64)	0.56 (0.48,0.65)	0.92 (0.80,1.04)	0.80 (0.57,0.69)	0.65 (0.48,0.82)	0.97 (0.78,1.15)
H5N2	0.42 (0.29,0.54)	0.29 (0.20,0.37)	0.35 (0.24,0.46)	0.49 (0.24,0.36)	0.23 (0.09,0.37)	0.41 (0.16,0.65)
H5N3	0.36 (0.22,0.49)	0.34 (0.20,0.48)	0.22 (0.13,0.31)	0.51 (0.19,0.35)	0.15 (0.02,0.29)	0.50 (0.15,0.84)
H6N1	0.76 (0.55,0.96)	0.60 (0.45,0.75)	0.57 (0.39,0.74)	0.64 (0.36,0.50)	0.73 (0.38,1.08)	0.83 (0.56,1.11)
H6N2	0.42 (0.29,0.54)	0.27 (0.20,0.35)	0.33 (0.24,0.41)	0.43 (0.21,0.32)	0.50 (0.25,0.75)	0.56 (0.32,0.80)
H6N5	0.53 (0.14,0.92)	0.46 (0.22,0.69)	0.41 (0.20,0.62)	0.77 (0.15,0.46)	0.04 (0.04,0.25)	0.72 (0.12,1.32)
H7N1	0.40 (0.21,0.59)	0.43 (0.19,0.67)	0.37 (0.18,0.56)	0.98 (0.24,0.61)	0.53 (0.08,1.14)	0.76 (0.12,1.40)
H7N3	0.73 (0.08,1.39)	0.57 (0.33,0.80)	0.57 (0.28,0.86)	0.95 (0.17,0.56)	1.05 (0.21,1.89)	1.01 (0.07,2.43)
H7N7	0.40 (0.19,0.61)	0.45 (0.20,0.71)	0.29 (0.16,0.42)	0.52 (0.08,0.30)	0.03 (0.05,0.21)	0.35 (0.03,0.84)
H9N2	0.55 (0.46,0.64)	0.59 (0.52,0.67)	0.64 (0.56,0.72)	0.80 (0.56,0.68)	0.57 (0.40,0.74)	0.92 (0.72,1.12)

Table 3.6 dN/dS ratio of internal segments of Eurasian AIV

[a] Estimated overall dN/dS ratio per internal segment. Mean ratio of each internal segments with 95% (CI) of all MCMC samples shown in brackets.

[b] Estimated dN/dS ratio per H-N combined subtype in each internal segment. Mean dN/dS of each subtype on MCMC tree nodes with 95% CI of all MCMC samples are shown in brackets.

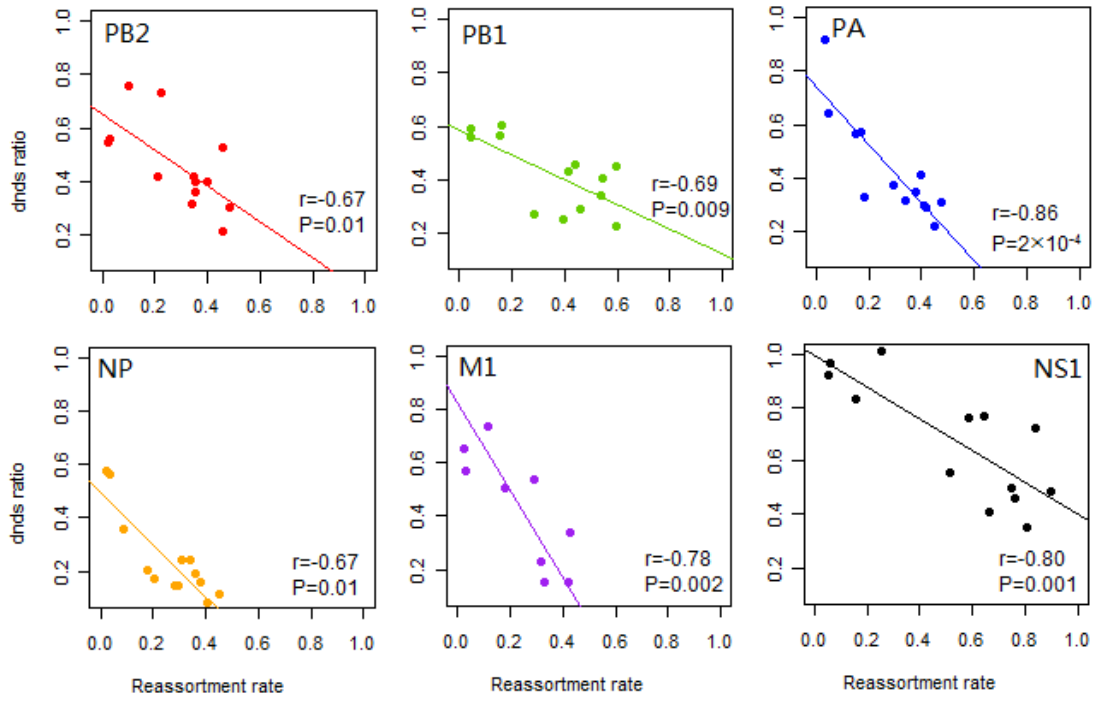


Figure 3.16 Correlation between selection and reassortment rate per subtype in 6 segments.

The scatter plot of reassortment rate of HA-NA combined subtype against dN/dS ratio of 6 internal segments were represented by different colours: PB2 (red); PB1 (green); PA (blue); NP (orange); M1 (purple); NS1 (black).

In order to confirm the negative correlation between the d_N/d_S ratio and inter-subtype reassortment rate seen in BEAST, I used the independent approach of

Single Likelihood Ancestor Counting (SLAC) in the HyPhy package (http://www.datamonkey.org/) to estimate dN/dS per subtype. I found a strong negative correlation between inter-subtype reassortment and d_N/d_S ratio in all six internal segments using SLAC as well (Figure 3.17). SLAC estimates the numbers of codons subject to positive and negative selection and I was thus able to identify the nature of the differences in selection. Table 3.7 showed the number of negative selected sites detected by SLAC for all six internal segments of each subtype, with the correlation coefficient (R) between the number of negatively selected sites and reassortment rate per subtype in each segment and the P-value in the row below. There was no evidence for positive selection contributing to the variation among subtypes, however there was substantial variation in the number of negatively selected sites, which was strongly negatively correlated with reassortment rate with $R < -0.65$, $p \leq 0.01$, for all segments except NS1 for which $R = -0.55$, $p = 0.05$. Therefore the intensity of purifying selection, or selective constraint, on the internal segments is strongly associated with the rate of reassortment.

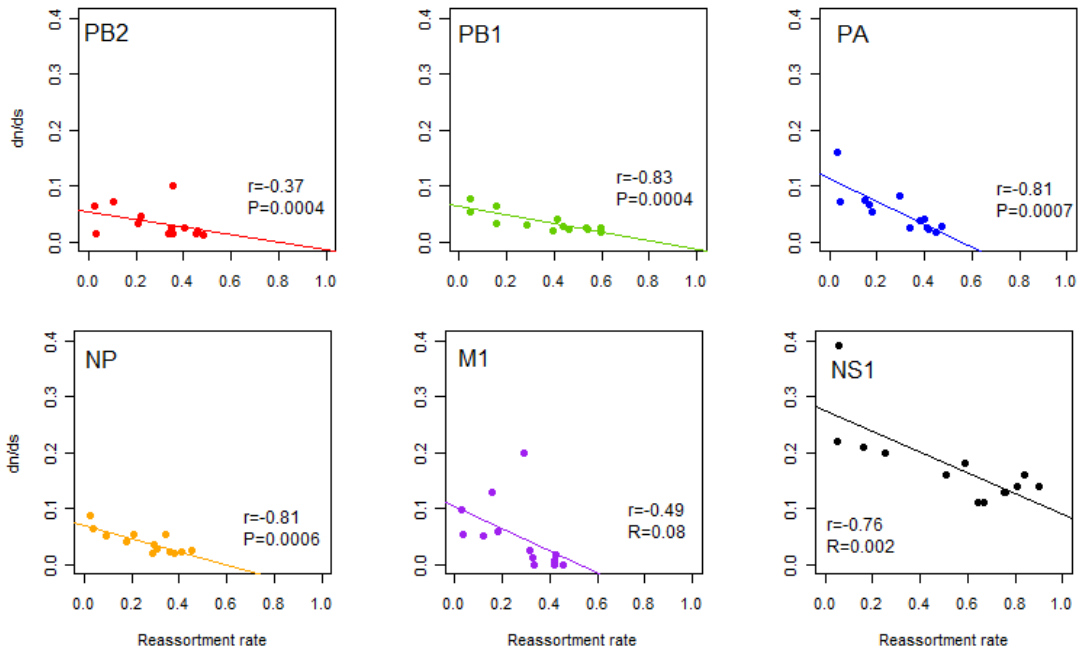


Figure 3.17 Correlation between inter-subtype reassortment rate and dN/dS (using SLAC).

		Number of negative selected sites						Number of sequences
Segment		**PB2**	**PB1**	**PA**	**NP**	**M1**	**NS1**	
	H1N1	53	21	29	23	8	12	24
	H3N8	214	145	183	112	32	33	22
	H4N6	195	115	119	69	17	16	12
	H5N1	337	521	327	246	252	34	234
	H5N2	324	295	209	163	55	46	24
	H5N3	129	87	153	79	42	22	14
	H6N1	336	305	289	183	62	47	66
	H6N2	484	458	374	239	83	75	140
	H6N5	38	38	13	20	1	0	12
	H7N1	4	29	7	1	2	6	40
	H7N3	25	129	26	16	6	10	33
	H7N7	40	66	70	27	11	7	12
	H9N2	492	491	423	266	97	72	114
R		-0.7	-0.77	-0.67	-0.79	-0.69	-0.55	
P-value		0.007	0.002	0.01	0.001	0.008	0.05	

Table 3.7 Number of selective sites of 13 subtypes and the correlation with inter-subtype reassortment rate in each internal segment.

In addition, I also found a strong negative correlation between the proportion of wild birds and the d_N/d_S ratio of different subtypes per segment (PB2: $p = 0.04$, $R = -0.58$; PB1: $p = 0.01$, $R = -0.46$; PA: $p = 0.04$, $R = -0.57$; NP: $p = 0.02$, $R = -0.64$; M: $p = 0.007$, $R = -0.71$; NS: $p = 0.002$, $R = -0.77$). Joint estimates of the partial correlation coefficients among d_N/d_S, host species and reassortment rates were also performed (See Methods 3.2.8). The results showed that reassortment rate was still significantly negatively correlated with d_N/d_S independent of the proportion of wild birds as host in all six internal segments (R: -0.79 to −0.44, $P < 0.01$; Table 3.8 column 5 and 7); and independently positively correlated with proportion of wild birds as host (R:-0.66 to −0.29, $P < 0.07$; Table 3.8 column 8 and 10) with borderline significance, thus both host and selective constraint are associated with reassortment rate.

	R_{AB}	R_{BC}	R_{AC}	$R_{AB.C}$	$t_{AB.C}$	$P_{AB.C}$	$R_{AC.B}$	$t_{AC.B}$	$P_{AC.B}$
PB2	-0.67	0.58	-0.64	-0.47	2.95	0.01	-0.42	2.61	0.02
PB1	-0.69	0.46	-0.74	-0.59	3.87	0.003	-0.66	4.75	0.001
PA	-0.86	0.57	-0.68	-0.79	7.67	0.0001	-0.46	2.83	0.02
NP	-0.67	0.64	-0.64	-0.44	2.75	0.02	-0.37	2.35	0.04
M	-0.78	0.71	-0.62	-0.62	4.19	0.002	0.30	1.34	0.07
NS	-0.80	0.77	-0.73	-0.55	3.51	0.005	-0.29	1.97	0.07

Table 3.8 Partial correlation between reassortment and domestic proportion or dN/dS ratio. A: dNdS; B: reassortment rate; C: proportion of wild bird.

3.4 Discussion

In this chapter, the evolution and reassortment patterns of internal segments of avian influenza viruses have been analysed using a comprehensive avian influenza A virus dataset. Employing Bayesian analysis of phylogenetic discrete traits within the integrated statistical framework in the BEAST software package, I quantified evolutionary parameters including reassortment rate, evolutionary relaxed clock rate and selective constraint. Notably, the quantified pattern of reassortment illustrates how and to what extent the subtypes of AIV interact in the internal segments in bird populations. Firstly, the overall inter-subtype reassortment rates are similar in five segments, while NS has higher rates overall, and secondly, in each segment, reassortment rates of different subtypes showed significant diversity and were asymmetric.

The reassortment rate of the six internal segments of Eurasian AIV is significantly associated with gene diversity, which is probably due to the phylogenetic history and the biological function of the segments during virus infection (Rambaut *et al.*, 2008). Here I identified that NS has a much more diverse phylogeny and higher inter-subtype reassortment rate than the other five segments of the avian influenza viruses of Eurasian lineage. This result is in agreement with those of a previous study on AIV from the wild bird population of North America, which found extensive phylogenetic incongruence in the NS gene (Perez *et al.*, 2008). The NS phylogeny of avian influenza viruses is characterized by a deep divergence between the A and B alleles suggesting that the two alleles are subject to some form of balancing selection

(Dugan *et al.*, 2008; Ludwig *et al.*, 1991; Suarez & Perdue, 1998; Treanor *et al.*, 1989b; Wang *et al.*, 2005).

While I found that reassortment rate was not associated with evolutionary rate, the segment-wide dN/dS ratio per subtype was strongly negatively correlated with the inter-subtype reassortment. The variation in dN/dS ratio was revealed by two independent methods: robust counting in BEAST (Lemey *et al.*, 2012), and SLAC in HyPhy (Kosakovsky Pond & Frost, 2005). The latter gives direct estimates of the numbers of selected sites from which it was seen that the variation between subtypes in d_N/d_S appears to arise from differences in the extent of purifying selection, as indicated by differences in the number of negatively selected sites for each subtype, which itself yielded a strongly negative association with reassortment rate (SLAC identified only 3 positively selected sites; data not shown). This suggests that the overall level of purifying selection, or selective constraint, differs between subtypes: more negatively selected sites indicating higher overall constraints leading to lower dN/dS.

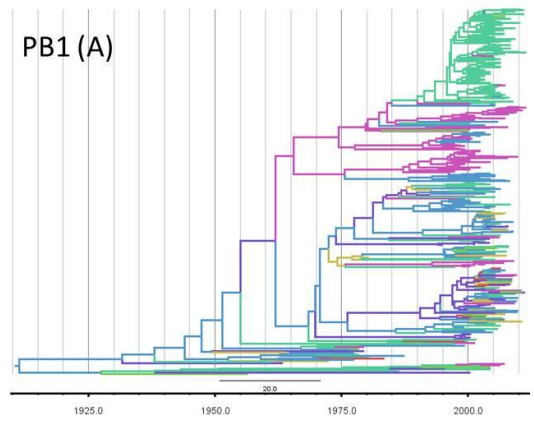
As mentioned above, the evolutionary relaxed clock rates (of the internal gene segments) in Eurasian AIV show little difference among viral subtypes and also between the internal segments. In these results therefore, evolutionary rates were not correlated with inter-subtype reassortment rates and also not affected much by host species. This agrees with the conclusions of a previous study which indicated that the evolutionary rates in wild aquatic birds are not radically lower than those seen in poultry species (Bahl *et al.*, 2009; Chen *et al.*, 2006a). Pathogenicity status might also be an explanation of the lower reassortment rates of H5N1: HPAI H5N1 is lethal to poultry (Cauthen *et al.*, 2000) thus could have had less chance to undergo reassortment in the host due to the short duration of infection of the birds as compared to the less symptomatic infection, typically of longer duration, in low pathogenicity strains. However, the results indicate that pathogenicity status (which is defined in chickens) is not the major factor affecting the overall reassortment rate in all species; of those subtypes with high pathogenicity, such as H5N1, H7N3, H7N1 and H7N7, only H5N1 has low reassortment rate; and there are other AIV subtypes with low pathogenicity that have low reassortment rate, most notably H9N2 (Li *et al.*, 2005).

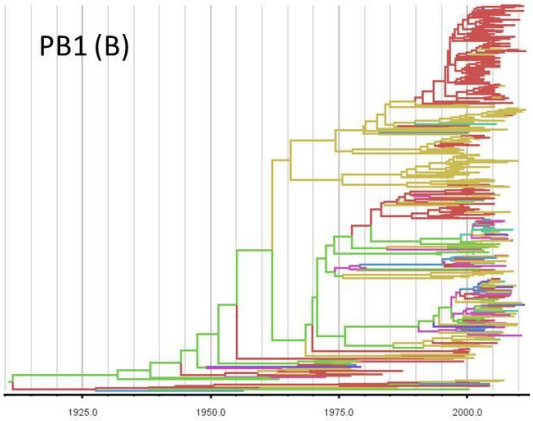
This study showed that the reassortment rates among viral subtypes are correlated with their prevalence in different types of host. High reassortment rates could be associated with high prevalence of certain subtypes in wild birds and particularly among Anseriformes host species. In contrast, the lower reassortment rate of certain subtypes could be explained by their high prevalence in domestic birds (mainly in Galliformes). Wild bird and domestic bird populations represent two different ecosystems: natural and artificial respectively, in which influenza viruses face different ecologic constraints such as host population structure density, and opportunities for virus transmission (Lebarbenchon *et al.*, 2010). For instance, transmission by faces provides a way for wild birds to spread viruses to other wild birds as they migrate through an area (Webster, 2002), while AIV in domestic birds, especially domestic birds of the order Galliformes may exist in a more isolated population (within a physically confined flock) than might occur in wild birds, thus with less possibility of reassortment since the potential for dual infections is reduced.

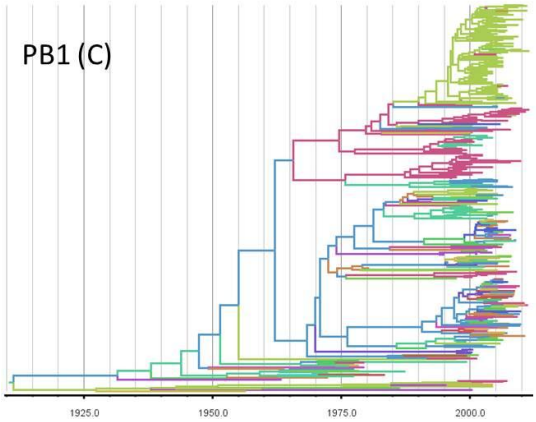
In comparison to the AIV in domestic Galliformes, AIV at high prevalence in the domestic duck (the H6 subtype in this study) has a higher reassortment rate. Domestic ducks or geese farming occurs in high-density but free-range settings in Asia (Gu *et al.*, 2011). This farming style creates an environment in which migratory birds and domestic ducks are in close contact, and provide more opportunities for AIV to reassort (Deng *et al.*, 2013). In addition, different responses to virus infection may also influence the opportunity of AIV reassortment in certain host species. Avian influenza virus in chickens (Galliformes) often leads to severe respiratory symptoms and high mortality rates of up to 100%. In contrast, AIV infection generally causes no major clinical signs and lack of immunological cross-protection in Anseriformes (Kida *et al.*, 1980). Different influenza virus subtypes can also infect wild ducks concomitantly, creating the opportunity for genetic mixing (Sharp *et al.*, 1997). Those subtypes showing the high reassortment rates in this study, especially H1N1, H3N8 and H4N6 represent the most common subtype combinations that are found in wild ducks during periods of peak prevalence and have a maximum detectable subtype diversity in the late summer and fall (Olsen *et al.*, 2006; Perez *et al.*, 2008).

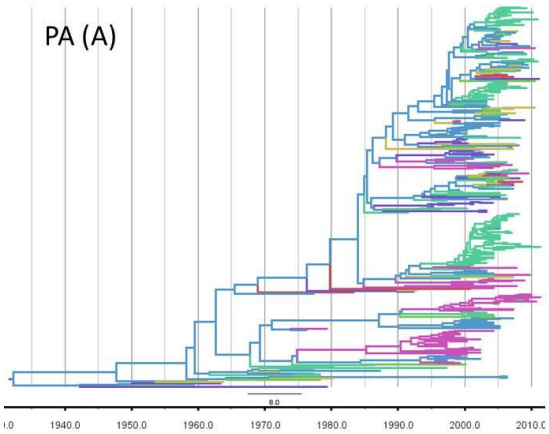
In summary, I quantified the inter-subtype reassortment rates using discrete trait modelling upon time resolved phylogenies with respect to all six internal segments of Eurasian AIV. The overall reassortment rates for the external protein coding segments with respect to internal protein coding segments were highly correlated with the internal segment genetic diversity (divergence time) and negatively correlated with d_N/d_S ratio and selective constraint. In addition, I found that different subtypes are characterized by reassortment rates that are strongly dependent on host type (wild *v.* domestic). These results are consistent with the hypothesis that lack of immunological cross-protection at the subtype level in avian species allows frequent mixed infections in wild bird populations and provide novel insights into the evolutionary dynamics of avian influenza viruses.

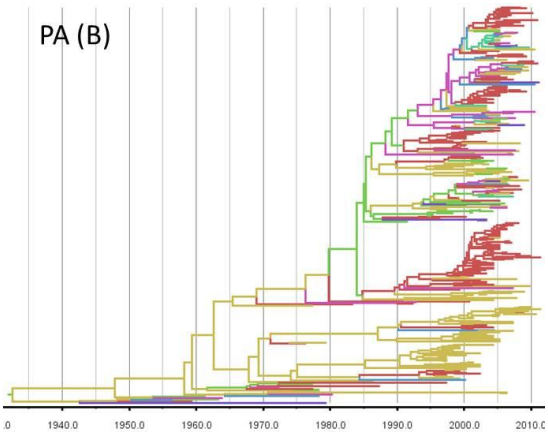
Appendix

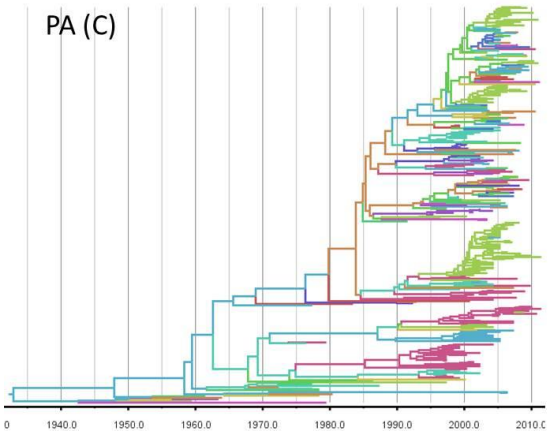

PB1 (A)
20.0
1925.0
1950.0
1975.0
2000.0

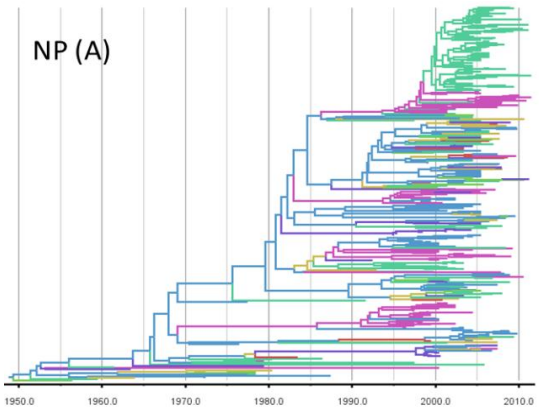

PB1 (B)
1925.0
1950.0
1975.0
2000.0

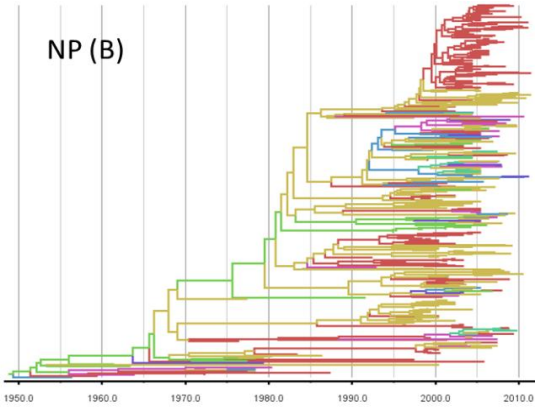

PB1 (C)
1925.0
1950.0
1975.0
2000.0

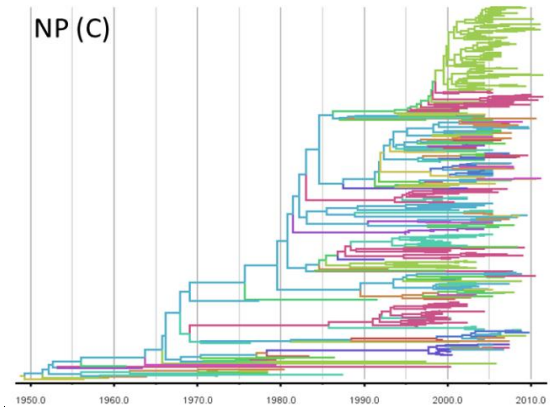

PA (A)
1940.0
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0

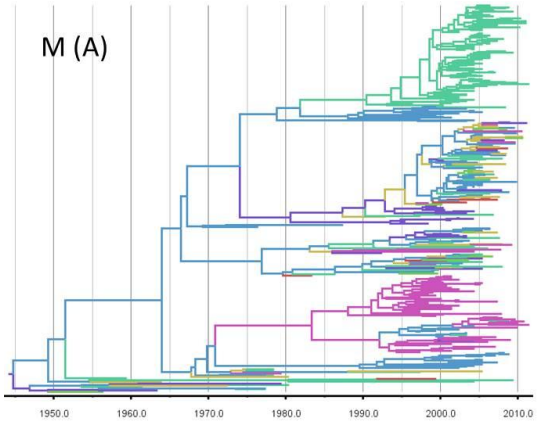

PA (B)
1940.0
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0

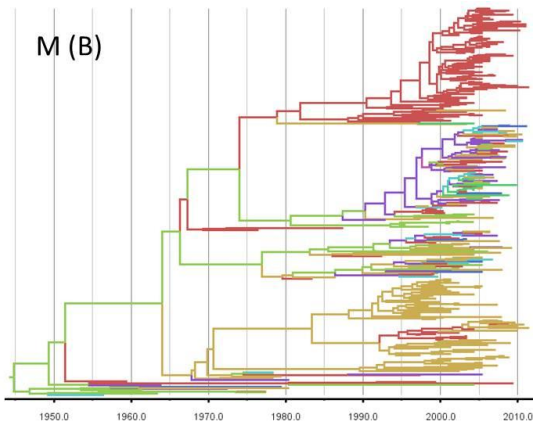

PA (C)
1940.0
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0

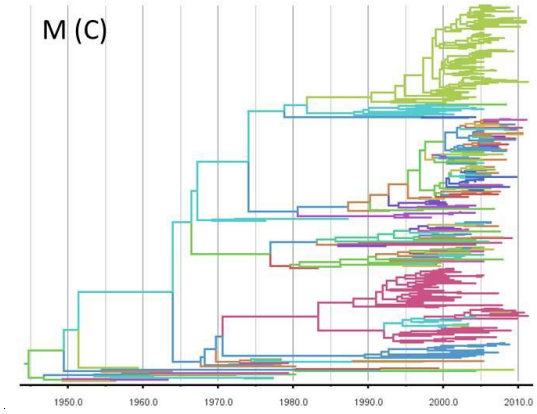

NP (A)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0


NP (B)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0


NP (C)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0


M (A)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0


M (B)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0


M (C)
1950.0
1960.0
1970.0
1980.0
1990.0
2000.0
2010.0

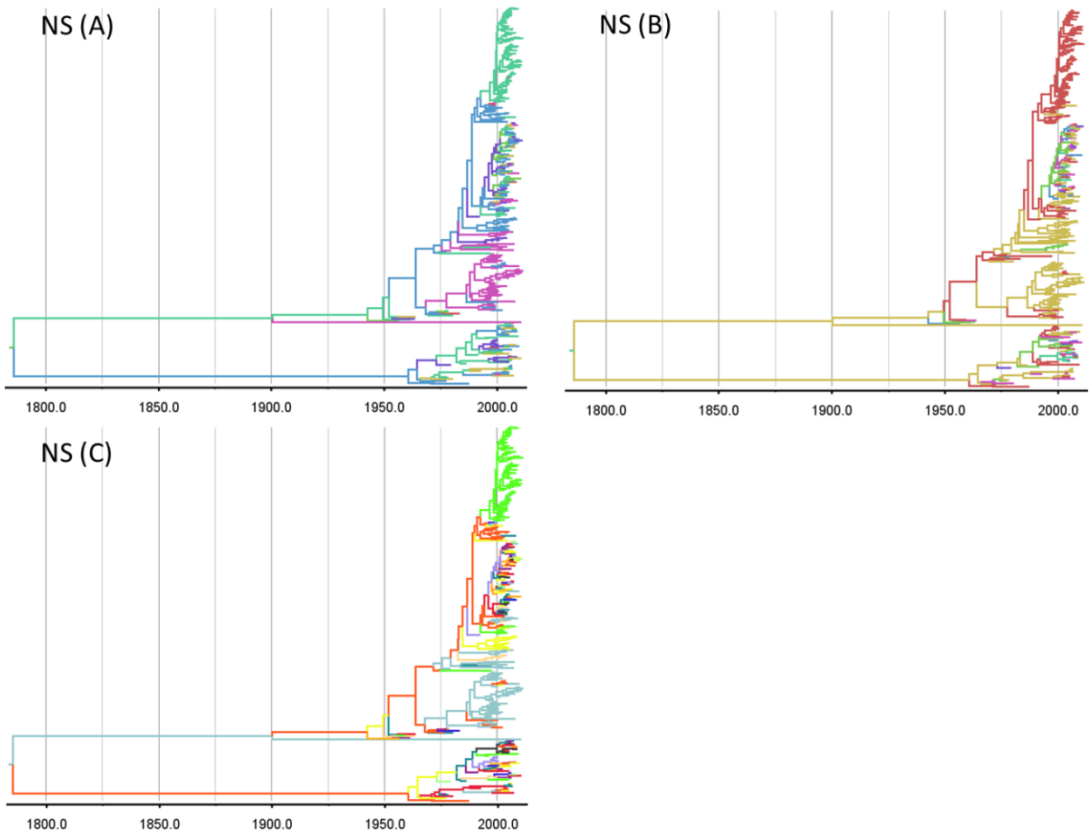


Figure A 3.1 Bayesian MCC phylogenies for the 5 internal segments (PB, PA, NP, M and NS)
(A): HA subtype, (B): NA subtype, (C): Combined HA-NA subtype of Eurasian Avian influenza viruses. Branches were coloured according to the different subtypes of their descendent nodes. The colour markers are the same as Figure 3.4.

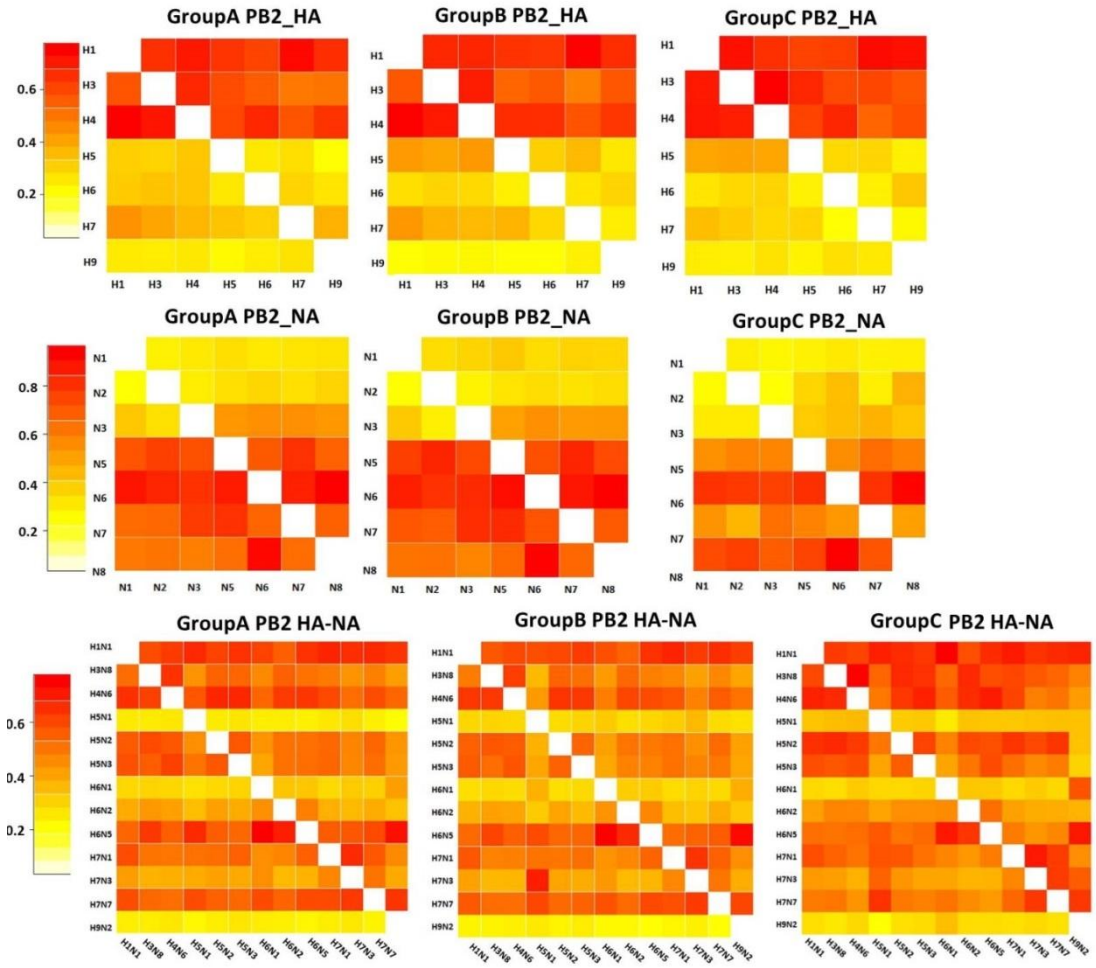


Figure A 3.2 Heat map of inter-subtype reassortment rates with different sample sizes
The reassortment rates of HA, NA and combined HA-NA subtypes of PB2 segment with the AIV datasets of Group A, B and C in Table 3.1. Subtypes labelled on left side are as donors, subtypes labelled in the bottom are as recipients. The values of rates from low to high are represented colours from yellow to red. The scale of rates is on the left hand side of each plot.

				HA			
PB2	H1	H3	H4	H5	H6	H7	H9
H1		0.55	0.51	0.46	0.45	0.50	0.33
H3	0.34		0.42	0.42	0.40	0.26	0.24
H4	0.53	0.63		0.39	0.46	0.32	0.30
H5	0.08	0.09	0.11		0.04	0.03	0.02
H6	0.08	0.15	0.10	0.05		0.05	0.08
H7	0.22	0.28	0.16	0.16	0.11		0.05
H9	0.02	0.02	0.02	0.01	0.03	0.02	
PB1	H1	H3	H4	H5	H6	H7	H9
H1		0.63	0.62	0.40	0.68	0.59	0.30
H3	0.35		0.47	0.25	0.45	0.42	0.14
H4	0.46	0.64		0.37	0.62	0.50	0.19
H5	0.05	0.04	0.04		0.02	0.05	0.03
H6	0.11	0.13	0.13	0.10		0.07	0.03
H7	0.30	0.31	0.31	0.12	0.24		0.12
H9	0.05	0.05	0.05	0.03	0.06	0.04	
PA	H1	H3	H4	H5	H6	H7	H9
H1		0.40	0.39	0.28	0.50	0.37	0.41
H3	0.28		0.56	0.26	0.51	0.30	0.23
H4	0.32	0.83		0.56	0.81	0.36	0.28
H5	0.04	0.05	0.06		0.05	0.04	0.03
H6	0.06	0.20	0.23	0.09		0.05	0.04
H7	0.14	0.17	0.17	0.07	0.19		0.08
H9	0.04	0.04	0.04	0.03	0.05	0.03	
NP	H1	H3	H4	H5	H6	H7	H9
H1		0.46	0.49	0.38	0.52	0.44	0.42
H3	0.31		0.45	0.23	0.29	0.28	0.14
H4	0.40	0.58		0.42	0.42	0.40	0.25
H5	0.02	0.02	0.03		0.02	0.03	0.02
H6	0.08	0.16	0.14	0.05		0.07	0.02
H7	0.23	0.21	0.24	0.10	0.23		0.10
H9	0.05	0.04	0.04	0.04	0.03	0.03	
M	H1	H3	H4	H5	H6	H7	H9
H1		0.71	0.69	0.42	0.62	0.49	0.31
H3	0.38		0.75	0.33	0.58	0.26	0.16
H4	0.48	0.88		0.46	0.76	0.36	0.24
H5	0.03	0.03	0.03		0.03	0.03	0.02
H6	0.13	0.22	0.19	0.06		0.05	0.03
H7	0.19	0.19	0.18	0.09	0.15		0.05
H9	0.03	0.03	0.03	0.02	0.03	0.02	
NS	H1	H3	H4	H5	H6	H7	H9
H1		1.62	1.50	0.81	1.28	0.98	0.69
H3	0.92		1.14	0.46	0.79	0.55	0.32
H4	1.06	1.20		0.67	1.06	0.58	0.38
H5	0.09	0.09	0.09		0.05	0.04	0.03
H6	0.24	0.31	0.28	0.08		0.14	0.06
H7	0.38	0.43	0.43	0.13	0.23		0.09
H9	0.05	0.04	0.04	0.03	0.04	0.04	

(Continued on next page)

NA							
PB2	N1	N2	N3	N5	N6	N7	N8
N1		0.08	0.04	0.07	0.06	0.05	0.08
N2	0.05		0.03	0.06	0.13	0.06	0.13
N3	0.18	0.14		0.27	0.39	0.27	0.37
N5	0.44	0.67	0.56		0.58	0.52	0.58
N6	0.58	0.76	0.66	0.63		0.56	0.97
N7	0.39	0.54	0.63	0.57	0.53		0.59
N8	0.36	0.50	0.40	0.39	0.87	0.36	
PB1	N1	N2	N3	N5	N6	N7	N8
N1		0.06	0.05	0.04	0.09	0.09	0.06
N2	0.07		0.06	0.04	0.06	0.10	0.06
N3	0.18	0.22		0.25	0.35	0.36	0.31
N5	0.30	0.37	0.41		0.62	0.56	0.64
N6	0.38	0.53	0.53	0.53		0.74	0.79
N7	0.82	1.15	0.76	0.59	0.86		0.87
N8	0.25	0.33	0.41	0.41	0.58	0.56	
PA	N1	N2	N3	N5	N6	N7	N8
N1		0.09	0.04	0.08	0.08	0.06	0.07
N2	0.06		0.10	0.11	0.14	0.10	0.13
N3	0.17	0.23		0.30	0.35	0.30	0.36
N5	0.57	0.82	0.38		0.81	0.45	0.79
N6	0.59	0.76	0.45	0.89		0.49	0.93
N7	0.45	0.48	0.55	0.63	0.69		0.74
N8	0.39	0.44	0.34	0.66	0.72	0.39	
NP	N1	N2	N3	N5	N6	N7	N8
N1		0.05	0.03	0.04	0.04	0.04	0.06
N2	0.09		0.06	0.06	0.08	0.04	0.07
N3	0.12	0.20		0.19	0.26	0.18	0.22
N5	0.25	0.45	0.23		0.44	0.34	0.42
N6	0.26	0.44	0.31	0.43		0.37	0.44
N7	0.26	0.36	0.33	0.42	0.50		0.46
N8	0.18	0.37	0.22	0.29	0.34	0.26	
M	N1	N2	N3	N5	N6	N7	N8
N1		0.04	0.04	0.03	0.03	0.03	0.03
N2	0.04		0.43	0.69	0.09	0.06	0.11
N3	0.13	0.15		0.25	0.31	0.22	0.31
N5	0.42	0.52	0.42		0.76	0.36	0.78
N6	0.45	0.58	0.43	0.69		0.58	0.93
N7	0.26	0.27	0.35	0.40	0.44		0.43
N8	0.32	0.46	0.34	0.54	0.71	0.25	
NS	N1	N2	N3	N5	N6	N7	N8
N1		0.08	0.04	0.05	0.04	0.04	0.05
N2	0.05		0.13	0.16	0.17	0.13	0.20
N3	0.22	0.36		0.57	0.77	0.48	0.75
N5	0.64	0.82	1.06		1.14	0.92	1.13
N6	0.53	0.97	1.12	0.91		0.84	1.45
N7	0.64	0.81	1.01	0.98	1.05		1.12
N8	0.42	0.83	0.85	0.77	1.17	0.69	

(Continued on next page)

HA-NA													
PB2	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.64	0.62	0.31	0.67	0.69	0.25	0.47	0.48	0.52	0.28	0.53	0.34
H3N8	0.40		0.65	0.16	0.55	0.50	0.16	0.46	0.33	0.30	0.14	0.31	0.18
H4N6	0.57	0.68		0.24	0.78	0.73	0.20	0.58	0.47	0.42	0.19	0.44	0.26
H5N1	0.03	0.03	0.03		0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
H5N2	0.45	0.65	0.52	0.17		0.55	0.15	0.41	0.33	0.35	0.16	0.37	0.19
H5N3	0.48	0.57	0.60	0.20	0.59		0.12	0.41	0.34	0.35	0.15	0.36	0.18
H6N1	0.08	0.12	0.09	0.05	0.09	0.09		0.15	0.13	0.07	0.06	0.11	0.18
H6N2	0.21	0.35	0.27	0.10	0.32	0.24	0.15		0.30	0.17	0.08	0.20	0.12
H6N5	0.42	0.72	0.49	0.31	0.57	0.49	0.33	0.67		0.38	0.22	0.44	0.39
H7N1	0.48	0.49	0.41	0.21	0.52	0.52	0.15	0.35	0.37		0.28	0.41	0.21
H7N3	0.25	0.25	0.22	0.13	0.31	0.25	0.15	0.19	0.18	0.27		0.34	0.16
H7N7	0.48	0.54	0.44	0.26	0.56	0.50	0.24	0.45	0.42	0.42	0.26		0.33
H9N2	0.02	0.02	0.03	0.02	0.02	0.02	0.04	0.02	0.02	0.02	0.02	0.02	
PB1	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.78	0.80	0.27	0.75	0.87	0.47	0.85	0.61	0.58	0.24	0.73	0.24
H3N8	0.45		0.56	0.12	0.65	0.58	0.31	0.63	0.36	0.30	0.13	0.47	0.15
H4N6	0.61	0.85		0.20	0.80	0.81	0.47	0.85	0.48	0.39	0.19	0.66	0.20
H5N1	0.05	0.04	0.04		0.04	0.05	0.05	0.04	0.06	0.06	0.02	0.05	0.03
H5N2	0.53	0.68	0.67	0.15		0.72	0.37	0.76	0.38	0.31	0.16	0.58	0.22
H5N3	0.62	0.75	0.75	0.18	0.87		0.47	0.86	0.49	0.37	0.17	0.70	0.19
H6N1	0.13	0.15	0.20	0.07	0.25	0.20		0.31	0.13	0.11	0.07	0.21	0.08
H6N2	0.30	0.44	0.44	0.07	0.57	0.45	0.28		0.22	0.16	0.07	0.33	0.10
H6N5	0.48	0.59	0.57	0.22	0.55	0.53	0.41	0.57		0.39	0.22	0.49	0.24
H7N1	0.54	0.51	0.52	0.22	0.47	0.54	0.34	0.52	0.40		0.20	0.45	0.22
H7N3	0.12	0.14	0.13	0.38	0.15	0.13	0.15	0.15	0.10	0.12		0.15	0.17
H7N7	0.55	0.67	0.69	0.51	0.80	0.70	0.49	0.88	0.51	0.42	0.36		0.60
H9N2	0.05	0.05	0.05	0.03	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.05	
PA	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.47	0.48	0.27	0.51	0.52	0.37	0.37	0.48	0.39	0.26	0.42	0.36
H3N8	0.30		0.58	0.12	0.56	0.60	0.23	0.39	0.42	0.21	0.11	0.39	0.15
H4N6	0.33	0.72		0.25	0.86	0.76	0.40	0.51	0.74	0.26	0.15	0.42	0.26
H5N1	0.03	0.03	0.03		0.03	0.03	0.03	0.02	0.03	0.04	0.04	0.03	0.07
H5N2	0.27	0.51	0.66	0.22		0.68	0.32	0.40	0.65	0.22	0.12	0.30	0.18

H5N3	0.34	0.62	0.68	0.27	0.86		0.37	0.42	0.72	0.27	0.16	0.42	0.25
H6N1	0.16	0.20	0.21	0.06	0.17	0.19		0.16	0.16	0.13	0.09	0.15	0.12
H6N2	0.12	0.35	0.31	0.06	0.23	0.24	0.15		0.29	0.12	0.06	0.19	0.06
H6N5	0.27	0.58	0.54	0.24	0.48	0.44	0.45	0.88		0.22	0.16	0.30	0.19
H7N1	0.31	0.37	0.35	0.15	0.32	0.35	0.27	0.30	0.33		0.23	0.32	0.23
H7N3	0.16	0.17	0.18	0.13	0.17	0.17	0.15	0.17	0.15	0.19		0.15	0.21
H7N7	0.41	0.55	0.55	0.20	0.54	0.61	0.33	0.40	0.46	0.40	0.23		0.29
H9N2	0.04	0.04	0.05	0.03	0.05	0.05	0.06	0.05	0.05	0.04	0.03	0.04	

NP	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.51	0.63	0.31	0.59	0.49	0.35	0.48	0.52	0.37	0.24	0.54	0.38
H3N8	0.33		0.42	0.10	0.48	0.50	0.16	0.40	0.29	0.22	0.11	0.30	0.12
H4N6	0.43	0.53		0.16	0.56	0.50	0.24	0.56	0.53	0.27	0.14	0.41	0.21
H5N1	0.03	0.02	0.02		0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.03	0.03
H5N2	0.40	0.44	0.45	0.22		0.50	0.14	0.33	0.27	0.21	0.12	0.33	0.29
H5N3	0.42	0.49	0.46	0.27	0.62		0.20	0.36	0.29	0.26	0.16	0.41	0.38
H6N1	0.08	0.11	0.09	0.04	0.08	0.08		0.17	0.14	0.09	0.08	0.08	0.05
H6N2	0.18	0.36	0.27	0.05	0.28	0.25	0.11		0.24	0.13	0.06	0.16	0.05
H6N5	0.29	0.42	0.42	0.12	0.30	0.28	0.25	0.69		0.21	0.15	0.28	0.13
H7N1	0.34	0.32	0.40	0.29	0.35	0.36	0.30	0.36	0.38		0.20	0.36	0.41
H7N3	0.20	0.21	0.20	0.16	0.20	0.19	0.27	0.24	0.18	0.19		0.20	0.22
H7N7	0.49	0.45	0.57	0.28	0.51	0.43	0.29	0.46	0.47	0.32	0.23		0.41
H9N2	0.04	0.04	0.04	0.03	0.03	0.04	0.03	0.03	0.04	0.03	0.03	0.03	

M	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.56	0.61	0.26	0.63	0.57	0.46	0.47	0.53	0.39	0.27	0.45	0.31
H3N8	0.33		0.61	0.10	0.52	0.57	0.23	0.42	0.45	0.20	0.10	0.24	0.18
H4N6	0.42	0.76		0.16	0.60	0.62	0.37	0.51	0.66	0.27	0.16	0.31	0.25
H5N1	0.03	0.02	0.02		0.02	0.03	0.02	0.03	0.03	0.03	0.03	0.03	0.03
H5N2	0.35	0.51	0.54	0.11		0.47	0.27	0.37	0.45	0.19	0.11	0.27	0.14
H5N3	0.47	0.71	0.72	0.16	0.62		0.32	0.48	0.52	0.30	0.16	0.35	0.23
H6N1	0.13	0.15	0.15	0.06	0.14	0.16		0.15	0.17	0.07	0.07	0.11	0.06
H6N2	0.17	0.36	0.30	0.05	0.30	0.29	0.14		0.21	0.09	0.06	0.13	0.08
H6N5	0.39	0.62	0.63	0.23	0.51	0.51	0.46	0.62		0.27	0.21	0.34	0.27
H7N1	0.36	0.35	0.35	0.18	0.39	0.36	0.25	0.27	0.28		0.22	0.31	0.20
H7N3	0.16	0.17	0.17	0.17	0.18	0.17	0.19	0.13	0.15	0.13		0.14	0.10
H7N7	0.35	0.38	0.39	0.23	0.42	0.39	0.33	0.34	0.37	0.31	0.24		0.24

H9N2	0.03	0.03	0.03	0.02	0.03	0.04	0.04	0.03	0.03	0.03	0.02	0.03	
NS	H1N1	H3N8	H4N6	H5N1	H5N2	H5N3	H6N1	H6N2	H6N5	H7N1	H7N3	H7N7	H9N2
H1N1		0.95	0.94	0.15	0.92	0.90	0.65	0.99	0.81	0.71	0.37	0.82	0.68
H3N8	0.53		0.88	0.05	0.97	0.82	0.19	0.92	0.61	0.41	0.14	0.49	0.18
H4N6	0.66	1.12		0.07	1.08	1.06	0.26	0.98	0.75	0.50	0.20	0.66	0.26
H5N1	0.17	0.12	0.10		0.11	0.10	0.17	0.11	0.12	0.13	0.09	0.12	0.17
H5N2	0.56	1.01	0.90	0.05		1.00	0.20	0.93	0.62	0.42	0.16	0.53	0.17
H5N3	0.67	1.07	1.10	0.09	1.11		0.25	0.96	0.74	0.48	0.19	0.62	0.23
H6N1	0.16	0.13	0.12	0.05	0.12	0.11		0.14	0.14	0.12	0.09	0.15	0.12
H6N2	0.41	0.71	0.62	0.05	0.76	0.62	0.17		0.50	0.30	0.12	0.37	0.15
H6N5	0.76	1.00	0.94	0.11	0.93	0.94	0.43	0.94		0.63	0.32	0.73	0.39
H7N1	0.53	0.62	0.66	0.11	0.65	0.66	0.29	0.58	0.53		0.27	0.54	0.31
H7N3	0.30	0.26	0.25	0.11	0.26	0.26	0.16	0.26	0.25	0.23		0.27	0.13
H7N7	0.74	0.86	0.85	0.14	0.87	0.77	0.47	0.82	0.76	0.67	0.37		0.47
H9N2	0.05	0.04	0.04	0.03	0.06	0.04	0.04	0.04	0.04	0.05	0.04	0.04	

Table A 3.1 Matrix of reassortment rate between pairs of subtypes (HA, NA and HA-NA combined) of 6 internal segments.
Reassortment rates in this table are represented as heat map in Figure 3.5. Both forward and reverse rates are shown, with subtypes labelled on the left side acting as donors and subtypes labelled at the bottom acting as recipients. The values of the mean reassortment rates are the exchanges pre site per year. The transition rates with BF<3 (the estimated indicator for HA/NA subtype < 0.3, for HA-NA subtype< 0.2) are labelled in grey.

Isolate name	Subtype	Host species[a]	Date of collection	Decimal date
A/mallard/Marquenterre/Z237/1983	H1N1	ans	1983	1983.452
A/mallard/Bavaria/185-8/2008	H1N1	ans	22/09/2008	2008.723
A/goose/Italy/296426/2003	H1N1	ans	2003	2003.452
A/duck/Italy/281904/2006	H1N1	ans	2006	2006.452
A/duck/Italy/69238/2007	H1N1	ans	2007	2007.452
A/duck/Shimane/188/1999	H1N1	ans	1999	1999.452
A/commonteal/Netherlands/10/2000	H1N1	ans	23/11/2000	2000.893
A/Anasplatyrhynchos/Belgium/12827/2007	H3N8	ans	18/09/2007	2007.712
A/duck/Nanchang/1681/1992	H3N8	ans	01/12/1992	1992.915
A/duck/Beijing/40/04	H3N8	ans	2004	2004.452
A/duck/Beijing/61/05	H3N8	ans	2005	2005.452
A/mallard/Sanjiang/90/2007	H3N8	ans	2007	2007.452
A/mallard/CzechRepublic/13577-24K/2010	H3N8	ans	16/09/2010	2010.707
A/duck/Hokkaido/8/1980	H3N8	ans	1980	1980.452
A/mallard/Netherlands/3/2005	H3N8	ans	2005	2005.452
A/commoneider/Netherlands/1/2006	H3N8	ans	2006	2006.452
A/duck/Chabarovsk/1610/1972	H3N8	ans	1972	1972.452
A/gadwall/Altai/1326/2007	H3N8	ans	2007/09/	2007.704
A/mallard/Sweden/50/2002	H3N8	ans	14/11/2002	2002.868
A/duck/Ukraine/1/1963	H3N8	ans	12/12/1963	1963.945
A/duck/Vietnam/G119/2006	H3N8	ans	2006/11/	2006.871
A/duck/Zambia/04/2008	H3N8	ans	2008/06/	2008.452
A/goose/Zambia/06/2008	H3N8	ans	2008/07/	2008.534
A/turnstone/Netherlands/1/2007	H3N8	cha	2007	2007.452
A/chicken/Vietnam/G14/2008	H3N8	gal	2008/01/	2008.038
A/duck/Nanchang/4-165/2000	H4N6	ans	12/04/2000	2000.277
A/mallard/ZhaLong/88/2004	H4N6	ans	08/08/2004	2004.6
A/mallard/Yanchen/2005	H4N6	ans	2005/10/	2005.786
A/duck/CzechRepublic/1/1956	H4N6	ans	11/06/1956	1956.441
A/mallard/CzechRepublic/13579-84K/2010	H4N6	ans	13/09/2010	2010.699
A/duck/HongKong/365/1978	H4N6	ans	1978	1978.452
A/duck/Shiga/8/2004	H4N6	ans	2004	2004.452
A/mallard/Netherlands/1/1999	H4N6	ans	04/10/1999	1999.756
A/mallard/PT/35910-2/2006	H4N6	ans	30/10/2006	2006.827
A/mallard/Sweden/62/2003	H4N6	ans	05/09/2003	2003.677
A/goose/Zambia/07/2008	H4N6	ans	2008/09/	2008.704
A/dunlin/Sweden/1/2005	H4N6	cha	2005	2005.452
A/crestedeagle/Belgium/01/2004	H5N1	acc	2004	2004.452
A/mountainhawk-eagle/Kumamoto/1/07	H5N1	acc	2007	2007.452
A/duck/Cambodia/D14AL/2006	H5N1	ans	04/02/2006	2006.093
A/Goose/Guangdong/1/96	H5N1	ans	1996	1996.452

A/Duck/Anyang/AVL-1/2001	H5N1	ans	2001	2001.452
A/duck/Zhejiang/bj/2002	H5N1	ans	2002	2002.452
A/duck/China/E319-2/03	H5N1	ans	2003	2003.452
A/duck/Guangdong/173/04	H5N1	ans	2004	2004.452
A/domesticgreen-wingedteal/Hunan/79/2005	H5N1	ans	2005	2005.452
A/wildduck/Liaoning/8/2006	H5N1	ans	2006	2006.452
A/duck/Hunan/11/2007	H5N1	ans	2007	2007.452
A/duck/Hunan/3/2007	H5N1	ans	2007/10/	2007.786
A/duck/EasternChina/108/2008	H5N1	ans	15/12/2008	2008.953
A/duck/Zhejiang/2244/2011	H5N1	ans	2011/02/	2011.123
A/cygnusolor/Croatia/1/2005	H5N1	ans	2005	2005.452
A/duck/Egypt/2253-3/2006	H5N1	ans	2006	2006.452
A/duck/Egypt/D2br10/2007	H5N1	ans	2007/01/	2007.038
A/duck/Egypt/0871/2008	H5N1	ans	20/02/2008	2008.137
A/duck/France/080036/2008	H5N1	ans	2008/01/	2008.038
A/swan/Germany/R65/2006	H5N1	ans	2006/02/	2006.123
A/domesticduck/Germany/R1779/2007	H5N1	ans	2007	2007.452
A/goose/HongKong/485.3/2000	H5N1	ans	2000	2000.452
A/Duck/HongKong/380.5/2001	H5N1	ans	2001	2001.452
A/muteswan/Hungary/4571/2006	H5N1	ans	16/02/2006	2006.126
A/goose/Tripura/103596/2008	H5N1	ans	01/04/2008	2008.247
A/duck/EastJava/UT1046/2004	H5N1	ans	12/03/2004	2004.192
A/Cygnuscygnus/Iran/754/2006	H5N1	ans	2006	2006.452
A/mallard/Italy/3401/2005	H5N1	ans	2005	2005.452
A/mallard/Italy/835/2006	H5N1	ans	2006	2006.452
A/duck/Yokohama/aq10/2003	H5N1	ans	2003	2003.452
A/whooperswan/Akita/1/2008	H5N1	ans	2008	2008.452
A/mallard/Hokkaido/24/2009	H5N1	ans	2009	2009.452
A/duck/Fukushima/2/2011	H5N1	ans	2011/01/	2011.038
A/duck/Laos/25/2006	H5N1	ans	2006	2006.452
A/duck/Lao/981/2010	H5N1	ans	2010/03/	2010.2
A/whooperswan/Mongolia/244/2005	H5N1	ans	2005	2005.452
A/commongoldeneye/Mongolia/12/2006	H5N1	ans	2006/05/	2006.367
A/whooperswan/Mongolia/6/2009	H5N1	ans	2009/05/	2009.367
A/ruddyshelduck/Mongolia/X42/2009	H5N1	ans	2009/07/	2009.534
A/duck/Niger/914/2006	H5N1	ans	2006	2006.452
A/guineafowl/Nigeria/957-12/2006	H5N1	ans	2006	2006.452
A/goose/Krasnoozerskoye/627/2005	H5N1	ans	2005	2005.452
A/grebe/Tyva/Tyv06-1/2006	H5N1	ans	24/06/2006	2006.477
A/duck/Omsk/1822/2006	H5N1	ans	2006/07/	2006.534
A/Cygnuscygnus/Krasnodar/329/07	H5N1	ans	06/09/2007	2007.679
A/grebe/Tyva/3/2009	H5N1	ans	22/06/2009	2009.471
A/grebe/Tyva/2/2010	H5N1	ans	22/06/2010	2010.471
A/swan/Slovenia/760/2006	H5N1	ans	2006	2006.452

A/duck/Korea/ESD1/03	H5N1	ans	2003	2003.452
A/littlegrebe/Thailand/Phichit-01/2004	H5N1	ans	2004	2004.452
A/duck/Vietnam/7A/2004	H5N1	ans	2004	2004.452
A/duck/Vietnam/367/2005	H5N1	ans	2005	2005.452
A/duck/Vietnam/206/2005	H5N1	ans	07/12/2005	2005.932
A/muscovyduck/HaNam/07-47/2007	H5N1	ans	2007	2007.452
A/duck/Vietnam/G12/2008	H5N1	ans	2008/01/	2008.038
A/muscovyduck/Vietnam/OIE-559/2011	H5N1	ans	2011/06/	2011.452
A/commongull/Chany/P/2006	H5N1	cha	18/07/2006	2006.542
A/openbillstork/Thailand/VSMU-20-AYA/2004	H5N1	cic	2004	2004.452
A/openbilledstork/Nakhonsawan/BBD1821J/05	H5N1	cic	2005	2005.452
A/peregrinefalcon/HongKong/2142/2008	H5N1	fal	2008	2008.452
A/peregrinefalcon/Tochigi/15/2011	H5N1	fal	2011/02/	2011.123
A/sakerfalcon/Kuwait/0286/2007	H5N1	fal	2007	2007.452
A/falcon/SaudiArabia/D1795/2005	H5N1	fal	2005	2005.452
A/falcon/SaudiArabia/D1936/2007	H5N1	fal	2007	2007.452
A/chicken/Afghanistan/1573-47/2006	H5N1	gal	2006	2006.452
A/chicken/Bhutan/248009/2010	H5N1	gal	20/02/2010	2010.137
A/chicken/Cambodia/TKCMB5T/2010	H5N1	gal	28/01/2010	2010.074
A/chicken/Hubei/wj/1997	H5N1	gal	1997	1997.452
A/chicken/Hebei/718/2001	H5N1	gal	2001	2001.452
A/chicken/Hebei/108/02	H5N1	gal	2002	2002.452
A/chicken/Hubei/wo/2003	H5N1	gal	2003	2003.452
A/chicken/Jiangxi/25/2004	H5N1	gal	2004	2004.452
A/chicken/Guangdong/1/2005	H5N1	gal	2005	2005.452
A/chicken/Liaoning/A-11/2006	H5N1	gal	2006	2006.452
A/chicken/Liaoning/A-1/2007	H5N1	gal	2007	2007.452
A/chicken/Jiangsu/18/2008	H5N1	gal	2008	2008.452
A/chicken/Hunan/1/2009	H5N1	gal	2009	2009.452
A/chicken/IvoryCoast/1787-35/2006	H5N1	gal	2006	2006.452
A/turkey/IvoryCoast/4372-2/2006	H5N1	gal	2006/12/	2006.953
A/turkey/Egypt/2253-2/2006	H5N1	gal	2006	2006.452
A/chicken/Egypt/CL6/2007	H5N1	gal	2007/03/	2007.2
A/chicken/Egypt/0836/2008	H5N1	gal	15/01/2008	2008.038
A/chicken/Egypt/0891/2008	H5N1	gal	26/09/2008	2008.734
A/chicken/Germany/R3234/2007	H5N1	gal	2007	2007.452
A/Ck/HK/YU22/2002	H5N1	gal	2002	2002.452
A/chicken/India/NIV33487/06	H5N1	gal	2006	2006.452
A/chicken/India/WB-NIV2653/2008	H5N1	gal	21/01/2008	2008.055
A/chicken/WestBengal/170564/2009	H5N1	gal	09/03/2009	2009.184
A/chicken/WestBengal/239020/2010	H5N1	gal	12/01/2010	2010.03
A/quail/Yogjakarta/UT1023/2004	H5N1	gal	15/01/2004	2004.038
A/chicken/CentralJava/UT3091/2005	H5N1	gal	2005	2005.452
A/chicken/SouthKalimantan/UT6028/2006	H5N1	gal	2006/11/	2006.871

A/chicken/Oita/8/2004	H5N1	gal	2004	2004.452
A/chicken/Kuwait/KISR8/2007	H5N1	gal	2007	2007.452
A/chicken/Laos/P0130/2007	H5N1	gal	2007	2007.452
A/chicken/Lao/LH1/2010	H5N1	gal	2010/05/	2010.367
A/chicken/Nigeria/1047-62/2006	H5N1	gal	2006	2006.452
A/chicken/Nigeria/OG11/2007	H5N1	gal	2007	2007.452
A/chicken/Rawalakot/NARC2441A/2006	H5N1	gal	2006	2006.452
A/turkey/Poland/R3249/2007	H5N1	gal	2007	2007.452
A/chicken/Kurgan/3/2005	H5N1	gal	2005	2005.452
A/chicken/Reshoty/02/2006	H5N1	gal	2006/06/	2006.452
A/chicken/Russia/Krasnodar/2/2007	H5N1	gal	2007	2007.452
A/chicken/Krasnodar/300/07	H5N1	gal	05/09/2007	2007.677
A/turkey/SaudiArabia/6732-6/2007	H5N1	gal	2007	2007.452
A/chicken/Sudan/1784-10/2006	H5N1	gal	2006	2006.452
A/chicken/KohnKaen/NIAH330/2004	H5N1	gal	2004	2004.452
A/quail/NakhonPathom/NIAH7562/2005	H5N1	gal	2005	2005.452
A/chicken/Phichit/NIAH606988/2006	H5N1	gal	2006	2006.452
A/chicken/Sukhothai/NIAH114843/2008	H5N1	gal	2008	2008.452
A/chicken/Crimea/08/2005	H5N1	gal	2005	2005.452
A/chicken/Scotland/1959	H5N1	gal	1959	1959.452
A/chicken/Vietnam/921/2004	H5N1	gal	2004	2004.452
A/chicken/LangSon/200/2005	H5N1	gal	2005	2005.452
A/chicken/Vietnam/209/2005	H5N1	gal	08/12/2005	2005.934
A/chicken/NamDinh/07-32/2007	H5N1	gal	2007	2007.452
A/piedmagpie/Liaoning/7/2006	H5N1	pas	2006	2006.452
A/magpierobin/HongKong/1897/2008	H5N1	pas	2008	2008.452
A/crow/Assam/142119/2008	H5N1	pas	2008/12/	2008.953
A/crow/Osaka/102/2004	H5N1	pas	2004	2004.452
A/pigeon/Laos/P0022/2007	H5N1	pas	2007	2007.452
A/pigeon/Rostov-on-Don/6/2007	H5N1	pas	14/12/2007	2007.951
A/pigeon/Thailand/VSMU-7-NPT/2004	H5N1	pas	2004	2004.452
A/pigeon/Thailand/VSMU-11-KRI/2005	H5N1	pas	2005	2005.452
A/egret/HongKong/757.2/2003	H5N1	pel	2003	2003.452
A/greyheron/HongKong/3088/2007	H5N1	pel	2007	2007.452
A/greyheron/HongKong/1046/2008	H5N1	pel	2008	2008.452
A/ostrich/Suzhou/097/2003	H5N1	str	16/04/2003	2003.288
A/ostrich/Nigeria/1047-25/2006	H5N1	str	2006	2006.452
A/garganey/SanJiang/160/2006	H5N2	ans	2006	2006.452
A/duck/Denmark/65047/04	H5N2	ans	2004	2004.452
A/duck/France/080032/2008	H5N2	ans	2008/01/	2008.038
A/duck/Potsdam/1402-6/1986	H5N2	ans	1986	1986.452
A/duck/HongKong/342/78	H5N2	ans	1978	1978.452
A/teal/Italy/3931-38/2005	H5N2	ans	2005	2005.452
A/mallard/Italy/4223-2/2006	H5N2	ans	11/11/2006	2006.86

A/northernpintail/Akita/714/2006	H5N2	ans	2006/12/	2006.953
A/duck/Malaysia/F118-08-04/2004	H5N2	ans	2004	2004.452
A/mallard/Netherlands/3/1999	H5N2	ans	09/10/1999	1999.77
A/duck/Primorie/2621/2001	H5N2	ans	24/08/2001	2001.644
A/gadwall/Altai/1202/2007	H5N2	ans	2007/09/	2007.704
A/mallard/Sweden/74/2003	H5N2	ans	2003	2003.452
A/turkey/Italy/1980	H5N2	gal	1980	1980.452
A/poultry/Italy/382/1997	H5N2	gal	1997	1997.452
A/turkey/Italy/1258/2005	H5N2	gal	2005	2005.452
A/duck/France/02166/2002	H5N3	ans	2002	2002.452
A/duck/France/06436/2006	H5N3	ans	21/03/2006	2006.216
A/mallard/France/061054/2006	H5N3	ans	07/11/2006	2006.849
A/duck/HongKong/205/1977	H5N3	ans	1977	1977.452
A/goose/HongKong/23/1978	H5N3	ans	1978	1978.452
A/duck/Italy/775/2004	H5N3	ans	2004	2004.452
A/teal/Italy/3812/2005	H5N3	ans	2005	2005.452
A/mallard/Netherlands/65/2006	H5N3	ans	2006	2006.452
A/duck/Altai/1285/1991	H5N3	ans	15/08/1991	1991.619
A/duck/Primorie/2633/2001	H5N3	ans	24/08/2001	2001.644
A/spot-billedduck/Korea/KNUSYG06/2006	H5N3	ans	2006	2006.452
A/tern/SouthAfrica/1961	H5N3	cha	1961	1961.452
A/mallard/Jiangxi/7787/2003	H6N1	ans	2003	2003.452
A/duck/Shantou/1275/2004	H6N1	ans	2004	2004.452
A/mallard/Jiangxi/13228/2005	H6N1	ans	2005	2005.452
A/duck/EasternChina/1/2008	H6N1	ans	2008/11/	2008.871
A/duck/HongKong/d73/1976	H6N1	ans	1976	1976.452
A/Teal/HongKong/W312/97	H6N1	ans	1997	1997.452
A/greylaggoose/Netherlands/4/1999	H6N1	ans	01/03/1999	1999.162
A/mallard/Sweden/81/2002	H6N1	ans	03/12/2002	2002.921
A/mallard/Sweden/30/2005	H6N1	ans	2005	2005.452
A/duck/Taiwan/0526/72	H6N1	ans	1972	1972.452
A/duck/Taiwan/WB29/99	H6N1	ans	1999	1999.452
A/duck/Taiwan/29-3/00	H6N1	ans	2000	2000.452
A/duck/Taiwan/WB239/03	H6N1	ans	2003	2003.452
A/chicken/HongKong/17/1977	H6N1	gal	1977	1977.452
A/quail/HongKong/1721-30/99	H6N1	gal	1999	1999.452
A/quail/HongKong/SF595/00	H6N1	gal	2000	2000.452
A/pheasant/HongKong/SSP44/2002	H6N1	gal	2002	2002.452
A/chukkar/HongKong/SF126/2003	H6N1	gal	2003	2003.452
A/chicken/Taiwan/0824/97	H6N1	gal	1997	1997.452
A/chicken/Taiwan/na3/98	H6N1	gal	1998	1998.452
A/chicken/Taiwan/0705/99	H6N1	gal	1999	1999.452
A/chicken/Taiwan/SP1/00	H6N1	gal	2000	2000.452
A/chicken/Taiwan/1205/01	H6N1	gal	2001	2001.452

A/chicken/Taiwan/0208/02	H6N1	gal	2002	2002.452
A/chicken/Taiwan/1203/03	H6N1	gal	2003	2003.452
A/chicken/Taiwan/ch1006/04	H6N1	gal	2004	2004.452
A/chicken/Taiwan/A342/05	H6N1	gal	2005	2005.452
A/pigeon/HongKong/WF47/2003	H6N1	pas	2003	2003.452
A/partridge/Taiwan/LU1/99	H6N1	pas	1999	1999.452
A/wildduck/Shantou/1651/2000	H6N2	ans	2000	2000.452
A/duck/Shantou/2444/2001	H6N2	ans	2001	2001.452
A/duck/Shantou/2471/2002	H6N2	ans	2002	2002.452
A/mallard/Jiangxi/7376/2003	H6N2	ans	2003	2003.452
A/mallard/Jiangxi/8264/2004	H6N2	ans	2004	2004.452
A/duck/Shantou/7568/2005	H6N2	ans	2005	2005.452
A/mallard/HeiLongjiang/131/2006	H6N2	ans	2006	2006.452
A/duck/EasternChina/40/2007	H6N2	ans	2007/05/	2007.367
A/duck/EasternChina/46/2008	H6N2	ans	2008/04/	2008.285
A/duck/EasternChina/2/2008	H6N2	ans	2008/12/	2008.953
A/mallard/CzechRepublic/15902-17K/2009	H6N2	ans	30/09/2009	2009.745
A/duck/HongKong/d134/1977	H6N2	ans	1977	1977.452
A/whitefrontedgoose/Netherlands/2/1999	H6N2	ans	02/03/1999	1999.164
A/barnaclegoose/Netherlands/1/2005	H6N2	ans	2005	2005.452
A/mallard/Netherlands/71/2006	H6N2	ans	2006	2006.452
A/aquaticbird/Korea/CN17/2009	H6N2	ans	2009/11/	2009.871
A/mallard/Sweden/52/2003	H6N2	ans	30/10/2003	2003.827
A/duck/Kingmen/E322/04	H6N2	ans	2004	2004.452
A/Muscovyduck/Vietnam/G33/2007	H6N2	ans	2007/09/	2007.704
A/duck/Zambia/03/2008	H6N2	ans	2008/06/	2008.452
A/duck/Zambia/10/2009	H6N2	ans	2009/09/	2009.704
A/gull/Moscow/3100/2006	H6N2	cha	2006/10/	2006.786
A/chicken/Hunan/989/2005	H6N2	gal	2005	2005.452
A/chicken/EasternChina/42/2007	H6N2	gal	2007/06/	2007.452
A/duck/Hunan/5613/2003	H6N5	ans	2003	2003.452
A/mallard/Jiangxi/8346/2004	H6N5	ans	2004	2004.452
A/duck/Hunan/748/2005	H6N5	ans	2005	2005.452
A/duck/Yangzhou/013/2008	H6N5	ans	15/12/2008	2008.953
A/mallard/Netherlands/11/2007	H6N5	ans	2007	2007.452
A/aquaticbird/Korea/W69/2005	H6N5	ans	2005/11/	2005.871
A/aquaticbird/Korea/CN5/2009	H6N5	ans	2009/11/	2009.871
A/duck/Taiwan/WB459/04	H6N5	ans	2004	2004.452
A/duck/Nanchang/1904/1992	H7N1	ans	1992	1992.452
A/duck/Denmark/53-147-8/2008	H7N1	ans	25/04/2008	2008.312
A/guineafowl/Italy/155/2000	H7N1	ans	11/01/2000	2000.027
A/duck/Mongolia/47/2001	H7N1	ans	2001	2001.452
A/duck/Mongolia/867/2002	H7N1	ans	2002	2002.452
A/mallard/Netherlands/22/2007	H7N1	ans	2007	2007.452

A/chicken/Italy/1391/1999	H7N1	gal	16/04/1999	1999.288
A/chicken/Italy/2335/2000	H7N1	gal	31/03/2000	2000.244
A/turkey/Italy/1351/2001	H7N1	gal	15/02/2001	2001.123
A/ostrich/Italy/984/00	H7N1	str	2000	2000.452
A/mallard/Netherlands/12/2000	H7N3	ans	2000	2000.452
A/tuftedduck/PT/13771/2006	H7N3	ans	22/03/2006	2006.219
A/turkey/Italy/8534/2002	H7N3	gal	2002	2002.452
A/duck/Korea/BC10/2007	H7N3	ans	2007/03/	2007.216
A/duck/Shimane/137/2006	H7N3	ans	2006	2006.452
A/chicken/Pakistan/34669/1995	H7N3	gal	1995	1995.452
A/chicken/Pakistan/c1998/1998	H7N3	gal	1998	1998.452
A/chicken/Chakwal/NARC-35/2001	H7N3	gal	2001/05/	2001.367
A/chicken/Rawalpindi/NARC68/2002	H7N3	gal	2002	2002.452
A/chicken/Chakwal/NARC-148/2004	H7N3	gal	12/05/2004	2004.359
A/swan/CzechRepublic/5416/2011	H7N7	ans	06/04/2011	2011.26
A/mallard/Netherlands/9/2005	H7N7	ans	2005	2005.452
A/swan/Slovenia/53/2009	H7N7	ans	20/01/2009	2009.052
A/duck/Tsukuba/664/2007	H7N7	ans	2007/12/	2007.953
A/mallard/Sweden/95/2005	H7N7	ans	2005	2005.452
A/chicken/Germany/R28/03	H7N7	gal	2003	2003.452
A/chicken/Netherlands/1/03	H7N7	gal	2003	2003.452
A/duck/Korea/JSM/2010	H7N7	ans	2010/06/	2010.452
A/duck/Nanjing/2/97	H9N2	ans	1997	1997.452
A/wildduck/Nanchang/2-0480/2000	H9N2	ans	17/02/2000	2000.129
A/duck/Hubei/W1/2004	H9N2	ans	2004	2004.452
A/duck/Guangxi/51/2005	H9N2	ans	2005	2005.452
A/duck/Fujian/FQ107/2007	H9N2	ans	2007/12/	2007.953
A/duck/Tibet/S2/2009	H9N2	ans	10/09/2009	2009.69
A/mallard/France/090360/2009	H9N2	ans	2009/09/	2009.704
A/duck/HongKong/702/1979-quailadapted	H9N2	ans	1979	1979.452
A/duck/HongKong/784/1979	H9N2	ans	10/10/1979	1979.773
A/duck/Hokkaido/49/98	H9N2	ans	1998	1998.452
A/duck/Hokkaido/9/99	H9N2	ans	1999	1999.452
A/duck/Malaysia/2001	H9N2	ans	2001	2001.452
A/Eurasianwigeon/Netherlands/3/2005	H9N2	ans	2005	2005.452
A/gadwall/Netherlands/1/2006	H9N2	ans	2006	2006.452
A/Bewicksswan/Netherlands/5/2007	H9N2	ans	2007	2007.452
A/duck/Korea/KNUDPJ09/2009	H9N2	ans	2009/05/	2009.367
A/chicken/Shandong/7/96	H9N2	gal	1996	1996.452
A/chicken/Shenzhen/9/97	H9N2	gal	1997	1997.452
A/chicken/Beijing/8/98	H9N2	gal	1998	1998.452
A/chicken/Ningxia/4/99	H9N2	gal	1999	1999.452
A/quail/Nanchang/2-0460/2000	H9N2	gal	17/02/2000	2000.129
A/chicken/China/Guangxi14/2000	H9N2	gal	13/11/2000	2000.866

A/chicken/Henan/43/02	H9N2	gal	2002	2002.452
A/chicken/Guangxi/37/2005	H9N2	gal	2005	2005.452
A/chicken/Hebei/L1/2006	H9N2	gal	2006	2006.452
A/chicken/Hubei/C1/2007	H9N2	gal	2007	2007.452
A/chicken/Henan/1.2/2008	H9N2	gal	10/01/2008	2008.025
A/chicken/Hebei/Y2/2009	H9N2	gal	2009/03/	2009.2
A/chicken/Tibet/S1/2009	H9N2	gal	10/09/2009	2009.69
A/chicken/China/AH-10-01/2010	H9N2	gal	02/12/2010	2010.918
A/chicken/Guangdong/ZCY/2011	H9N2	gal	2011/07/	2011.534
A/chicken/Uchal/8286/2006	H9N2	gal	25/02/2006	2006.151
A/chicken/Israel/90658/2000	H9N2	gal	2000	2000.452
A/turkey/Israel/810/2001	H9N2	gal	2001	2001.452
A/turkey/Israel/619/2002	H9N2	gal	2002	2002.452
A/chicken/Israel/1475/2003	H9N2	gal	2003	2003.452
A/chicken/Israel/1966/2004	H9N2	gal	2004	2004.452
A/chicken/Israel/554/2005	H9N2	gal	2005	2005.452
A/chicken/Israel/178/2006	H9N2	gal	2006	2006.452
A/chicken/Israel/375/2007	H9N2	gal	06/03/2007	2007.175
A/chicken/Israel/310/2008	H9N2	gal	27/06/2008	2008.485
A/chicken/Israel/184/2009	H9N2	gal	15/02/2009	2009.123
A/chicken/Osaka/aq48/97	H9N2	gal	1997	1997.452
A/chicken/Osaka/aq19/2001	H9N2	gal	2001	2001.452
A/chicken/Yokohama/aq45/2002	H9N2	gal	2002	2002.452
A/quail/Lebanon/272/2010	H9N2	gal	2010/08/	2010.619
A/chicken/Pakistan/2/99	H9N2	gal	1999	1999.452
A/chicken/Pakistan/UDL-01/2005	H9N2	gal	18/05/2005	2005.375
A/chicken/Pakistan/UDL-03/2005	H9N2	gal	21/12/2005	2005.97
A/chicken/Pakistan/UDL-01/2007	H9N2	gal	07/06/2007	2007.43
A/chicken/Pakistan/UDL-03/2008	H9N2	gal	04/03/2008	2008.17
A/Korea/KBNP-0028/2000	H9N2	gal	2000	2000.452
A/chicken/Korea/S21/2004	H9N2	gal	2004	2004.452
A/chicken/Korea/GH2/2007	H9N2	gal	24/11/2007	2007.896
A/chicken/Korea/KNUSWR09/2009	H9N2	gal	2009/04/	2009.285
A/chicken/Dubai/338/2001	H9N2	gal	2001	2001.452
A/chicken/Emirates/R66/2002	H9N2	gal	2002	2002.452
A/pigeon/Nanchang/2-0461/2000	H9N2	pas	17/02/2000	2000.129
A/black-billedmagpie/Guangxi/31/2005	H9N2	pas	2005/05/	2005.367
A/parakeet/Chiba/1/97	H9N2	psi	1997	1997.452
A/parakeet/Narita/92A/98	H9N2	psi	1998	1998.452

Table A 3.2 Sequences for the discrete traits analysis in chapter 3.
Information for subsampling dataset A2 (n = 344, Table A 3.1) of Eurasian AIV sequences used in this discrete traits analysis. Isolate name, subtype, host species, date of collection and the estimated decimal date are listed. [a]: The abbreviations of bird orders being used here are described in Chapter 2 Table 2.1.

Chapter 4

Predicting the spatial diffusion of avian influenza virus in China

Chapter Abstract

Avian influenza virus (AIV) is of particular importance in China, where it causes endemics and severe outbreaks among poultry and sporadic human infections. It is still obscure how AIV transmits among different geographic locations especially the direction of transmission in a region which involves both virus import and virus export. I have reconstructed the phylogeographic diffusion network of AIV in China, and found strong virus clustering in geographic areas as well as the economic zones. Central area and Pan-Pearl River Delta are two main sources of the AIV diffusion while the East Coast areas especially the Yangtze River delta, are the major target of viral invasion. I have also integrated predictors (in terms of economic-agriculture, climate and natural environment) into viral diffusion, and investigated their potential impacts using a homogenous statistical methodology during the diffusion process. I have highlighted that the economic-agricultural predictors, especially the poultry population density and the number of farm product markets, are the determinants of the spatial diffusions of AIV in China; high human density and capability of transportation freight in the geographic locations are vital predictive of high rate of viral transmission; Climate features (temperature and humidity) are also correlated to the viral invasion in the destination in some degree. While little or no impacts are found from natural environment factors such as stock volume of forest, surface water resources and the coverage of nature reserves. This study will advance our understanding of the spatial dynamics of avian influenza virus in bird populations. In addition, the identification of the determinants correlated with such transmission will strengthen efforts to achieve disease control.

4.1 Introduction

China plays a central role in maintaining a source population for influenza diversity globally. Of the four human influenza pandemics between 1918 and 2009, the 1957 "Asian" H2N2 flu and H3N2 1968 "Hong Kong" flu were spread from China, and the origin viruses of these outbreaks are reassortants between human and avian strains (Kalthoff *et al.*, 2010). More than half of the provinces and municipalities in China have experienced avian influenza outbreaks, although the distribution and prevalence of AIV outbreaks vary among locations and the major regional differences in AIV prevalence are in ecological systems, husbandry practices, cultural behaviours and economic development (WHO, 2005). There is a consequential impact on the distribution of infectious diseases and maintenance. As a result, the disease control options are expanded (Martin *et al.*, 2011).

Two types of locations are considered as potential hotspots of influenza outbreaks and therefore of particular importance. One is the areas with a high ecological complexity: areas rich in nature reserves, water resources, and on the flyway route of migratory birds. This creates an environment in which domestic ducks and wild aquatic birds live together, sharing water, food, and habitat. Wild birds migrate via three major Asian flyway routes across different regions of China (see Chapter1 section 1.2.2.2); Qinghai Lake is one of the most important breeding and stopover sites for migratory birds along the Central Asian Flyway (Cui *et al.*, 2011). An outbreak of H5N1 virus in wild birds at Qinghai Lake in Northwest China occurred in 2005 in which the viruses carried by wild bird migration from Qinghai Lake contributed to the global H5N1 prevalence and to the increase in human H5N1 influenza virus infections (Chen *et al.*, 2006a; Fauci, 2006; Kandeel *et al.*, 2010; Oner *et al.*, 2006; Van Boeckel *et al.*, 2011). In addition, Poyang Lake in Jiangxi Province and Dongting Lake in Hunan Province locate in South Central China and situate within the East Asian Flyway. They both possess important national natural reserves and wetlands. These locations have been identified with extensive reassortment events of AIV and found close interactions between wild migratory birds and domestic poultry. Providing opportunities for outbreaks of

highly pathogenic avian influenza (HPAI) virus (Deng *et al.*, 2013; Takekawa *et al.*, 2010; Zhang *et al.*, 2013).

The southern coastal zone in China is one of the type of locations with great importance as highly developed in economy and trade, which has also been considered as a potential epicentre (Cox & Subbarao, 2000). Highly pathogenic avian influenza (HPAI) H5N1 was first encountered in 1996 in Guangdong province in domestic geese. By 2000, the host range had extended to domestic ducks, which played a key role in the genesis of the 2003/04 outbreaks, and subsequently spread throughout Asia and the western Palearctic in 2004–2006 (Martin *et al.*, 2011; Sims *et al.*, 2005; Smith *et al.*, 2006). The recent outbreak of human infections with an emerging avian influenza A (H7N9) virus occurred mostly in the *Yangtze* River Delta region (YRD) in eastern China since late March 2013. The outbreak caused 375 cases of human infections including 115 deaths up to date (WHO, 2014). This novel H7N9 virus is asymptomatic in birds, and is likely to spread silently in birds or other animal reservoirs. Human infection appears to be associated with exposure to infected live poultry or contaminated environments, including markets where live poultry are sold (Fang *et al.*, 2013; Lam *et al.*, 2013).

In addition, previous studies on the transmission and distribution of avian influenza virus are mainly focused on the highly pathogenic outbreaks, such as H5 and H7 (Li *et al.*, 2009; Martin *et al.*, 2011). However, phylogenetic studies have shown that the avian influenza with low pathogenicity in birds could transmit their gene segments (especially internal segments) to other strains. The gene invasions contribute to the creation of new viruses with high pathogenicity, and with the potential to cause zoonotic outbreaks in human and mammals. Co-circulating H5N1 and H9N2 subtype of influenza viruses were isolated from chickens, ducks and geese, and led to genetic exchange of internal genes identified between the human H5N1 subtype (isolated 1997) and the avian H5N1 and H9N2 subtypes (Guan *et al.*, 1999). Moreover, surveillance data provide direct evidence that H9N2 strains share similar segments and coexisted with the novel human-pathogenic H7N9 influenza virus in epidemiologically linked live poultry markets, indicating viruses of H9N2 subtype made a recent contribution to the evolution of the H7N9 virus (Cui *et*

al., 2014; Yu *et al.*, 2014b). Avian influenza viruses of the H6 subtype also make up (mainly H6N1 and H6N2) one of the most commonly recognized subtypes in domestic ducks. In addition, H6 viruses which are circulating in minor terrestrial poultry species (such as quail, pheasant, and chukar) are continually reassorting with H5N1 and H9N2 (Cheung *et al.*, 2007; Huang *et al.*, 2010; Huang *et al.*, 2012). Therefore, understanding the feature of the total interaction pattern of avian influenza viruses which are of different subtypes, with both low and high pathogenicity, sampled from both wild and domestic birds are of vital importance. There are still many questions that remain unknown:

1. What is the distribution and prevalence of AIV in bird reservoirs and how and where does it circulate?
2. What are the main routes of exposure, transmission and spread?
3. What are the determinants to explain the viral distribution and transmission?

Various methodologies have been applied to explore determinants of the spread and distribution of infectious diseases. For example, using statistical methods to explain the disease distribution (Martin *et al.*, 2011; Si *et al.*, 2013); or using mathematical and epidemic modelling to determine the transmission parameters (Chitnis *et al.*, 2008). In addition, there are studies on using spatial modelling in different agro-ecological systems, and studies on the statistical association with the presence of influenza A virus (especially HPAI H5N1) (Fang *et al.*, 2013; Fuller *et al.*, 2013; Gilbert *et al.*, 2010; Pavade *et al.*, 2011; Si *et al.*, 2013). However, none of these approaches take evolutionary parameters into consideration. Bayesian phylogenetics link patterns of genetic diversity to ecological processes and allow the quantitative studies on how and to what extent ecological features shape pathogen genetic diversity (Drummond & Rambaut, 2007; Lemey *et al.*, 2009; Lemey *et al.*, 2010; Lycett *et al.*, 2012). In this study, I adopted a recently developed method to reconstruct virus spatial transmission with a flexible Bayesian statistical framework. This method simultaneously tests and quantifies the contribution of multiple ecological and evolutionary drivers of the viral spatial diffusion by parameterizing

the rates of a stochastic discrete diffusion process as a generalized linear model (GLM) (Faria *et al.*, 2013; Lemey *et al.*, 2014).

The aim of the study is to describe the pattern of spatial diffusion AIV in bird population in different regions of China and to indicate the related risk factors across the diversity of economy, agriculture, environment and climate. I have explored the key sources and destinations of the AIV special diffusion in China by performing a Bayesian stochastic search variable selection procedure to quantify the significant viral spatial transmission among regions classified by geographic and economic parameters. I have also evaluated the impacts of potential predictors on viral diffusion by integrating agricultural and ecological data with viral genetics in the phylogeographic reconstruction. The results highlight the importance of the agro-economic impact on the evolution and spatial dissemination of the influenza virus which may inform future strategies for the surveillance and control of influenza.

4.2 Methods

4.2.1 Sequence Data

In total 835 full length avian influenza sequences isolated from China were downloaded from NCBI Flu Resources (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). For each internal segment, sequence alignments were created using MUSCLE (Edgar, 2004). The preliminary trees for each segment were generated by neighbour joining method (with 500 bootstraps) using MEGA 5.1.0. Sequences that might contain incorrect date information were detected in these preliminary phylogenies using path-o-gen V1.2 and were removed from the dataset. To account for potential sampling biases in space and time, sequences were subsampled to preserve diversity with respect to location (provinces), host, pathogenicity, subtype background and year of sampling using custom R scripts which resulted in 405 sequences. I further subsampled one over-represented clade (conserved in all six segments) which composed of sequences mainly isolated from the south region, resulting a final dataset composed of all the

eight gene segments of 320 AIV sequences, which cover an 18-year period from the 1996 to 2013. The dataset is summarized in Table A4.3.

4.2.2 Time-scaled phylogenetic analysis

Time-scaled trees were generated for each internal protein-coding segment with the 320 Chinese AIV dataset using BEAST version 1.7.3 (Drummond *et al.*, 2006). Different substitution models, clock models and tree models were evaluated for each segment by calculating the Bayes factor (BF), which is the ratio of the marginal likelihoods of the two models (Raftery, 1994) in Tracer v1.4 (http://tree.bio.ed.ac.uk/software/tracer/). Models being compared are described in Chapter 2. A general-time-reversal (GTR) model was chosen as the default nucleotide substitution model. A constant population size model and a relaxed uncorrelated log-normal molecular clock model were chosen for all eight segments. The MCMC was run for 10^8 steps and sampled every 10^4 steps. Two independent runs were used in each segment to confirm the convergence and then combined. The Maximum Clade Credibility (MCC) trees were obtained and the mean ages of each node and the corresponding HPD were calculated by using Tree Annotator in BEAST.

4.2.3 Spatial diffusion analysis

4.2.3.1 Geographic region classification

To investigate the spatial diffusion pattern of (AIV) in China, the locations of 320 Chinese AIV (including 27 provinces) were categorized into larger regions according to two different schemes: (1) Traditional regions (TR) are the 31 provincial-level divisions grouped by the 6 former administrative areas from 1949 to 1952: East, North, Northeast, Northwest, South Central and Southwest; (2) Economic region (ER) - are categorized by the per capita income basis and composed of 4 areas: East Coast, Central, Northeast and Western. This scheme showed that provinces in the coastal regions of China tend to be more industrialized, while regions in the hinterland are less developed. Economic regions are further divided into 7 smaller economic zones (ED) according to the result of the trait-phylogeny test described below. To be specific, the East Coast area (which

contains the highest AIV prevalence) is divided into three smaller economic zones areas: Yangtze River Delta (YRD), Pan-Pearl River Delta (PRD) and Bohai Economic Rim (BER); The vast and less developed West areas are divided into Northwest and Southwest, which resulted in total of 7 zones in total. The detailed classification and distribution of provinces into regions can be seen in Table 4.1.

Region	Area	Provinces
TR	East	Shandong, Jiangsu, Anhui, Jiangxi, Zhejiang, Fujian, Shanghai
	South Central	Guangdong, Guangxi, Henan, Hubei, Hunan
	North	Beijing, Hebei, Shanxi
	Northeast	Heilongjiang, Jilin, Liaoning
	Northwest	Gansu, Ningxia, Qinghai, Xinjiang, Shaanxi
	Southwest	Guizhou, Sichuan, Tibet, Yunnan
ER	EastCoast	Beijing, Hebei, Shandong, Jiangsu, Zhejiang, Fujian, Shanghai, Guangdong
	Central	Shanxi, Anhui, Henan, Hubei, Hunan, Jiangxi
	Northeast	Heilongjiang, Jilin, Liaoning
	Western	Gansu, Ningxia, Qinghai, Xinjiang, Shaanxi, Guizhou, Sichuan, Tibet, Yunnan
ED	YangtzeRD (YRD)	Shanghai, Jiangsu, Zhejiang
	Pan-PearlRD (PRD)	Guangdong, Fujian
	Bohai Economic Rim (BER)	Beijing, Tianjin, Hebei, Shandong
	Central	Shanxi, Anhui, Henan, Hubei, Hunan, Jiangxi
	Northeast	Heilongjiang, Jilin, Liaoning
	Northwest	Gansu, Ningxia, Qinghai, Xinjiang, Shaanxi
	Southwest	Guizhou, Sichuan, Tibet, Yunnan, Guangxi

Table 4.1 Classification of provinces into areas in 3 region types

4.2.3.2 Distribution of region traits on tree phylogeny

Bayesian Tip-association Significance testing (BaTS) is an analytical tool which can be used for testing hypothesis regarding AIV evolution and associations between phylogeny and epidemiologically relevant traits including host range and temporal–spatial distribution, by comparing the observed distribution of states at the tips with a null distribution under which states are randomly distributed across the tips of the trees (Girard *et al.*, 2012). By integrating over the credible set of topologies, three statistical tests (including Parsimony score (PS), association index (AI) and the monophyletic clade ('MC') size) on significant phylogeny-trait correlations are conducted whilst uncertainties arising from phylogenetic errors are taken into account. However, the BaTS program takes into account only tree topology, not branch length, meaning that monophyletic clades consisting of nearly identical sequences are weighted the same as those consisting of distantly related sequences (Parker *et al.*, 2008). In this study, associations between phylogenies and location traits for each region type (TR, ER, or ED) were tested using BaTS. To test the significance of the observed data, 1000 subsampled trees with 1000 state randomizations were generated to create the null distributions.

4.2.3.3 Discrete trait mapping of the spatial diffusion

I inferred rates of viral migration between locations across different regions in China. I used asymmetric discrete traits model (being fully explained in Chapter 2) to reconstruct the ancestral subtype states of the internal nodes in time-scaled tree phylogeny of each segment. The significant transition rates between discrete traits were identified with the Bayesian stochastic search variable selection (BSSVS) extension of the discrete model. For the rates calculated from BSSVS, a BF test was adopted to identify significant non-zero transition rates and a cutoff of BF=5 was applied (Kass & Raftery, 1995) .

4.2.4 Predictors of spatial diffusion

4.2.4.1 Predictor data

With the aim to identify the determinants of AIV diffusion in China, I selected possible factors in different categories including Agro-economics, environment and climate from the nationwide statistics which are published annually by the National Statistics Bureau (NSB) and the Ministry of Agriculture's Animal Husbandry Bureau (AHB). Data were referenced from: (i) China statistical year book 2013 (Sen, 2013) and (ii) China agriculture yearbook 2012 (Liu, 2012).

In the agro-economic categories, I focused on the standard statistics related to the production and sale of poultry productions (chickens, ducks and geese). The poultry density is the annual counts of each poultry type collected from farms and households at the township level and reported up through county, prefecture, and provincial administrative units with final submission to the national level. In addition, given that live bird markets (LBMs) play a crucial role in the maintenance, amplification and dissemination of AIV (Murhekar *et al.*, 2013; Yu *et al.*, 2014a), I have included the number of farm product markets per provinces as a potential predictor obtained from the national navigation map (Fang *et al.*, 2013). In addition, the agro-economic variables including human population density and the total capacity of transportation (by means of railway, highway and waterway) per province are also considered. The physical environmental variables comprise the coverage of forest, surface water and natural resource. This may indicate a higher chance of wild birds interacting with poultry which is expected to be positively correlated to disease outbreaks; climatic factors including the average annual temperature and the average relative humidity are also being considered.

There were initially 22 predictors which were collected from 27 provinces (including autonomous regions and municipalities) in mainland China. However, the large number of possible predictors may reduce the power of the association test in this study. Therefore correlations between pairs of predictors were tested, predictors with significantly high correlations or having very similar meanings are removed, in this way I selected 9 predictors. In order to test the impact of sampling effects on the

diffusion process, I considered the number of AIV sequences being included in discrete locations as a separate predictor. Then the predictor data of all provinces were gathered into regions according to the traditional division and economic divisions (Table 4.1).The final 10 predictors of each region type are summarized in Table 4.2. All selected predictors were transformed to log space and normalised prior to their incorporation in the GLM approach. These predictors are further split into subgroups according to the correlation test (Table 4.3) to ensure the predictors in the same GLM models are independent to each other. Independent predictors are jointly analysed in GLM models, and the impact on the overall transmission, as well as the source and recipient are tested.

TR	East	North	Northeast	Northwest	South Central	Southwest
1	14132.8	3766.9	3793.7	539.1	10115.2	1514.8
2	29912	5601	8604	5823	17588	7067
3	128885.7	17056.33	256744.1	87692.69	145843.83	563511.33
4	4.28	6.34	13.22	15.06	5.66	16.33
5	17.00	12.49	5.76	8.83	18.50	13.86
6	69.83	52.28	66.22	55.85	74.47	65.75
7	6222.31	201.61	1575.44	2366.17	7041.57	9751.36
8	6688.97	1790.22	529.06	357.86	1974.90	485.01
9	1363443.08	389900.12	326827.45	295949.22	1003544.75	296865.77
10	119	14	16	16	134	21

Table 4.2 A Predictors for traditional regions (TR) and economic divided regions (ER)
The data of 10 selected predictors per area for the 2 types of regions.
The unit of each predictors are listed below:
1 Poultry density (10000unit/km^2)
2 Number of farm product markets (region) (unit)
3 Stock Volume of Forest (10000 cu.m)
4 Percentage of Nature Reserves (%)
5 Average Temperature of Major Cities (℃)
6 Average Relative Humidity (%)
7 Surface Water Resources (100 million cu.m)
8 population density (10000 persons/km^2)
9 Freight by transportation total (10000 tons)
10 Sample size (unit)

ED	BER	Central	Northeast	Northwest	PRD	Southwest	YRD
1	7176.94	9959.04	3793.74	539.12	2751.70	2796.32	6845.58
2	6638	12094	8604	5823	10653	8808	21975
3	15751	129714	256744	87693	78620	610387	20826
4	5.43	5.73	39.67	15.06	4.92	14.26	3.62
5	13.72	15.76	17.28	8.83	20.92	15.37	16.65
6	53.61	68.72	66.22	55.85	78.38	68.55	69.44
7	317.88	5800.06	1575.44	2366.17	3527.56	11837.72	1733.64
8	2231.64	2061.80	529.06	357.86	897.94	683.14	5064.57
9	578895	1170352	326827	295949	340422	458222	505863
10	30	98	16	16	65	49	46

Table 4.2 B Predictors for economic divided regions (ER)

TR	1	2	3	4	5	6	7	8	9	10
1 Poultry population density	0.96	-0.3	-0.84	0.7	0.66	**0.28**	0.89	0.98	0.91	
2 Number of farm product markets		**-0.18**	-0.69	0.65	0.67	0.39	0.93	0.97	0.87	
3 Stock Volume of Forest			0.62	**-0.02**	0.35	0.7	-0.32	**-0.28**	**-0.21**	
4 Percentage of Nature Reserves				-0.65	**-0.26**	**0.08**	-0.76	-0.8	-0.73	
5 Average Temperature					0.52	0.65	0.6	0.76	0.82	
6 Average Relative Humidity						0.69	0.36	0.65	0.77	
7 Surface Water Resources							0.2	0.37	0.48	
8 Human population density								0.9	0.7	
9 Freight by transportation									0.94	
10 Sample size										

ED	1	2	3	4	5	6	7	8	9	10
1 Poultry population density		0.45	-0.319	-0.38	0.21	0.03	**-0.12**	0.63	0.86	0.60
2 Number of farm product markets				-0.25	-0.37	0.41	0.49	-0.08	0.87	**0.22**
3 Stock Volume of Forest				0.39	0.05	0.20	0.86	-0.46	-0.11	**-0.005**
4 Percentage of Nature Reserves						**-0.03**	-0.11	-0.06	-0.53	-0.40
5 Average Temperature						0.86	0.09	**0.15**	0.04	0.45
6 Average Relative Humidity								0.360	0.0803	**0.05**
7 Surface Water Resources								-0.29	0.20	0.42
8 Human population density									0.32	0.23
9 Freight by transportation										0.79
10 Sample size										

Table 4.3 Correlation test among predictors

The values in the cell represented the correlation coefficient of the two predictors in the column and row, which >0.5 may indicate a supported positive correlation and <-0.5 indicate a supported negative correlation. The predictors which have weak correlation with others are put into the same GLM analysis, with the correlation coefficient values marked in bold.

In addition, the possible impacts of the distribution of host population based on the AIV dataset are being considered as explanatory variables in the GLM model in this study. Two variables of host population were estimated in the regional level: (1) the proportion of domestic birds that shows the difference between domestic birds and wild birds; (2) the proportion of domestic Galliformes (mainly chickens), which describe the differences between chickens and ducks/geese (Figure 4.1).

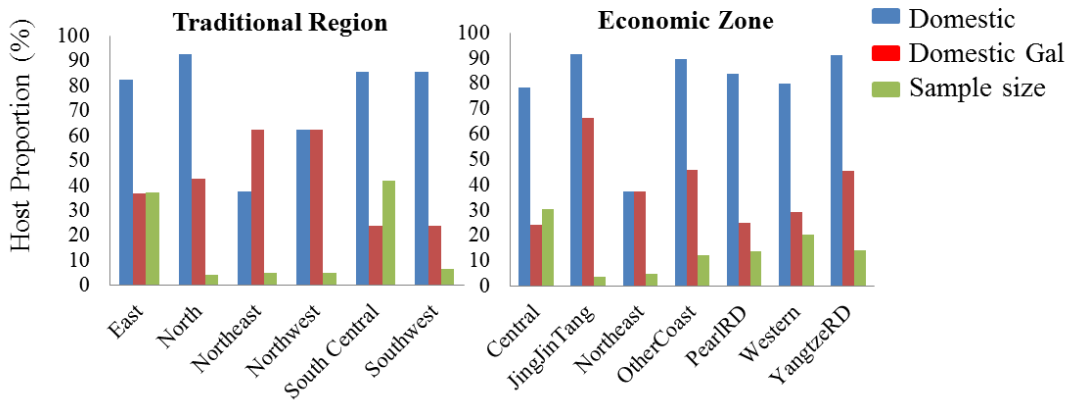


Figure 4.1 The host distribution of 320 Chinese AIV sequences in traditional and economic region types.
The proportion of domestic birds per region trait is shown in blue; domestic Galliformes per region trait is shown in red and the proportion of samples per region trait is shown in green.

4.2.4.2 Hypothesis test of predictors

To test the contribution of potential predictors for the CTMC transition rates among pairs of location traits (Λ_{ij}), an extension of the phylogenetic diffusion model was used to parameterize these rates of between-location movement in the phylogeographic model as a log-linear function (GLM) of a set of predictors (Lemey *et al.*, 2014).

The GLM specifies coefficients (β_k) for each predictor p_k, allowing the estimation of their contribution to the diffusion process and the indicator variables (δ_i) to model the inclusion or exclusion of each predictor. The Λ_{ij} are modelled by k predictors according to:

$$\log \Lambda_{ij} = \beta_1 \delta_1 \log(p_1) + \beta_2 \delta_2 \log(p_2) + \ldots + \beta_k \delta_i \log(p_k) \qquad (4.1)$$

The analysis involves averaging over all potential predictors for the discrete trait diffusion process (Drummond & Suchard, 2010; Lemey *et al.*, 2014). The prior and posterior expectation for δ_i can be considered as the inclusion probability for a predictor p_k,

which is expressed with a Bayes factor support. In this way, it was possible to simultaneously reconstruct the spatial transition processes in viral evolutionary history while identifying the variables that contribute significantly to this process, and compare candidate predictors while accounting for phylogenetic uncertainty (Faria *et al.*, 2013; Lemey *et al.*, 2014).

In this study, I used only pairs of uncorrelated predictors in separate analysis. To compare the results of separate analysis, I did a single GLM-diffusion analysis including all 10 predictors (in total/donor/recipient directions). I also did one with all predictors but excluding sample sizes to examine the possible sample size effect. In the analysis with many predictors, a binomial prior that prefers the inclusion of a small number of predictors was used and a multivariate normal updater for the coefficients which considers their correlation.

4.3 Results

4.3.1 Phylogeographic evolution of AIV in China

Of the 320 full genome AIV sequences that were analysed in this study, sequences are distributed more evenly in the last 14 years from 2000 to 2013, only 6 samples were collected from the early years (<2000). In this dataset, the earliest sequence in this study was the HPAI H5N1 sample isolated from Guangdong province in 1996 (Figure 4.2). The 320 AIV are classified according to different host types (domestic or wild bird) and are also categorized into seven different bird orders (shown in Table S1). The majority of AIV is sampled from domestic birds (81.25%, mainly chicken, duck and geese), and the most represented bird order is Anseriformes, represent the water birds (e.g., duck and geese) which takes up 58.1% within the domestic birds and 80% within the wild birds, respectively. The second largest order proportion (41% of domestic birds) is Galliformes (land bird) which only includes chickens.

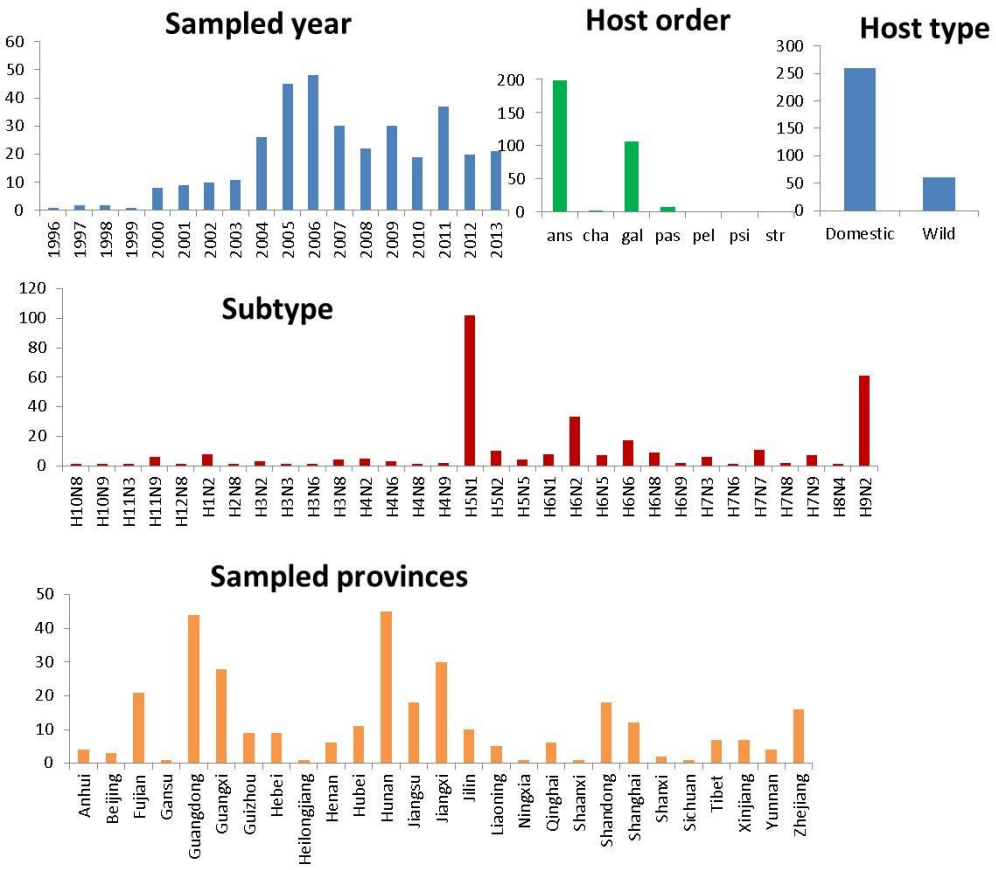


Figure 4.2 Distributions of 320 Chinese AIV sequences in the sampled time, host order, subtype and sampled provinces.

The host distribution in the study data set corresponds to the distribution of the subtypes of AIV. Among the 12 HA subtypes being analysed in this study (subtype H1 to H12), H5, H6, H7 and H9 are mainly circulating in domestic birds and are more prevalent than the other subtypes. The dominant viral samples are of the H5N1 subtype, which is the most important highly pathogenic AIV in China, and is monitored across China since the global spread begins at the year 2005. The second largest population is H9N2 subtype, which is mainly identified in chickens and is endemic over all in China. It is being recognized as the donor of internal segments of HP H5N1 and the human H7N9 in 2013. These sequences were collected from 27 out of total 32 provinces in China (including 22 provinces, 5 autonomous regions, four municipalities, and 3 special administrative regions), of

which the Guangdong, Fujian, Hunan, Jiangsu and Jiangxi provinces had the highest number of sequences Figure 4.1.

Bayesian phylogenetic trees of the 6 internal segments of the 320 Chinese AIV were generated. The evolutionary rates for five segments (PB2, PB1, PA, NP and M) are from 2.47×10^{-03} to 4.36×10^{-03} substitutions per site per year. The estimated most recent common ancestors (TMRCA) of the five internal segments are dated back from 1955 to 1960.4. Due to its deep divergent alleles A and B, the NS segment has a significantly lower substitution rate (1.66×10^{-03}) and earlier TMRCA (1553.8) (Treanor *et al.*, 1989a).

To further explore the spatial diffusion patterns of AIV across different regions in China, I used a discrete trait procedure on the empirical trees sampled from the original phylogenies of the six internal segments, with each taxon being classified and coloured according to different geographic region types, which were classified based on the geographic distance (TR) or on the economic relationship (ED). Co-circulation of multiple lineages and diverse gene flows of avian influenza virus between different areas of China were detected. For the TR areas, I found that South Central is the source of the majority of lineages (Figure 4.3A). Multiple introductions are observed among the economic divided region type, with the Pan-Pearl River Delta (including Guangdong and Fujian provinces) as the main ancestor of all AIV, and the Central area represented as the source of two main clades (Figure 4.3B and Figure A4.1-4.5).

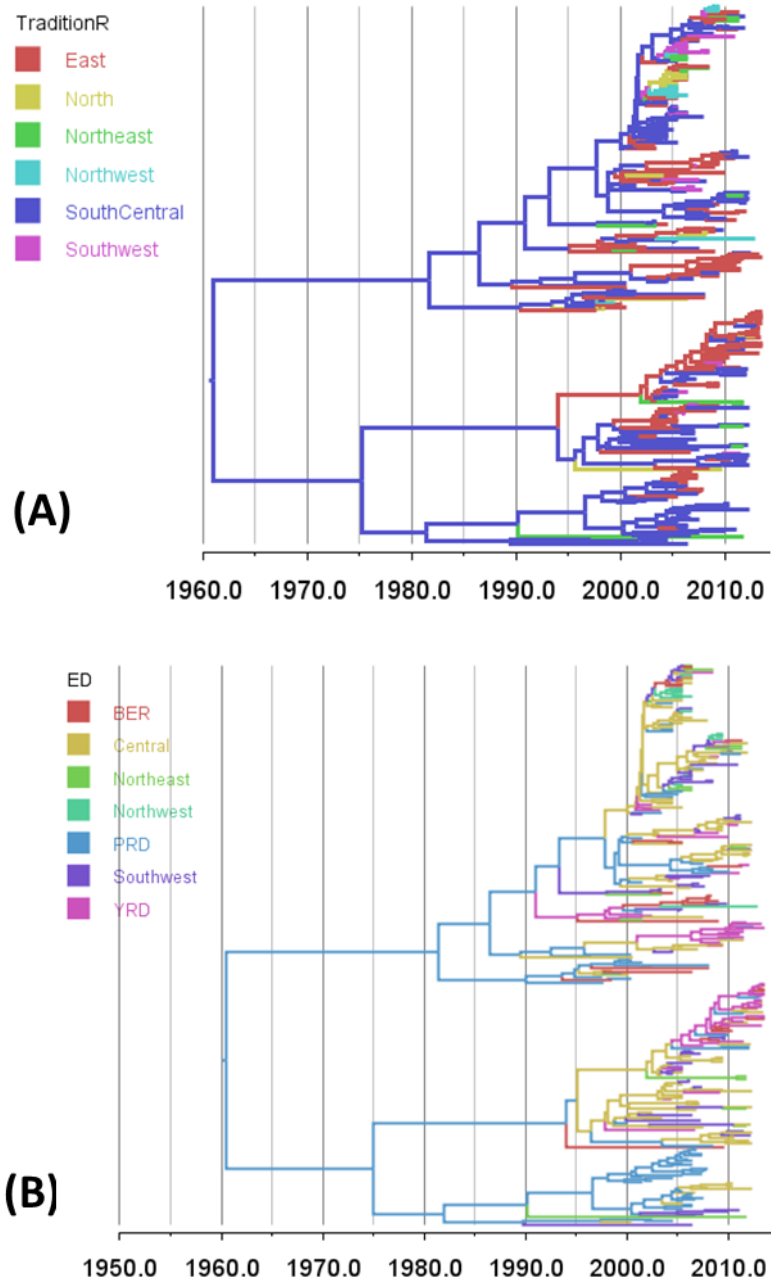


Figure 4.3 Bayesian MCC phylogenies and between regions diffusion networks on PB2 gene segment of AIV in China.
The sequences are classified according to their variant designation in (1) TR: Traditional geographic region; (2) ED: Economic divide region. Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left.
A: Tree phylogeny with branches coloured according to the Tradition regions
B: Tree phylogeny with branches coloured according to the Economic zones

The strength of phylogeny-trait association was further tested on different regions types. The results were summarized in Table 4.4, in which column 1 is the types of region classification; Column 2 shows the three statistics: Association Index (AI), Parsimony Score (PS) and Maximum monophyletic clade (MC) size; Column 3 indicates the area traits in the region type which has the corresponding MC size; Column 5-7 correspond to posterior estimates of observed and expected values and the p value for the AI, PS and MC metricsLower AI and PS values represent strong phylogeny–trait association; while larger values of mean MC indicate increased phylogeny-trait association and the significance indicated by p-value (<0.05). Statistical analysis of AI and PS of the full phylogeny revealed that for each type of region, the total number of migration events occurring in each category was less than expected under the null hypothesis (Table 4.4) (P <0.005). This indicates that there is a sufficient correlation between these geographical characters and the phylogeny and suggested that geography (both according to geographic direction and economic area) is a key factor determining gene flow of AIV in China. I found that the economic region model has a stronger clustering (lower AI and PS) in the tree phylogeny than the traditional classifications.

The significance of the clustering of areas within each region type was indicated by the estimation of the maximum monophyletic clade (MC) size. There is evidence of clustering of AIV belonging to each of the six areas in the traditional region; two areas (Northeast and Central) that belong to the economic region also have a strong clustering support. However, the East Coast area is the most dominant group (141/320 sequences) among economic regions with a significantly higher value (mean MC=13.95) than the other areas (mean MC=3 to 6). This shows a very strong phylogeny-trait association thus indicating strong gene flow within the areas rather than to the other areas. To further identify the pattern of gene flow within east coast region, I further divided it into 3 economic zones: Pan-Pearl River Delta (PRD), Yangtze River Delta (YRD) and Bohai Economic Rim (BER). The three smaller economic zones showed evidence of trait-phylogeny correlation with the MC value from 2 to 4, with strong supported. In addition, the Western area showed poor statistical support, indicating the AIV are randomly distributed alone the tree phylogeny. This can be explained by the limited number of sequences (65/320)

sampled from 10 provinces divided into a Northwest and a Southwest area with supported clustering for both with mean MC=2.

Region	Statistic		Observed mean	Null mean	P-value
TR	AI		14.10 (13.13, 15.17)	22.56 (20.68, 24.42)	<0.005
	PS		114.70 (112, 118)	149.31 (142.70, 156.17)	<0.005
	MC (6)	East	8.89 (8, 11)	3.59 (2.63, 5.11)	0.002
		North	2 (2, 2)	1.16 (1, 2)	0.05
		Northeast	1.99 (1.9, 2)	1.21 (1, 2)	0.06
		Northwest	1.98 (1.9, 2)	1.19 (1, 2)	0.06
		SouthCentral	12.5 (12, 13)	4 (3, 6)	0.002
		Southwest	3 (2.9, 3)	1.34 (1, 2)	0.006
ER	AI		12.67 (12.10, 13.17)	27.42 (24, 29.20)	<0.005
	PS		103.40 (103, 104.1)	148.41 (148, 149.40)	<0.005
	MC (4)	**East Coast**	13.95 (13.9, 14)	4.11 (3.01, 6)	0.001
		Central	6 (5.9, 6)	3.1 (2.07, 4.17)	0.008
		Northeast	4.27(3, 7)	1.97 (1.25.3)	0.001
		Western	3 (3, 3)	2.37 (1.95, 3.18)	0.14
ED	AI		15.54 (14.46, 16.68)	27.42 (25.67, 29)	<0.005
	PS		137.29 (134, 140)	192.83 (185.83, 199. 37)	<0.005
	MC (7)	BER	2 (2, 2)	1.12(1, 1.99)	0.03
		PRD	4.27 (3, 7)	1.97 (1.25, 3)	0.001
		YRD	4.05 (3.9, 4)	2 (1.25, 3)	0.004
		Central	6 (6, 6)	3.1 (2.07, 4. 17)	0.008
		Northeast	1.98(1.9, 2)	1.2 (1, 2)	0.06
		Northwest	1.98 (1.9, 2)	1.19 (1, 2)	0.06
		Southwest	3 (2.9, 3)	2.9 (2, 3)	0.06

Table 4.4 Results of trait-phylogeny association of different region types

I also compared the trait-phylogeny interaction among the 6 internal segments using the traits of economic regions (Table 4.5). The last column represented the number of sequences of certain states on the tree phylogeny. In the full phylogenies, AI and PS varied among 6 internal segments, ranging from PB2 (AI mean=11.6; PS=96.9) to NS (mean=17.7, PS=117.6), indicating an increase in gene flow from PB2 and NS. Similar results are found in the MC size of the East coast area, which is the area with the majority of AIV, the MC size of PB2 (mean=20) is larger than the other five segments, while NS has the lowest value (mean=7.7),

which indicated it is more randomly distributed, which could be attributed to the Allele A and B divergence of NS segment (Ludwig *et al.*, 1991).

Statistic			Observed mean	Null mean	P-value	Sample size
AI		PB2	11.57 (10.8, 12.36)	22.8 (20.6, 24.8)	<0.001	
		PB1	14.35 (13.4, 15.25)	23.21 (21.26, 25.22)	<0.001	
		PA	13.53 (12.64, 14.34)	22.8 (20.6, 24.8)	<0.001	
		NP	12.67 (11.62, 13.74)	22.42 (20.67, 25)	<0.001	
		M	15.76 (14.5, 16.71)	22.95(21.14, 24.71)	<0.001	
		NS	17.67 (16.43, 18.95)	23.3 (21.46, 25.29)	<0.001	
PS		PB2	96.87 (95, 99)	148.6 (141.7, 155.3)	<0.001	
		PB1	106.60 (105, 109)	148.5 (141.1, 155.1)	<0.001	
		PA	105.06 (102, 108)	148.6 (141.7, 155.3)	<0.001	
		NP	103.4 (101, 106)	148.4 (141.45, 154.75)	<0.001	
		M	113.2 (110, 116)	148.4 (141.52, 154.9)	<0.001	
		NS	117.61 (114, 121)	148.4 (141.96, 154.9)	<0.001	
MC	EastCoast	PB2	20 (19.98, 20)	4.14 (3, 6)	0.001	141
		PB1	12.97 (12.9, 13)	4.2 (3, 6)	0.001	
		PA	10.09 (10, 11)	4.15 (3, 6)	0.002	
		NP	13.95 (13.9, 14)	4.11 (3, 6)	0.001	
		M	9.11 (8, 12)	4.3 (3, 6)	0.005	
		NS	7.68 (5, 10)	4.2 (3, 6)	0.002	
	Central	PB2	4 (3.9, 4)	3.07 (2, 4)	0.12	98
		PB1	6.42 (6, 7)	3.11 (2, 4)	0.01	
		PA	4.11 (3.9, 4)	3.11 (2, 4.2)	0.13	
		NP	6 (6, 6)	3.1 (2.07, 4.17)	0.008	
		M	5 (4.9, 5)	3.12 (2.23, 4.21)	0.01	
		NS	6.12 (6, 7)	3.13 (2.23, 4.24)	0.004	
	Western	PB2	6.99 (6.9, 7)	2.38 (2, 3)	0.1	16
		PB1	3 (3, 3)	2.38 (2, 3)	0.2	
		PA	4 (3.9, 4)	2.38 (2, 3)	0.17	
		NP	3 (2.9, 3)	2.37 (2, 3)	0.14	
		M	3 (3, 3)	2.37 (2, 3)	0.15	
		NS	3 (3, 3)	2.38 (2, 3)	0.14	
	Northeast	PB2	3 (2.9, 3)	1.22 (1, 2)	0.001	65
		PB1	2.02 (1.9, 2)	1.22 (1, 2)	0.1	
		PA	3 (3, 3)	1.22 (1, 2)	0.003	
		NP	4.27 (3, 7)	1.97 (1.25, 2)	0.001	
		M	2.97 (2.9, 3)	1.21 (1, 2)	0.003	
		NS	1.34 (1, 2)	1.22 (1, 2)	0.1	

Table 4.5 Results of trait-phylogeny association of different segments

4.3.2 Quantified spatial diffusion patterns

I quantified the pattern of spatial diffusion of Chinese AIV under a BSSVS procedure. The transmission network among traditional regions (TR) as well as economic zones (ED) is summarized on the maps in Figure 4.4 (A, B). The diffusion rate and BF support for each region and each internal segment are summarized in

Table 4.7 and Table 4.8. In general, BSSVS procedure identified supported spatial transition pairs (with BF>5) among segments in both regions types (2 to 3 pairs in TR; 4 to 9 pairs in ED), but not all the strongly supported BF rates (BF>100) actually correspond to a high magnitude in host transition rates. The average BF supported diffusion rate of six internal segments also vary (For TR: PB2=0.31; PB1=0.3; PA=0.46; NP=0.24; M=1.18 and NS=0.91; For ED: PB2=0.35; PB1=0.42; PA=0.5; NP=0.42; M=0.63 and NS=0.70).

Among the 15 possible viral transmission pairs among the traditional regions, 3 pairs are found with significant BF support in PB2, PA and NP segments, while 2 pairs are found in the other segments. The South Central area (composed of Guangdong, Guangxi, Henan, Hubei, Hunan) plays a role as a major source region while the AIV are transmitted to East (composed of Shandong, Jiangsu, Anhui, Jiangxi, Zhejiang, Fujian, Shanghai), Northwest (Gansu, Ningxia, Qinghai, Xinjiang, Shaanxi) and Southwest areas (Guizhou, Sichuan, Tibet, Yunnan) (Figure 4.4A). This also corresponds to the region-annotated MCC tree and shows South Central occupied the majority of ancestral nodes (Figure 4.3 A). The most significant gene flow is from South Central to East area (mean rate=0.47, BF=3703). The Northwest area is also likely to accept AIV virus from other areas. Invasions were also found from Northeast or from South Central to other areas, but only with moderate support (3<BF<6) (Table 4.6).

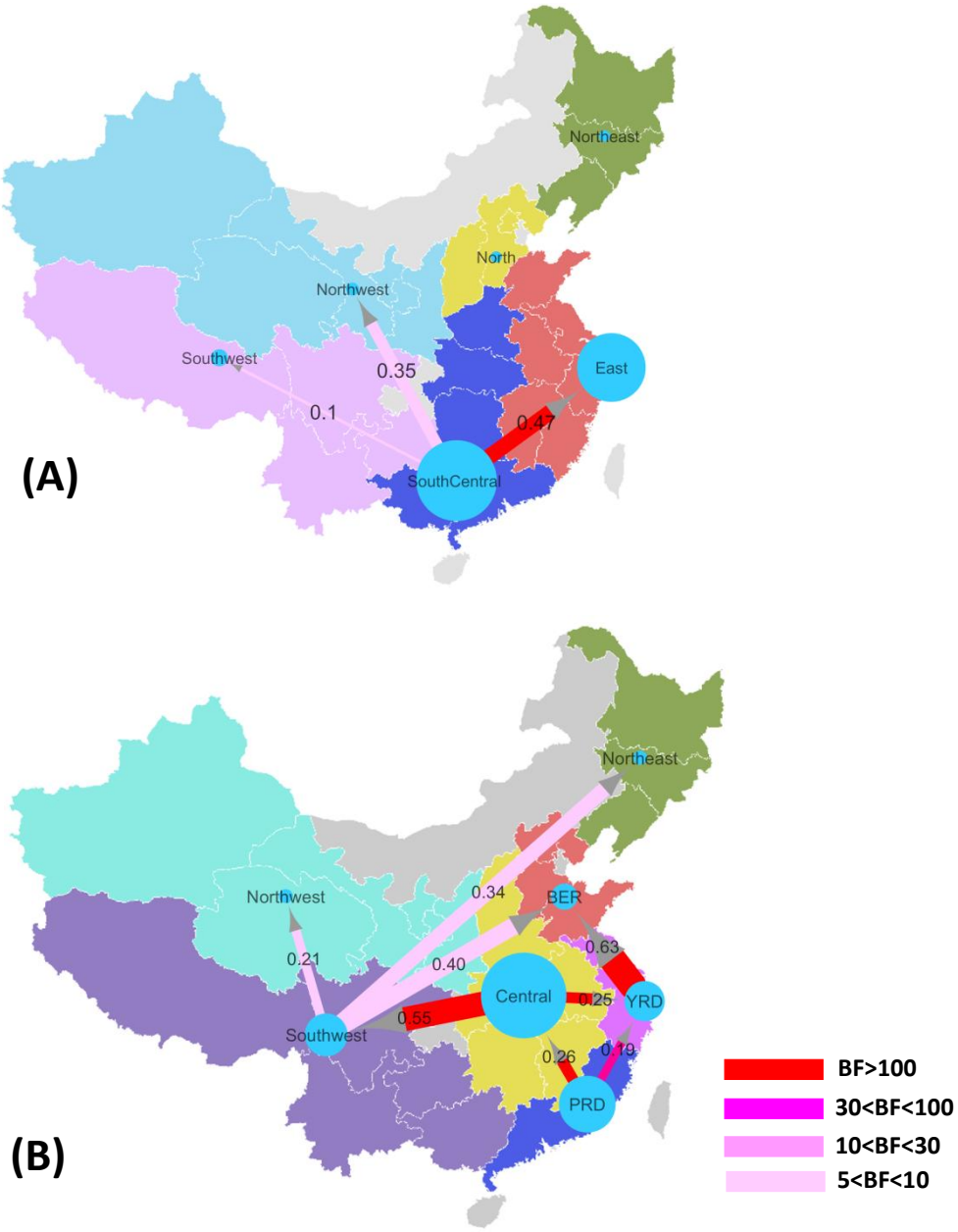


Figure 4.4 Diffusion networks with quantified diffusion rate and BF support.
Quantified diffusion rate between regions and BF support were estimated by a irreversible discrete trait model on the phylogeny of PB2 segment.
Areas in each region types are labelled by a certain colour which is the same colour in the region annotated phylogenetic trees in Figure 4.3. The size of blue circles represented the AIV sample size in each area.. The diffusion rate and statistical support for all 6 internal segments are summarized in Table 4.7 and 4.8.
A: Diffusion network of AIV among traditional regions
B: Diffusion network of AIV among economic zones

Different viral diffusion patterns are identified among different economic zones (Figure 4.4B and Table 4.7). Two major sources of AIV migration are found. One is the Central area (composed of Shanxi, Anhui, Henan, Hubei, Hunan, Jiangxi). Highly supported virus transmissions (BF>100) are identified from Central area to southwest area and YRD (Shanghai, Jiangsu, Zhejiang); the other source is PRD (Guangdong and Fujian) with high support transmit to YRD (main=0.19, BF=43) and Central areas (mean=0.26, BF=158). In addition, the Guangxi province might also play an important role as the source location when it is included in the southwest area in ED type. There are migration routes from southwest to Northeast, Northwest and BER. These routes have weak support (BF<10) (Figure4.4 B), and do not appear in the TR region grouping. In comparison, AIV are mainly transmitted to two economic zones: YRD and BER (Beijing, Hebei and Shandong). Specifically, YRD mainly receives AIV from Central (mean= 0.19 to 0.28, BF>100) and moderate supported invasion from PRD (mean= 0.18 to 0.21, BF = 10 to 43). A high transmission rate is identified from YRD to BER (mean rate= 0.62 to 0.97), which is consistent in all six segments with high support (BF= 573 to 5172).

PB2	**Origin**	SouthCentral	SouthCentral	SouthCentral
	Destination	East	Northwest	Southwest
	Mean rate	0.47	0.35	0.1
	Indicator	1	0.84	0.84
	BF	3703	6	6
PB1	**Origin**	SouthCentral	SouthCentral	
	Destination	East	Southwest	
	Mean rate	0.49	0.1	
	Indicator	1	0.81	
	BF	1233	5	
PA	**Origin**	SouthCentral	Northeast	SouthCentral
	Destination	East	Northwest	Northwest
	Mean rate	0.33	0.96	0.1
	Indicator	1	0.86	0.89
	BF	426	8	10
NP	**Origin**	SouthCentral	SouthCentral	SouthCentral
	Destination	East	Northwest	Southwest
	Mean rate	0.42	0.13	0.14
	Indicator	0.91	0.88	0.87
	BF	410	9	8
M	**Origin**	SouthCentral	Southwest	
	Destination	East	Northwest	
	Mean rate	0.44	1.92	
	Indicator	1	0.88	
	BF	1233	9	
NS	**Origin**	SouthCentral	Southwest	
	Destination	East	Northwest	
	Mean rate	0.49	1.32	
	Indicator	1	0.8	
	BF	1233	5	

Table 4.6. Transmission rate and statistical support between areas belong to traditional regions

States=6; only presents the transitions with BF support > 5.

PB 2	**Origin**	Southwest	YRD	PRD	Central	Central	Southwest	Southwest	PRD
	Destination	BER	BER	Central	Southwest	YRD	Northeast	Northwest	YRD
	mean	0.40	0.63	0.26	0.55	0.25	0.34	0.21	0.19
	Indicator	0.84	1.00	0.99	1.00	0.99	0.85	0.85	0.96
	BF	9	3102	158	815	169	10	10	43
PB 1	**Origin**	YRD	Central	Central	Central	Northwest	PRD	PRD	
	Destination	BER	Northwest	Southwest	YRD	Northeast	Southwest	YRD	
	mean	0.73	0.14	0.60	0.19	0.92	0.20	0.18	
	Indicator	1.00	0.78	1.00	0.93	0.99	0.73	0.86	
	BF	7758	6	1409	21	286	5	10	
PA	**Origin**	YRD	Central	Central	Central	Northwest	PRD	PRD	
	Destination	BER	Northwest	Southwest	YRD	Northeast	Southwest	YRD	
	mean	1.17	0.20	0.61	0.23	0.84	0.26	0.21	
	Indicator	1.00	0.86	1.00	0.81	0.96	0.80	0.91	
	BF	1722	10	815	7	39	7	18	
NP	**Origin**	Southwest	YRD	Central	Central	Central	Southwest	PRD	
	Destination	BER	BER	Northwest	Southwest	YRD	Northeast	YRD	
	mean	0.26	0.62	0.23	0.50	0.24	0.84	0.24	
	Indicator	0.84	1.00	0.98	1.00	0.97	0.79	0.98	
	BF	9	1938	76	774	50	7	76	
M	**Origin**	YRD	PRD	Central					
	Destination	BER	Central	Southwest					
	mean	0.74	0.41	0.73					
	Indicator	0.98	0.89	1.00					
	BF	91	15	418					
NS	**Origin**	YRD	Central	PRD	Southwest				
	Destination	BER	Southwest	Southwest	YRD				
	mean	0.66	0.91	0.39	0.83				
	Indicator	0.75	0.99	0.94	0.84				
	BF	5	121	26	9				

Table 4.7. Transmission rate and statistical support between areas belong to economic zones

States=7; Only presents the transitions with BF support > 5.

To identify the possible bidirectional movement of AIV between regions, asymmetric diffusion models were applied. The only bidirectional transmission is

found between South Central and East in PB2 segment, with transmission rate=0.43, BF=202 from South Central to East and rate=0.24, BF=6 conversely. Bidirectional transitions are found between YRD and PRD but with weak support, with the transition from PRD to YRD mean= 0.39, BF= 3, and the reverse direction with mean rate= 0.3, BF= 4.

4.3.3 Hypothesis-based spatial analysis

To identify the possible factors that determine the viral diffusion of avian influenza among different regions of AIV in China, I employed a generalized linear model (GLM), which is an extension of a Bayesian phylogenetic diffusion model on 6 internal genes independently.

I selected 10 potential predictors over categories including economic (agricultural) activity, environment, climate as well as the sample size of AIV (Table 4.2). The identified determinants of spatial AIV (PB2 gene segment as an example) diffusion are summarized in Figure 4.5, which shows the mean and 95% HPD of the BF support for each predictor as well as the conditional effect sizes (coefficient) on a log scale. The conditional effect sizes are the estimated effect sizes with indicator =1 ($\beta i|\delta i = 1$), which summarize the coefficients conditional on the predictor being included in the GLM model. The results of all six internal segments and all region types are summarized in Figure 4.6 and Table A4.2.

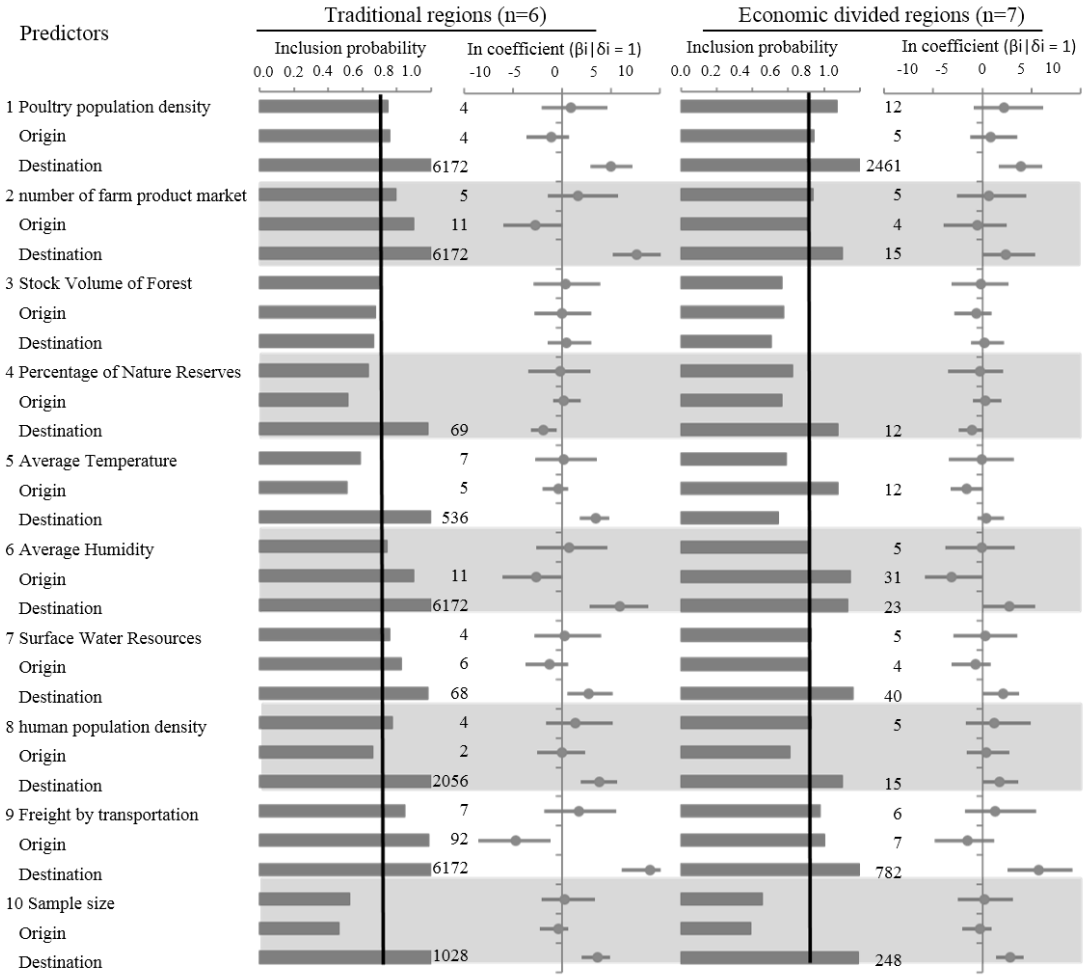


Figure 4.5 Predictors of Chinese AIV spatial diffusion.
The inclusion probabilities (IP) shown in the bar plots are defined by the indicator expectations which reflect the frequency at which the predictor is included in the model and therefore represent the support for the predictor; The predictors with Bayes factor (BF) are indicated with the correspondent BF value (those with BF<5 are not shown); The contribution of each predictor, when included in the model is represented by the conditional effect sizes (cES) and 95% HPD on a log scale (ß|δi = 1). The zero line not included in the credible intervals indicated the positive correlation (on the right, above 0) or the negative correlation (on the left, below 0).

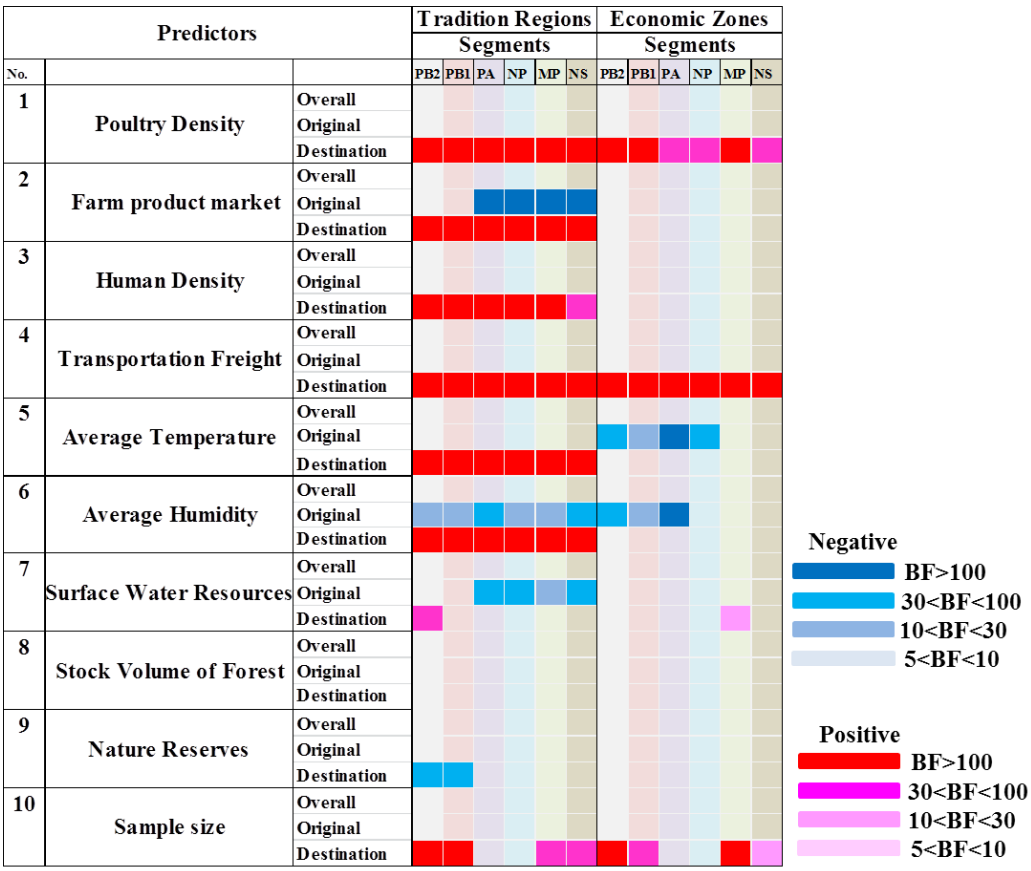


Figure 4.6 Predictors of the Chinese AIV spatial diffusion on 6 internal genes segments. For the 10 predictors, their correlations to the AIV diffusion in two region types(tradition regions and economic zones) were summarized in the cells, with rows represented certain predictor and the columns represented six internal segments (PB2, PB1, PA, NP, M, NS) of Chinese AIV dataset. The markers of the correlation with statistical supports are on the right. Colour in red indicating positive correlation. Colour in blue indicating negative correlation. The rank of colours indicating the Bayes factor (BF) support from low to high; Cell in blank indicating either no correlation or no BF support was detected. The inclusion probabilities and the log coefficients for each predictor for each region type in all six internal segments are summarized in Table A4.2.

In general, I found all the potential predictors have strongly directional impacts on the diffusion of AIV in China, which means that the predictors have significant impacts on the sources and/or the recipients rather than the overall transmission. Furthermore, the impacts are not always consistent among the six internal segments, which can be reflected in both the effected sample size and the BF support. In addition, there are more statistical supported determinants on the diffusion between traditional regions than the economic divided regions, but same trends (positive or negative) are found among those predictors with impacts in both

region types. Moreover, the variability within each predictor is not correlated with the GLM results.

Specifically, I found the agro-economic predictors played decisive roles in viral spread in both traditional regions and economic zones in China. To be specific, poultry population density and productivity have a significant positive correlation in geographic divided region and economic zones, consistent in all six segments with high estimated size of the effect (mean =3.5 to 5). The statistical support for its inclusion in the model (BF=256 to 6172) indicated that the locations with high poultry density and productivity are more likely to accept the viral migration. The GLM diffusion model also revealed that the number of farm product markets (selling poultry product and also the live poultry) is also a driver of the AIV dissemination in China. It has a negative correlation (mean=-4.9 to -5, BF=166 to 293) in the source (in PA, NP, M and NS segments) and positive impact in the destination regions (mean= 5.9 to 7.6, BF= 6172) in all six segments, indicating the direction of high viral transmission rate is towards the location with high demands of poultry products. The productivity and sale of poultry is also linked with the impact of human actions. Human population density in destination location, which has varied support (mean=2.8 to 3.8, BF=38 to 2056) in different segments and may suggest that commuting is more likely to occur into high dense subpopulations. Interestingly, the capacity of freight transportation has the most significant impact on the viral diffusion (mean=7.7 to 8.9, BF=6172), indicating the movement of avian influenza virus is mainly dependent on the human transportation, not through wild bird migration. In addition, climate factors (temperature and relative humidity) are highly possible to have impact on the diffusion pattern. Locations having comparably higher temperature (mean=2.9 to 3.4, BF=122 to 6172) and higher humidity (mean=4.2 to 5.9, BF=122 to 6172) are more likely to accept the virus than those with lower values.

In contrast, the environmental factors (coverage of forest, water and natural reserves) impact on the viral diffusion in some degree. Surface water resources has a negative impact on the source regions in PA, NP, M and NS segments, suggested that AIV are not likely to be transmitted from region that is rich in water resources;

The coverage of natural reserves has slight negative correlation with the recipient location of the viral diffusion in only PB2 and PB1 segments (mean= -1.95 and -1.85, BF=56 and 69), while forest coverage does not yield noticeable support in any segments.

The possible impacts of the distribution of host population based on the AIV dataset (the proportion of domestic birds and the proportion of domestic Galliformes) are being considered as explanatory variables in the GLM model in this study (Figure 4.1). Neither of these have any statistical supported for impact on viral diffusion among any type of regions.

4.4 Discussion

I investigated the spatial patterns of influenza A virus in bird population in mainland China. Clustering was found in each area within the geographic regions and economic zones, indicating the evolution of AIV is strongly geographically correlated. In addition, by applying phylogeographical diffusion models, the key source/sink locations are captured and the rates of movement are simultaneously informed from the evolution history of viral transmission.

I further incorporated the ecological and agriculture data in the regional categories into the viral phylogeography models and simultaneously identified determinants of the between region diffusion with quantified coefficient and the correlated Bayes factor support. I highlighted that AIV in China are mainly exported from YRD (including Guangdong, Fujian provinces) and the South Central area (including Hunan, Hubei, Jiangxi provinces). YRD has historically been referred as a hypothetical influenza epicentre (Martin *et al.*, 2011), while the multiple viral exposures from South Central areas are probably due to its location along the Central Asian Flyway. For example, H7 influenza viruses which were identified from East Asian migratory waterfowl were introduced into domestic ducks in China on several occasions during the past decade (Lam *et al.*, 2013). In addition, H7 viruses with

multiple NA subtypes were found in ducks in Jiangxi province (in South Central), suggesting an epidemiological bridge from migratory birds to sentinel farm ducks and then to market birds. Epidemiological studies also suggested that HPAI H5N1 outbreaks in poultry mostly occurred in areas which overlapped with habitats for wild birds, whereas outbreaks in wild birds were mainly found in areas where food and shelter are available (Si *et al.*, 2013). The bidirectional movement of wild birds among the provinces along the flyway route in south central areas are identified in studies tracking the movements of migrating wild birds, and the transmission of HPAI H5N1 to Qinghai lake (Northwest) have also been identified (Newman *et al.*, 2012; Prosser *et al.*, 2011b; Zhang *et al.*, 2013). The HA and NA genes of the recent human H7N9 virus may originate from duck avian influenza viruses, which might have obtained the viral genes from wild ducks along the east Asian flyway a year before (Liu *et al.*, 2013).

In contrast, the East Coast areas especially YRD are a major target for receiving the viral strains. The importance of the east coastal region as the hotspot of avian influenza outbreaks is supported by previous studies: 1) The potential hotspots for H5N1 reassortment being identified in previous studies are the coastal provinces bordering the South China Sea, East China Sea (Guangdong, Jiangsu, Shanghai, and Zhejiang Provinces) and Central China (Hunan and Sichuan Provinces) (Fuller *et al.*, 2013). 2) HPAIV H5N1 risk distributions are identified in both south and north of the YRD (Martin *et al.*, 2011). In addition, the recent H7N9 outbreaks occurred initially in the YRD, and then spread along both south and north of the East Coast lines (Fang *et al.*, 2013).

The transmission between Southwest and Northwest is supported by the studies that identified the outbreaks reported in poultry or wild birds on the Qinghai-Tibet Plateau. The initial outbreaks occurred in poultry then occurred in wild birds afterwards. The locations of outbreaks started in Lasa (in Xinjiang province) then were transmitted to Tibet (Prosser *et al.*, 2011a). No significant viral exposure has been identified from the Northeast China. Northeast China has a rapid development of intensive chicken production with intensive poultry production

systems. It has invested in disease prevention measures and applies mass vaccination of their flocks in order to prevent viral infection and spread (Martin *et al.*, 2011).

Most of China's economic production and growth originate in the coastal regions. The PRD and YRD have served as the commercial arteries for China's economic development and major manufacturing and transportation hubs of China's international trade since the Chinese economic reform in the late 1970s. BER rivals PRD and YRD in recent years and is rising as a vital economic region in North. Cities in the three economic zones have built a complete network for water, land and air transportation, and are interconnected by highways and railways. Therefore, the strong transport connections among the three economic zones can explain the strong viral flow linkages being found in our study, especially between PRD and YRD, which showed a trend of bidirectional transmission. In addition, the connection can also be the explanation of the extremely strong positive correlation between the capability of transportation and the gene flow that identified in the GLM results.

My results indicate that poultry density is strongly and positively correlated with viral diffusion in the AIV sink regions. There are more than 13 billion chickens in China, of which 60% are raised on small farms (WHO, 2005). Domestic ducks are positioned at the interface between wild aquatic birds and terrestrial poultry and therefore play an important role in the ecology of AIV (Huang *et al.*, 2010). Currently, more than 75% of the world's domestic duck population is bred in China (Van Boeckel *et al.*, 2011). Previous studies indicated that the spread of HPAIV H5N1 and occurrence of clinical disease outbreaks is facilitated in regions where the density of chickens is particularly high (Martin *et al.*, 2011). Regions with high chicken densities are found across much of eastern China, particularly the Yellow River Basin and high duck densities occur in southeastern China and the Sichuan Basin; while geese were least abundant, but exhibited consistent patterns with highest densities in Sichuan and parts of Guangdong (Prosser *et al.*, 2011c). Shandong and Hebei (belong to the Bohai economic Rim) are the major sources of the output of poultry meat (mainly chickens) and have the highest poultry density, whereas the PRD has much less poultry being raised in the commercial farms with more being raised as free range and backyard.

I found the regions with higher number of live bird markets are more likely to accept the AIV gene flow from other regions. Most of the live poultry markets in China and Southeast Asia have been characterized by poor sanitation where the domestic ducks and geese are being housed together with chickens (Deng *et al.*, 2013). Surveillance and phylogenetic studies showed that the current human H7N9 outbreak virus has spread over a large geographic region and is prevalent in poultry (mainly chickens) and in live poultry markets in the outbreak areas, which are thought to be the immediate source of human infections (Lam *et al.*, 2013). Human population density is also found to be strongly associated with AIV diffusion. AIV epidemiological processes are more likely to occur in highly-populated areas, and there are higher likelihood of outbreak detections and higher possibilities of HPAIV H5N1 transmission through trade and farming-related activities (Martin *et al.*, 2011).

The climate and natural factors have little or no impact on AIV diffusion among the economic regions, but have varied impact on the viral spatial diffusion among the geographic regions. Firstly, I found that AIV are highly likely to be transmitted from areas with lower temperature and relative humidity to those with higher temperature and humidity, which also can be explained by the warmer regions as a rest place for migrating wild birds. A previous study has shown that the aerosol spread of influenza virus is dependent upon both relative humidity and temperature, and high temperature and humidity will limit the transmission of influenza between humans (Lowen *et al.*, 2007). However, the transmission of AIV is mainly by faeces and contact between bird populations (sometimes by faecal-oral route and fomite transmissions), and the pathway also differs between wild birds and terrestrial birds, thus this is not expected to be influenced by the same factors as airborne transmission in humans. I found that the surface water resources and the coverage of natural reserves on the exposure of virus have a negative impact on the AIV transmission, although only mild support was detected. Nevertheless, this is consistent with the suggestion that AIV are likely to be transmitted to areas rich in water resources and natural reserves such as wetlands and lakes that are important stopover, breeding, or wintering sites for migratory water birds (Boender *et al.*, 2007; Xiao *et al.*, 2007). However, this is not the single most important factor for AIV transmission. Finally since the agricultural and economics statistics are updated each

year, it would be interesting to take into consideration the change of the agricultural and economic factors through the same years as the phylogenetic time scales to explain the AIV evolution for future studies.

It is crucial to recognize that if variables in the GLM are correlated, the results can be misleading. Previous similar study also showed the sample size absorbed effect of other predictors in a GLM model (Lemey *et al.*, 2014). Therefore, I used pairs of uncorrelated predictors in separate analysis in this study. To compare the results of separate analysis, I did a single GLM-diffusion analysis including all 10 predictors. The results showed no predictor expects sample size has correlation with viral spatial diffusion. In addition, I did analysis with all predictors but excluding sample sizes, which gave weaker but still significant coefficient and BF support comparing to those estimated by using separate analysis, indicating the inclusion/exclusion of sample size with other predictors in the same GLM analysis have strong impact on the results.

In summary, this study describes the AIV diffusion pattern among geographic regions and economic zones in China by looking at the evolution of internal segments of AIV sequences. I found that transmission patterns of avian influenza virus in China are mainly driven by exposure from the South region and the central region, which are composed of the areas along the East Asian migration routes also with high density of poultry industry; while virus tends to be imported into the east coast areas that are economically developed. I highlight that the transmission between regions of AIV in China is largely driven by human activity including transportation of the poultry product and trade rather than the natural and ecology conditions. The impact of economic development on the transportation of domestic animals and the agricultural practices may enhance spread and evolution of influenza in certain areas, and good biosecurity measures as adopted in the North East region may limit viral transmission.

Appendix

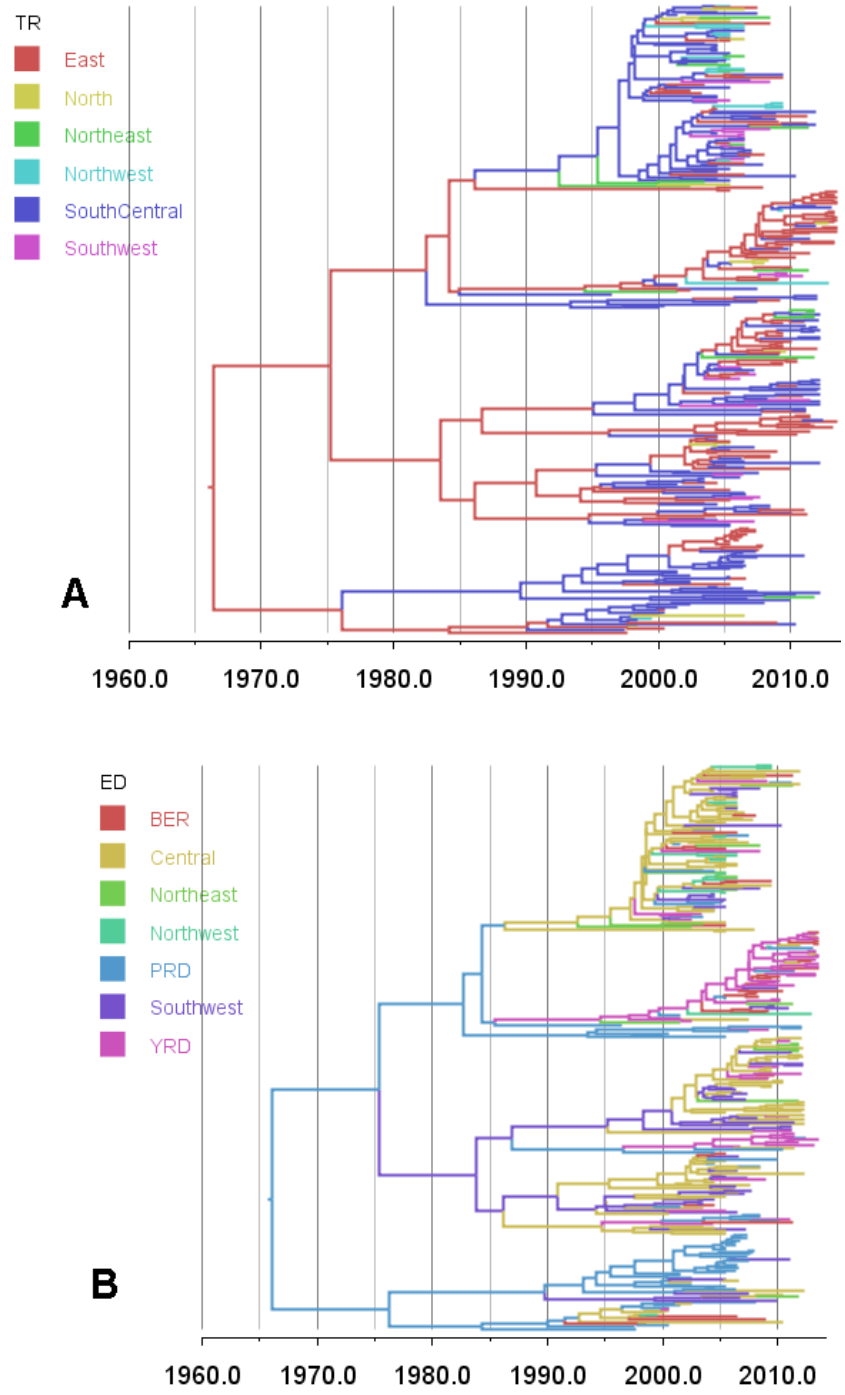


Figure A4.1 Bayesian MCC phylogenies of PB1 segment.
A: PB1 phylogeny, coloured by traditional regions
B: PB1 phylogeny coloured by economic zones
Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left.

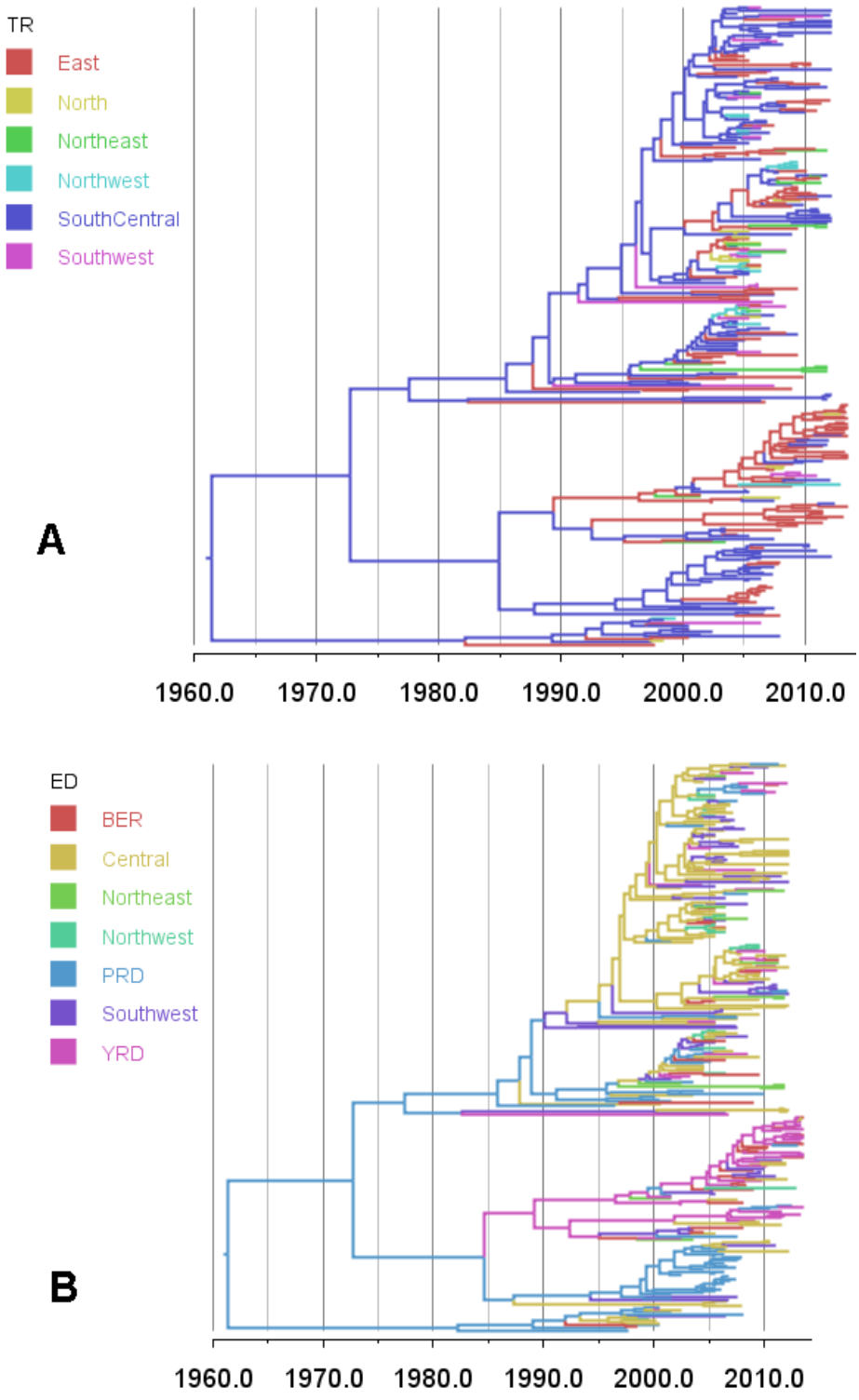


Figure A4.2 Bayesian MCC phylogenies of PA segment.
A: PA phylogeny, coloured by traditional regions
B: PA phylogeny coloured by economic zones
Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left.

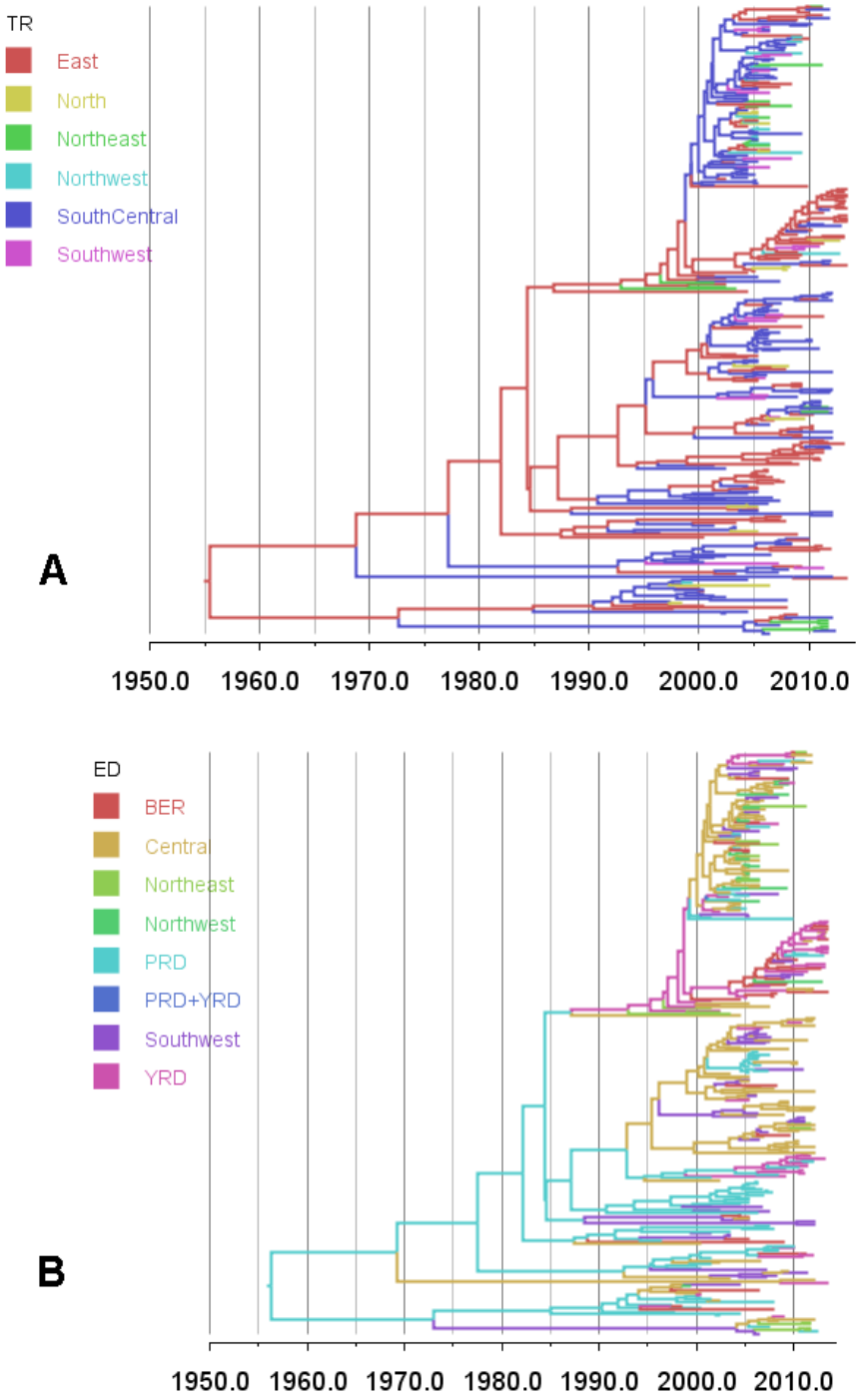


Figure A4.3 Bayesian MCC phylogenies of NP segment.
A: NP phylogeny, coloured by traditional regions
B: NP phylogeny coloured by economic zones
Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left.

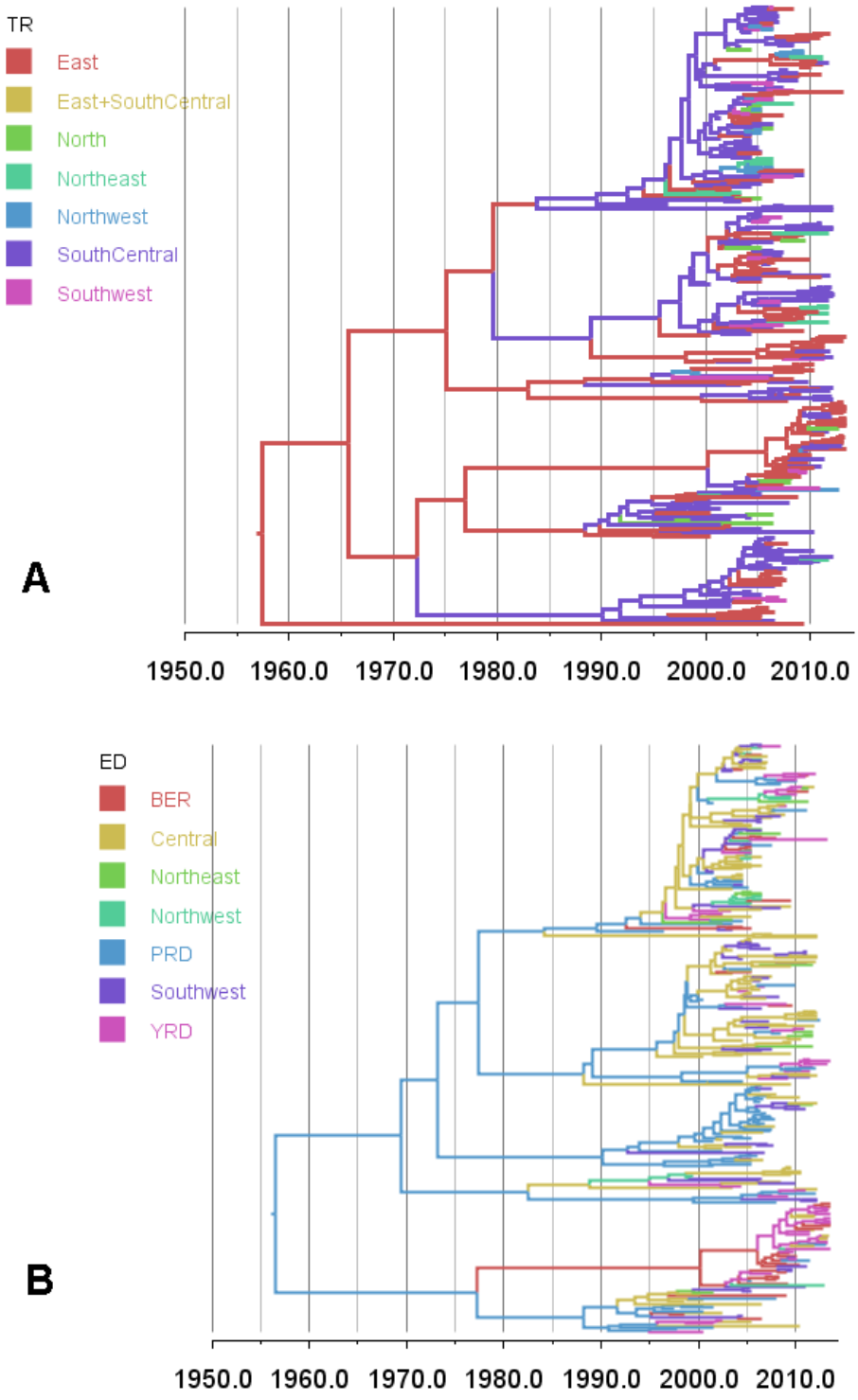


Figure A4.4 Bayesian MCC phylogenies of M segment.
A: M phylogeny, coloured by traditional regions
B: M phylogeny coloured by economic zones
Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left.

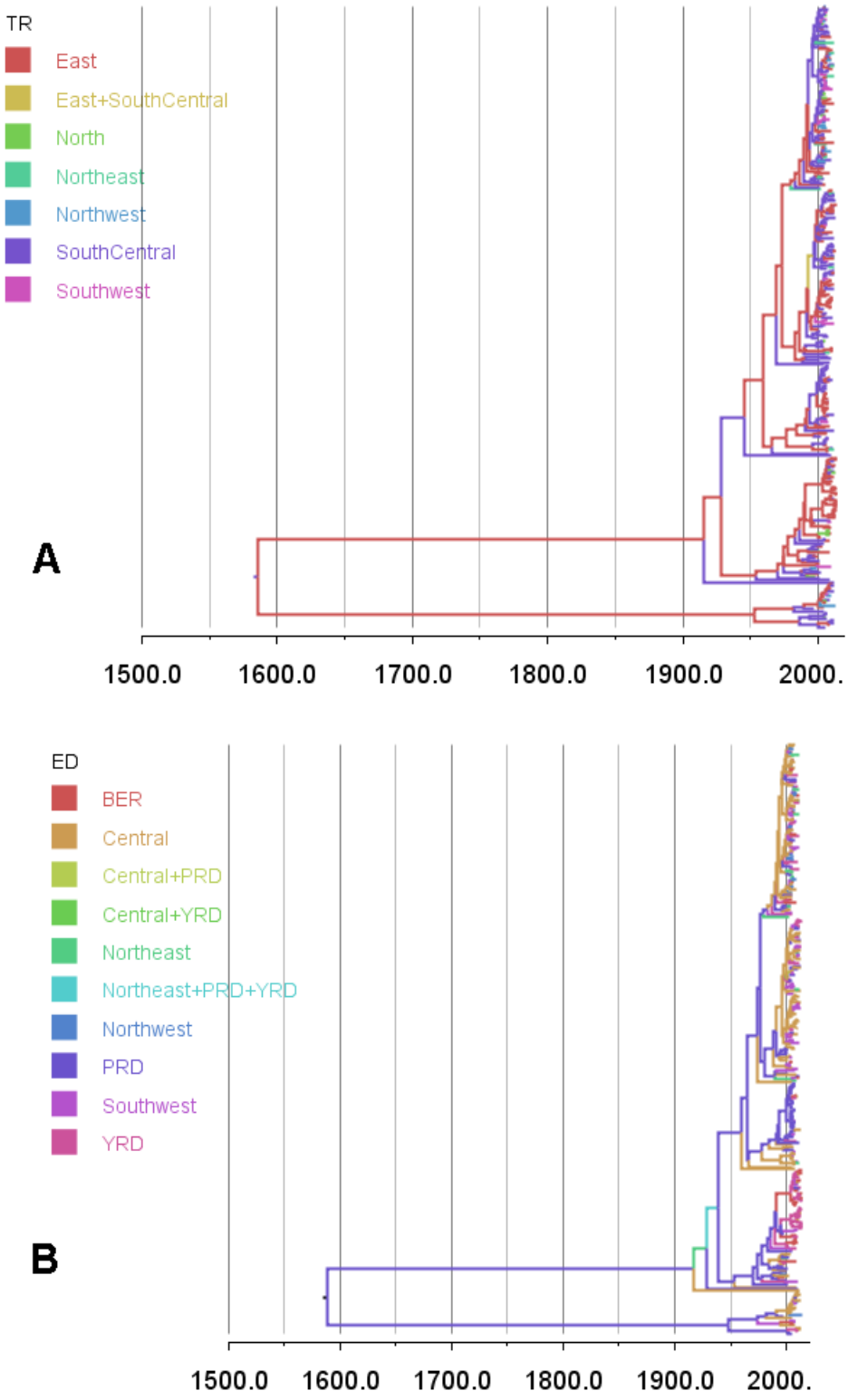


Figure A4.5 Bayesian MCC phylogenies of NS segment.
A: NS phylogeny, coloured by traditional regions
B: NS phylogeny coloured by economic zones
Branches are coloured according to their descendent nodes annotated by the different sampled areas within the certain region type, with the key for colours shown on the left. .

Provinces	1	2	3	4	5	6	7	8	9
Beijing	1584.34	1700	1038.58	7.97	12.85	51.00	17.95	1231.21	26161.91
Hebei	1877.28	2466	8374.08	3.61	13.96	54.67	117.76	383.55	219130.28
Shandong	3715.32	2472	6338.53	4.7	14.34	55.17	182.2	616.88	333602.6
Jiangsu	3386.15	9486	3501.75	4.1	15.98	68.08	279.144	771.93	220007.48
Shanghai	2239.75	4043	100.95	5.22	16.87	69.50	27.35	3754.62	94038.29
Zhejiang	1219.69	8446	17223.14	1.53	17.11	70.75	1427.15	538.02	191817.32
Guangdong	2057.78	8836	30183.37	6.73	21.67	81.58	2017.49	589.21	256076.69
Fujian	693.92	1817	48436.28	3.11	20.18	75.17	1510.07	308.73	84345.093
Anhui	1712.46	1992	13755.41	3.76	16.49	73.08	640.64	428.94	312436.77
Henan	3870.78	2463	12936.12	4.4	15.46	53.17	172.652	563.23	272114.94
Jiangxi	1165.46	1656	39529.64	7.55	18.03	77.08	2155.79	269.86	127195.53
Hubei	1603.67	2217	20942.49	5.14	16.37	81.42	783.763	310.87	122945.35
Hunan	1301.41	2331	34906.67	6.07	17.59	76.42	1981.31	313.45	191051.75
Shanxi	305.25	1435	7643.67	7.4	10.65	51.17	65.9	175.45	144607.9
Heilongjiang	289.98	2192	152104.96	14.85	4.62	67.17	695.687	81.06	65230.656
Jilin	816.58	1818	84412.29	12.43	5.23	63.00	387.33	146.77	54808.118
Liaoning	2687.18	4594	20226.85	12.39	7.43	68.50	492.419	301.24	206788.68
Gansu	80.80	1322	19363.83	16.17	7.54	57.42	258.95	56.65	45831.668
Ningxia	133.21	1126	492.14	10.34	9.81	48.00	8.45	97.47	41113.313
Qinghai	3.36	561	3915.64	30.21	5.22	59.00	879.21	7.96	13483.922
Xinjiang	17.81	1479	30100.54	12.95	7.39	53.08	851.602	13.41	58793.527
Shaanxi	303.94	1335	33820.54	5.7	14.17	61.75	368.0	182.37	136726.8
Guizhou	436.90	1128	24007.96	5.41	13.65	84.58	974.023	197.73	52654.917
Sichuan	773.60	4151	159572.37	18.54	15.86	78.08	2891.21	166.52	174349.34
Tibet	1.00	131	224550.91	33.91	9.61	33.50	4196.35	2.50	1126.6275
Yunnan	303.29	1657	155380.09	7.45	16.33	66.83	1689.77	118.25	68734.883
Guangxi	1281.53	1741	46875.18	5.98	21.42	79.75	2086.36	198.14	161356.01

Table A4.1 The predictor data per province

The 9 predictors per province (including Autonomous Region and Special Municipalities
The correspondent name and unit are listed below:

1 Poultry density (10000unit/km ^{2})
2 Number of farm product markets (region) (unit)
3 Stock Volume of Forest (10000 cu.m)
4 Percentage of Nature Reserves (%)
5 Average Temperature of Major Cities (℃)
6 Average Relative Humidity (%)
7 Surface Water Resources (100 million cu.m)
8 population density (10000 persons/km ^{2})
9 Freight by transportation total (10000 tons)

Predictor		Segment	TR			ED		
			Coeffient	Indicator	BF	Coeffient	Indicator	BF
1 Poultry population density	Overall	PB2	0.94 (-2.88, 4.78)	0.75	4	0.87 (-2.02, 6.2)	0.75	4
		PB1	0.75 (-3.29, 4.49)	0.74	3	0.81 (-2.97, 4.65)	0.69	3
		PA	0.21 (-3.65, 3.96)	0.77	4	0.65 (-3.06, 4.47)	0.70	3
		NP	0.21 (-3.65, 3.96)	0.77	4	0.72 (-3.21, 4.53)	0.69	3
		M	0.4 (-3.58, 4.01)	0.75	4	0.48 (-3.43, 4.13)	0.68	3
		NS	0.47 (-3.43, 4.38)	0.75	4	0.44 (-3.28, 4.22)	0.68	3
	Origin	PB2	-1.09 (-4.08, 2.15)	0.76	4	0.74 (-2.49, 4.02)	0.76	4
		PB1	-1.27 (-4.18, 1.78)	0.79	5	-0.18 (-3.28, 3.11)	0.58	2
		PA	-2.44 (-4.96, 0.2)	0.95	24	-0.36 (-3.64, 2.78)	0.62	2
		NP	-2.44 (-4.96, 0.2)	0.95	24	-0.25 (-3.55, 3.05)	0.62	2
		M	-2.04 (-4.79, 0.6)	0.91	13	-0.6 (-3.6, 2.85)	0.64	2
		NS	-1.85 (-4.62, 1.67)	0.88	9	-0.65 (-3.79, 2.64)	0.65	2
	Destination	PB2	4.99 (2.88, 7.12)	1	6172	1.8 (1.7, 6.09)	1	6172
		PB1	4.62 (2.28, 6.78)	1	3085	3.07 (0.93, 5.3)	0.99	122
		PA	3.49 (1.27, 5.82)	0.9952	256	2.84 (0.69, 5.04)	0.98	60
		NP	3.49 (1.27, 5.82)	1	256	2.42 (0.06, 4.94)	0.95	23
		M	3.76 (1.26, 5.96)	0.99	236	3.34 (1.07, 5.39)	0.99	122
		NS	4.01 (1.52, 6.62)	0.99	227	3.08 (0.7, 5.28)	0.98	60
2 Number of farm product markets	Overall	PB2	1.64 (-2.57, 5.72)	0.79	5	0.74 (-3.37, 4.52)	0.79	5
		PB1	1.02 (-2.8, 5.32)	0.73	3	0.24 (-3.69, 4.06)	0.61	2
		PA	0.41 (-3.45, 4.27)	0.71	3	0.09 (-3.74, 3.85)	0.67	3
		NP	0.41 (-3.45, 4.27)	0.71	3	0.35 (-3.46, 4.5)	0.62	2
		M	0.34 (-3.62, 4.31)	0.69	3	0.3 (-3.41, 4.2)	0.69	3
		NS	0.22 (-3.63, 4.19)	0.7	3	-0.01 (-4, 3.73)	0.70	3
	Origin	PB2	-2.71 (-6.33, 1.24)	0.9	11	0.71 (-4.06, 3.3)	0.9	11
		PB1	-3.57 (-6.86, 0.02)	0.96	27	-0.68 (-4.16, 3.3)	0.64	2
		PA	-4.88 (-7.97, -1.64)	0.99	166	-1.31 (-4.86, 2.57)	0.74	4
		NP	-4.88 (-7.97, -1.64)	0.99	166	-0.44 (-4.22, 3.12)	0.60	2
		M	-4.89 (-7.94, -1.83)	0.99	195	-1.17 (-4.77, 2.67)	0.74	4
		NS	-5.09 (-8.01, -2.13)	1	293	-2.07 (-5.82, 1.7)	0.85	7

	Destination	PB2	7.63 (5.13, 10.21)	1	6172	0.9 (-0.93, 5.84)	1	6172
		PB1	6.88 (4.4, 9.57)	1	6172	1.28 (-2.6, 4.59)	0.72	3
		PA	6.17 (3.65, 8.64)	1.00	6172	1.7 (-2.06, 4.97)	0.80	5
		NP	6.17 (3.65, 8.64)	1	6172	1.22 (-2.63, 4.46)	0.70	3
		M	6.08 (3.51, 8.63)	1	6172	2.32 (-1.09, 5.58)	0.89	10
		NS	5.93 (3.38, 8.6)	1	6172	2.27 (-1.26, 5.59)	0.88	9
3 Stock Volume of Forest	Overall	PB2	0.35 (-3.38, 4.17)	0.7	3	0.56 (-3.59, 3.74)	0.7	3
		PB1	-0.07 (-3.98, 3.58)	0.68	3	0.04 (-3.6, 3.66)	0.59	2
		PA	-0.46 (-4.18, 3.32)	0.71	3	-0.37 (-3.83, 3.38)	0.63	2
		NP	-0.46 (-4.18, 3.32)	0.71	3	-0.08 (-3.83, 3.52)	0.59	2
		M	-0.12 (-4.07, 3.57)	0.69	3	0.02 (-3.76, 3.68)	0.41	1
		NS	-0.11 (-4.11, 3.64)	0.69	3	0.15 (-3.59, 4.04)	0.31	1
	Origin	PB2	-0.04 (-3.49, 3.36)	0.67	3	0.57 (-3.55, 3.09)	0.67	3
		PB1	-0.22 (-3.75, 3.02)	0.65	2	-0.45 (-3.47, 2.99)	0.56	2
		PA	-0.76 (-4.05, 2.65)	0.70	3	-0.9 (-3.81, 2.53)	0.69	3
		NP	-0.76 (-4.05, 2.65)	0.7	3	-0.44 (-3.4, 2.96)	0.57	2
		M	-0.26 (-3.74, 3.11)	0.64	2	-0.43 (-3.62, 3.3)	0.43	1
		NS	-0.23 (-3.57, 2.96)	0.63	2	-0.01 (-3.87, 3.57)	0.23	0
	Destination	PB2	0.48 (-2.86, 3.44)	0.66	2	0.5 (-3.12, 3.44)	0.66	2
		PB1	0.08 (-3.24, 3.13)	0.62	2	0.4 (-3.17, 3.16)	0.56	2
		PA	0.02 (-3.17, 3.12)	0.62	2	0.17 (-3.15, 3.17)	0.57	2
		NP	0.02 (-3.17, 3.12)	0.62	2	0.32 (-3.16, 3.29)	0.54	1
		M	0.06 (-3.27, 3.16)	0.61	2	0.38 (-3.41, 3.45)	0.41	1
		NS	0.01 (-3.1, 3.32)	0.61	2	0.11 (-3.54, 3.74)	0.23	0
4 Percentage of Nature Reserves	Overall	PB2	-0.17 (-3.97, 3.56)	0.63	2	0.62 (-3.64, 3.5)	0.63	2
		PB1	-0.13 (-3.82, 3.49)	0.64	2	0.07 (-3.32, 3.5)	0.64	2
		PA	0.07 (-3.91, 3.74)	0.77	4	0.01 (-3.54, 3.35)	0.66	2
		NP	0.07 (-3.91, 3.74)	0.77	4	0.21 (-3.37, 3.89)	0.60	2
		M	0.09 (-3.88, 3.85)	0.74	4	-0.1 (-3.81, 3.29)	0.62	2
		NS	0.05 (-3.84, 3.99)	0.7	3	0.32 (-3.35, 4.05)	0.63	2
	Origin	PB2	0.24 (-3.33, 3.14)	0.51	1	0.56 (-3.06, 3.22)	0.51	1
		PB1	0.28 (-2.95, 3.52)	0.53	1	0.74 (-2.48, 2.88)	0.72	3

		PA	0.82 (-2.22, 3.11)	0.76	4	0.65 (-2.58, 3.12)	0.70	3
		NP	0.82 (-2.22, 3.11)	0.76	4	0.77 (-2.48, 3.15)	0.73	3
		M	0.81 (-2.25, 3.16)	0.73	3	0.48 (-2.89, 3.27)	0.60	2
		NS	0.77 (-2.36, 3.18)	0.71	3	1.31 (-1.25, 3.48)	0.88	9
	Destination	PB2	-1.95 (-3.12, -0.63)	0.98	69	0.88 (-2.96, 1.16)	0.98	69
		PB1	-1.85 (-2.95, -0.4)	0.98	56	-0.69 (-2.83, 2.7)	0.73	3
		PA	-1.39 (-3.06, 0.48)	0.92	13	-0.76 (-2.81, 2.51)	0.77	4
		NP	-1.39 (-3.06, 0.48)	0.92	13	-0.57 (-2.99, 2.75)	0.69	3
		M	-1.39 (-3.01, 0.84)	0.9	11	-1.01 (-3.04, 1.59)	0.84	6
		NS	-1.36 (-3.06, 1.25)	0.87	8	-0.34 (-3.07, 3.08)	0.59	2
5 Average Temperature	Overall	PB2	0.17 (-3.63, 4.06)	0.59	2	-0.09 (-3.99, 3.69)	0.59	2
		PB1	0.11 (-3.83, 3.76)	0.58	2	-0.39 (-4.18, 3.53)	0.56	2
		PA	-0.08 (-4.12, 4.03)	0.75	4	-0.9 (-4.83, 3.21)	0.67	3
		NP	0 (-4.3, 4.04)	0.68	3	-0.85 (-4.87, 3.53)	0.64	2
		M	-0.02 (-3.72, 4.2)	0.67	3	-0.11 (-3.93, 3.93)	0.56	2
		NS	-0.06 (-3.87, 4.06)	0.68	3	-0.29 (-4.28, 3.67)	0.59	2
	Origin	PB2	-0.34 (-3.17, 3.4)	0.51	1	-1.59 (-3.76, 1.07)	0.88	9
		PB1	-0.41 (-3.47, 3.13)	0.53	1	-1.88 (-4.08, -0.05)	0.94	19
		PA	-1.08 (-3.36, 1.65)	0.82	6	-2.47 (-4.67, -0.6)	0.99	122
		NP	-0.9 (-3.73, 2.13)	0.72	3	-2.59 (-5.19, -0.64)	0.98	60
		M	-0.79 (-3.27, 2.53)	0.7	3	-0.79 (-3.17, 2.59)	0.7	3
		NS	-1.02 (-3.79, 2.11)	0.75	4	-1 (-3.42, 1.92)	0.79	5
	Destination	PB2	3.4 (1.8, 4.8)	1	6172	0.42 (-3.07, 3.35)	0.54	1
		PB1	3.46 (1.83, 4.89)	1	6172	0.05 (-3.33, 3.47)	0.43	1
		PA	2.55 (0.82, 4.25)	0.99	122	-0.49 (-3.46, 3.04)	0.58	2
		NP	2.9 (0.95, 4.69)	0.99	122	-0.56 (-3.77, 2.91)	0.57	2
		M	2.93 (1.19, 4.44)	1	6172	0.49 (-2.98, 2.98)	0.6	2
		NS	2.88 (1.02, 4.6)	0.99	122	0.26 (-3.23, 3.2)	0.55	2
6 Average Relative Humidity	Overall	PB2	0.68 (-3.28, 4.59)	0.74	4	-0.08 (-3.9, 3.91)	0.72	3
		PB1	0.45 (-3.5, 4.31)	0.75	4	-0.27 (-4.19, 3.69)	0.7	3
		PA	0.05 (-3.74, 4.19)	0.76	4	-0.48 (-4.52, 3.35)	0.67	3
		NP	0.4 (-3.59, 4.17)	0.76	4	-0.39 (-4.32, 3.61)	0.67	3

		M	0.25 (-3.64, 4.3)	0.75	4	0.13 (-3.97, 3.93)	0.69	3
		NS	0.28 (-3.79, 4.03)	0.75	4	-0.15 (-3.91, 3.92)	0.71	3
	Origin	PB2	-2.61 (-6.26, 1.28)	0.9	11	-3.11 (-6.44, 0.09)	0.95	23
		PB1	-3.46 (-6.75, -0.25)	0.95	23	-3.61 (-7.02, -0.34)	0.96	30
		PA	-4.04 (-7.35, -0.62)	0.98	60	-4.14 (-7.11, -1.02)	0.98	60
		NP	-3.34 (-6.93, -0.38)	0.95	23	-3.52 (-6.84, 0.11)	0.96	30
		M	-3.48 (-7.09, 0.14)	0.96	30	-2.21 (-5.55, 1.65)	0.86	8
		NS	-3.96 (-7.3, -0.75)	0.98	60	-3.12 (-6.48, 0.26)	0.94	19
	Destination	PB2	5.92 (2.81, 8.77)	1	6172	2.77 (-0.44, 5.84)	0.93	16
		PB1	5.63 (2.7, 8.81)	1	6172	1.76 (-1.73, 5.13)	0.82	6
		PA	4.21 (1.22, 7.42)	0.99	122	1.01 (-2.53, 4.23)	0.7	3
		NP	5.13 (2.1, 8.19)	1	6172	1.42 (-2.02, 4.91)	0.76	4
		M	4.73 (1.76, 7.81)	1	6172	3.13 (0.08, 6.14)	0.96	30
		NS	5.1 (2.02, 8.05)	1	6172	2.32 (-1.15, 5.68)	0.88	9
7 Surface Water Resources	Overall	PB2	0.27 (-3.33, 4.15)	0.76	4	0.73 (-3.49, 3.95)	0.76	4
		PB1	0.1 (-3.68, 4.07)	0.72	3	0.14 (-3.59, 3.74)	0.60	2
		PA	-0.2 (-4.04, 3.72)	0.73	3	-0.08 (-3.79, 3.65)	0.67	3
		NP	-0.2 (-4.04, 3.72)	0.73	3	0.1 (-3.84, 3.8)	0.62	2
		M	-0.21 (-4.17, 3.59)	0.73	3	0.27 (-3.43, 4.06)	0.58	2
		NS	-0.33 (-4.28, 3.57)	0.73	3	0.21 (-3.55, 3.92)	0.64	2
	Origin	PB2	-1.31 (-4.13, 1.99)	0.83	6	0.71 (-3.46, 2.48)	0.83	6
		PB1	-1.5 (-4.58, 1.86)	0.82	6	-0.58 (-3.42, 2.88)	0.61	2
		PA	-2.63 (-5, -0.18)	0.97	38	-1.46 (-3.96, 1.66)	0.82	6
		NP	-2.63 (-5, -0.18)	0.97	38	-0.86 (-3.56, 2.62)	0.68	3
		M	-2.7 (-5.3, 0.04)	0.96	28	-0.43 (-3.51, 2.98)	0.53	1
		NS	-3.06 (-5.71, -0.35)	0.97	43	-0.72 (-3.65, 2.77)	0.64	2
	Destination	PB2	2.75 (0.57, 5)	0.98	68	0.96 (-0.04, 4.08)	0.98	68
		PB1	2.28 (-0.31, 4.71)	0.93	17	1.26 (-1.65, 3.61)	0.82	6
		PA	1.41 (-1.55, 4.17)	0.84	7	1.35 (-1.78, 3.91)	0.80	5
		NP	1.41 (-1.55, 4.17)	0.84	7	1.65 (-1.37, 4.04)	0.86	8
		M	1.51 (-1.42, 4.37)	0.85	7	2.18 (0.05, 4.22)	0.94	19
		NS	1.45 (-1.79, 4.34)	0.83	6	2.14 (-0.3, 4.31)	0.93	16

8 Human population density	Overall	PB2	1.34 (-2.76, 5.33)	0.77	4	0.73 (-2.82, 5.14)	0.77	4
		PB1	0.85 (-3.21, 4.67)	0.75	4	0.51 (-3.08, 4.26)	0.57	2
		PA	0.41 (-3.55, 4.05)	0.74	4	0.32 (-3.27, 4.02)	0.64	2
		NP	0.41 (-3.55, 4.05)	0.74	4	0.35 (-3.38, 4.01)	0.59	2
		M	0.3 (-3.58, 3.92)	0.74	3	0.19 (-3.63, 3.89)	0.59	2
		NS	0.2 (-3.59, 4.04)	0.74	3	-0.09 (-3.75, 3.66)	0.61	2
	Origin	PB2	-0.07 (-3.09, 3.41)	0.66	2	0.61 (-2.87, 3.6)	0.66	2
		PB1	-0.38 (-3.64, 2.81)	0.68	3	-0.13 (-3.57, 3.14)	0.49	1
		PA	-1.34 (-4.38, 2.04)	0.81	5	-0.52 (-3.44, 2.81)	0.63	2
		NP	-1.34 (-4.38, 2.04)	0.81	5	-0.3 (-3.22, 3.19)	0.55	2
		M	-1.28 (-4.46, 2.08)	0.79	5	-0.59 (-3.49, 2.85)	0.61	2
		NS	-1.45 (-4.98, 1.68)	0.81	5	-1.3 (-3.93, 2.05)	0.77	4
	Destination	PB2	3.84 (1.91, 5.62)	1	2056	0.9 (-0.72, 4.3)	1	2056
		PB1	3.48 (1.39, 5.22)	1	513	0.87 (-2.57, 3.44)	0.67	3
		PA	3.15 (1.09, 5.06)	0.99	170	1.3 (-1.57, 3.58)	0.84	6
		NP	3.15 (1.09, 5.06)	0.99	170	0.82 (-2.58, 3.53)	0.68	3
		M	3.1 (0.91, 4.97)	0.99	126	1.27 (-1.67, 3.66)	0.80	5
		NS	2.82 (0.37, 5.02)	0.97	38	0.93 (-2.6, 3.42)	0.71	3
9 Freight by transportation	Overall	PB2	1.69 (-2.38, 5.89)	0.85	7	0.78 (-2.64, 5.61)	0.85	7
		PB1	1.12 (-2.93, 5.16)	0.81	5	0.87 (-3.12, 4.87)	0.74	4
		PA	0.58 (-3.52, 4.33)	0.77	4	0.55 (-3.45, 4.36)	0.77	4
		NP	0.58 (-3.52, 4.33)	0.77	4	0.63 (-3.42, 4.59)	0.75	4
		M	0.51 (-3.43, 4.39)	0.76	4	0.72 (-3.14, 4.7)	0.76	4
		NS	0.35 (-3.76, 4.26)	0.75	4	0.62 (-3.2, 4.56)	0.75	4
	Origin	PB2	-4.71 (-8.32, -1.12)	0.99	92	0.8 (-5.17, 2.26)	0.99	92
		PB1	-5.59 (-8.94, -2.27)	1	342	-1.93 (-5.69, 2.07)	0.82	6
		PA	-6.2 (-9.77, -2.75)	1	424	-2.97 (-6.65, 0.52)	0.93	16
		NP	-6.2 (-9.77, -2.75)	1	424	-2.43 (-6.17, 1.68)	0.87	8
		M	-6.45 (-9.78, -3.17)	1	1542	-3.51 (-6.96, -0.05)	0.96	30
		NS	-7.03 (-10.31, -4.03)	1	6172	-3.45 (-6.95, 0)	0.96	30
	Destination	PB2	8.94 (6.03, 11.81)	1	6172	3.1 (2.57, 9.17)	1	12344
		PB1	8.61 (5.62, 11.43)	1	6172	4.99 (1.51, 8.61)	0.99	122

		PA	7.74 (4.8, 10.47)	1	6172	4.86 (1.6, 8.56)	0.99	122
		NP	7.74 (4.8, 10.47)	1	6172	4.62 (1.12, 8.36)	0.99	122
		M	7.87 (5, 10.76)	1	6172	5.72 (2.61, 9.02)	0.99	6172
		NS	8 (4.97, 10.86)	1	6172	5.3 (2.03, 8.64)	1.00	6172
10 Sample size	Overall	PB2	0.26 (-3.23, 4.12)	0.52	1	0.46 (-3.34, 4.34)	0.52	1
		PB1	0.21 (-3.36, 4.05)	0.49	1	-0.02 (-3.61, 3.9)	0.50	1
		PA	-0.07 (-3.88, 3.83)	0.69	3	-0.03 (-3.78, 3.97)	0.61	2
		NP	-0.07 (-3.88, 3.83)	0.69	3	-0.09 (-4.06, 3.85)	0.59	2
		M	-0.03 (-3.96, 3.76)	0.64	2	0.04 (-3.55, 4.09)	0.51	1
		NS	0.05 (-3.65, 4.09)	0.63	2	-0.09 (-3.86, 3.88)	0.56	2
	Origin	PB2	-0.31 (-3.32, 3.57)	0.46	1	0.39 (-3.74, 3.32)	0.46	1
		PB1	-0.36 (-3.47, 3.23)	0.46	1	-0.44 (-3.55, 3.3)	0.48	1
		PA	-1.31 (-3.89, 1.54)	0.84	6	-1.23 (-3.81, 1.94)	0.75	4
		NP	-1.31 (-3.89, 1.54)	0.84	6	-1.17 (-3.9, 2.25)	0.71	3
		M	-1.14 (-3.75, 2.17)	0.78	4	-0.47 (-3.7, 3.15)	0.48	1
		NS	-1.07 (-4.02, 2.31)	0.71	3	-0.81 (-3.61, 2.72)	0.61	2
	Destination	PB2	3.65 (2, 4.88)	1	1028	0.99 (1.42, 4.23)	1	1028
		PB1	3.56 (1.94, 4.76)	1	1121	2.56 (0.81, 3.99)	0.98	60
		PA	-0.13 (-3.2, 2.92)	0.84	70	1.76 (-1.01, 3.95)	0.87	8
		NP	2.62 (-0.85, 4.43)	0.98	70	1.84 (-0.71, 4.2)	0.88	9
		M	2.69 (0.81, 4.34)	0.98	68	2.82 (1.03, 4.2)	0.99	122
		NS	2.82 (0.79, 4.6)	0.98	56	2.27 (0.07, 3.92)	0.94	19

Table A4.2 Correlations of predictors to the spatial diffusion of Chinese AIV.

For the 10 predictors, their correlations to the AIV diffusion in four region types(TR, ER,ED and EL) were summarized and the columns represented six internal segments.

[a]: IP is the abbreviation of the inclusion probabilities, ranging between 0 to 1, representing the less precise values by showing only 2 decimal places.

[b]: the Bayes Factor estimates for the inclusion probabilities with BF>3 indicating a statistical significant (in bold).

[c]: the correlation Coefficient with mean and 95% HPD interval. The mean and intervals above 0 indicates a positive correlation (in red), while the value below 0 indicates a negative correlation (in blue).

Struthioniformesain Name	Subtype	Host		Provinces
		Order	Type	
A/chicken/Henan/26/00	H9N2	Galliformes	Domestic	Henan
A/duck/Shantou/1275/2004	H6N1	Anseriformes	Domestic	Guangdong
A/wild duck/Shantou/867/2002	H6N2	Anseriformes	Wild	Guangdong
A/little grebe/Xianghai/429/2011	H5N2	Anseriformes	Wild	Jilin
A/duck/Guangxi/2281/2007	H6N6	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/GXd-1/2009	H6N5	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/GXd-7/2011	H6N6	Anseriformes	Domestic	Guangxi
A/duck/Fujian/378/2007	H6N6	Anseriformes	Domestic	Fujian
A/duck/Guangxi/3574/2006	H6N2	Anseriformes	Domestic	Guangxi
A/wild duck/Shantou/5769/2004	H6N2	Anseriformes	Wild	Guangdong
A/wild duck/Shantou/865/2002	H6N2	Anseriformes	Wild	Guangdong
A/duck/Fujian/2018/2007	H6N6	Anseriformes	Domestic	Fujian
A/wild duck/Shantou/311/2001	H6N9	Anseriformes	Wild	Guangdong
A/duck/Fujian/5813/2007	H6N6	Anseriformes	Domestic	Fujian
A/duck/Shantou/10776/2006	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Fujian/331/2007	H6N6	Anseriformes	Domestic	Fujian
A/duck/Fujian/958/2006	H6N6	Anseriformes	Domestic	Fujian
A/duck/Fujian/11339/2005	H6N2	Anseriformes	Domestic	Fujian
A/duck/Fujian/8807/2006	H6N6	Anseriformes	Domestic	Fujian
A/duck/Shantou/7568/2005	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Hunan/S1661/2012	H6N6	Anseriformes	Domestic	Hunan
A/duck/Hubei/5/2010	H6N6	Anseriformes	Domestic	Hubei
A/duck/Hunan/4056/2006	H6N8	Anseriformes	Domestic	Hunan
A/duck/Shantou/1080/2007	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Shantou/14841/2006	H6N6	Anseriformes	Domestic	Guangdong
A/duck/Shantou/5808/2005	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Shantou/2472/2005	H6N2	Anseriformes	Domestic	Guangdong
A/wild duck/Shantou/9466/2006	H6N6	Anseriformes	Wild	Guangdong
A/duck/Shantou/9689/2006	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Shantou/9395/2006	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Guangdong/W12/2011	H3N2	Anseriformes	Domestic	Guangdong
A/duck/Hebei/0908/2009	H5N2	Anseriformes	Domestic	Hebei
A/duck/Guangdong/E1/2012	H10N8	Anseriformes	Domestic	Guangdong
A/chicken/Jiangsu/1001/2013	H5N2	Galliformes	Domestic	Jiangsu
A/chicken/Guangdong/LG1/2013	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Rizhao/1339/2013	H9N2	Galliformes	Domestic	Shandong
A/silkie chicken/Wenzhou/812/2013	H9N2	Galliformes	Domestic	Zhejiang
A/chicken/Rizhao/55/2013	H9N2	Galliformes	Domestic	Shandong
A/pigeon/Shanghai/JC1/2013	H9N2	Passerine	Wild	Shanghai
A/chicken/Wenzhou/253/2013	H9N2	Galliformes	Domestic	Zhejiang
A/chicken/Shanghai/020/2013	H9N2	Galliformes	Domestic	Shanghai
A/chicken/Guangdong/ZHJ/2011	H9N2	Galliformes	Domestic	Guangdong

A/chicken/Zhejiang/607/2011	H9N2	Galliformes	Domestic	Zhejiang
A/chicken/Wenzhou/89/2013	H7N7	Galliformes	Domestic	Zhejiang
A/chicken/Shandong/03/2010	H9N2	Galliformes	Domestic	Shandong
A/duck/Shanghai/C163/2009	H9N2	Anseriformes	Domestic	Shanghai
A/chicken/Shandong/H/2009	H9N2	Galliformes	Domestic	Shandong
A/chicken/Shandong/02/2010	H9N2	Galliformes	Domestic	Shandong
A/chicken/Hunan/12/2011	H9N2	Galliformes	Domestic	Hunan
A/chicken/Jiangsu/Q3/2010	H9N2	Galliformes	Domestic	Jiangsu
A/chicken/Shanghai/S1080/2013	H7N9	Galliformes	Domestic	Shanghai
A/brambling/Beijing/16/2012	H9N2	Passerine	Wild	Beijing
A/chicken/Shanghai/017/2013	H7N9	Galliformes	Domestic	Shanghai
A/chicken/Zhejiang/SD007/2013	H7N9	Galliformes	Domestic	Zhejiang
A/chicken/Hunan/1/2012	H9N2	Galliformes	Domestic	Hunan
A/chicken/Zhejiang/611/2011	H9N2	Galliformes	Domestic	Zhejiang
A/chicken/Shandong/01/2009	H9N2	Galliformes	Domestic	Shandong
A/chicken/Tibet/S1/2009	H9N2	Galliformes	Domestic	Tibet
A/spot-billed duck/Xianghai/427/2011	H5N2	Anseriformes	Wild	Jilin
A/baikal teal/Xianghai/426/2011	H5N2	Anseriformes	Wild	Jilin
A/duck/Jiangxi/23005/2009	H7N3	Anseriformes	Domestic	Jiangxi
A/duck/Jiangxi/8028/2009	H7N3	Anseriformes	Domestic	Jiangxi
A/chicken/Zhejiang/HJ/2007	H9N2	Galliformes	Domestic	Zhejiang
A/duck/Hunan/2110/2006	H6N2	Anseriformes	Domestic	Hunan
A/duck/Guangxi/2736/2006	H6N8	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/1533/2007	H6N2	Anseriformes	Domestic	Guangxi
A/mallard/Jiangxi/7787/2003	H6N1	Anseriformes	Wild	Jiangxi
A/duck/Guangxi/141/2005	H6N2	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/3459/2005	H6N2	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/1455/2004	H6N5	Anseriformes	Domestic	Guangxi
A/chicken/Fujian/G9/2009	H9N2	Galliformes	Domestic	Fujian
A/sparrow/Guangxi/GXs-1/2012	H1N2	Passerine	Wild	Guangxi
A/duck/Hunan/S11313/2012	H4N2	Anseriformes	Domestic	Hunan
A/duck/Guizhou/1078/2011	H11N9	Anseriformes	Domestic	Guizhou
A/duck/Jiangxi/21980/2010	H7N7	Anseriformes	Domestic	Jiangxi
A/duck/Hunan/S1256/2012	H3N8	Anseriformes	Domestic	Hunan
A/wild goose/Dongting/C1037/2011	H12N8	Anseriformes	Wild	Hunan
A/wild waterfowl/Dongting/C2383/2012	H1N2	Anseriformes	Wild	Hunan
A/wild duck/Jiangxi/8462/2006	H6N1	Anseriformes	Wild	Jiangxi
A/canvasback/Xianghai/428/2011	H5N2	Anseriformes	Wild	Jilin
A/duck/Shanghai/C84/2009	H3N2	Anseriformes	Domestic	Shanghai
A/chicken/Hubei/2856/2007	H5N1	Galliformes	Domestic	Hubei
A/chicken/Hunan/3157/2006	H5N1	Galliformes	Domestic	Hunan
A/duck/Hubei/2911/2007	H5N1	Anseriformes	Domestic	Hubei
A/duck/Hunan/1590/2007	H6N9	Anseriformes	Domestic	Hunan
A/wild duck/Jiangxi/9157/2005	H7N8	Anseriformes	Wild	Jiangxi

A/wild duck/Jiangxi/10179/2005	H7N3	Anseriformes	Wild	Jiangxi
A/duck/Hunan/S1824/2012	H3N8	Anseriformes	Domestic	Hunan
A/duck/Hunan/3748/2004	H6N8	Anseriformes	Domestic	Hunan
A/duck/Guizhou/2492/2007	H6N1	Anseriformes	Domestic	Guizhou
A/duck/Jiangsu/26/2004	H3N2	Anseriformes	Domestic	Jiangsu
A/mallard/Jiangxi/12147/2005	H6N2	Anseriformes	Wild	Jiangxi
A/duck/Guizhou/888/2006	H6N5	Anseriformes	Domestic	Guizhou
A/duck/Guizhou/1084/2006	H6N2	Anseriformes	Domestic	Guizhou
A/duck/Shanghai/Y20/2006	H4N6	Anseriformes	Domestic	Shanghai
A/garganey/SanJiang/160/2006	H5N2	Anseriformes	Wild	Guangxi
A/green-winged teal/Xianghai/430/2011	H5N2	Anseriformes	Wild	Jilin
A/wild goose/Dongting/PC0360/2012	H7N7	Anseriformes	Wild	Hunan
A/mallard/Jiangxi/6845/2003	H6N1	Anseriformes	Wild	Jiangxi
A/duck/Hunan/5613/2003	H6N5	Anseriformes	Domestic	Hunan
A/duck/Hunan/1469/2002	H6N8	Anseriformes	Domestic	Hunan
A/chicken/Shaanxi/11/2012	H9N2	Galliformes	Domestic	Shaanxi
A/chicken/Henan/43/02	H9N2	Galliformes	Domestic	Henan
A/chicken/Guangdong/V/2008	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Guangdong/47/01	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Guangdong/10/00	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Shandong/1/2008	H9N2	Galliformes	Domestic	Shandong
A/chicken/Fujian/25/00	H9N2	Galliformes	Domestic	Fujian
A/chicken/Shijiazhuang/2/98	H9N2	Galliformes	Domestic	Hebei
A/chicken/Guangdong/5/97	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Gansu/2/99	H9N2	Galliformes	Domestic	Gansu
A/chicken/Hebei/L1/2006	H9N2	Galliformes	Domestic	Hebei
A/chicken/Shandong/02/2008	H9N2	Galliformes	Domestic	Shandong
A/chicken/Shandong/KD/2009	H9N2	Galliformes	Domestic	Shandong
A/chicken/Hebei/7/2008	H9N2	Galliformes	Domestic	Hebei
A/chicken/Hebei/C4/2008	H9N2	Galliformes	Domestic	Hebei
A/chicken/Shandong/LY-1/2008	H9N2	Galliformes	Domestic	Shandong
A/chicken/Guangdong/56/01	H9N2	Galliformes	Domestic	Guangdong
A/chicken/Guangxi/37/2005	H9N2	Galliformes	Domestic	Guangxi
A/chicken/Heilongjiang/48/01	H9N2	Galliformes	Domestic	Heilongjiang
A/chicken/Hubei/C1/2007	H9N2	Galliformes	Domestic	Hubei
A/chicken/Jiangsu/7/2002	H9N2	Galliformes	Domestic	Jiangsu
A/duck/Wenzhou/771/2013	H7N3	Anseriformes	Domestic	Zhejiang
A/chicken/Jiangsu/RD5/2013	H10N9	Galliformes	Domestic	Jiangsu
A/goose/Guangdong/7472/2012	H7N7	Anseriformes	Domestic	Guangdong
A/duck/Jiangsu/10-d4/2011	H11N3	Anseriformes	Domestic	Jiangsu
A/duck/Zhejiang/0224-6/2011	H1N2	Anseriformes	Domestic	Zhejiang
A/duck/Shanghai/28-1/2009	H4N2	Anseriformes	Domestic	Shanghai
A/duck/Jiangxi/21714/2011	H11N9	Anseriformes	Domestic	Jiangxi
A/duck/Zhejiang/5/2011	H3N3	Anseriformes	Domestic	Zhejiang

A/chicken/Guangxi/12/2004	H5N1	Galliformes	Domestic	Guangxi
A/duck/Anhui/56/2005	H5N1	Anseriformes	Domestic	Anhui
A/duck/Hunan/3340/2006	H5N1	Anseriformes	Domestic	Hunan
A/chicken/Shandong/A-1/2009	H5N1	Galliformes	Domestic	Shandong
A/chicken/Hunan/1/2009	H5N1	Galliformes	Domestic	Hunan
A/wild duck/Shantou/1737/2000	H6N8	Anseriformes	Wild	Guangdong
A/duck/Shandong/093/2004	H5N1	Anseriformes	Domestic	Shandong
A/duck/Guangxi/12/2003	H5N1	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/27/2003	H5N1	Anseriformes	Domestic	Guangxi
A/duck/Guangdong/wy24/2008	H5N5	Anseriformes	Domestic	Guangdong
A/duck/Guangdong/wy19/2008	H5N5	Anseriformes	Domestic	Guangdong
A/duck/Jiangsu/m234/2012	H5N2	Anseriformes	Domestic	Jiangsu
A/wild duck/Shandong/628/2011	H5N1	Anseriformes	Wild	Shandong
A/duck/Jiangsu/1-15/2011	H4N2	Anseriformes	Domestic	Jiangsu
A/goose/Guangdong/k0103/2010	H5N5	Anseriformes	Domestic	Guangdong
A/duck/Shantou/728/2001	H6N2	Anseriformes	Domestic	Guangdong
A/quail/Jiangsu/k0104/2010	H5N5	Galliformes	Domestic	Jiangsu
A/goose/Guangdong/xb/2001	H5N1	Anseriformes	Domestic	Guangdong
A/duck/Jiangxi/5748/2006	H6N2	Anseriformes	Domestic	Jiangxi
A/duck/Jiangxi/21669/2009	H7N7	Anseriformes	Domestic	Jiangxi
A/wild duck/Jiangxi/19831/2009	H7N7	Anseriformes	Wild	Jiangxi
A/duck/Jiangxi/23008/2009	H7N7	Anseriformes	Domestic	Jiangxi
A/duck/Hunan/S1607/2012	H11N9	Anseriformes	Domestic	Hunan
A/chicken/Guangxi/GXc-1/2011	H1N2	Galliformes	Domestic	Guangxi
A/duck/Guangxi/GXd-4/2011	H1N2	Anseriformes	Domestic	Guangxi
A/duck/Guangxi/GXd-1/2011	H1N2	Anseriformes	Domestic	Guangxi
A/duck/Guizhou/1560/2007	H6N8	Anseriformes	Domestic	Guizhou
A/chicken/Jiangxi/2369/2010	H7N7	Galliformes	Domestic	Jiangxi
A/duck/Nanjing/1102/2010	H4N8	Anseriformes	Domestic	Jiangsu
A/domestic green-winged teal/Hunan/3450/2006	H5N1	Anseriformes	Domestic	Hunan
A/duck/Guizhou/1073/2007	H6N8	Anseriformes	Domestic	Guizhou
A/duck/Yunnan/1282/2007	H11N9	Anseriformes	Domestic	Yunnan
A/duck/Guangxi/912/2008	H4N2	Anseriformes	Domestic	Guangxi
A/mallard/Jiangxi/6285/2004	H6N8	Anseriformes	Wild	Jiangxi
A/mallard/Jiangxi/10071/2005	H6N1	Anseriformes	Wild	Jiangxi
A/duck/Jiangxi/22041/2008	H4N9	Anseriformes	Domestic	Jiangxi
A/duck/Guangdong/4323/2007	H11N9	Anseriformes	Domestic	Guangdong
A/duck/Beijing/61/05	H3N8	Anseriformes	Domestic	Beijing
A/duck/Beijing/40/04	H3N8	Anseriformes	Domestic	Beijing
A/duck/Hunan/8-19/2009	H4N2	Anseriformes	Domestic	Hunan
A/duck/Guangxi/GXd-2/2012	H1N2	Anseriformes	Domestic	Guangxi
A/duck/Hunan/S11547/2012	H4N9	Anseriformes	Domestic	Hunan
A/egret/Hunan/1/2012	H9N2	Pelecaniformes	Wild	Hunan
A/baikal teal/Xianghai/421/2011	H9N2	Anseriformes	Wild	Jilin

A/duck/Hunan/S11200/2012	H4N6	Anseriformes	Domestic	Hunan
A/chicken/Jilin/xw/2003	H5N1	Galliformes	Domestic	Jilin
A/chicken/Guangdong/178/04	H5N1	Galliformes	Domestic	Guangdong
A/duck/Guangdong/23/2004	H5N1	Anseriformes	Domestic	Guangdong
A/duck/Hunan/70/2004	H5N1	Anseriformes	Domestic	Hunan
A/goose/Fujian/bb/2003	H5N1	Anseriformes	Domestic	Fujian
A/duck/Hunan/69/2004	H5N1	Anseriformes	Domestic	Hunan
A/chicken/Xinjiang/16/2005	H5N1	Galliformes	Domestic	Xinjiang
A/chicken/Anhui/39/2004	H5N1	Galliformes	Domestic	Anhui
A/shrike/Tibet/13/2006	H5N1	Passerine	Wild	Tibet
A/wild duck/Hunan/021/2005	H5N1	Anseriformes	Wild	Hunan
A/duck/Yunnan/5310/2006	H5N1	Anseriformes	Domestic	Yunnan
A/duck/Hubei/Hangmei01/2006	H5N1	Anseriformes	Domestic	Hubei
A/duck/Hubei/49/2005	H5N1	Anseriformes	Domestic	Hubei
A/chicken/Hunan/21/2005	H5N1	Galliformes	Domestic	Hunan
A/duck/Hubei/xn/2007	H5N1	Anseriformes	Domestic	Hubei
A/goose/Hubei/65/2005	H5N1	Anseriformes	Domestic	Hubei
A/chicken/Sichuan/81/2005	H5N1	Galliformes	Domestic	Sichuan
A/chicken/Hunan/41/2004	H5N1	Galliformes	Domestic	Hunan
A/chicken/Xinjiang/17/2005	H5N1	Galliformes	Domestic	Xinjiang
A/duck/Jiangxi/80/2005	H5N1	Anseriformes	Domestic	Jiangxi
A/duck/Guangdong/173/04	H5N1	Anseriformes	Domestic	Guangdong
A/chicken/Xinjiang/54/2005	H5N1	Galliformes	Domestic	Xinjiang
A/chicken/Jiangxi/25/2004	H5N1	Galliformes	Domestic	Jiangxi
A/ostrich/Suzhou/097/2003	H5N1	Struthioniformes	Domestic	Jiangsu
A/chicken/Xinjiang/67/2005	H5N1	Galliformes	Domestic	Xinjiang
A/chicken/Xinjiang/28/2006	H5N1	Galliformes	Domestic	Xinjiang
A/chicken/Xinjiang/78/2005	H5N1	Galliformes	Domestic	Xinjiang
A/duck/Guangxi/xa/2001	H5N1	Anseriformes	Domestic	Guangxi
A/chicken/Xinjiang/68/2005	H5N1	Galliformes	Domestic	Xinjiang
A/parrot/Guangdong/C99/2005	H5N1	Psittaciformes	Domestic	Guangdong
A/duck/Shanghai/xj/2002	H5N1	Anseriformes	Domestic	Shanghai
A/duck/Zhejiang/bj/2002	H5N1	Anseriformes	Domestic	Zhejiang
A/duck/Hubei/wp/2003	H5N1	Anseriformes	Domestic	Hubei
A/duck/Anhui/1/06	H5N1	Anseriformes	Domestic	Anhui
A/chicken/Jiangsu/18/2008	H5N1	Galliformes	Domestic	Jiangsu
A/chicken/Sheny/0606/2008	H5N1	Galliformes	Domestic	Liaoning
A/chicken/Shandong/A-5/2006	H5N1	Galliformes	Domestic	Shandong
A/chicken/Shanxi/2/2006	H5N1	Galliformes	Domestic	Shanxi
A/chicken/Shanxi/10/2006	H5N1	Galliformes	Domestic	Shanxi
A/chicken/Shandong/A-10/2006	H5N1	Galliformes	Domestic	Shandong
A/chicken/Liaoning/A-11/2006	H5N1	Galliformes	Domestic	Liaoning
A/chicken/Hebei/102/2005	H5N1	Galliformes	Domestic	Hebei
A/chicken/Henan/A-7/2006	H5N1	Galliformes	Domestic	Henan

A/chicken/Hebei/126/2005	H5N1	Galliformes	Domestic	Hebei
A/chicken/Hebei/326/2005	H5N1	Galliformes	Domestic	Hebei
A/duck/Guangxi/GXd-5/2010	H6N1	Anseriformes	Domestic	Guangxi
A/duck/Hunan/S4234/2011	H5N1	Anseriformes	Domestic	Hunan
A/wild duck/Jilin/HF/2011	H5N1	Anseriformes	Wild	Jilin
A/wild duck/Jilin/ZF/2011	H5N1	Anseriformes	Wild	Jilin
A/chicken/Jiangsu/k0402/2010	H5N1	Galliformes	Domestic	Jiangsu
A/wild duck/Shandong/2/2011	H5N1	Anseriformes	Wild	Shandong
A/goose/Jiangsu/k0403/2010	H5N1	Anseriformes	Domestic	Jiangsu
A/great crested-grebe/Qinghai/1/2009	H5N1	Anseriformes	Wild	Qinghai
A/brown-headed gull/Qinghai/9/2009	H5N1	Charadriiformes	Wild	Qinghai
A/great black-headed gull/Qinghai/8/2009	H5N1	Charadriiformes	Wild	Qinghai
A/chicken/Tibet/6/2008	H5N1	Galliformes	Domestic	Tibet
A/chicken/Guizhou/7/2008	H5N1	Galliformes	Domestic	Guizhou
A/duck/Hunan/8/2008	H5N1	Anseriformes	Domestic	Hunan
A/swan/Shanghai/10/2009	H5N1	Anseriformes	Wild	Shanghai
A/bar-headed goose/Qinghai/1-HVRI/2006	H5N1	Anseriformes	Wild	Qinghai
A/bar-headed goose/Qinghai/3/2005	H5N1	Anseriformes	Wild	Qinghai
A/wild duck/Liaoning/8/2006	H5N1	Anseriformes	Wild	Liaoning
A/chicken/Liaoning/23/2005	H5N1	Galliformes	Domestic	Liaoning
A/pied magpie/Liaoning/7/2006	H5N1	Passerine	Wild	Liaoning
A/bar-headed goose/Tibet/8/2006	H5N1	Anseriformes	Wild	Tibet
A/great cormorant/Tibet/12/2006	H5N1	Anseriformes	Wild	Tibet
A/chicken/Tibet/LZ01/2010	H5N2	Galliformes	Domestic	Tibet
A/chicken/Hunan/8/2008	H5N1	Galliformes	Domestic	Hunan
A/chicken/Hunan/3/2007	H5N1	Galliformes	Domestic	Hunan
A/duck/Hunan/3/2007	H5N1	Anseriformes	Domestic	Hunan
A/duck/Hunan/11/2007	H5N1	Anseriformes	Domestic	Hunan
A/chicken/Fujian/1/2007	H5N1	Galliformes	Domestic	Fujian
A/duck/Hunan/29/2006	H5N1	Anseriformes	Domestic	Hunan
A/goose/Yunnan/3798/2006	H5N1	Anseriformes	Domestic	Yunnan
A/goose/Yunnan/4371/2006	H5N1	Anseriformes	Domestic	Yunnan
A/goose/Guangdong/1/1996	H5N1	Anseriformes	Domestic	Guangdong
A/duck/Fujian/5476/2008	H7N7	Anseriformes	Domestic	Fujian
A/mallard/Jiangxi/10668/2005	H6N1	Anseriformes	Wild	Jiangxi
A/black-billed magpie/Guangxi/29/2005	H9N2	Passerine	Wild	Guangxi
A/chicken/Jilin/hd/2002	H5N1	Galliformes	Domestic	Jilin
A/chicken/Guangdong/1/2005	H5N1	Galliformes	Domestic	Guangdong
A/duck/Fujian/7794/2007	H6N6	Anseriformes	Domestic	Fujian
A/duck/Shantou/494/2007	H6N6	Anseriformes	Domestic	Guangdong
A/chicken/Rizhao/1436/2013	H9N2	Galliformes	Domestic	Shandong
A/duck/Guangxi/890/2007	H6N8	Anseriformes	Domestic	Guangxi
A/chicken/Henan/13/2004	H5N1	Galliformes	Domestic	Henan
A/chicken/Henan/01/2004	H5N1	Galliformes	Domestic	Henan

A/duck/Hunan/908/2005	H6N2	Anseriformes	Domestic	Hunan
A/chicken/Wenzhou/606/2013	H9N2	Galliformes	Domestic	Zhejiang
A/duck/Guangxi/13/2004	H5N1	Anseriformes	Domestic	Guangxi
A/pigeon/Nanchang/2-0461/2000	H9N2	Passerine	Wild	Jiangxi
A/wild waterfowl/Dongting/PC2562/2012	H9N2	Anseriformes	Wild	Hunan
A/duck/Shantou/1522/2001	H6N2	Anseriformes	Domestic	Guangdong
A/duck/Fujian/3193/2005	H6N2	Anseriformes	Domestic	Fujian
A/duck/Fujian/3354/2006	H6N6	Anseriformes	Domestic	Fujian
A/duck/Hubei/2/2010	H6N6	Anseriformes	Domestic	Hubei
A/duck/Jiangxi/7348/2007	H6N2	Anseriformes	Domestic	Jiangxi
A/duck/Fujian/5643/2005	H6N2	Anseriformes	Domestic	Fujian
A/duck/Shantou/10124/2006	H6N2	Anseriformes	Domestic	Guangdong
A/chicken/Rizhao/723/2013	H9N2	Galliformes	Domestic	Shandong
A/chicken/Guangxi/G8/2009	H9N2	Galliformes	Domestic	Guangxi
A/chicken/Jiangxi/SD001/2013	H7N9	Galliformes	Domestic	Jiangxi
A/duck/Jiangxi/25186/2009	H7N6	Anseriformes	Domestic	Jiangxi
A/duck/Guangxi/1248/2006	H6N2	Anseriformes	Domestic	Guangxi
A/mallard/Jiangxi/8346/2004	H6N5	Anseriformes	Wild	Jiangxi
A/duck/Hunan/748/2005	H6N5	Anseriformes	Domestic	Hunan
A/duck/Jiangxi/16769/2010	H7N7	Anseriformes	Domestic	Jiangxi
A/duck/Hunan/573/2002	H6N2	Anseriformes	Domestic	Hunan
A/Muscovy duck/Fujian/CL/1997	H9N2	Anseriformes	Wild	Fujian
A/chicken/Jiangsu/1/00	H9N2	Galliformes	Domestic	Jiangsu
A/chicken/Hebei/4/2008	H9N2	Galliformes	Domestic	Hebei
A/wild duck/Hunan/211/2005	H5N1	Anseriformes	Wild	Hunan
A/duck/Zhejiang/2/2011	H7N3	Anseriformes	Domestic	Zhejiang
A/chicken/Guangdong/191/04	H5N1	Galliformes	Domestic	Guangdong
A/duck/Guizhou/5302/2007	H6N2	Anseriformes	Domestic	Guizhou
A/wild waterfowl/Dongting/C2032/2011	H9N2	Anseriformes	Wild	Hunan
A/chicken/Hubei/489/2004	H5N1	Galliformes	Domestic	Hubei
A/wild duck/Fujian/2/2011	H5N1	Anseriformes	Wild	Fujian
A/duck/Anhui/SC702/2013	H7N9	Anseriformes	Domestic	Anhui
A/mallard/SanJiang/151/2006	H6N2	Anseriformes	Wild	Guangxi
A/duck/Fujian/10774/2006	H6N6	Anseriformes	Domestic	Fujian
A/duck/Fujian/4276/2006	H6N2	Anseriformes	Domestic	Fujian
A/duck/Zhejiang/0607-13/2011	H1N2	Anseriformes	Domestic	Zhejiang
A/chicken/Wenzhou/645/2013	H7N7	Galliformes	Domestic	Zhejiang
A/duck/Hunan/S4111/2011	H9N2	Anseriformes	Domestic	Hunan
A/brown-headed gull/Qinghai/19/2009	H5N1	Charadriiformes	Wild	Qinghai
A/duck/Tibet/S2/2009	H9N2	Anseriformes	Domestic	Tibet
A/duck/Fujian/6388/2010	H7N3	Anseriformes	Domestic	Fujian
A/duck/Yangzhou/02/2005	H8N4	Anseriformes	Domestic	Jiangsu
A/duck/Guizhou/2773/2006	H6N2	Anseriformes	Domestic	Guizhou
A/chicken/Guangdong/174/04	H5N1	Galliformes	Domestic	Guangdong

A/duck/Hunan/S11893/2012	H4N6	Anseriformes	Domestic	Hunan
A/duck/Yangzhou/013/2008	H6N5	Anseriformes	Domestic	Jiangsu
A/wild duck/Jiangxi/19615/2009	H7N8	Anseriformes	Wild	Jiangxi
A/chicken/Henan/5/98	H9N2	Galliformes	Domestic	Henan
A/chicken/Shanghai/10/01	H9N2	Galliformes	Domestic	Shanghai
A/duck/Zhejiang/2245/2011	H5N1	Anseriformes	Domestic	Zhejiang
A/quail/Nanchang/7-026/2000	H3N6	Galliformes	Domestic	Jiangxi
A/wild duck/Shantou/992/2000	H2N8	Anseriformes	Wild	Guangdong
A/duck/Jiangxi/3190/2009	H7N9	Anseriformes	Domestic	Jiangxi
A/Baikal teal/Hongze/14/2005	H11N9	Anseriformes	Wild	Jiangsu
A/chicken/Ningxia/24/2006	H5N1	Galliformes	Domestic	Ningxia
A/duck/Hunan/S4150/2011	H5N1	Anseriformes	Domestic	Hunan
A/duck/Shantou/6847/2004	H6N2	Anseriformes	Domestic	Guangdong
A/pigeon/Shanghai/S1421/2013	H7N9	Passerine	Wild	Shanghai

Table A4.3 Taxon of the 320 AIV in Chapter 4
Sequence name, subtype, host order, host type and region types are summarized in each column.

Chapter 5
Determining the phylogenetic and phylogeographic origin of highly pathogenic avian influenza (H7N3) in Mexico

Published as Lu Lu, Samantha J. Lycett, and Andrew J. Leigh Brown. Determining the genetic and phylogeographic origin of highly pathogenic avian influenza H7N3 in Mexico.*PLos one.* 2014, 9, doi: 10.1371/journal.pone.0107330

Chapter Abstract

Highly pathogenic (HP) avian influenza virus (AIV) H7N3 outbreaks occurred 3 times in the Americas in the past 10 years and caused severe economic loss in the affected regions. In June/July 2012, new HP H7N3 outbreaks occurred at commercial farms in Jalisco, Mexico. Outbreaks continued to be identified in neighbouring states in Mexico till August 2013. To explore the origin of this outbreak, time resolved phylogenetic trees were generated from the eight segments of full-length AIV sequences in North America using BEAST. Location, subtype, avian host species and pathogenicity were modelled as discrete traits upon the trees using continuous time Markov chains. A further joint analysis among segments was performed using a hierarchical phylogenetic model (HPM) which allowed trait rates (location, subtype, host species) to be jointly inferred across different segments. The complete spatial diffusion process was visualised through virtual globe software. My results indicate that the Mexico HP H7N3 originated from the large North America low pathogenicity AIV pool through complicated reassortment events. Different segments were contributed by wild waterfowl from different N. American flyways. Five of the eight segments (HA, NA, NP, M, NS) were introduced from wild birds migrating along the central North American flyway, and PB2, PB1 and PA were introduced via the western North American flyway. These results highlight a potential role for Mexico as a hotspot of virus reassortment as it is where wild birds from different migration routes mix during the winter.

5.1 Introduction

Migratory birds are major candidates for long-distance dispersal of zoonotic pathogens and low pathogenicity (LP), avian-origin influenza A viruses (AIVs) are widely distributed in free-ranging water birds (Schnebel *et al.*, 2007). Wild birds spread their viruses to other wild as well as domestic birds as they migrate through an area, allowing extensive reassortment (Webster, 2002). Once introduced into poultry (especially chickens and turkeys), LPAI may switch to high pathogenic viruses (HPAI) with the introduction of basic amino acid residues into the haemagglutinin cleavage site, which is associated with a high mortality rate in poultry (Chen *et al.*, 1998; Stienekegrober *et al.*, 1992). In Chapter 3, and my associated publication (Lu *et al.*, 2014), I have recently shown that a higher inter-subtype reassortment rate can be found in wild Anseriformes than domestic Galliformes in the internal segments of Eurasian AIV, indicating the wild bird population was the source of the new reassortants, rather than domestic.

Migrating wild birds have been implicated in the spread and emergence of HPAI such as HP H5N1 and H7N3. Viral transmission between wild birds and domestic poultry, and consequent genetic exchange, has contributed to genomic reassortment which confounded disease control efforts (Campitelli *et al.*, 2004; Duan *et al.*, 2007). Although predictors of such outbreaks have long been sought, surveillance in wild birds in North America has failed to provide a clear early warning signal. Three H7N3 HPAI events in poultry have occurred in North Americas since 2000, and, in one case, it was reported that the outbreak H7N3 AIV were transmitted from poultry to humans (Tweed *et al.*, 2004). Phylogenetic analyses indicated that each of these H7N3 HPAI strains had a close relationship with LPAI isolated from wild birds sampled in neighbouring provinces (Berhane *et al.*, 2009; Hirst *et al.*, 2004; Suarez *et al.*, 2004).

In June 2012, H7N3 HPAI outbreaks were found in poultry farms in Jalisco state in Mexico, a region of high poultry density (Wainwright, 2012) and concurrent infections of humans with this HPAI A (H7N3) virus (2 cases) have been confirmed (Lopez-Martinez *et al.*, 2013). The outbreak has been affecting broilers, breeders,

layers and backyard poultry in the Mexican States of Jalisco, Aguascalientes, Guanajuato and Puebla: the latest outbreak reported by the World Organization for Animal Health (OIE) was on 19th May 2014. Ongoing epidemiological investigations have implicated contact with wild birds as a factor in the outbreaks (Wainwright, 2012). However, the specific origin of the novel outbreak strain and its relationship to the previous outbreak strains is not known.

The aim of this study is to investigate the origin of the precursor strain of the Mexico H7N3, using a Bayesian phylogeographic inference framework by reconstructing the spatiotemporal spread of AIV from wild birds in North America.

5.2 Methods

5.2.1 Data preparation

The complete genome of three outbreak strains of H7N3 from Mexico (A/chicken/Jalisco/CPA1/2012; A/chicken/Jalisco/12283/2012; A/Mexico/InDRE7218/2012) and all previously published influenza A virus sequences of North American lineage (complete genome only) were downloaded from GenBank on 1st March 2013. Sequences of each gene segment were aligned using MUSCLE v3.5 (Edgar, 2004). Maximum likelihood (ML) phylogenetic trees for each segment were generated using RAxML v7.04 (Stamatakis *et al.*, 2005), each employing a GTR GAMMA substitution model with 500 bootstraps. I established a full genome dataset which was composed of the same 2343 North American strains for each segment. The HA and NA segments have extremely high divergence between different subtypes, therefore, I used all available H7 and N3 to generate the raw trees of HA and NA segment respectively. Since there are diverse reassortment and interactions among the six internal segments, I constructed ML trees for the internal segments from all subtypes in order to identify the outbreak strains related strains. Background sequences for further study were selected from the closest clades to the novel H7N3 HPAI viruses on the maximum likelihood tree of each segment. The final dataset of 427 AIV strains collected over a 12 year period (2001 to 2012) is displayed in Table A5.1. In this table the segments selected for analysis for each

strain are indicated. For the majority of strains, only one segment is selected (n=289), but for the remainder of isolates more than one segment is included. There are 131 HA (H7) sequences included in the analysis based on their relationship to the Mexico H7N3 strain (1698 nt); other H7 sequences included in the joint analysis are related to the outbreak strain in other segments. For the other segments the distribution is as follows: NA (N3), n=100 (1410 nt); PB2, n = 86 sequences (alignment length of 2277 nucleotides); PB1, n=67 (2271 nt); PA, n= 89 (2148 nt); NP, n=79 (1494 nt); MP, n=39 (982 nt); and NS, n=42 (838 nt). The trait information (host bird order; host species; location; flyway; state; subtype) of the background AIV sequences are also provided in Table A5.1.

5.2.2 Time-scaled phylogeny reconstruction

To estimate the origin in time and space of the HPAI H7N3 outbreak strain in Mexico, models in BEAST v.1.7.3 (Drummond & Rambaut, 2007; Drummond *et al.*, 2012) were applied independently to each gene segment (each segment has a different number of AIV sequences). Different combinations of substitution models: general time-reversible (GTR) substitution model+ Γ distributed site-site rate variation (Lanave *et al.*, 1984) and SRD06 (Shapiro *et al.*, 2006); clock models: strict and uncorrelated relaxed lognormal; and population size models: constant size, exponential, skyride models were evaluated by Bayes Factor test. The best fitting model - incorporating a GTR substitution model+ Γ with uncorrelated lognormal relaxed molecular clocks and a constant-population coalescent process prior over the phylogenies was selected. Parameters were estimated using the Bayesian Monte-Carlo Markov Chain (MCMC) approach implemented in BEAST. MCMC chains were run for 100 million states, sampled every 10,000 states with 10% burn-in. MCMC convergence, and effective sample size of parameter estimates were evaluated using Tracer 1.5 (http://beast.bio.ed.ac.uk). Trees were summarized by using Tree Annotator and visualized by using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

A graphical representation of the origin of HPAI H7N3 Mexico was obtained by spatial reconstruction using a Bayesian framework. The SPREAD

application (Bielejec *et al.*, 2011) was used to convert the estimated divergence times and the spatially-annotated time-scaled phylogeny (by associating each location with a particular latitude and longitude) to a spatiotemporal movement. The mapped objects were exported to keyhole markup language (KML) files and then were visualized by geographic information systems software: ARCGIS (http://www.esri.com/software/arcgis). The map source is OpenStreetMap (http://www.openstreetmap.org/).

5.2.3 Joint analysis of the transition rates between discrete traits

The dataset I used in the Bayesian phylogenetic analysis is a sub-sampled dataset based on the H7N3 Mexico AIV in the Maximum Likelihood trees composed of all available North America AIV sequences, therefore the number of sequences comprising the selected lineage is distributed unevenly among segments. A study improved the resolution of the phylogeographic history of the Dengue virus by combining information from multiple phylogeographic datasets in a hierarchical setting using a hierarchical phylogenetic model (HPM) framework (Suchard *et al.*, 2003). Virus sequences were partitioned by gene and through pooling information regarding the evolutionary history across genes, then the spread of the virus through the continent was estimated (Cybis *et al.*, 2013). Consequently, I decided to use a similar approach, and to reduce the effects of sampling bias and identify reassortment events in each individual segment, a joint analysis of all segments was performed using a hierarchical phylogenetic model (HPM) in this chapter (Drummond & Rambaut, 2007; Edo-Matas *et al.*, 2011).

In this hierarchical structure the prior on a parameter may be shared between partitioned datasets in order to increase the efficiency of its estimation (Suchard *et al.*, 2003). Here, I use joint analysis for discrete trait models where each segment is treated as an independent dataset with an individual relaxed clock model and tree prior, but shares the discrete trait model (subtype, location and host) with all other segments resulting in a joint transition matrix for each discrete trait. Note that is it not necessary to have the same taxa in each of the data partitions within this framework, since the transition rate matrix is estimated using the tip patterns and

trees from all of the partitions, and when there are missing states and transitions in one partition the estimates of those rates comes from the other partitions where the states and transitions are present.

Bayesian stochastic search variable selection (BSSVS) was employed to reduce the number of parameters to those with significantly non-zero transition rates (Lemey *et al.*, 2009). A Bayes Factor support (BF) > 100 was considered to indicate decisive support, whereas $3 \leq BF \leq 30$, and $30 \leq BF \leq 100$ indicate substantial, and strong statistical support, respectively. The Bayes Factor threshold was estimated from Rate indicators using the indicator BF tool in the BEAST v1.6.2 package. The default Poisson distribution (with mean equal to the number of states) was used as a prior on the number of rate indicators.

Three factors: host, geographic location and subtype were analysed using discrete trait models in BEAST (Goldberg & Igic, 2008). Specifically, host population was discretized in terms of bird orders and bird species, while geographic location was discretized in terms of states and migration flyway routes in North America. The hosts of AIV in this study were first categorized into five different bird orders: Anseriformes (ans), Charadriiformes (cha) and Galliformes (gal), Gruiformes (gru) and Passeriformes (pas); The majority of this AIV data set were collected from Anseriformes (366/427), which were further classified as five major species groups: mallard (Anas platyrhynchos), northern pintail (Anasacuta), northern shoveller (Anas clypeata), blue-winged teal (Anas discors), green-winged teal (Anas carolinensis) and other Anseriformes. Each AIV sequence was assigned a discrete geographical state according to its province / state of isolation (in Table A5.1). In total 26 US states and Canadian provinces were considered in this study. In addition, each state and province was categorised into a specific North American flyway: the Atlantic, Mississippi, Central, or Pacific flyway, following a previous study (Lam *et al.*, 2012). Viral sequences were also labelled according to their host species for the analysis of the species contribution to AIV gene flow. The distributions of sequences for each discrete trait (host order, host species, flyway and location) are summarized in Table 5.1-5.4.

Host Type	Number of Sequences
Anseriformes-Wild (ans-wild)	366
Charadriiformes-Wild (cha-wild)	45
Passeriformes-Wild (pas-wild)	2
Gruiformes-Wild (gru-wild)	3
Galliformes-Domestic (gal-domestic)	8
Mexico-gal-domestic	3

Table 5.1. Host order states and the sequences distributions

Host species	Number of Sequences
Mallard	163
Northern shoveler	67
Northern pintail	24
Green winged teal	39
Blue winged teal	29
American black duck	9
Others Anseriformes	35

Table 5.2. Host species (Wild Anseriformes only) states and the sequences distributions

State	Flyway	Numbers of Sequences
Alaska	Pacific	44
Alberta	Central	28
British Columbia	Pacific	5
California	Pacific	114
Delaware	Atlantic	26
Illinois	Mississippi	28
Iowa	Mississippi	5
Jalisco	Outbreak	3
Louisiana	Mississippi	2
Manitoba	Mississippi	3
Maryland	Atlantic	4
Minnesota	Mississippi	11
Mississippi	Mississippi	14
Missouri	Mississippi	19
Nebraska	Central	2
New Brunswick	Atlantic	10
New Jersey	Atlantic	15
North Carolina	Atlantic	3
Nova Scotia	Atlantic	3
Ohio	Mississippi	18
Pennsylvania	Atlantic	1
Quebec	Atlantic	12

Saskatchewan	Central	3
Texas	Central	10
Washington	Pacific	9
Wisconsin	Mississippi	35

Table 5.3. Location (State/province) and the sequences distributions

Number	Flyway	Numbers
1	Atlantic	76
2	Central	44
3	Mississippi	132
4	Pacific	172
5	Outbreak	3

Table 5.4. Flyways and the sequences distributions

These traits were jointly analysed across all eight segments using irreversible substitution models and strict clocks in the discrete trait models. An exponential distribution with mean equal to 1.0 was chosen for the discrete trait clock rate prior in order to favour smaller numbers of transitions across the phylogeny of each segment (especially PB2, M and NS). The parameter values in each trait model were examined using Tracer v1.6. Significant transition rate estimates between discrete traits were calculated using the same methods as in Chapter 3 (Lu *et al.*, 2014) and plotted as a network in Cytoscape v2.8.0 (http://www.cytoscape.org/).

5.2.4 Flyway-restricted transmission and model comparisons

The location trait initially had 26 states, resulting in 325 possible transition rates, however to avoid over-parameterising the model, I reduced the number of transitions in the matrix to 98 to increase the power of the analysis. To demonstrate that it was not the case that the reduced flyway matrix performed best only because of the lower number of parameters, I performed a randomization test with the same number of parameters: I randomly switched off (meaning the indicator of a transition pair was forced from 1 to 0, ensuring that it was not included in the model during sampling) the same number of transitions as were disallowed in the between flyway states matrix. I did this for 10 replicates. I then compared the performance to that of the reduced flyway matrix and I found the latter has the best performance, indicating the model with reduced flyway transition is the best model to explain the spatial transmission with location state.

I further modified the discrete transition model specification in the xml configuration files to reflect the hypothesis the gene flow of AIV carried by migration birds are restricted between flyways but can mix in Mexico (and further south). By default, all possible transitions between states are permitted, but I modified the settings to disallow transitions between locations on different flyways (Table A5.2). The output from this restricted model can then be compared to the original model. The corresponding indicators for BF thresholds were recalculated based on the reduced number of transitions. Eight discrete trait models were further compared based on 3 settings: 1) with BSSVS or without BSSVS implemented; 2) Asymmetric or symmetric transitions model; 3) with reduced or original matrix (The names of the models are: 1. Reduce_BSSVS_asym; 2. Reduce_nonBSSVS_asym; 3. Reduce_BSSVS_sym; 4. Reduce_nonBSSVS_sym; 5. Original_BSSVS_asym; 6. Original_nonBSSVS_asym; 7. Original_BSSVS_sym; 8. Original_nonBSSVS_sym).

The fits of the discrete models can be compared by Bayes Factor tests, however, these employ estimations of the marginal likelihood, and there are different ways to perform this estimation. Consequently, the applicability and differences in marginal likelihood estimators were considered and, The three model comparison methods HME, AICM and PS/SS are described in Chapter 2, Section 2.3.

I compared the 8 different discrete trait models (states and provinces as discrete trait) via the three methods, using Tracer V 1.6 with a further subsampled log file of each model (>1000 states after burn-in). Nevertheless, I applied the different estimators where possible. For the joint analysis (across all segments) of the discrete trait models the comparisons are using AICM and HME. However, PS/SS sampling could not be applied on the joint analysis (over all segments) performed in this study because it is not yet incorporated within BEAST. The problem might lie with the implementation of the particular operators (Gibbs samplers) that are used in the joint analysis, and it is not yet known how to integrate them with the PS/SS code (Baele *et al.*, 2012; Baele *et al.*, 2013); Also PS (or SS) methods of estimating marginal likelihoods are important when fundamentally different models are being compared; (e.g. two clock models or two population growth models: Simple

parametric models available include constant size N(t) = Ne (1 parameter), while exponential growth N(t) = Nee-gt (2 parameters) and logistic growth (3 parameters) (Drummond & Rambaut, 2007).), which contain different priors and algorithms for the likelihood calculations as in (Drummond *et al.*, 2006; Drummond & Rambaut, 2007)). However, the models I compared, e.g. the model of reduced flyway matrix and the other 10 models with random reduced matrix have the same form differing only in which transition is fixed on or off, not in numbers of parameters, nor priors, nor the likelihood calculation algorithm.

Additionally since I was unable to perform PS/SS in the joint analysis situation, I performed it for the discrete traits models on the HA segment only. To be specific, in the joint analysis of the discrete trait models, I compared results given by HME (Table 5.5A) with those given by AICM (Table 5.5B), and also those of the flyway reduced models with random reduced models estimated by HME (Table 5.6A) and AICM (Table 5.6B). The comparisons were conducted in Tracer V 1.6 and I subsampled the large log files (100 million) to 1000 rows for each model due to the restricted capacity of Tracer. Both the AICM and HME estimators showed that the best discrete trait model is the symmetric BSSVS with reduced matrix (Table 5.5A, B), and the best reduced model is that with between-flyway pairs switched off (Table 5.6 A, B).

I also generated a set of models based on the HA segment only instead of the joint analysis of all 8 segments. For HA, there are 22 location traits, the initial 231 non-reversible pairs, which fell to 76 non-reversible pairs after switching off between flyway diffusions. I compared discrete trait models with the same parameter settings in the joint analysis, using the results estimated by stepping stone (SS) (Table 5.7A), AICM (Table 5.7B) and HME (Table 5.7C). Similar to the joint analysis, the results also showed the discrete trait model with BSSVS and reduced matrix was preferred over the other models.

Then I compared the reduced-flyway model with 10 random models that had the same number of transitions by HME, AICM and SS (Table 5.8). I found HME

failed to identify the difference between reduced BSSVS symmetric model (Mod1) and reduced symmetric model without BSSVS (Mod2) (Table 5.7A), and also failed to the difference between flyway reduced model (Flyway) and random reduced model1 (ran1) (Table 58A), as both showed a BF <3. However AICM (Table 5.7B, Table 5.8B) and SS showed more obvious differences between the models that are being compared (Table 5.7C, Table 5.8C), which confirmed that the flyway reduced model performed better than random reduced models. Therefore, I concluded that the symmetric discrete trait model with a reduced matrix in BSSVS is preferred over all the other models, and the model with between flyway diffusion pairs switched off is favoured over a randomly reduced matrix.

Joint Discrete trait models	HME	S.E.	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6	Mod7	Mod8
Mod1 Reduce_BSSVS_sym[a]	**-54589**	**+/- 0.328**	**-**	**25**	**161**	**144**	**219**	**173**	**550**	**653**
Mod2 Reduce_nonBSSVS_sym	-54615	+/- 0.054	-25	-	135	119	193	148	525	627
Mod3 Reduce_BSSVS_asym	-54750	+/- 0.1	-161	-135	-	-16	58	13	389	492
Mod4 Reduce_nonBSSVS_asym	-54734	+/- 0.048	-144	-119	16	-	74	29	406	508
Mod5 Original_BSSVS_sym	-54808	+/- 0.042	-219	-193	-58	-74	-	-45	331	434
Mod6 Original_nonBSSVS_sym	-54763	+/- 0.084	-173	-148	-13	-29	45	-	377	479
Mod7 Original_BSSVS_asym	-55139	+/- 0.053	-550	-525	-389	-406	-331	-377	-	103
Mod8Original_nonBSSVS_asym	-55242	+/- 0.764	-653	-627	-492	-508	-434	-479	-103	-

Table 5.5A HME estimates for the fitness of models in joint analysis

The names of the 8 models (Mod1-8) in the comparison test;
The estimated harmonic marginal likelihood of posterior (HME), lower (absolute) values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;
The Bayes factor comparisons are shown in the matrix being composed by column 4 to 11. In each row of the matrix, the positive value (log BF) in a cell represented the support of one model (in column 1) over the other (the title of the matrix). The difference of logBF=5 is considered a strong preference of one model over the others.

Joint discrete trait	AIC	S.E.	Mod	Mod	Mod	Mod	Mod	Mod	Mod	Mod
Mod1	**1110**	**+/-**	**-**	**107**	**490**	**712**	**863**	**110**	**408**	**658**
Mod2	1111	+/-	-107	-	384	605	756	995	397	647
Mod3	1115	+/-	-490	-384	-	221	372	611	359	609
Mod4	1117	+/-	-712	-605	-221	-	151	390	337	587
Mod5	1118	+/-	-863	-756	-372	-151	-	239	322	571
Mod6	1121	+/-	-110	-995	-611	-390	-239	-	298	548
Mod7	1151	+/-	-408	-397	-359	-337	-322	-298	-	249
Mod8	1176	+/-	-658	-647	-609	-587	-571	-548	-249	-

Table 5.5B AICM estimates for the fitness of models in joint analysis

The names of the 8 models (Mod1-8) in the comparison test;

The estimated AICM score of posterior, lower values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

The AICM comparisons are shown in the matrix being composed by column 4 to 11. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix). The difference of AICM=10 is considered a strong preference of one model over the others

Joint reduced models	HME	S.E.	Flyway	Ran1	Ran2	Ran3	Ran4	Ran5	Ran6	Ran7	Ran8	Ran9	Ran10
Flyway	**-54589**	**+/- 0.398**	**-**	**30**	**6**	**67**	**18**	**24**	**19**	**14**	**38**	**341**	**2651**
Ran1	-54619	+/- 0.083	-30	-	-24	37	-11	-6	-11	-16	8	311	2621
Ran2	-54595	+/- 0.108	-6	24	-	61	13	18	13	8	32	335	2645
Ran3	-54656	+/- 0.568	-67	-37	-61	-	-48	-43	-48	-53	-29	274	2584
Ran4	-54608	+/- 0.14	-18	11	-13	48	-	5	0	-5	19	323	2633
Ran5	-54613	+/- 0.041	-24	6	-18	43	-5	-	-5	-10	14	318	2628
Ran6	-54608	+/- 0.099	-19	11	-13	48	0	5	-	-5	19	323	2633
Ran7	-54603	+/- 0.541	-14	16	-8	53	5	10	5	-	24	327	2637
Ran8	-54627	+/- 0.063	-38	-8	-32	29	-19	-14	-19	-24	-	304	2613
Ran9	-54930	+/- 0.708	-341	-311	-335	-274	-323	-318	-323	-327	-304	-	2310
Ran10	-57240	+/- 0.043	-2651	-2621	-2645	-2584	-2633	-2628	-2633	-2637	-2613	-2310	-

Table 5.6A HME estimates for the fitness of models with reduced number of transitions.

The names of the 11 models (Flyway model and Ran1-10) in the comparison test: Model1 is the non-reversible BSSVS model without between flyway transitions (Flyway); Ran1-10 are models randomly switch off same number of transitions (Ran1-10);

The estimated harmonic marginal likelihood of posterior (HME), lower (absolute) values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

The Bayes factor comparisons are shown in the matrix being composed by column 4 to 13. In each row of the matrix, the positive value (log BF) in a cell represented the support of one model (in column 1) over the other (the title of the matrix). The difference of logBF=5 is considered a strong preference of one model over the others.

Joint reduced models	AICM	S.E.	Flyway	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Flyway	**111031**	**+/- 5.2**	**-**	**18**	**70**	**108**	**129**	**157**	**158**	**163**	**171**	**521**	**407822**
Run1	111049	+/- 4.4	-18	-	52	90	111	139	139	145	153	502	407804
Run2	111101	+/- 5.9	-70	-52	-	38	59	87	88	93	101	451	407752
Run3	111138	+/- 14.9	-108	-90	-38	-	21	50	50	55	63	413	407714
Run4	111160	+/- 6.5	-129	-111	-59	-21	-	28	28	34	42	391	407693
Run5	111188	+/- 5.6	-157	-139	-87	-50	-28	-	0	6	13	363	407664
Run6	111188	+/- 4.3	-158	-139	-88	-50	-28	0	-	6	13	363	407664
Run7	111194	+/- 12.8	-163	-145	-93	-55	-34	-6	-6	-	8	357	407659
Run8	111201	+/- 8.6	-171	-153	-101	-63	-42	-13	-13	-8	-	350	407651
Run9	111551	+/- 9.9	-521	-502	-451	-413	-391	-363	-363	-357	-350	-	407301
Run10	518853	+/- 3495.6	-407822	-407804	-407752	-407714	-407693	-407664	-407664	-407659	-407651	-407301	-

Table 5.6B AICM estimates for the fitness of models with reduced number of transitions.

The names of the 11 models (Flyway model and Run1-10) in the comparison test: Model1 is the non-reversible BSSVS model without between flyway transitions (Flyway); Run1-10 are models randomly switch off same number of transitions (Run1-10).

The estimated AICM score of posterior, lower values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

Differences of models are shown in the matrix being composed by column 4 to 13. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix). The difference of AICM=10 is considered a strong preference of one model over the others

HA discrete trait models	HME	S.E.	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6	Mod7	Mod8
Mod1 Reduce_BSSVS_sym	-8814	+/- 0.07	-	-1	110	88	175	188	441	478
Mod2 Reduce_nonBSSVS_sym	**-8813**	**+/- 0.018**	**1**	**-**	**111**	**89**	**176**	**189**	**442**	**479**
Mod3 Reduce_BSSVS_asym	-8924	+/- 0.05	-110	-111	-	-22	65	78	331	368
Mod4 Reduce_nonBSSVS_asym	-8901	+/- 0.032	-88	-89	22	-	88	100	353	391
Mod5 Original_BSSVS_sym	-8989	+/- 0.511	-175	-176	-65	-88	-	13	266	303
Mod6 Original_nonBSSVS_sym	-9001	+/- 0.068	-188	-189	-78	-100	-13	-	253	291
Mod7 Original_BSSVS_asym	-9255	+/- 0.054	-441	-442	-331	-353	-266	-253	-	37
Mod8 Original_nonBSSVS_asym	-9292	+/- 0.975	-478	-479	-368	-391	-303	-291	-37	-

Table 5.7A HME estimates for the fitness of models in HA segment

The names of the 8 models (Mod1-8) in the comparison test;

The estimated harmonic marginal likelihood of posterior (HME), lower (absolute) values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

The Bayes factor comparisons are shown in the matrix being composed by column 4 to 11. In each row of the matrix, the positive value (log BF) in a cell represented the support of one model (in column 1) over the other (the title of the matrix), logBF =5 is considered a strong preference of one model over the others.

HA Discrete trait model	AICM	S.E.	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6	Mod7	Mod8
Mod1 Reduce_BSSVS_sym	**18494**	**+/- 3.624**	**-**	**35**	**116**	**411**	**165**	**754**	**1681**	**1834**
Mod2 Reduce_nonBSSVS_sym	18530	+/- 9.741	-35	-	81	376	130	719	1645	1799
Mod3 Reduce_BSSVS_asym	18610	+/- 1.583	-116	-81	-	295	49	638	1564	1718
Mod4 Reduce_nonBSSVS_asym	18905	+/- 5.353	-411	-376	-295	-	-245	344	1270	1423
Mod5 Original_BSSVS_sym	18660	+/- 5.817	-165	-130	-49	245	-	589	1515	1669
Mod6 Original_nonBSSVS_sym	19249	+/- 4.442	-754	-719	-638	-344	-589	-	926	1080
Mod7 Original_BSSVS_asym	20175	+/- 8.931	-1681	-1645	-1564	-1270	-1515	-926	-	154
Mod8 Original_nonBSSVS_asym	20329	+/- 4.468	-1834	-1799	-1718	-1423	-1669	-1080	-154	-

Table 5.7B AICM estimates for the fitness of models in HA segment

The names of the 8 models (Mod1-8) in the comparison test;

The estimated AICM of posterior, lower values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

The AICM comparisons are shown in the matrix being composed by column 4 to 11. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix). AICM score =10 is considered a strong preference of one model over the others.

HA discrete trait models	SS	S.E.	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6	Mod7	Mod8
Mod1 Reduce_BSSVS_sym	**-8905.8**	**-**	**-**	**4**	**5**	**7**	**13**	**12**	**12**	**18**
Mod2 Reduce_nonBSSVS_sym	-8910.0	-	-4	-	1	2	8	8	8	14
Mod3 Reduce_BSSVS_asym	-8911.2	-	-5	-1	-	1	7	6	7	13
Mod4 Reduce_nonBSSVS_asym	-8912.3	-	-7	-2	-1	-	6	5	6	12
Mod5 Original_BSSVS_sym	-8918.3	-	-13	-8	-7	-6	-	-1	0	6
Mod6 Original_nonBSSVS_sym	-8917.6	-	-12	-8	-6	-5	1	-	1	7
Mod7 Original_BSSVS_asym	-8918.2	-	-12	-8	-7	-6	0	-1	-	6
Mod8 Original_nonBSSVS_asym	-8924.3	-	-18	-14	-13	-12	-6	-7	-6	-

Table 5.7C SS estimates for the fitness of models in HA segment

The names of the 8 models (Mod1-8) in the comparison test;

The estimated log marginal likelihood using steeping stone sampling (SS), lower (absolute) values of marginal likelihood indicate a better fit to the data.

The model with the best performance is indicated in bold;

The standard error of SS estimated using 1000 bootstrap replicates;

The comparisons are shown in the matrix being composed by column 4 to 13. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix).

HA reduced models	HME	S.E.	Flyway	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Flyway	**-8814**	**+/- 0.167**	**-**	**1**	**17**	**18**	**25**	**25**	**26**	**29**	**35**	**40**	**43**
Run1	-8815	+/- 0.076	-1	-	16	17	24	24	25	28	34	39	42
Run2	-8830	+/- 0.042	-17	-16	-	2	9	9	10	12	18	24	27
Run3	-8832	+/- 0.113	-18	-17	-2	-	7	7	8	11	17	22	25
Run4	-8839	+/- 0.092	-25	-24	-9	-7	-	0	1	4	10	15	18
Run5	-8839	+/- 0.047	-25	-24	-9	-7	0	-	1	4	10	15	18
Run6	-8840	+/- 0.044	-26	-25	-10	-8	-1	-1	-	3	9	14	17
Run7	-8843	+/- 0.15	-29	-28	-12	-11	-4	-4	-3	-	6	11	14
Run8	-8849	+/- 0.039	-35	-34	-18	-17	-10	-10	-9	-6	-	5	8
Run9	-8854	+/- 0.038	-40	-39	-24	-22	-15	-15	-14	-11	-5	-	3
Run10	-8857	+/- 0.08	-43	-42	-27	-25	-18	-18	-17	-14	-8	-3	-

Table 5.8A HME estimates for the fitness of models with reduced number of transitions

The names of the 11 models (Flyway model and Run1-10) in the comparison test: Model1 is the non-reversible BSSVS model without between flyway transitions (Flyway); Run1-10 are models randomly switch off same number of transitions (Run1-10);

The estimated harmonic marginal likelihood of posterior (HME), lower (absolute) values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

The standard error of HME estimated using 1000 bootstrap replicates;

The Bayes factor comparisons are shown in the matrix being composed by column 4 to 13. In each row of the matrix, the positive value (log BF) in a cell represented the support of one model (in column 1) over the other (the title of the matrix), logBF=5 is considered a strong preference of one model over the others.

HA reduced models	AICM	S.E.	Flyway	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Flyway	**17014**	**+/- 0.852**	**-**	**64**	**73**	**128**	**151**	**197**	**113**	**132**	**117**	**183**	**86**
Run1	17078	+/- 1.718	-64	-	10	64	87	133	50	69	54	119	22
Run2	17087	+/- 1.181	-73	-10	-	54	77	124	40	59	44	110	12
Run3	17142	+/- 0.963	-128	-64	-54	-	23	69	-15	5	-10	55	-42
Run4	17165	+/- 1.256	-151	-87	-77	-23	-	46	-38	-18	-34	32	-65
Run5	17211	+/- 2.2	-197	-133	-124	-69	-46	-	-84	-64	-80	-14	-111
Run6	17127	+/- 3.191	-113	-50	-40	15	38	84	-	19	4	70	-28
Run7	17146	+/- 2.114	-132	-69	-59	-5	18	64	-19	-	-15	51	-47
Run8	17131	+/- 1.559	-117	-54	-44	10	34	80	-4	15	-	66	-32
Run9	17197	+/- 1.593	-183	-119	-110	-55	-32	14	-70	-51	-66	-	-97
Run10	17100	+/- 2.163	-86	-22	-12	42	65	111	28	47	32	97	-

Table 5.8B AICM estimates for the fitness of models with reduced number of transitions

The names of the 11 models (Flyway model and Run1-10) in the comparison test: Model1 is the non-reversible BSSVS model without between flyway transitions (Flyway); Run1-10 are models randomly switch off same number of transitions (Run1-10);

The estimated AICM of posterior, lower values of marginal likelihood indicate a better fit to the data. The model with the best performance is indicated in bold;

Differences of models are shown in the matrix being composed by column 4 to 11. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix). AICM score=10 is considered a strong preference of one model over the others.

HA reduced models	SS	S.E.	Flyway	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Flyway	**-8906**	**-**	**-**	**15**	**36**	**55**	**62**	**66**	**59**	**62**	**68**	**61**	**43**
Run1	-8921	-	-15	-	21	40	47	51	44	47	53	46	28
Run2	-8941	-	-36	-21	-	20	27	31	23	27	32	25	7
Run3	-8961	-	-55	-40	-20	-	7	11	4	7	13	6	-13
Run4	-8968	-	-62	-47	-27	-7	-	4	-4	0	6	-1	-20
Run5	-8972	-	-66	-51	-31	-11	-4	-	-7	-4	2	-5	-24
Run6	-8965	-	-59	-44	-23	-4	4	7	-	3	9	2	-16
Run7	-8968	-	-62	-47	-27	-7	0	-4	-3	-	6	-1	-20
Run8	-8974	-	-68	-53	-32	-13	-6	2	9	-6	-	-7	-25
Run9	-8967	-	-61	-46	-25	-6	1	-5	-2	1	7	-	-18
Run10	-8949	-	-43	-28	-7	13	20	-24	16	20	25	18	-

Table 5.8C SS estimates for the fitness of models with reduced number of transitions

The names of the 11 models (Flyway model and Run1-10) in the comparison test: Model1 is the non-reversible BSSVS model without between flyway transitions (Flyway); Run1-10 are models randomly switch off same number of transitions (Run1-10);
The estimated log marginal likelihood using steeping stone sampling (SS), lower (absolute) values of marginal likelihood indicate a better fit to the data.
The model with the best performance is indicated in bold; the comparisons are shown in the matrix being composed by column 4 to 13. In each row of the matrix, the positive value in a cell represented the support of one model (in column 1) over the other (the title of the matrix).

5.3 Results

5.3.1 Phylogenetics of the HPAI H7N3 Mexico with north America AIV

To investigate the origin of the AIV causing the HPAI H7N3 outbreak in Mexico in 2012, an initial phylogenetic analysis using Maximum likelihood was performed for each segment of both the outbreak sequences and a background dataset which comprised all available AIV of North American AIV lineages (Figure 5.1). The phylogenetic trees of all available H7 segments in North America showed that AIV isolated in recent years have diverged from those before 1990 (Figure 5.1A). In addition, in the HA segment a sub-lineage mainly composed of H7N2 AIV from domestic birds in New York state is clearly separate from the recent lineage composed of AIV from wild birds, which indicates extensive diversity of LP AIV in wild and domestic birds. Since 2000, the N3 NA segment of North American AIV has split into two separate lineages (Figure 5.1B). The mechanism for maintenance of this divergence remains unknown as viruses from both lineages co-circulate in geographically overlapping host populations, mainly wild waterfowl.

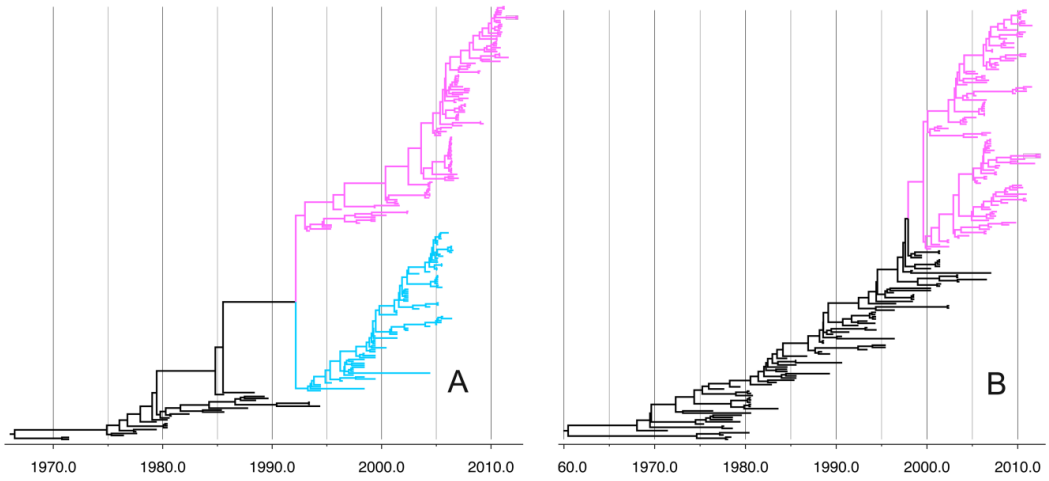


Figure 5.1 Phylogenies of the H7 and N3 segments of North American AIV.
A: HA. Sequences in grey are from before 1990; the clade coloured blue is composed of H7N2 AIV isolated from a single surveillance in poultry in New York; the clade coloured pink was selected for time-scaled phylogenetic analysis. B: NA tree. The uncoloured sequences are from before 2000; the AIV clade in pink was selected for time-scaled phylogenetic analysis.

Diverse reassortment events involving the six internal segments can be inferred from the maximum likelihood phylogenies of 2343 North American AIV. Clades identified in the phylogeny for one segment (e.g., PB2) are not maintained in the phylogenies of other internal segments (Figure 5.2). In addition, internal segments of AIV viruses isolated in distant locations can be closely related to each other within the same time period, which suggests not just frequent reassortment but also rapid movement of influenza viruses across North America.

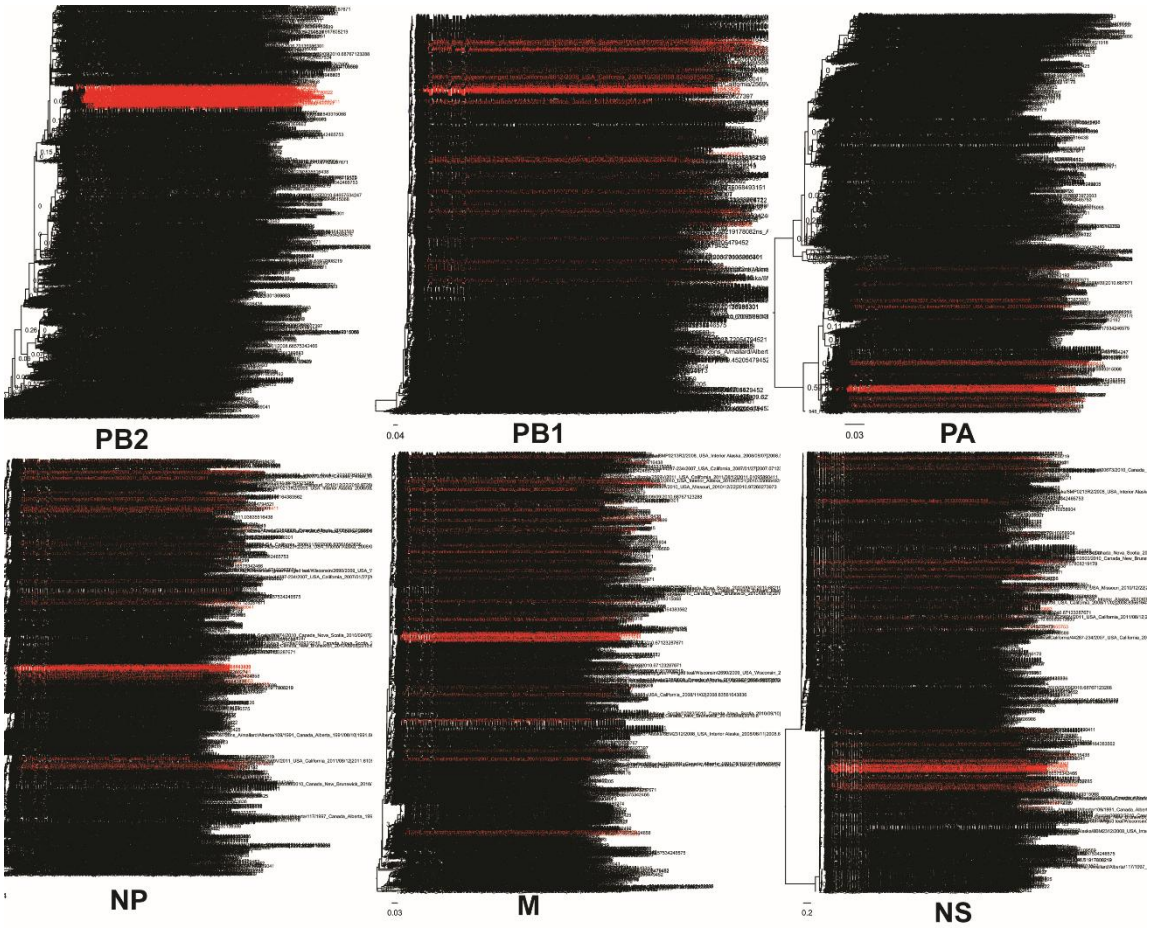


Figure 5.2 ML phylogenies of the six internal segments of North American AIV.
Phylogenetic trees are generated with the same 2343 strains in six internal segments. In PB2 segment, 86 AIV sequences in the same clade with H7N3 Mexico strains are labelled in red, these strains are diversely scattered in the phylogenetic trees of other internal segments.

Sequences for time-scaled phylogenetic analysis were selected from the closest clades to the novel H7N3 HPAI viruses on the maximum likelihood tree of each segment. This dataset comprised 427 AIV strains collected over a 12 year period (2001 to 2012). The time to the most recent common ancestor (TMRCA) for

each segment of the novel HPAI H7N3 in Mexico was estimated from the time-scaled phylogenetic trees. The HPAI H7N3 strains sampled in Mexico shared similar common ancestors among different genes between October 2011 and March 2012, i.e. during the winter of 2011-2012 (Table 5.9). The common ancestor of the HPAI H7N3 Mexico outbreak and the closest related avian influenza strains existed between 1.1 to 3.9 years ago, which varied among their different genomic segments (Table 5.9). The difference in unsampled diversity among gene segments suggested that the reassortment of North American AIV lineages which led to the H7N3 Mexico outbreak may have involved several events spread over this time period. This can be seen by comparing the closest related strains in the phylogenetic tree for any segment: they can be quite distant from the H7N3 Mexico strain in the other segments. This result supports the hypothesis of the occurrence of multiple reassortment events.

Co-circulation of multiple H7 clades was observed in HA across North America. Interestingly, the HPAI H7N3 Mexico strains are not related in HA to the HP H7N3 outbreak in British Columbia in 2004 and 2007, but instead are closely related to a subgroup of H7 AIV (H7N3, H7N8 and H7N9) from wild waterfowl isolated from Nebraska, Illinois, Missouri and Mississippi in 2010 and 2011. The mean estimate of the date of the common ancestor is February 2010 (Figure 5.3). On the other hand, the picture in NA is different: the closest related strain to that of the Mexico outbreak is a subtype H2N3 AIV isolated from a green winged teal in Illinois in 2010 (Figure 5.S1).

Gene	TMRCA[a]	Closely related strain[b]
HA	20/3/2012 (29/11/2011, 23/5/2012)	A/north shoveler/Missouri /2010 (H7N3)
NA	4/1/2012 (10/8/2011, 23/4/2012)	A/American green-winged teal/Illinois/2010 (H2N3)
PB2	30/10/2011 (8/6/2011, 25/2/2012)	A/mallard/California/198/2012 (H11N9)
PB1	10/10/2011 (10/10/2010, 28/1/2012)	A/American green-winged teal/California/123/2012 (H1N1)
PA	18/11/2011 (3/7/2011, 26/3/2012)	A/American green-winged teal/California/123/2012 (H1N1)
NP	21/1/2012 (16/8/2011, 5/5/2012)	A/American green-winged teal/Mississippi/2012 (H11N9)
M	2/11/2011 (7/0/2011, 2/5/2012)	A/American green-winged teal/Illinois/2008 (H10N7)
NS	25/10/2011 (25/3/2011, 17/1/2012)	A/American green-winged teal/Illinois/2010 (H10N7)

Table 5.9. Time of the most recent common ancestors for the Mexico H7N3 virus

[a] Time of the most recent common ancestors (TMRCA) of each segment of the novel Mexico H7N3 virus is represented in the order of date/month/year. The values in parentheses represent the 95% HPD intervals.

[b] The strains are identified are those most closely related to the outbreak strains in each tree phylogeny in this study.

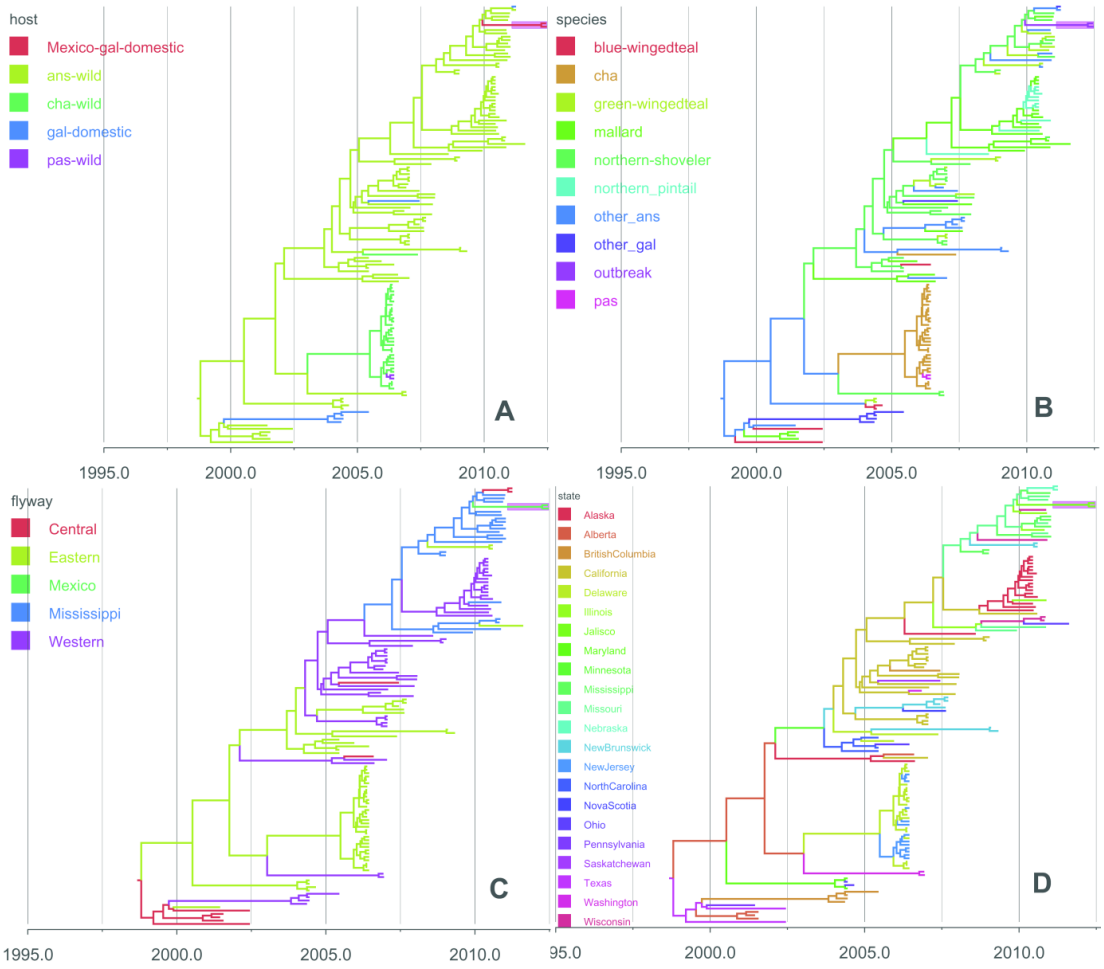


Figure 5.3 MCC phylogenies for the HA segment.
Branches are coloured according to the 4 discrete traits (host order, host species, flyway and location) on internal nodes. Mexican outbreak strains are highlighted with pink. A: Host order; B: Host species; C: Flyway; D: State.
The trees for the other 7 segments without taxa names can be found in Figure A5.1 to A5.7.

In contrast, the three polymerase encoding gene segments PB2, PB1 and PA of the Mexico outbreak strain belong to lineages composed mainly of AIV found in wild waterfowl in California from the beginning of 2012 (Figure A5.2-A5.4), with segments PB1 and PA having the same most closely related strain: A/American green-winged teal/California/123/2012 (H1N1). The other internal segments of the Mexico H7N3 strains have a different origin. From the Bayesian phylogenetic tree of the NP segment, the closest AIV strain to H7N3 Mexico is an H11N9 strain isolated from Mississippi in 2012 (Table 5.9). The NP segment of the H7N3 Mexico outbreak strain is also in the same lineage as a small number of H7N7 AIV strains carried by northern teal in Illinois/Missouri in the fall of 2010; these strains belong to the same

lineage in the HA segment as well (see above and Figures 3 and A5.5). The NS segment was derived from an H10N7 AIV which was also circulating in the same region at that time (Figure A5.7). In the M segment, however, Mexico H7N3 strains are distinct from all currently available AIV in North America, suggesting a surveillance gap (Figure A5.6).

These results indicate that the HPAI H7N3 virus that caused the outbreaks in Mexico is not related to any of the previous H7N3 HPAI outbreaks in North America, nor related to other AI outbreaks (HP H5N2 outbreaks in Mexico in 1994-1995, LP H7 outbreaks in Canada in 2009) in domestic birds in recent years (Horimoto *et al.*, 1995b; Pasick *et al.*, 2009). In addition, no clear pattern of association among the segments of the Mexico H7N3 strains was observed, indicating multiple segment exchange events occurred among North American influenza strains to give rise to it.

5.3.2 Gene flow of the precursor of the HPAI H7N3 outbreak in Mexico

To further explore the origin of the Mexico outbreak strain, a joint analysis of discrete trait models was performed to estimate the overall genetic transmission process. In this the phylogenetic tree space was sampled independently for each segment, while the transition pattern was jointly estimated in a single analysis as the diffusion parameters being applied in the discrete trait models were assumed to be the same (see methods). Four major factors including seven specific traits were tested by implementing Bayesian stochastic search variable selection (BSSVS): i) host population of AIV (order/species); ii) geographic location of sampled AIV (bird migration flyways/ provinces and states of North America); iii) subtype of AIV and iv) virulence (pathogenicity/ cleavage sites). For each trait, the evolving process of the HPAI H7N3 in Mexico and closely related AIV can be seen from the reconstructed time-scaled phylogeny of each segment independently (with exception of pathogenicity and cleavage sites which only applied to the HA segment), with the

branches coloured by the specific trait according to the ancestor trait in the internal nodes (Figure 5.3 and Figure A5.1 to A5.7).

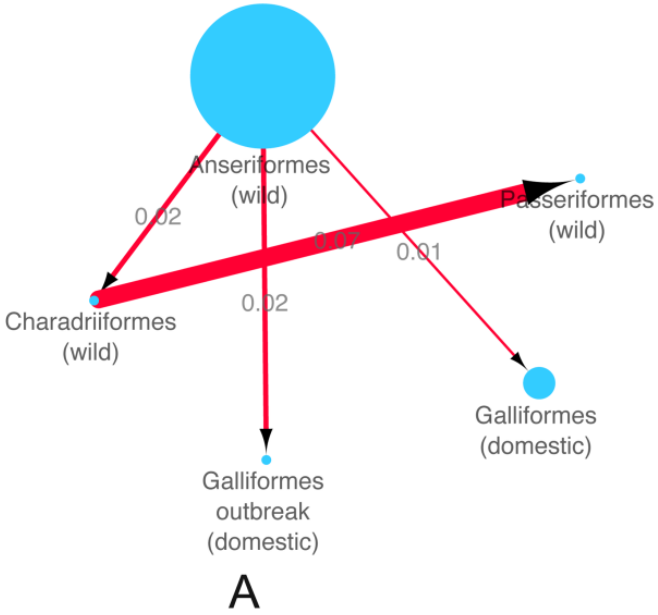
In Figure 5.3, five host orders are labelled on HA tree: wild birds of the order Anseriformes (ans-wild); wild birds of the order Charadriiformes (cha-wild); wild birds of the order Passeriformes (pas-wild); domestic birds of the order Galliformes and Mexico H7N3 outbreak in the order Galliformes (gal-domestic-Mexico); For host species, Wild Anseriformes are classified into the five main species and a group comprising the other rarer species of Anseriformes in this study: mallard (Anas platyrhynchos), northern pintail (Anasacuta), northern shoveller, blue-winged teal, green-winged teal and other Anseriformes (other ans); The order Galliformes are shown as "outbreak" (the H7N3 Mexico outbreak) and "other_gal"; The other orders are shown as: Charadriiformes (cha) and Galliformes (gal), Gruiformes (gru) and Passeriformes (pas); In addition, four specific North American flyways are labelled on the HA tree: the Atlantic, Mississippi, Central, and Pacific; And finally, 22 states and provinces of the viral sample locations are labelled on the HA tree. The trees for the other 7 segments without taxa names can be found in Figure A5.1 to A5.7.

Host populations of AIV in this study were first analysed by Order: wild Charadriiformes (cha) such as gulls, wild Gruiformes (gru) such as cranes, wild Passeriformes (pas) such as sparrows, domestic Galliformes (gal) such chickens and wild Anseriformes (ans) such as mallards, with Anseriformes comprising the majority of the AIV data set (n= 366/427), see Table 5.1. Among all four strongly supported transitions with Bayes Factor (BF) >100 with mean diffusion rate (R) between 0.01 and 0.07, the highest diffusion rate was found between Charadriiformes and Passeriformes. The other three were found between Anseriformes and other bird orders, and the HPAI H7N3 outbreak in poultry (labelled as Galliformes Mexico) is linked to Anseriformes with strong support (R=0.02, BF>100) (Figure 5.4A and Table 5.10).

Transition		Mean rate	Indicator	BF
Charadriiformes (wild)	Passeriformes (wild)	0.07	1	>100
Anseriformes	Charaiformes (wild)	0.02	1	>100
Anseriformes (wild)	Galliformes outbreak	0.02	0.99	>100
Anseriformes (wild)	Galliformes (domestic)	0.01	0.99	>100

Table 5.10. Transmission rates of host orders and the Bayes Factor support
States=6 (host)
Indicator cutoff (for BF = 3.0) = 0.65

In Figure 5.4A, node labels in the nodes are host orders identified following the abbreviations used in the coloured phylogenetic trees (Figure 5.3 and Figure A5.1 to A5.7): wild birds of the order Anseriformes (ans-wild); wild birds of the order Charadriiformes (cha-wild); wild birds of the order Passeriformes (pas-wild); domestic birds of the order Galliformes and Mexico H7N3 outbreak in the order Galliformes (gal-domestic-Mexico). Arrows show the direction of transmission between two host orders; the arrow weight and the number above each arrow indicates the transmission rate. Node size reflects the number of AIV for each host order (Table 5.1). Line colours indicate the overall Bayes Factor test support for epidemiological linkage between host orders, Red lines indicate statistical support with BF>100 (very strong support), dark pink lines indicate support with 30<BF<100(strong support), pink lines indicate support with 3<BF<30. The results confirm that there has been mixing of influenza A virus between different orders of birds, both wild and domestic.



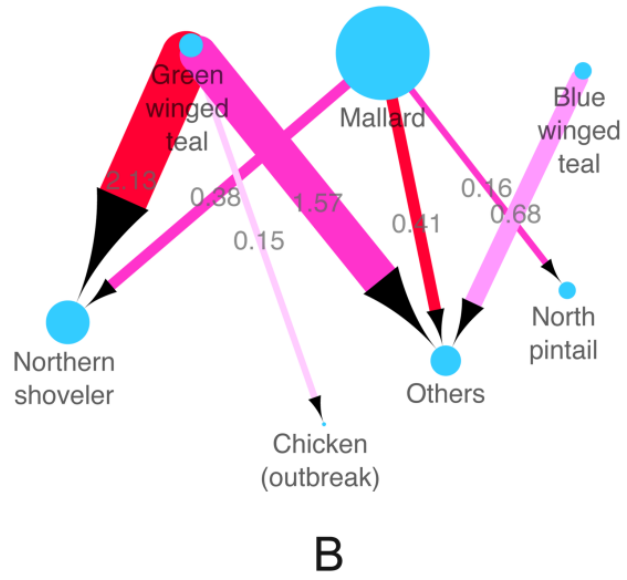


Figure 5.4 Inferred host transmission networks of Mexican outbreak AIV.
A: Host order. Node labels in the nodes are host orders identified following the abbreviations used in the coloured phylogenetic trees (Figure 5.3 and Figure A5.1 to A5.7). B: Host species (Anseriformes only). Wild Anseriformes are further classified into the five main species and a group comprising the other rarer species of Anseriformes in this study: mallard (Anas platyrhynchos), northern pintail (Anasacuta), northern shoveller, blue-winged teal, green-winged teal and other Anseriformes (other ans).

To explore which host species might have been the direct donor of the Mexico outbreak strains, AIV belonging to Anseriformes were further divided into the five predominant species: mallard (Anas platyrhynchos), northern pintail (Anasacuta), northern shoveller (Anas clypeata), blue-winged teal (Anas discors) and green-winged teal (Anas carolinensis) (Table 5.2). Species that were sampled at relatively low levels were combined as "other Anseriformes" (Table 5.2). Multiple statistically supported transitions (with R from 0.15 to 2.13, BF from 6 to over 100) were identified among different host species within this Order, and both mallard and green winged teal are linked to 3 other host species (Figure 5.3B and Table 5.11). Similar to Figure 5.4A, arrows in Figure 5.4B show the direction of transmission between two host species; the arrow weight and the number above each arrow indicates the per capita transmission rate. Node size reflects the number of AIV for each host species. Line colours indicate the overall Bayes Factor test support for epidemiological linkage between host species, Red lines indicate statistical support with BF>100 (very strong support), dark pink lines indicate support with 30<BF<100 (strong support), pink lines indicate support with 3<BF<30. This analysis indicates the Mexican outbreak strains were most likely to have been transmitted from green winged teals (R=0.15 and BF =6).

Transition		Mean rate	Indicator	BF
other_ans	blue-wingedteal	0.68	0.85	19
northern-shoveler	green-wingedteal	2.13	1	>100
other_ans	green-wingedteal	1.57	0.99	>100
outbreak	green-wingedteal	0.15	0.67	6
northern-shoveler	Mallard	0.38	0.99	>100
northern_pintail	Mallard	0.16	0.99	>100
other_ans	Mallard	0.41	1	>100

Table 5.11. Transmission rates of host species and the Bayes Factor support
State=9 (host species)
Indicator cutoff (for BF = 3.0) = 0.49

The phylogeographic analysis for each segment of the Mexico HPAI H7N3 strain was summarised by a MCC tree in a geographic context. However, to visualize the evolution process in a spatiotemporal mode I converted the spatial annotated time-scaled phylogeny to an annotated map (Figure 5.5 and Figure A5.8). In Figure 5.5, the plotted lines represent the branches of the MCC trees for different segments,

distinguished by colour; the size of each circle represents the number of lineages with that location state. The map source for this figure was OpenStreetMap (http://www.openstreetmap.org/). The spatial diffusions of other five segments (PB1, PA NP, M and NS) on the map are shown in Figure A5.8.

Five segments (HA, NA, NP, M, and NS) of the Mexico HPAI H7N3 strain were introduced directly from different states in central US, while PB2, PB1 and PA were introduced from states in the western region. The introductions of segments from several different geographic locations indicate multiple reassortment events were likely to have been involved in the generation of the novel H7N3 Mexico AIV.

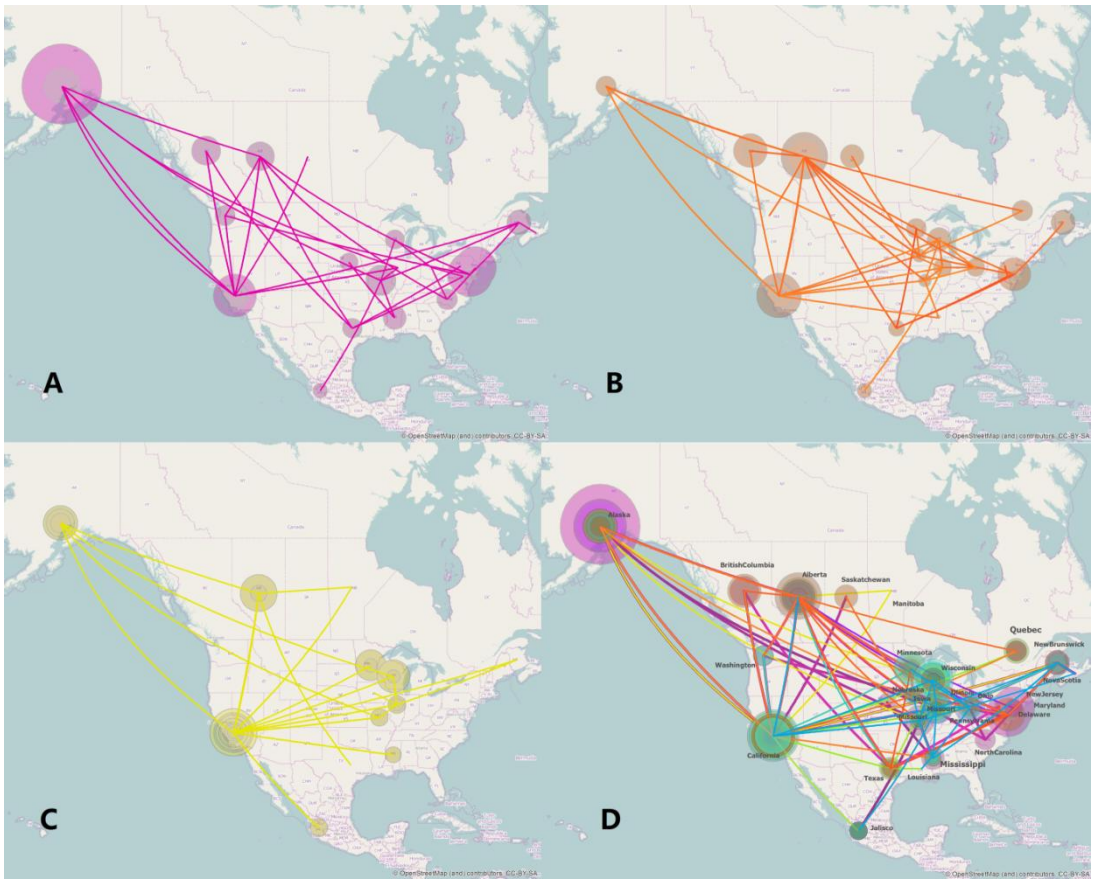


Figure 5.5 Spatial diffusion of AIV segments of the Mexico outbreak AIV.
The first three panels represent three segments separately (A: HA, B: NA, C: PB2) and D represents the spatial transmission of all 8 segments jointly

Joint discrete trait analysis of all eight segments indicated frequent gene transfer among locations (states and provinces) where the background AIV sequences were isolated. However, in this initial analysis, no significant support was found between Jalisco (the outbreak state) and any other location, probably due to the large number of possible transitions (26 states, 325 irreversible transition pairs) and the limited number of outbreak strains (3; see Table 5.3). Previous studies have shown that combining locations into larger areas greatly improves phylogeographic descriptions of the pattern of virus gene flow (Lam *et al.*, 2012; Wilke *et al.*, 2011). Therefore, I enhanced the statistical power of the analysis in two ways: first by decreasing the number of locations by combining states and provinces into major regions (flyways) and secondly by reducing the number of pairwise transitions possible (see Methods 5.2.4).

To aggregate locations I used the known migration routes, or "flyways": Atlantic, Mississippi, Central, or Pacific (Figure 5.6). The distributions of avian influenza virus for each flyway are summarized in Table 5.4, showing a wide range in rate and statistical support (R=0.02 to 0.25; BF=3 to >100). Highly significant links [BF>100, Indicator (I) =1] were found between major North American flyways, particularly between Atlantic and Mississippi (R=0.17 exchange/year); Mississippi and Pacific (R=0.22 exchange/year), Central and Atlantic (R=0.25 exchange/year) and Central and Pacific (R=0.18 exchange/year) (Figure 5.7A and Table 5.12). Linkages between Atlantic and Pacific flyways, Mississippi and Central flyways were also identified although with weaker support (3<BF<6). The transitions between flyways and the H7N3 HPAI in Mexico are not that strongly supported compared to the between flyway transitions, probably due to the limited number of sequences available from the outbreak AIV, but three direct donors among these flyways to the predecessor of the Mexico outbreak were identified: the Pacific, Central and Mississippi flyways. Among these flows, the transition rate from Mississippi (R=0.05 exchange/year) is the most strongly supported with BF=86. In comparison, the link with the Pacific (R=0.02 exchange/year) and Central (R=0.04 exchange/year) were weaker (BF = 4; Figure 5.7A and Table 5.12). There was no supported link between the Atlantic flyway and Mexico. The results indicated that

the HPAI H7N3 in Mexico probably originated from AIV transmitted by wild birds from three different flyways.

Figure 5.6 North America states and provinces categorized according to flyways. Provinces and states which belong to Pacific flyways are coloured in purple; Central flyways are coloured in red; Mississippi flyways are coloured in blue; Atlantic flyways are coloured in yellow; and Mexico is coloured in green. Those colours on the map are the same as the coloured branches in the flyway discrete phylogenetic tree in Figure 5.3 C.

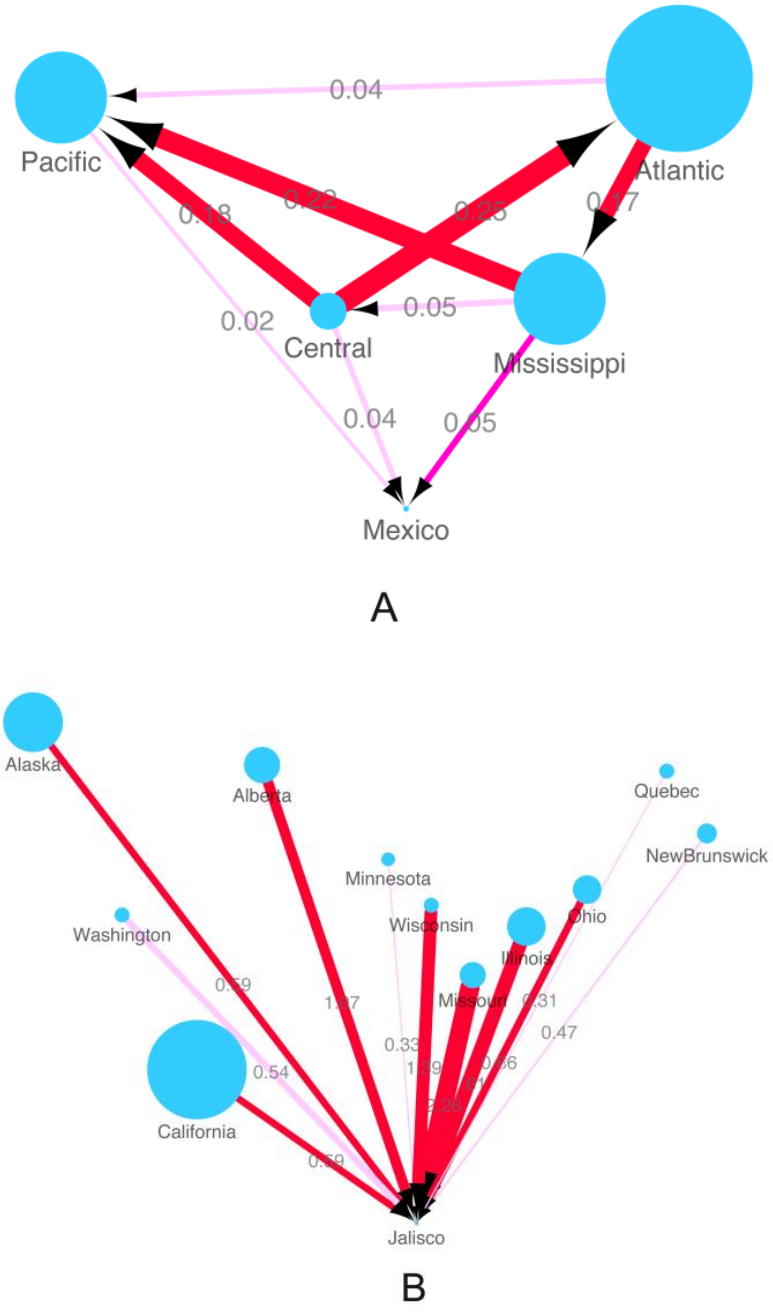


Figure 5.7 Inferred phylogeographic transmission networks of Mexican outbreak AIV. A: Flyway. AIV transmission between 4 flyways and links to the Mexican outbreak strains is shown. Arrows show the direction of transmission between two flyways; Node size reflects the number of AIV for each flyway. B: Location. AIV transmission between states/provinces in North America and Jalisco (the Mexican state where the outbreak strains were isolated) is shown.

Transition		Mean rate	Indicator	BF
Atlantic	Central	0.25	1	>100
Pacific	Central	0.18	1	>100
Mississippi	Atlantic	0.17	1	>100
Mexico	Mississippi	0.05	0.98	86
Pacific	Mississippi	0.22	1	>100
Mexico	Pacific	0.02	0.73	4
Mexico	Central	0.04	0.73	4
Central	Mississippi	0.05	0.73	4
Pacific	Atlantic	0.04	0.72	3

Table 5.12. Transmission rates of flyways and the BF support
States=5 (flyway)
Indicator cutoff (for BF = 3.0) = 0.72

Given these findings it appeared likely that the precursor strains were generated somewhere near Mexico as it is the place where birds from the different migration routes meet during winter. To test this hypothesis, I further reduced the number of transitions in the locations transition matrix: transitions between two flyway regions were switched off (by forcing the initial indicator of a given transition pair from 1 to 0, so that this transition pair will not be counted), and those for within flyway transitions and transitions linked to Mexico were maintained. There are 98 non-reversible transition pairs in the new reduced matrix (Table A5.2). AICM tests (see Methods) revealed that the non-reversible BSSVS model with reduced number of transitions was significantly favoured over the other models with the original matrix (Table 5.1B), indicating the number of transitions has an effect on the performance of discrete trait models. This reduced model has better support than a randomly reduced model with the same number of transition pairs and the same non-reversible BSSVS setting (Table 5.2B); I conclude that gene transitions within flyways and Mexico alone better explain the gene flow of North America AIV than a model incorporating the between flyway transitions.

Considering individual locations, I found 11 locations were linked to the HPAI H7N3 in Mexico, among which 7 showed significantly strong links (BF>100). These were: Alaska (R=0.59), Alberta (R=1.07), California (R=0.59), Illinois (R=1.61), Missouri (R=2.26), Ohio (R=0.66) and Wisconsin (R=1.39), belonging to the Pacific,

Central and Mississippi flyways. AIV from other four states/provinces (Minnesota, New Brunswick, Quebec and Washington) are also linked to Mexico H7N3 AIV but with weaker support (BF= 21 to 30) and lower rate (0.7 to 0.84) (Figure 5.7B and Table 5.13). This result indicates the donor locations of the Mexico outbreak are spread widely across North America.

Transition		Mean rate	Indicator	BF
Jalisco	Alaska	0.59	0.99	>100
Jalisco	Alberta	1.07	1	>100
Jalisco	California	0.59	1	>100
Jalisco	Illinois	1.61	1	>100
Jalisco	Minnesota	0.33	0.7	21
Jalisco	Missouri	2.26	1	>100
Jalisco	New Brunswick	0.47	0.82	26
Jalisco	Ohio	0.66	1	>100
Jalisco	Quebec	0.31	0.84	30
Jalisco	Washington	0.54	0.78	21
Jalisco	Wisconsin	1.39	1	>100

Table 5.13. Transmission rates of location (state/province) and the BF support
States=26 (locations)
Indicator cutoff (for BF = 3.0) = 0.34

In addition, an extremely complex pattern of linkage between the 52 ancestral subtypes (Table A5.1) was identified, which confirmed the extent of the reassortment events which had occurred between, especially in the internal segments. However, as for locations, no significant linkage between H7N3 in Mexico and other subtypes was identified due to the large number of candidate donor subtypes.

Mexico H7N3 is confirmed in the phylogenetic trees of the HA segment (Figure 5.8) to have mutated from a LPAI (low pathogenic avian influenza) virus to HPAI after the ancestral virus was introduced into poultry from wild birds (Figure 5.8A), rather than being associated with previous HP outbreaks. The virulence of HP avian influenza viruses is associated with the appearance of an insertion of multiple basic amino acids at the cleavage site of the HA protein (Steinhauer, 1999). Categorizing the cleavage sites in this dataset into three types: 1) Insertion: which are multibasic insertions (insertion), only found in HP AIV, including HP H7N3 outbreaks strains in Mexico 2012 and in Canada 2004 and 2007; 2) Partial insertion: LP AIV which has a partial insertion of 2 amino acids; 3) No insertion: other LP AIV with no insertion in the cleavage site (Figure 5.8C). I find that the H7N3 Mexico strain has a unique insertion - DRKSRHRR - compared to other HPAI strains (Figure 5.8 C). I conclude the Mexico H7N3 strains originated from a lineage composed of LP strains with partial insertions in the cleavage sites. Similarly, the H7N3 strains which caused an outbreak of HPAI in Canada in 2004 were also derived from LP strains with the same partial insertions, but from a completely separate HA clade, indicating parallel evolution with respect to the acquisition of the multibasic cleavage site, starting from different lineages (Figure 5.8 B).

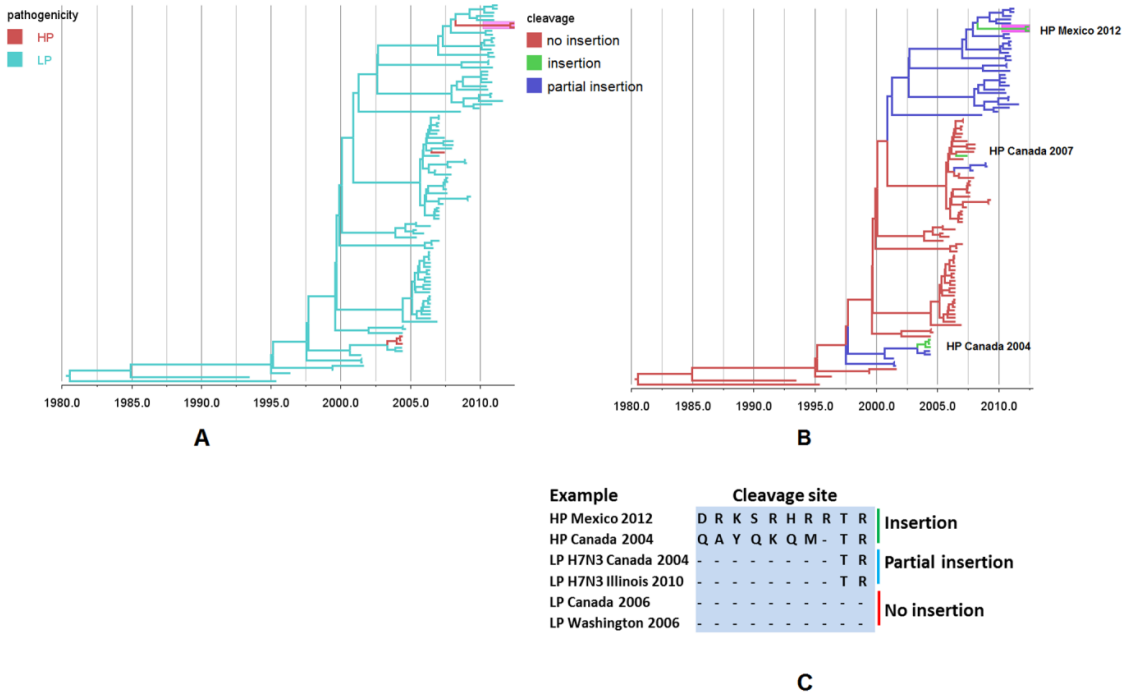


Figure 5.8 Virulence evolving on the MCC phylogeny of the HA segment.
(A): Pathogenicity changes (according to high pathogenicity (HP) and low pathogenicity (LP)) are indicated by colour changes at tree; Cleavage site changes (with insertion, partial insertions and no insertions (C) are indicated by colour changes at tree nodes.

5.4 Discussion

I have investigated the origin of the recent highly pathogenic H7N3 outbreaks in Mexico. The analysis showed that the progenitor of the HPAI H7N3 was a reassortant virus with several different origins among the eight segments. I also found that gene segments of AIV in North American wild birds are exchanged at a very high frequency, with no evidence of any restriction which might imply linkage of segments.

This study confirmed the assumptions of earlier studies based on the HA segment that the outbreak strain derived from wild birds (Kapczynski *et al.*, 2013; Maurer-Stroh *et al.*, 2013; Sherrilyn Wainwright, 2012). I have now shown using powerful Bayesian phylogenetic methods that the origin of the HA segments of the

Mexico H7N3 strains can be dated to March 2012, and that they fall into a subgroup of H7 AIV (H7N3, H7N8 and H7N9) from wild waterfowl isolated from Nebraska, Illinois, Missouri and Mississippi with a common ancestor around February 2010. I have extended this analysis to all eight segments, thus obtaining the complete evolutionary history of the outbreak avian influenza viruses.

The predecessors of HPAI H7N3 in Mexico were transmitted from migrating waterfowl in North America. Previous cases of periodic transmission of H7N3 viruses from wild birds to gallinaceous poultry in the Americas suggests that these viruses continuously circulate in wild birds, and their propensity to become highly pathogenic after transmission suggests that they have a gene constellation conducive to generating pathogenic variants (Krauss & Webster, 2012). H7N3 has been responsible for all lethal influenza outbreaks in poultry in the Americas over the past decade (Krauss & Webster, 2012). Experimental studies have also indicated that H7 influenza viruses from the North American lineage have acquired sialic acid-binding properties that more closely resemble those of human influenza viruses and have the potential to spread to naive animals (Belser *et al.*, 2008). In parallel, H7 influenza viruses from East Asian migratory waterfowl were introduced into domestic ducks in China on several occasions during the past decade and subsequently reassorted with enzootic H9N2 viruses to generate a novel H7N9 influenza A virus, resulting in 44 human deaths in China (WHO reported in Dec.3rd 2013) since its first detection in March 2013 (Lam *et al.*, 2013). These results indicated that AIV of H7 subtypes carried by wild birds are potential threat to mammalian hosts.

Earlier studies showed that shorebirds and gulls in the Americas are more frequently the source of the potential precursors to HP H5 and H7 avian influenza viruses, while in Eurasia, the precursors of HP influenza viruses are usually from duck species (Krauss *et al.*, 2007; Liu *et al.*, 2013). However, I found that wild Anseriformes (ducks and geese) were the origin of the precursor of HPAI H7N3 in Mexico. Anseriformes showed substantial diversity of AIV in North America, and were divided into five avian species groups in this dataset – mallards, northern pintails, northern shovelers, blue-winged teals, green-winged teals, which have

specific ranges for breeding, migration and wintering (Lam *et al.*, 2012). I found the green winged teal was the species most strongly supported as the direct donor of the predecessor HPAI H7N3 in Mexico. This species nests as far north as Alaska, and migrates along all four flyways. However, the genetic transitions of different segments showed a complicated interaction involving different bird species.

Migratory birds are major candidates for long-distance dispersal of zoonotic pathogens. Studies have been made to evaluate the dispersive potential of HPAI H5N1 viruses by wildfowl, by monitoring the movement range and movement rate of birds in relation to the apparent asymptomatic infection duration by satellite telemetry in 2006–2009, over HPAI H5N1 affected regions of Asia, Europe and Africa. Their results indicate that individual migratory wildfowl have the potential to disperse HPAI H5N1 over extensive distances (Gaidet *et al.*, 2011; Gaidet *et al.*, 2010). In addition, a monitoring surveillance in wild birds and poultry along the same migration pathway on the Qinghai-Tibet Plateau found a temporal lag between peak seasons for poultry and wild bird outbreaks. Poultry outbreaks were found during the winter and the spring and subsequently during the spring and the breeding seasons for wild birds. In contrast, fewer outbreaks than expected occurred in the breeding and the fall for poultry, and the fall and the winter for wild birds (Prosser *et al.*, 2011b).

Migrating birds may exchange viruses with other populations at staging, stopover or wintering sites (Gunnarsson *et al.*, 2012). Many studies have been performed on AIV gene flows in North America during wild bird migration: One such revealed that avian influenza virus exhibits a strongly spatially structured population in North America, and the intra-continental spread of AIV by migratory birds is subject to major ecological barriers, including spatial distance and avian flyway (Lam *et al.*, 2012). Earlier studies suggested that AIV exhibits a strongly spatially structured population in North America, with relatively infrequent gene flow among localities and especially between those that are spatially distant or belong to different flyways using phylogeographic analysis (Munster *et al.*, 2007). This hypothesis was supported by studies showed that AIV isolates from mallard were linked by migration between sites in central Canada and Maryland but limited

reassortment occurred along the inter-migratory flyway routes (Dugan *et al.*, 2011). However, more recently, the opposite was seen in a another study, which emphasized that the long-term persistence of the influenza A virus gene pool in North American wild birds may be independent of migratory flyways, and the short-term evolutionary consequences of these ecological barriers may be rapidly erased by East-West virus migration (Bahl *et al.*, 2013).

I also found there are genetic interactions between flyways, using a similar discrete trait model. However, to find the strongest link between the Mexican outbreak and potential precursors I found the model had more power when I switched off the between flyway transitions, keeping only the links within flyways and with Mexico. I found that gene flow from three flyways (Pacific, Central and Mississippi) generated the reassortants which acted as the predecessor of HPAI H7N3 in Mexico, and it is possible that the reassortment events occurred in Mexico or farther afield. Flyway boundaries are not sharply defined and both in the northern breeding grounds and the southern wintering grounds there is overlap to some degree. For example, in Panama parts of all four flyways merge into one (http://www.birdnature.com/flyways.html). Birds that are long-distance migrants typically have ranges that extend from the United States and Canada in the summer to Mexico and further south in the winter and nearly all of the migratory birds of the eastern United States, as well as many western species, use the western Mexican Gulf during migration (Smith, 2005).

While the resolution and detection of migration events has been enhanced through increased surveillance in recent years, critical information for wild bird surveillance remains sparse. Only one AIV in Mexico has been published (A/cinnamon teal/Mexico/2817/2006) and it is not related to the new outbreak virus in any of the eight segments (data not shown). I found the origin of the M segment of the H7N3 outbreak gene was distinct from other North American AIV in the dataset, as seen in the phylogenetic trees. In addition, the relatively greater length of branches preceding the outbreak group in other gene segments (Figures 2 and S2-S8) suggests there may be missing intermediates, possibly through insufficient AIV surveillance

in Central and South America. Together with previous phylogenetic studies which also mentioned the importance of filling gaps in viral sampling (with record of sampling time and location) in these regions (Gonzalez-Reiche & Perez, 2012; Smith *et al.*, 2009c), this highlights the need for increased surveillance in those regions.

Overall, by combining the phylogenetic history of AIV with the host distribution and ecology in this analysis, I show the origin of HPAI H7N3 AIV that caused a series of poultry outbreaks in Mexico is an novel reassortant carried by migration of wild waterfowls from different migration flyways in North America throughout the time period studied, and, more importantly, Central America might be a potential hotspot for AIV reassortment events. Our results are useful for identifying the threat of AIV in wild birds and indicate comprehensive surveillance in South and Central America is highly desirable.

Appendix

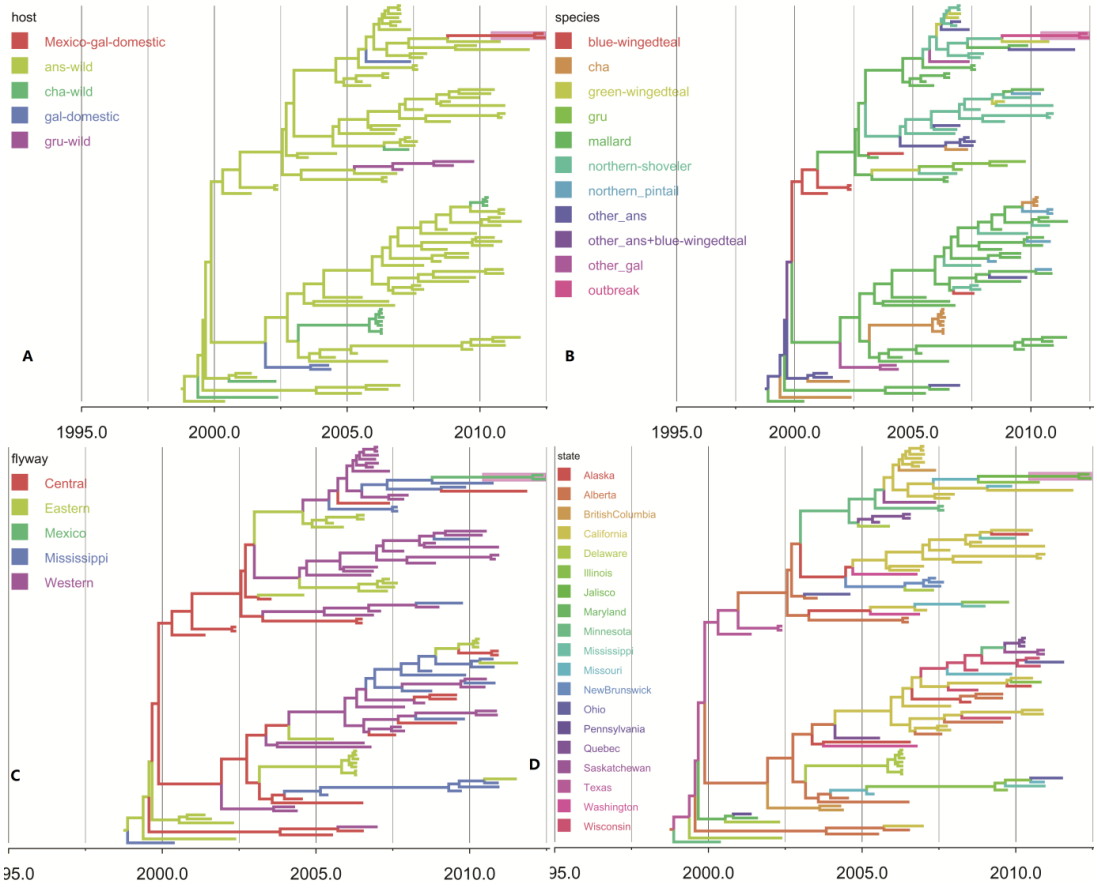


Figure A5.1 MCC phylogenies for the NA segment
Temporally structured maximum clade credibility (mcc) time-scaled phylogenetic tree showing the evolution of NA gene of avian influenza A virus isolated from North American wild birds for each individual gene dataset. Ancestral state (A: host order; B: host species; location; C: flyway; D: state) changes recovered from the discrete trait analyses are indicated by colour changes at tree nodes. Mexican outbreak strains are highlighted with pink.

A: Host order Five host orders are labelled on HA tree: wild birds of the order Anseriformes (ans-wild); wild birds of the order Charadriiformes (cha-wild); wild birds of the order Passeriformes (pas-wild); domestic birds of the order Galliformes and Mexico H7N3 outbreak in the order Galliformes (gal-domestic-Mexico) .

B: Host species Wild Anseriformes are classified into the five main species and a group comprising the other rarer species of Anseriformes in this study: mallard (Anas platyrhynchos), northern pintail (Anasacuta), northern shoveller, blue-winged teal, green-winged teal and other Anseriformes (other ans); The order Galliformes are shown as "outbreak" (the H7N3 Mexico outbreak) and "other_gal"; The other orders are shown as: Charadriiformes (cha) and Galliformes (gal), Gruiformes (gru) and Passeriformes (pas).

C: Flyway Four specific North American flyways are labelled on HA tree: the Atlantic (Eastern), Mississippi, Central, and Pacific (Western).

D: State 22 states and provinces of the viral sample locations are labelled on HA tree.

The same colour code also applies to figures A5.2-5.7.

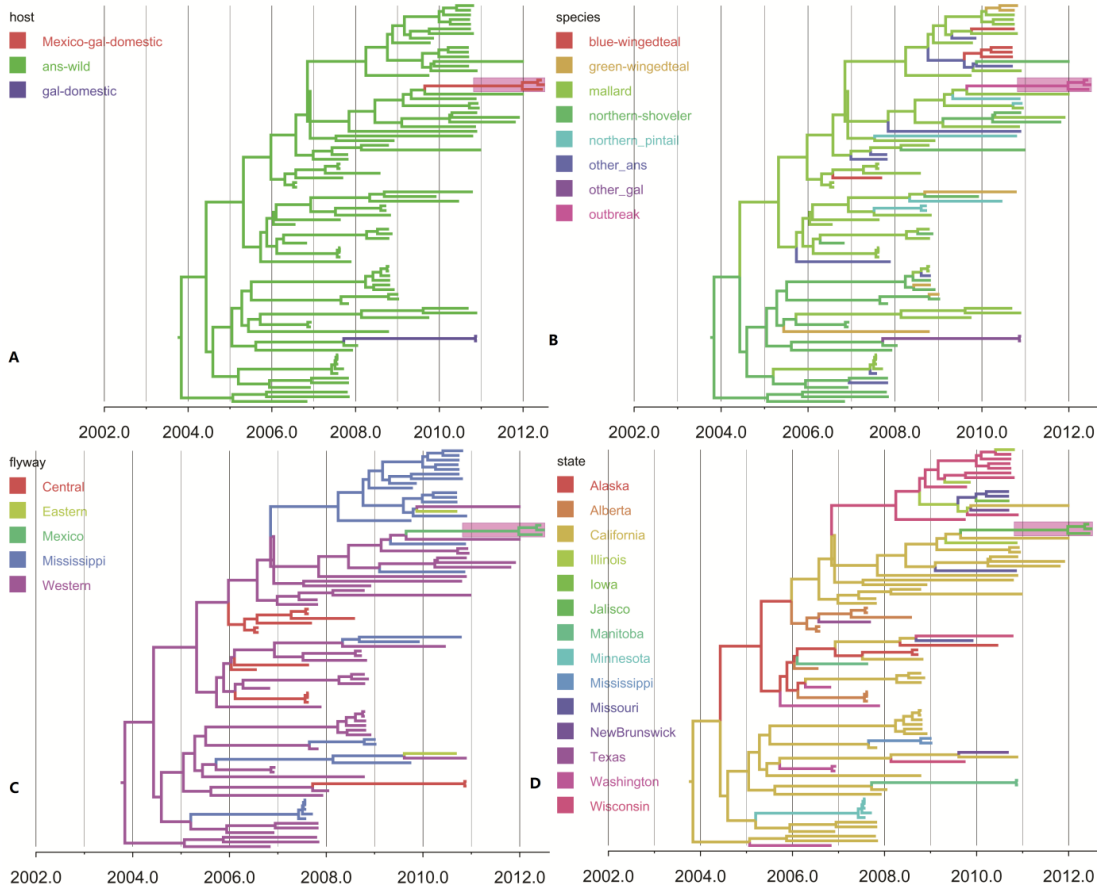


Figure A5.2 MCC phylogenies for the PB2 segment

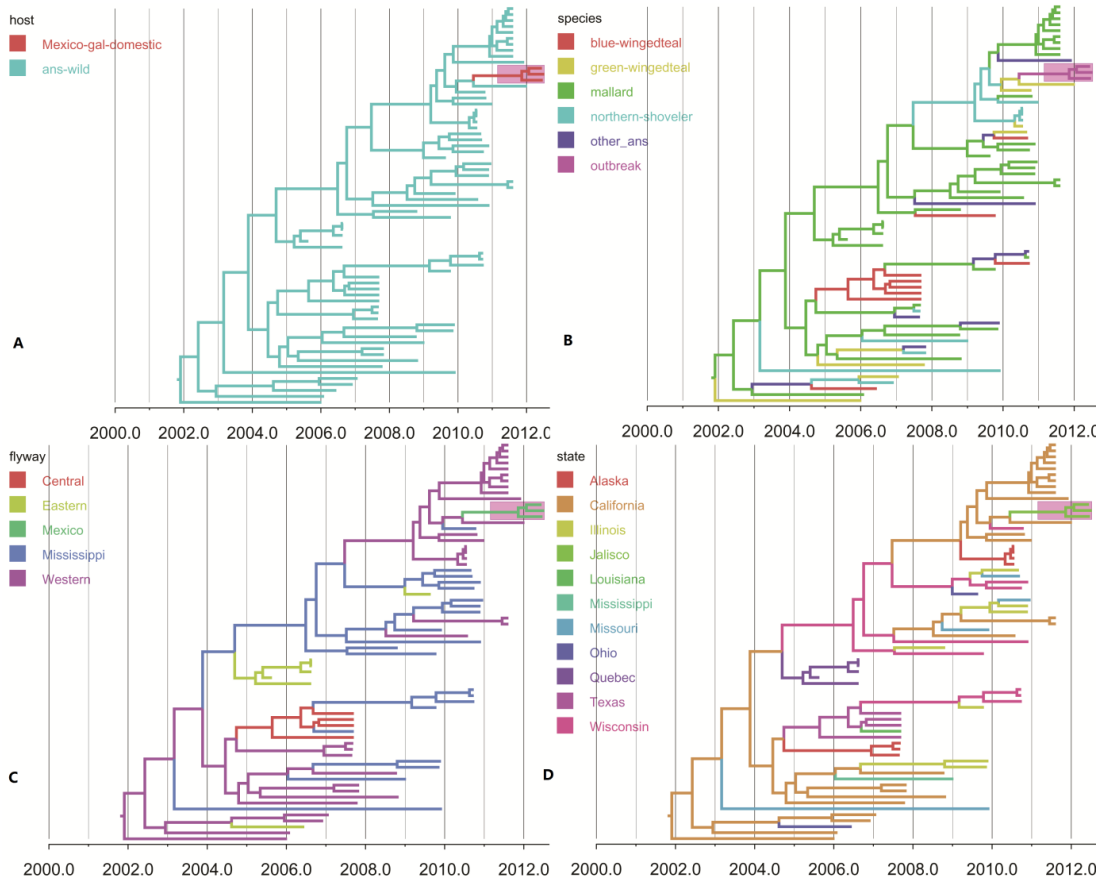


Figure A5.3 MCC phylogenies for the PB1 segment

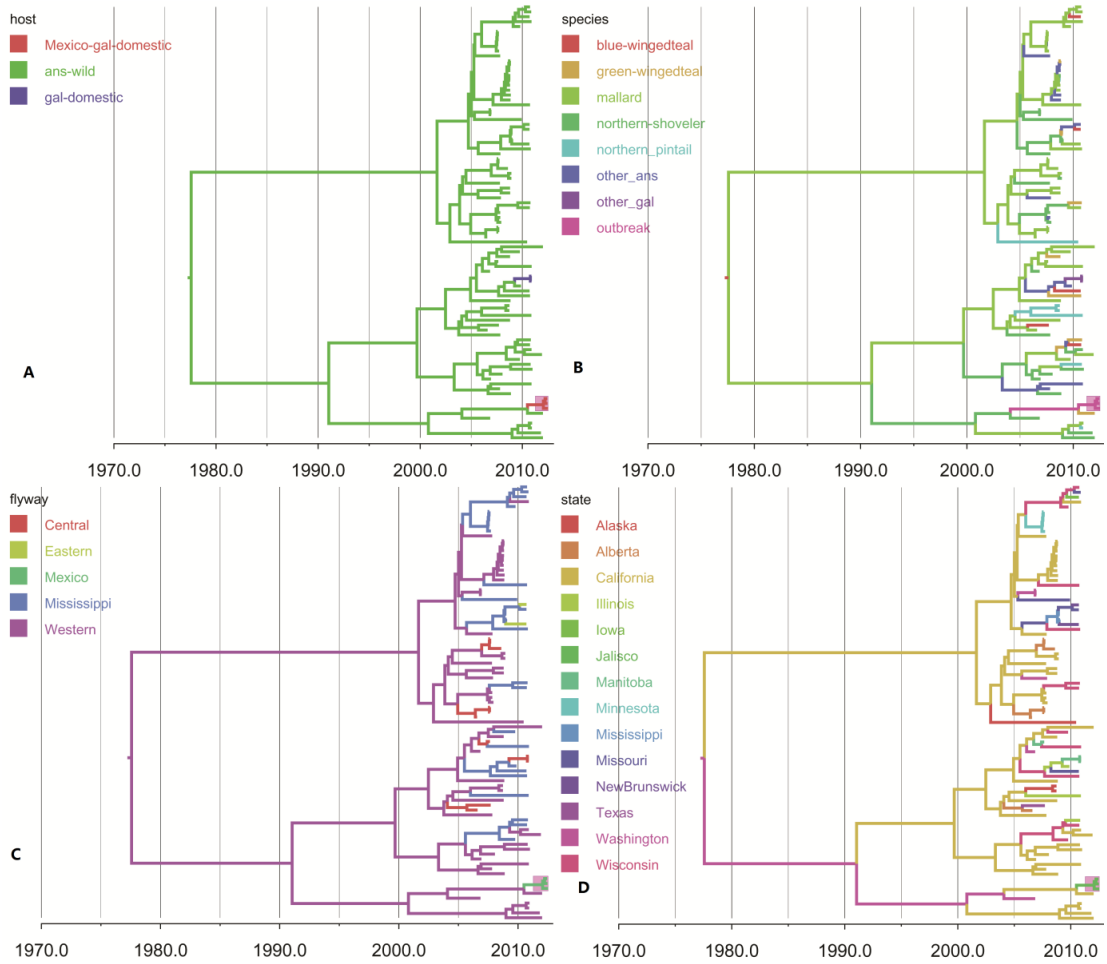


Figure A5.4 MCC phylogenies for the PA segment

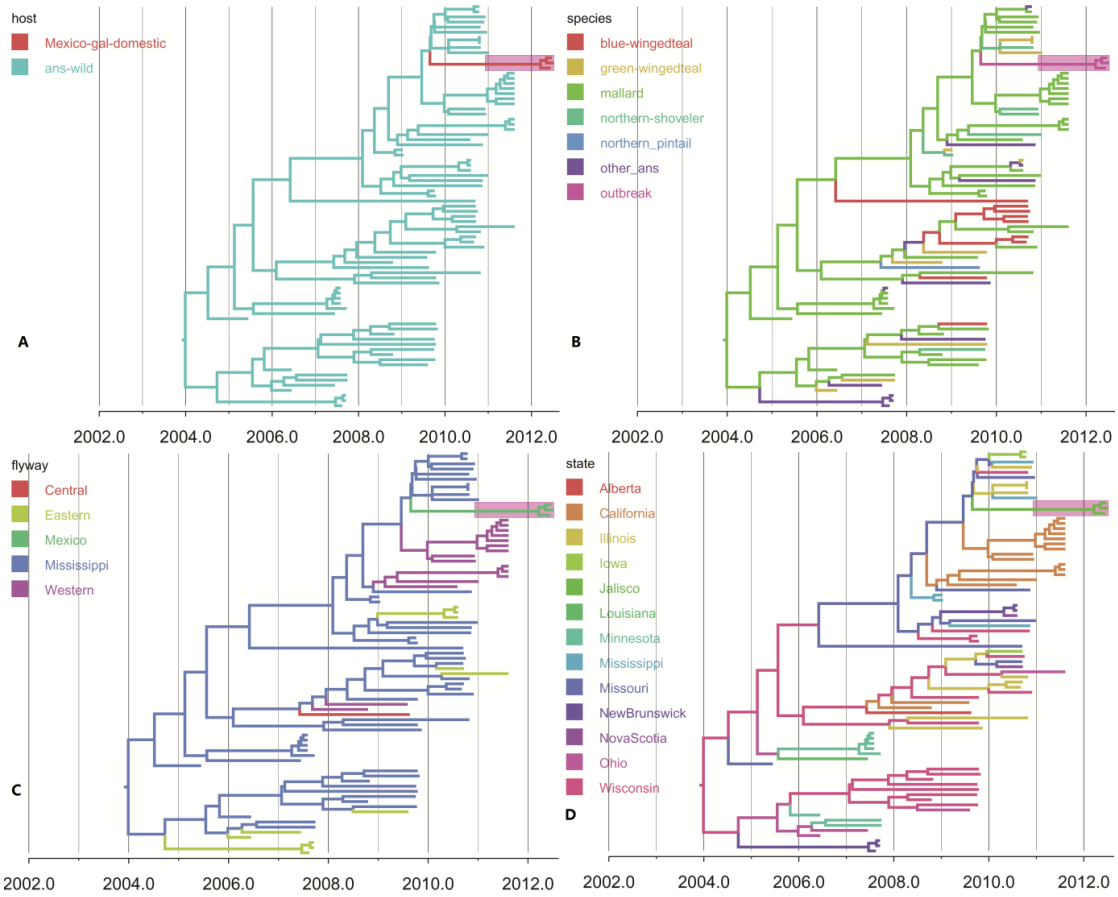


Figure A5.5 MCC phylogenies for the NP segment

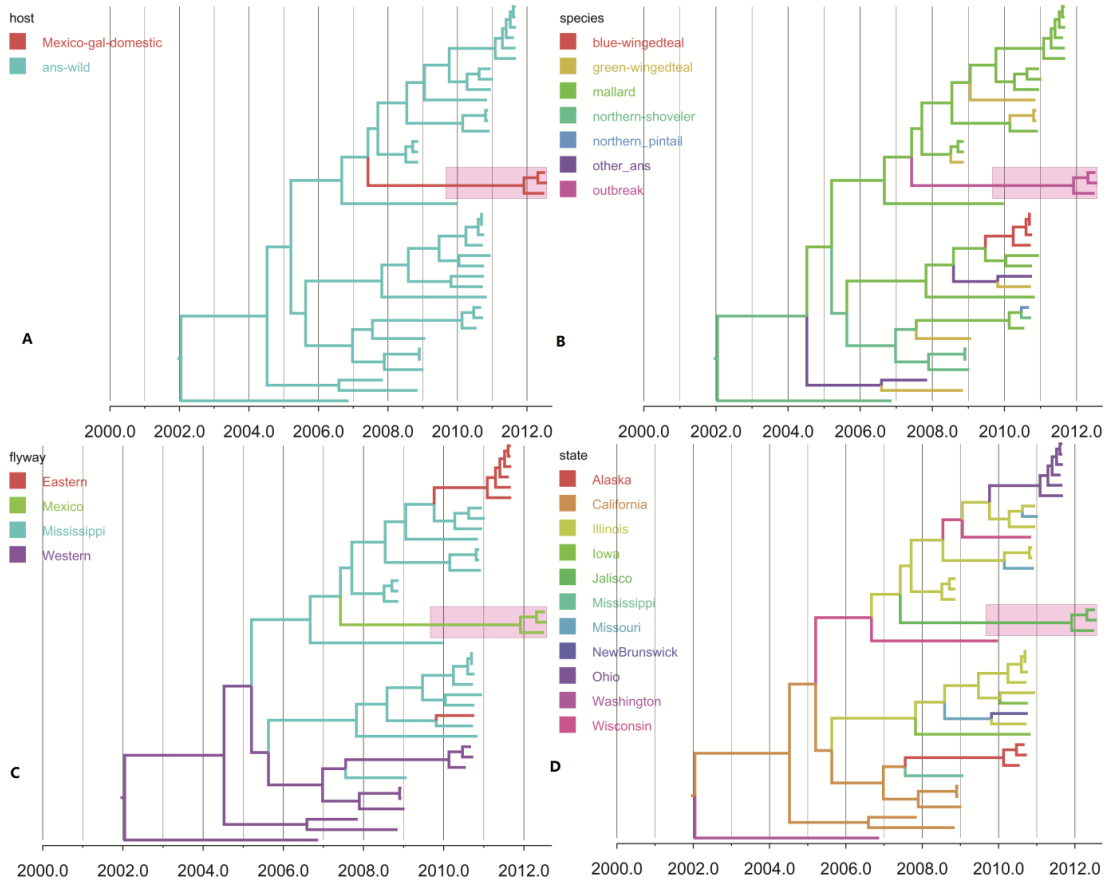


Figure A5.6 MCC phylogenies for the M segment

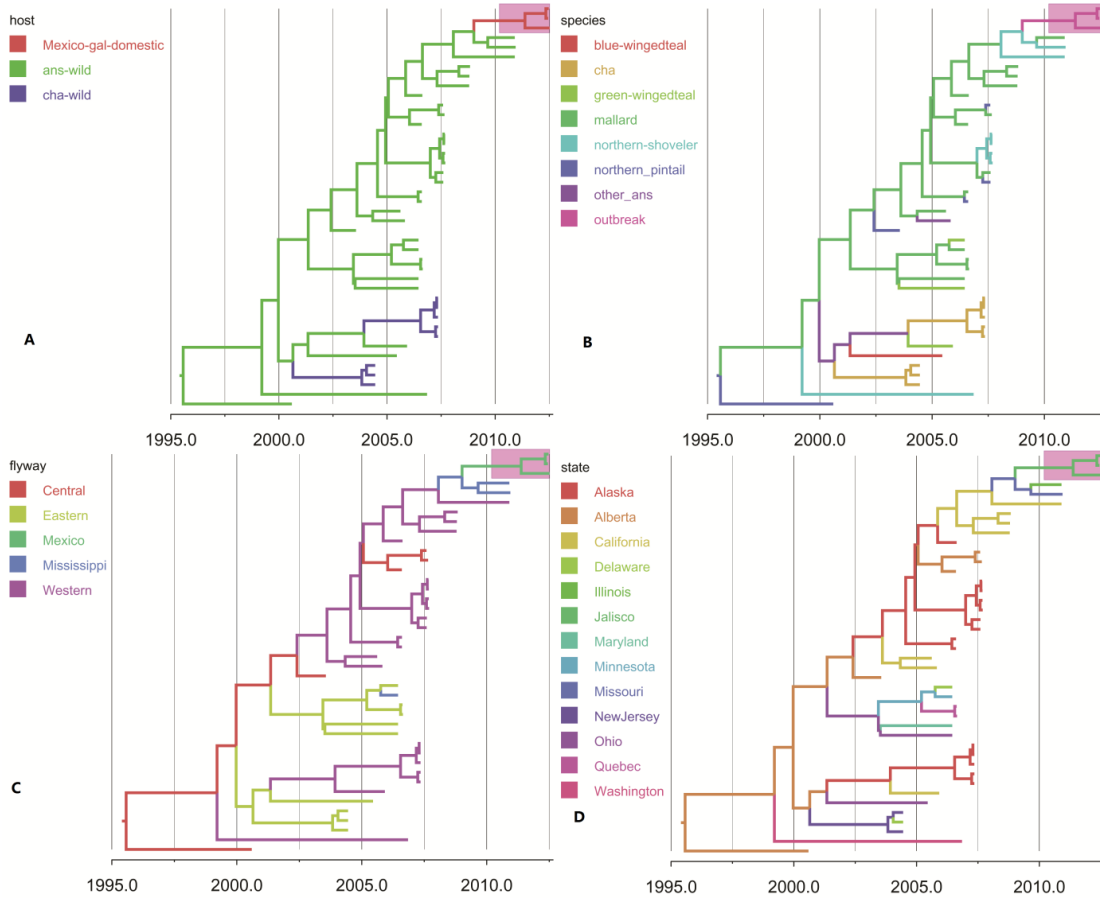


Figure A5.7 MCC phylogenies for the NS segment

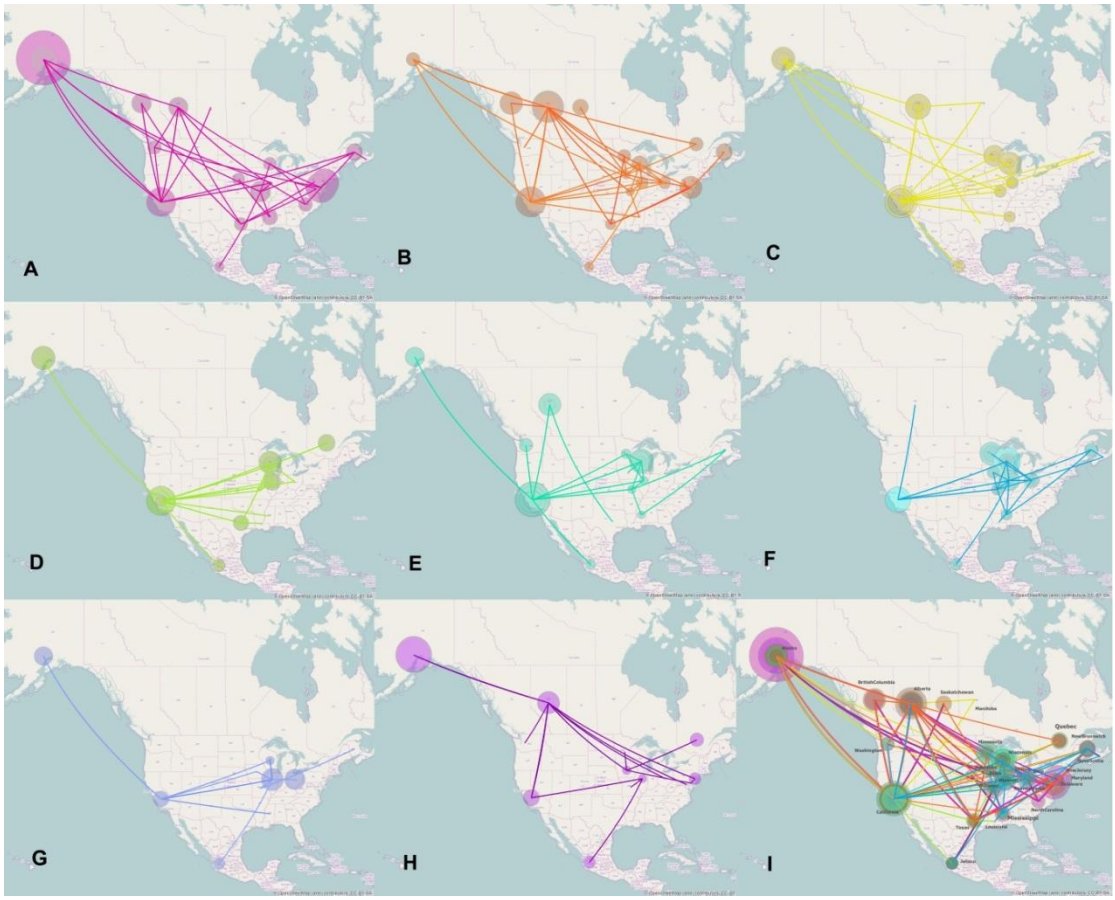


Figure A5.8 Spatial diffusion of 8 segments of the Mexico outbreak AIV.
Panels represent eight segments separately (A: HA, B: NA, C: PB2, D: PB1, E: PA, F: NP, G: M, H: NS) and I represents the spatial transmission of all 8 segments jointly. The Map source for this figure was OpenStreetMap (http://www.openstreetmap.org/). Plotted lines are the branches of the MCC trees of different segments, distinguished by colour; the size of each circle represents the number of lineages with that location state.

Name	Subtype	Host	State	Flyway	Segment						
A/mallard/Missouri/10MO0253/2010	H1N2	ans-wild	Missouri	Mississippi	PB2		PA				
A/northernshoveler/California/2810/2011	H11N2	ans-wild	California	Pacific	PB2		PA				
A/mallard/California/2961/2011	H2N3	ans-wild	California	Pacific	PB2		PA				
A/northernshoveler/California/3183/2010	H11N2	ans-wild	California	Pacific	PB2		PA				NS
A/Mexico/InDRE7218/2012	H7N3	Mexico-gal-domestic	Jalisco	Mexico	PB2	PB1	PA		NP	NA	M	NS
A/chicken/Jalisco/CPA1/2012	H7N3	Mexico-gal-domestic	Jalisco	Mexico	PB2	PB1	PA	HA	NP	NA	M	NS
A/chicken/Jalisco/12283/2012	H7N3	Mexico-gal-domestic	Jalisco	Mexico	PB2	PB1	PA	HA	NP		M	NS
A/northernpintail/Illinois/10OS3959/2010	H7N3	ans-wild	Illinois	Mississippi	PB2		PA	HA		NA		
A/northernpintail/California/3466/2010	H1N3	ans-wild	California	Pacific	PB2		PA			NA		
A/mallard/California/3569/2010	H1N3	ans-wild	California	Pacific	PB2		PA			NA		
A/mallard/California/198/2012	H11N9	ans-wild	California	Pacific	PB2		PA					
A/green-wingedteal/California/8326/2008	H1N2	ans-wild	California	Pacific	PB2		PA		NP		M	
A/northernshoveler/California/HKWF96/2007	H10N7	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/California/9680/2008	H6N2	ans-wild	California	Pacific	PB2		PA					
A/green-wingedteal/California/8612/2008	H6N1	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/California/8673/2008	H6N1	ans-wild	California	Pacific	PB2		PA					
A/greaterwhite-frontedgoose/California/6365/2008	H6N1	ans-wild	California	Pacific	PB2		PA					
A/mallard/California/8322/2008	H6N1	ans-wild	California	Pacific	PB2		PA					
A/mallard/California/8462/2008	H6N1	ans-wild	California	Pacific	PB2							
A/northernshoveler/California/44363-082/2007	H11N9	ans-wild	California	Pacific	PB2		PA					
A/Americangreen-wingedteal/Mississippi/09OS046/2009	H7N7	ans-wild	Mississippi	Mississippi	PB2		PA	HA	NP		M	
A/northernshoveler/Mississippi/09OS643/2009	H7N7	ans-wild	Mississippi	Mississippi	PB2		PA	HA	NP			
A/Americanwigeon/California/3180/2010	H5N2	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/California/JN587/2006	H10N3	ans-wild	California	Pacific	PB2	PB1	PA			NA		
A/Americanwigeon/California/HKWF450/2007	H4N7	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/California/HKWF392sm/2007	H10N7	ans-wild	California	Pacific	PB2		PA					
A/mallard/Minnesota/Sg-00056/2007	H10N7	ans-wild	Minnesota	Mississippi	PB2		PA		NP			
A/mallard/Minnesota/Sg-00057/2007	H10N7	ans-wild	Minnesota	Mississippi	PB2		PA		NP			
A/ring-neckedduck/Minnesota/Sg-00068/2007	H10N7	ans-wild	Minnesota	Mississippi	PB2		PA		NP			
A/mallard/Minnesota/Sg-00065/2007	H10N7	ans-wild	Minnesota	Mississippi	PB2		PA		NP			
A/mallard/Minnesota/Sg-00195/2007	H10N3	ans-wild	Minnesota	Mississippi	PB2		PA		NP	NA		

A/turkey/MB/FAV11/2010	H5N2	gal-domestic	Manitoba	Mississippi	PB2		PA				
A/turkey/MB/FAV10/2010	H5N2	gal-domestic	Manitoba	Mississippi	PB2		PA				
A/northernshoveler/California/HKWF2031/2008	H7N3	ans-wild	California	Pacific	PB2		PA	HA		NA	
A/northernpintail/California/2548/2010	H11N2	ans-wild	California	Pacific	PB2		PA				
A/northernshoveler/California/138/2012	H1N1	ans-wild	California	Pacific	PB2		PA				
A/northernpintail/InteriorAlaska/8BM3669/2008	H4N6	ans-wild	Alaska	Pacific	PB2		PA				
A/northernpintail/InteriorAlaska/8BM3736/2008	H12N5	ans-wild	Alaska	Pacific	PB2		PA				
A/northernshoveler/California/HKWF608/2007	H10N7	ans-wild	California	Pacific	PB2		PA				
A/northernshoveler/California/3628/2011	H10N3	ans-wild	California	Pacific	PB2	PB1	PA			NA	
A/northernshoveler/Washington/44249-664/2006	H7N3	ans-wild	Washington	Pacific	PB2		PA	HA		NA	
A/mallard/California/6524/2008	H12N5	ans-wild	California	Pacific	PB2		PA				
A/mallard/Wisconsin/2549/2009	H3N2	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/mallard/California/3188/2010	H6N8	ans-wild	California	Pacific	PB2		PA				
A/mallard/NewBrunswick/00854/201	H4N6	ans-wild	NewBrunswick	Atlantic	PB2		PA				
A/mallard/Manitoba/23912/2007	H4N7	ans-wild	Manitoba	Mississippi	PB2		PA				
A/mallard/Wisconsin/2575/2009	H3N2	ans-wild	Wisconsin	Mississippi	PB2		PA		NP		
A/mallard/Wisconsin/10OS2889/2010	H11N9	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/Americangreen-wingedteal/Wisconsin/10OS2955/2010	H5N2	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/Americangreen-wingedteal/Illinois/10OS3662/2010	H11N2	ans-wild	Illinois	Mississippi	PB2		PA				
A/mallard/Wisconsin/10OS2538/2010	H3N2	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/mallard/Wisconsin/10OS2909/2010	H3N8	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/mallard/Wisconsin/2653/2009	H4N6	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/mallard/Wisconsin/10OS3169/2010	H4N9	ans-wild	Wisconsin	Mississippi	PB2		PA				
A/Americanblackduck/Illinois/3854/2009	H11N9	ans-wild	Illinois	Mississippi	PB2		PA		NP		
A/blue-wingedteal/Wisconsin/10OS3092/2010	H3N6	ans-wild	Wisconsin	Mississippi	PB2		PA		NP		
A/mallard/Wisconsin/10OS4194/2010	H11N9	ans-wild	Wisconsin	Mississippi	PB2	PB1	PA		NP		
A/Americanblackduck/NewBrunswick/00998/2010	H12N6	ans-wild	NewBrunswick	Atlantic	PB2		PA				M
A/blue-wingedteal/Missouri/10MO003/2010	H4N6	ans-wild	Missouri	Mississippi	PB2		PA		NP		
A/blue-wingedteal/Missouri/10MO0011/2010	H3N6	ans-wild	Missouri	Mississippi	PB2	PB1	PA				
A/blue-wingedteal/Iowa/10OS2639/2010	H3N8	ans-wild	Iowa	Mississippi	PB2		PA		NP		
A/northernshoveler/Washington/44249-783/2006	H7N3	ans-wild	Washington	Pacific	PB2		PA	HA			
A/northernshoveler/Washington/44249-752/2006	H7N3	ans-wild	Washington	Pacific	PB2		PA				
A/northernshoveler/California/HKWF1325/2007	H8N4	ans-wild	California	Pacific	PB2		PA				
A/westerngrebe/Washington/20569-004/2007	H1N2	ans-wild	Washington	Pacific	PB2		PA				
A/Americanwigeon/California/HKWF295/2007	H6N5	ans-wild	California	Pacific	PB2		PA				

A/mallard/California/11100/2008	H11N2	ans-wild	California	Pacific	PB2		PA					
A/mallard/Alberta/107/2007	H3N8	ans-wild	Alberta	Central	PB2		PA					
A/mallard/Alberta/116/2007	H3N8	ans-wild	Alberta	Central	PB2							
A/mallard/Alberta/130/2007	H3N8	ans-wild	Alberta	Central	PB2		PA					
A/blue-wingedteal/Texas/Sg-00157/2007	H4N6	ans-wild	Texas	Central	PB2		PA					
A/mallard/California/8457/2008	H6N2	ans-wild	California	Pacific	PB2		PA					NS
A/mallard/California/8834/2008	H5N9	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/California/9140/2008	H1N9	ans-wild	California	Pacific	PB2		PA					
A/northernshoveler/Washington/44249-645/2006	H5N2	ans-wild	Washington	Pacific	PB2		PA					
A/mallard/Alberta/76/2006	H1N3	ans-wild	Alberta	Central	PB2		PA			NA		
A/mallard/California/8212/2008	H6N1	ans-wild	California	Pacific	PB2	PB1	PA					NS
A/mallard/Alberta/254/2006	H4N6	ans-wild	Alberta	Central	PB2		PA					
A/mallard/Alberta/297/2006	H4N6	ans-wild	Alberta	Central	PB2							
A/bufflehead/California/HKWF205/2007	H4N8	ans-wild	California	Pacific	PB2		PA					
A/mallard/Alberta/121/2008	H4N6	ans-wild	Alberta	Central	PB2		PA					
A/mallard/Alberta/114/2007	H4N6	ans-wild	Alberta	Central	PB2		PA					
A/mallard/Alberta/160/2007	H4N6	ans-wild	Alberta	Central	PB2		PA					
A/northernshovelerMissouri/196/2009	H10N3	ans-wild	Missouri	Mississippi	PB2	PB1	PA			NA		
A/northernpintail/InteriorAlaska/10BM02539R0/2010	H7N3	ans-wild	Alaska	Pacific	PB2		PA	HA		NA		
A/Americangreen-wingedteal/Wisconsin/10OS3127/2010	H5N2	ans-wild	Wisconsin	Mississippi	PB2	PB1	PA				M	
A/green-wingedteal/California/K481/2006	H1N3	ans-wild	California	Pacific		PB1						
A/mallard/California/2556P/2011	H3N8	ans-wild	California	Pacific		PB1						
A/mallard/California/2556V/2011	H3N8	ans-wild	California	Pacific		PB1						
A/mallard/California/1390/2010	H7N5	ans-wild	California	Pacific		PB1		HA	NP			
A/mallard/Missouri/350/2009	H11N9	ans-wild	Missouri	Mississippi		PB1						
A/mallard/Missouri/10MO0391/2010	H10N7	ans-wild	Missouri	Mississippi		PB1					M	
A/mallard/Illinois/10OS3786/2010	H10N7	ans-wild	Illinois	Mississippi		PB1					M	NS
A/mallard/Illinois/10OS4111/2010	H10N7	ans-wild	Illinois	Mississippi		PB1					M	
A/mallard/California/6471/2008	H3N8	ans-wild	California	Pacific		PB1						
A/ringneckedduck/California/HKWF402/2007	H6N1	ans-wild	California	Pacific		PB1						
A/northernshoveler/California/44363-062/2007	H9N2	ans-wild	California	Pacific		PB1						
A/blue-wingedteal/Ohio/566/2006	H7N9	ans-wild	Ohio	Mississippi		PB1		HA				
A/Americangreen-wingedteal/California/44287-373/2007	H8N4	ans-wild	California	Pacific		PB1						
A/commongoldeneye/Wisconsin/10OS4202/2010	H7N6	ans-wild	Wisconsin	Mississippi		PB1		HA				
A/blue-wingedteal/Texas/Sg-00074/2007	H4N8	ans-wild	Texas	Central		PB1						

A/green-wingedteal/California/123/2012	H1N1	ans-wild	California	Pacific	PB1	PA			
A/mallard/California/2589P/2011	H4N1	ans-wild	California	Pacific	PB1				
A/mallard/California/2533V/2011	H4N6	ans-wild	California	Pacific	PB1				
A/mallard/California/2555V/2011	H11N9	ans-wild	California	Pacific	PB1				
A/mallard/California/2563P/2011	H4N6	ans-wild	California	Pacific	PB1				
A/mallard/California/2595V/2011	H12N5	ans-wild	California	Pacific	PB1				
A/mallard/California/2536V/2011	H5N1	ans-wild	California	Pacific	PB1				
A/mallard/California/2527V/2011	H5N1	ans-wild	California	Pacific	PB1				
A/mallard/California/2531V/2011	H5N1	ans-wild	California	Pacific	PB1		NP		
A/mallard/California/2559P/2011	H5N8	ans-wild	California	Pacific	PB1				
A/Americangreen-wingedteal/InteriorAlaska/10BM05376R0/2010	H3N8	ans-wild	Alaska	Pacific	PB1				
A/northernshoveler/InteriorAlaska/10BM05491R0/2010	H3N8	ans-wild	Alaska	Pacific	PB1				
A/northernshoveler/InteriorAlaska/10BM05487R0/2010	H3N8	ans-wild	Alaska	Pacific	PB1				
A/northernshoveler/InteriorAlaska/10BM05382R0/2010	H3N8	ans-wild	Alaska	Pacific	PB1				
A/bufflehead/California/3118/2011	H4N8	ans-wild	California	Pacific	PB1				
A/mallard/California/2396/2010	H5N2	ans-wild	California	Pacific	PB1				
A/Americangreen-wingedteal/Illinois/10OS1598/2010	H4N8	ans-wild	Illinois	Mississippi	PB1				M
A/mallard/Wisconsin/10OS3066/2010	H6N2	ans-wild	Wisconsin	Mississippi	PB1				
A/mallard/Ohio/1695/2009	H4N6	ans-wild	Ohio	Atlantic	PB1				
A/duck/InteriorAlaska/7MP1550/2007	H4N6	ans-wild	Alaska	Pacific	PB1				
A/mallard/InteriorAlaska/5/2007	H3N8	ans-wild	Alaska	Pacific	PB1				
A/northernshoveler/InteriorAlaska/2/2007	H3N8	ans-wild	Alaska	Pacific	PB1				
A/mallard/California/K752/2006	H10N7	ans-wild	California	Pacific	PB1				
A/northernshoveler/Mississippi/09OS025/2009	H12N5	ans-wild	Mississippi	Mississippi	PB1				
A/Americangreen-wingedteal/California/44363-002/2007	H11N5	ans-wild	California	Pacific	PB1				
A/mallard/Illinois/3974/2009	H5N2	ans-wild	Illinois	Mississippi	PB1				
A/blue-wingedteal/Wisconsin/10OS2862/2010	H3N2	ans-wild	Wisconsin	Mississippi	PB1				
A/woodduck/Wisconsin/10OS2778/2010	H3N8	ans-wild	Wisconsin	Mississippi	PB1				
A/mallard/Wisconsin/10OS2773/2010	H3N8	ans-wild	Wisconsin	Mississippi	PB1				
A/blue-wingedteal/Texas/Sg-00085/2007	H3N6	ans-wild	Texas	Central	PB1				
A/blue-wingedteal/Texas/Sg-00079/2007	H3N8	ans-wild	Texas	Central	PB1				
A/blue-wingedteal/Louisiana/Sg-00224/2007	H3N8	ans-wild	Louisiana	Mississippi	PB1				
A/blue-wingedteal/Texas/Sg-00188/2007	H4N8	ans-wild	Texas	Central	PB1				
A/mallard/Illinois/3747/2009	H6N1	ans-wild	Illinois	Mississippi	PB1				
A/bufflehead/Illinois/4016/2009	H4N8	ans-wild	Illinois	Mississippi	PB1				

A/mallard/Quebec/10969/2006	H2N3	ans-wild	Quebec	Atlantic	PB1			NA	
A/mallard/Quebec/16566/2005	H11N2	ans-wild	Quebec	Atlantic	PB1				
A/mallard/Quebec/11247/2006	H3N2	ans-wild	Quebec	Atlantic	PB1				
A/mallard/Quebec/11194/2006	H3N2	ans-wild	Quebec	Atlantic	PB1				
A/mallard/Quebec/11093/2006	H3N2	ans-wild	Quebec	Atlantic	PB1				
A/mallard/Illinois/08OS2315/2008	H4N6	ans-wild	Illinois	Mississippi	PB1				
A/blue-wingedteal/Wisconsin/2649/2009	H6N1	ans-wild	Wisconsin	Mississippi	PB1		NP		
A/mallard/California/8416/2008	H6N1	ans-wild	California	Pacific		PA			
A/greaterwhite-frontedgoose/California/6548/2008	H6N1	ans-wild	California	Pacific		PA			
A/Americanwigeon/California/9044/2008	H6N1	ans-wild	California	Pacific		PA			
A/northernpintail/California/8470/2008	H6N1	ans-wild	California	Pacific		PA			
A/gadwall/California/8708/2008	H6N1	ans-wild	California	Pacific		PA			
A/green-wingedteal/MD/648/2004	H7N3	ans-wild	Maryland	Atlantic			HA		
A/green-wingedteal/Ohio/648/2004	H7N3	ans-wild	Ohio	Mississippi			HA		
A/blue-wingedteal/MD/658/2004	H7N3	ans-wild	Maryland	Atlantic			HA		
A/blue-wingedteal/Ohio/658/2004	H7N3	ans-wild	Ohio	Mississippi			HA	NA	
A/chicken/BritishColumbia/NS-01827-4/2004	H7N3	gal-domestic	BritishColumbia	Pacific			HA	NA	
A/chicken/BritishColumbia/CN-7/2004	H7N3	gal-domestic	BritishColumbia	Pacific			HA		
A/chicken/Canada/314514-1/2005	H7N3	gal-domestic	BritishColumbia	Pacific			HA		
A/chicken/BritishColumbia/04	H7N3	gal-domestic	BritishColumbia	Pacific			HA	NA	
A/blue-wingedteal/Texas/578575/2002	H7N1	ans-wild	Texas	Central			HA		
A/blue-wingedteal/Texas/578585/2002	H7N3	ans-wild	Texas	Central			HA		
A/mallard/Alberta/24/01	H7N3	ans-wild	Alberta	Central			HA		
A/mallard/Alberta/34/2001	H7N1	ans-wild	Alberta	Central			HA		
A/mallard/Alberta/22/2001	H7N3	ans-wild	Alberta	Central			HA		
A/northernshoveler/California/HKWF1026/2007	H7N3	ans-wild	California	Pacific			HA	NA	
A/green-wingedteal/California/11275/2008	H7N3	ans-wild	California	Pacific			HA	NA	
A/green-wingedteal/California/1841/2009	H7N3	ans-wild	California	Pacific			HA		
A/blackscoter/NewBrunswick/00014/2009	H7N6	ans-wild	NewBrunswick	Atlantic			HA		
A/blackscoter/NewBrunswick/00003/2009	H7N6	ans-wild	NewBrunswick	Atlantic			HA		
A/chicken/SK/HR-00011/2007	H7N3	gal-domestic	Saskatchewan	Central			HA	NA	
A/mallard/InteriorAlaska/6MP0984/2006	H7N3	ans-wild	Alaska	Pacific			HA	NA	NS
A/cinnamonteal/California/JN1310/2007	H7N3	ans-wild	California	Pacific			HA	NA	
A/mallard/Alberta/243/2006	H7N3	ans-wild	Alberta	Central			HA	NA	
A/northernpintail/InteriorAlaska/8MP0262R2/2008	H7N3	ans-wild	Alaska	Pacific			HA	NA	

A/mallard/Ohio/11OS2010/2011	H7N8	ans-wild	Ohio	Mississippi	HA			
A/northernshoveler/Wisconsin/10OS3226/2010	H7N3	ans-wild	Wisconsin	Mississippi	HA		NA	
A/mallard/Wisconsin/10OS3171/2010	H7N3	ans-wild	Wisconsin	Mississippi	HA		NA	
A/mallard/InteriorAlaska/10CH00016R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM09015R0/2010	H7N6	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM04564R1/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM10829R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM09528R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM10476R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM09563R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM10166R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM11208R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM07399R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM07469R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM07072R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM12534R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/mallard/InteriorAlaska/10BM07085R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/northernpintail/InteriorAlaska/10BM06524R0/2010	H7N3	ans-wild	Alaska	Pacific	HA			
A/gadwall/Missouri/10OS4731/2010	H7N3	ans-wild	Missouri	Mississippi	HA			
A/northernshoveler/Mississippi/11OS289/2011	H7N3	ans-wild	Mississippi	Mississippi	HA			
A/guineafowl/Nebraska/17096-1/2011	H7N9	gal-domestic	Nebraska	Central	HA			
A/goose/Nebraska/17097-4/2011	H7N9	ans-wild	Nebraska	Central	HA			
A/Americangreen-wingedteal/Illinois/10OS4014/2010	H7N3	ans-wild	Illinois	Mississippi	HA			
A/Americangreen-wingedteal/Illinois/10OS3329/2010	H7N7	ans-wild	Illinois	Mississippi	HA	NP		M
A/northernshoveler/Mississippi/11OS202/2011	H7N7	ans-wild	Mississippi	Mississippi	HA			
A/northernshoveler/Missouri/10OS4632/2010	H7N7	ans-wild	Missouri	Mississippi	HA			
A/Americangreen-wingedteal/Mississippi/11OS250/2011	H7N3	ans-wild	Mississippi	Mississippi	HA			
A/Americangreen-wingedteal/Mississippi/11OS255/2011	H7N7	ans-wild	Mississippi	Mississippi	HA			
A/Americanblackduck/Wisconsin/10OS3949/2010	H7N8	ans-wild	Wisconsin	Mississippi	HA			
A/northernshoveler/Missouri/10OS4750/2010	H7N3	ans-wild	Missouri	Mississippi	HA			
A/northernshoverl/Mississippi/11OS145/2011	H7N9	ans-wild	Mississippi	Mississippi	HA			
A/mallard/Missouri/10MO0551/2010	H7N7	ans-wild	Missouri	Mississippi	HA	NP		
A/mallard/Missouri/220/2009	H7N3	ans-wild	Missouri	Mississippi	HA		NA	
A/mallard/Missouri/10MO053/2010	H7N4	ans-wild	Missouri	Mississippi	HA			M
A/Americanblackduck/NewBrunswick/00344/2010	H7N7	ans-wild	NewBrunswick	Atlantic	HA	NP		

A/green-wingedteal/NewBrunswick/00392/2010	H7N7	ans-wild	NewBrunswick	Atlantic	HA	NP	
A/northernshoveler/California/JN1447/2007	H7N2	ans-wild	California	Pacific	HA		
A/ruddyturnstone/DelawareBay/124/2007	H7N3	cha-wild	Delaware	Atlantic	HA		
A/mallard/California/HKWF1971/2007	H7N7	ans-wild	California	Pacific	HA		
A/greenwingedteal/California/AKS1370/2008	H7N3	ans-wild	California	Pacific	HA		
A/mallard/NovaScotia/02286/2007	H7N4	ans-wild	NovaScotia	Atlantic	HA		
A/Americanblackduck/NewBrunswick/04388/2007	H7N3	ans-wild	NewBrunswick	Atlantic	HA	NP	NA
A/Americanblackduck/NB/2538/2007	H7N3	ans-wild	NewBrunswick	Atlantic	HA		NA
A/Americanblackduck/NewBrunswick/02493/2007	H7N3	ans-wild	NewBrunswick	Atlantic	HA	NP	NA
A/Americanblackduck/NewBrunswick/02490/2007	H7N3	ans-wild	NewBrunswick	Atlantic	HA	NP	
A/Canadagoose/BC/3752/2007	H7N3	ans-wild	BritishColumbia	Pacific	HA		NA
A/northernshoveler/California/28327/2007	H7N3	ans-wild	California	Pacific	HA		NA
A/northernshoveler/California/44287-364/2007	H7N3	ans-wild	California	Pacific	HA		NA
A/cinnamonteal/California/JN611/2006	H7N3	ans-wild	California	Pacific	HA		
A/northernshoveler/California/27820/2007	H7N3	ans-wild	California	Pacific	HA		
A/Americangreen-wingedteal/California/44287-084/2007	H7N3	ans-wild	California	Pacific	HA		
A/Americangreen-wingedteal/California/44242-906/2007	H7N3	ans-wild	California	Pacific	HA		NA
A/northernshoveler/California/44287-164/2007	H7N7	ans-wild	California	Pacific	HA		
A/Northernshoveler/NC/6412-052/2005	H7N6	ans-wild	NorthCarolina	Atlantic	HA		
A/mallard/Delaware/418/2005	H7N3	ans-wild	Delaware	Atlantic	HA		NA
A/northernshoveler/NorthCarolina/6412-050/2005	H7N3	ans-wild	NorthCarolina	Atlantic	HA		
A/northernshoveler/NorthCarolina/674-516/2005	H7N3	ans-wild	NorthCarolina	Atlantic	HA		
A/Americangreen-wingedteal/California/28228/2007	H7N6	ans-wild	California	Pacific	HA		
A/Americangreen-wingedteal/California/44287-305/2007	H7N6	ans-wild	California	Pacific	HA		
A/northernshoveler/California/27985/2007	H7N6	ans-wild	California	Pacific	HA		
A/northernshoveler/California/44287-179/2007	H7N6	ans-wild	California	Pacific	HA		
A/northernshoveler/California/HKWF1372C/2007	H7N3	ans-wild	California	Pacific	HA		NA
A/ruddyturnstone/NewJersey/604/2006	H7N4	cha-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/NewJersey/490/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/NewJersey/215/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/sanderling/NewJersey/369/2006	H7N3	pas-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/Delaware/752/2006	H7N7	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/NewJersey/612/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/sanderling/NewJersey/355/2006	H7N3	pas-wild	NewJersey	Atlantic	HA		
A/laughinggull/DelawareBay/46/2006	H7N3	cha-wild	Delaware	Atlantic	HA		NA

A/laughinggull/DelawareBay/42/2006	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/ruddyturnstone/Delaware/890/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/NewJersey/176/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/laughinggull/DelawareBay/50/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/NewJersey/204/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/DelawareBay/281/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/DelawareBay/283/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/DelawareBay/262/2006	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/ruddyturnstone/NewJersey/224/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/DelawareBay/290/2006	H7N4	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/NewJersey/589/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/ruddyturnstone/Delaware/892/2006	H7N7	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/NewJersey/576/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/laughinggull/DelawareBay/6/2006	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/shorebird/DelawareBay/513/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/Delaware/789/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/ruddyturnstone/Delaware/779/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/shorebird/DelawareBay/332/2006	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/ruddyturnstone/NewJersey/562/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/shorebird/Delaware/22/06	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/shorebird/DelawareBay/560/2006	H7N3	cha-wild	Delaware	Atlantic	HA	NA	
A/avian/DelawareBay/226/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/redknot/NewJersey/AI06-096/2006	H7N3	cha-wild	NewJersey	Atlantic	HA		
A/shorebird/DelawareBay/555/2006	H7N3	cha-wild	Delaware	Atlantic	HA		
A/northernshoveler/Washington/44249-749/2006	H7N3	ans-wild	Washington	Pacific	HA	NA	
A/mallard/Illinois/10OS3599/2010	H6N1	ans-wild	Illinois	Mississippi	NP		
A/mallard/Louisiana/476670-4/2007	H5N2	ans-wild	Louisiana	Mississippi	NP		
A/blue-wingedteal/Illinois/10OS2988/2010	H4N6	ans-wild	Illinois	Mississippi	NP		M
A/blue-wingedteal/Illinois/10OS1546/2010	H3N6	ans-wild	Illinois	Mississippi	NP		M
A/Americangreen-wingedteal/Wisconsin/2530/2009	H6N2	ans-wild	Wisconsin	Mississippi	NP		
A/northernpintail/Alberta/8/2009	H3N8	ans-wild	Alberta	Central	NP		
A/blue-wingedteal/NovaScotia/01009/2010	H4N6	ans-wild	NovaScotia	Atlantic	NP		
A/mallard/California/5351/2009	H1N1	ans-wild	California	Pacific	NP		
A/mallard/Ohio/11OS1992/2011	H3N8	ans-wild	Ohio	Mississippi	NP		
A/mallard/Illinois/10OS3676/2010	H3N8	ans-wild	Illinois	Mississippi	NP		

A/Americangreen-wingedteal/Wisconsin/2743/2009	H1N1	ans-wild	Wisconsin	Mississippi	NP		
A/green-wingedteal/Minnesota/Sg-00131/2007	H3N2	ans-wild	Minnesota	Mississippi	NP		
A/mallard/Ohio/1688/2009	H12N5	ans-wild	Ohio	Mississippi	NP		
A/mallard/Wisconsin/2712/2009	H3N6	ans-wild	Wisconsin	Mississippi	NP		
A/northernshoveler/Wisconsin/2508/2009	H4N2	ans-wild	Wisconsin	Mississippi	NP		
A/mallard/Minnesota/Sg-00133/2007	H4N6	ans-wild	Minnesota	Mississippi	NP		
A/mallard/Missouri/MO130/2005	H11N3	ans-wild	Missouri	Mississippi	NP	NA	
A/wildbird/Minnesota/460613/2006	H5N2	ans-wild	Minnesota	Mississippi	NP		
A/mallard/Wisconsin/08OS2271/2008	H11N9	ans-wild	Wisconsin	Mississippi	NP		
A/mallard/Wisconsin/08OS2841/2008	H2N3	ans-wild	Wisconsin	Mississippi	NP	NA	
A/greenwingedteal/Ohio/468160/2006	H5N2	ans-wild	Ohio	Mississippi	NP		
A/blue-wingedteal/Wisconsin/3060/2009	H3N2	ans-wild	Wisconsin	Mississippi	NP		
A/Americanblackduck/Wisconsin/2542/2009	H4N2	ans-wild	Wisconsin	Mississippi	NP		
A/mallard/Wisconsin/3165/2009	H1N1	ans-wild	Wisconsin	Mississippi	NP		
A/blue-wingedteal/Missouri/10MO013/2010	H3N1	ans-wild	Missouri	Mississippi	NP		
A/mallard/California/2540P/2011	H3N8	ans-wild	California	Pacific	NP		
A/mallard/California/2550V/2011	H3N8	ans-wild	California	Pacific	NP		
A/mallard/California/2549V/2011	H3N8	ans-wild	California	Pacific	NP		
A/northernshoveler/California/4020/2011	H4N3	ans-wild	California	Pacific	NP	NA	
A/mallard/NovaScotia/00372/2010	H7N7	ans-wild	NovaScotia	Atlantic	NP		
A/long-tailedduck/Wisconsin/10OS3915/2010	H3N6	ans-wild	Missouri	Mississippi	NP		
A/mallard/Wisconsin/10OS3845/2010	H5N2	ans-wild	Wisconsin	Mississippi	NP		
A/mallard/Wisconsin/2719/2009	H4N2	ans-wild	Wisconsin	Mississippi	NP		
A/long-tailedduck/Wisconsin/10OS3919/2010	H10N6	ans-wild	Mississippi	Mississippi	NP		
A/mallard/Mississippi/10OS4494/2010	H1N1	ans-wild	Mississippi	Mississippi	NP		
A/mallard/Missouri/10MO0333/2010	H11N9	ans-wild	Missouri	Mississippi	NP		
A/northernshoveler/California/3676/2010	H8N2	ans-wild	California	Pacific	NP		
A/mallard/Wisconsin/10OS3144/2010	H6N1	ans-wild	Wisconsin	Mississippi	NP		
A/northernshoveler/California/3483/2010	H12N5	ans-wild	California	Pacific	NP		
A/Americangreen-wingedteal/Illinois/10OS3368/2010	H7N7	ans-wild	Illinois	Mississippi	NP		M
A/northernshoveler/Illinois/10OS3619/2010	H11N9	ans-wild	Illinois	Mississippi	NP		
A/mallard/California/2590V/2011	H4N6	ans-wild	California	Pacific	NP		
A/mallard/California/2578P/2011	H3N8	ans-wild	California	Pacific	NP		
A/mallard/California/2566P/2011	H4N8	ans-wild	California	Pacific	NP		
A/mallard/California/2555P/2011	H11N9	ans-wild	California	Pacific	NP		

A/mallard/California/2563V/2011	H4N6	ans-wild	California	Pacific	NP	
A/mallard/California/2595P/2011	H12N5	ans-wild	California	Pacific	NP	
A/Americanwigeon/Iowa/10OS2748/2010	H2N2	ans-wild	Iowa	Mississippi	NP	
A/mallard/Iowa/10OS2721/2010	H2N2	ans-wild	Iowa	Mississippi	NP	
A/mallard/Illinois/10OS4179/2010	H11N2	ans-wild	Illinois	Mississippi	NP	
A/Americangreen-wingedteal/Mississippi/11OS90/2011	H11N9	ans-wild	Mississippi	Mississippi	NP	
A/mallard/Washington/44256-522/2006	H11N3	ans-wild	Washington	Pacific		NA
A/lesserscaup/Wisconsin/3964/2009	H10N3	ans-wild	Wisconsin	Mississippi		NA
A/mallard/Alberta/319/2009	H2N3	ans-wild	Alberta	Central		NA
A/mallard/Alberta/152/2006	H1N3	ans-wild	Alberta	Central		NA
A/mallard/Ohio/11OS1966/2011	H10N3	ans-wild	Ohio	Mississippi		NA
A/mallard/Missouri/10MO0550/2010	H11N3	ans-wild	Missouri	Mississippi		NA
A/mallard/Illinois/3051/2009	H11N3	ans-wild	Illinois	Mississippi		NA
A/mallard/Mississippi/11OS34/2011	H1N3	ans-wild	Mississippi	Mississippi		NA
A/mallard/California/1505/2010	H4N3	ans-wild	California	Pacific		NA
A/mallard/InteriorAlaska/10BM05347R0/2010	H7N3	ans-wild	Alaska	Pacific		NA
A/blue-wingedteal/Alberta/346/2007	H4N3	ans-wild	Alberta	Central		NA
A/northernshoveler/California/HKWF1005/2007	H10N3	ans-wild	California	Pacific		NA
A/northernshoveler/California/HKWF1370/2007	H10N3	ans-wild	California	Pacific		NA
A/mallard/Quebec/16334/2005	H5N3	ans-wild	Quebec	Atlantic		NA
A/mallard/Alberta/242/2004	H2N3	ans-wild	Alberta	Central		NA
A/mallard/Alberta/35/2009	H2N3	ans-wild	Alberta	Central		NA
A/mallard/Alberta/417/2009	H2N3	ans-wild	Alberta	Central		NA
A/ring-billedgull/Quebec/G066/2010	H1N3	cha-wild	Quebec	Atlantic		NA
A/ring-billedgull/Quebec/G139/2010	H1N3	cha-wild	Quebec	Atlantic		NA
A/northernpintail/SK/4612/2010	H5N3	ans-wild	Saskatchewan	Central		NA
A/northernpintail/SK/4628/2010	H5N3	ans-wild	Saskatchewan	Central		NA
A/ring-billedgull/Quebec/G018/2010	H1N3	cha-wild	Quebec	Atlantic		NA
A/mallard/Ohio/11OS2149/2011	H2N3	ans-wild	Ohio	Mississippi		NA
A/mallard/Wisconsin/08OS2844/2008	H2N3	ans-wild	Wisconsin	Mississippi		NA
A/mallard/Alberta/12017/2005	H2N3	ans-wild	Alberta	Central		NA
A/Americangreen-wingedteal/Illinois/10OS3343/2010	H2N3	ans-wild	Illinois	Mississippi		NA
A/northernshoveler/California/3046/2010	H4N3	ans-wild	California	Pacific		NA
A/northernshoveler/California/9781/2008	H1N3	ans-wild	California	Pacific		NA
A/ruddyturnstone/DelawareBay/108/2007	H7N3	cha-wild	Delaware	Atlantic		NA

A/northernshoveler/Mississippi/397/2010	H1N3	ans-wild	Mississippi	Mississippi		NA		
A/mallard/California/1438/2010	H2N3	ans-wild	California	Pacific		NA		
A/gadwall/California/44287-137/2007	H5N3	ans-wild	California	Pacific		NA		
A/Americanwigeon/California/2930/2011	H10N3	ans-wild	California	Central		NA		
A/Americancoot/Illinois/3405/2009	H10N3	gru-wild	Illinois	Mississippi		NA		
A/Americancoot/Mississippi/09OS615/2009	H10N3	gru-wild	Mississippi	Pacific		NA		
A/mallard/Alberta/80/2006	H1N3	ans-wild	Alberta	Central		NA		
A/redknot/Delaware/441/2002	H2N3	cha-wild	Delaware	Atlantic		NA		
A/shorebird/DelawareBay/53/2002	H7N3	cha-wild	Delaware	Atlantic		NA		
A/blue-wingedteal/TX/2/01	H7N3	ans-wild	Texas	Central		NA		
A/bluewingedteal/TX/75/2002	H1N3	ans-wild	Texas	Central		NA		
A/bluewingedteal/TX/34/2002	H1N3	ans-wild	Texas	Central		NA		
A/mallard/MN/479/2000	H5N3	ans-wild	Minnesota	Mississippi		NA		
A/blackduck/Maryland/415/2001	H7N3	ans-wild	Maryland	Atlantic		NA		
A/duck/PA/143585/2001	H7N3	ans-wild	Pennsylvania	Atlantic	HA	NA		
A/mallard/Minnesota/Sg-00221/2007	H10N3	ans-wild	Minnesota	Mississippi		NA		
A/Americancoot/California/20181-006/2007	H10N3	gru-wild	California	Pacific		NA		
A/mallard/Alberta/254/2003	H3N3	ans-wild	Alberta	Central		NA		
A/mallard/Quebec/11063/2006	H2N3	ans-wild	Quebec	Atlantic		NA		NS
A/mallard/Quebec/16485/2005	H3N3	ans-wild	Quebec	Atlantic		NA		
A/northernshoveler/California/HKWF979/2007	H3N3	ans-wild	California	Pacific		NA		
A/Americangreen-wingedteal/California/44287-713/2007	H7N3	ans-wild	California	Pacific		NA		
A/cinnamonteal/California/44287-659/2007	H10N3	ans-wild	California	Pacific		NA		
A/northernshoveler/California/JN770/2006	H10N3	ans-wild	California	Pacific		NA		
A/northernshoveler/Washington/44249-603/2006	H6N1	ans-wild	Washington	Pacific			M	
A/mallard/InteriorAlaska/10BM02980R0/2010	H9N2	ans-wild	Alaska	Pacific			M	
A/northernpintail/InteriorAlaska/10BM14807R2/2010	H9N2	ans-wild	Alaska	Pacific			M	
A/northernshoveler/InteriorAlaska/10BM16764R0/2010	H9N2	ans-wild	Alaska	Pacific			M	
A/northernshoveler/California/9228/2008	H4N6	ans-wild	California	Pacific			M	
A/northernshoveler/California/9267/2008	H4N6	ans-wild	California	Pacific			M	
A/northernshoveler/California/10024/2008	H4N4	ans-wild	California	Pacific			M	
A/mallard/Illinois/10OS4334/2010	H10N7	ans-wild	Illinois	Mississippi			M	
A/blue-wingedteal/Illinois/10OS1563/2010	H4N6	ans-wild	Illinois	Mississippi			M	
A/blue-wingedteal/Illinois/10OS1561/2010	H4N6	ans-wild	Illinois	Mississippi			M	
A/mallard/Iowa/10OS2420/2010	H4N6	ans-wild	Iowa	Mississippi			M	

A/mallard/Iowa/10OS2692/2010	H4N2	ans-wild	Iowa	Mississippi	M	
A/Americanwigeon/California/HKWF42/2007	H6N1	ans-wild	California	Pacific	M	
A/mallard/Missouri/10MO053/2010	H7N4	ans-wild	Missouri	Mississippi	M	
A/mallard/Illinois/08OS2711/2008	H10N7	ans-wild	Illinois	Mississippi	M	
A/mallard/Illinois/08OS2710/2008	H10N7	ans-wild	Illinois	Mississippi	M	
A/Americangreen-wingedteal/Illinois/08OS2713/2008	H10N7	ans-wild	Illinois	Mississippi	M	
A/mallard/Ohio/11OS2229/2011	H5N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Ohio/11OS2239/2011	H5N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Ohio/11OS2006/2011	H5N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Ohio/11OS1961/2011	H5N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Ohio/11OS2119/2011	H5N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Ohio/11OS2141/2011	H3N2	ans-wild	Ohio	Mississippi	M	
A/mallard/Wisconsin/4230/2009	H10N1	ans-wild	Wisconsin	Mississippi	M	
A/pintail/Alberta/84/2000	H11N9	ans-wild	Alberta	Central		NS
A/ruddyturnstone/DE/167/2004	H10N7	cha-wild	Delaware	Atlantic		NS
A/ruddyturnstone/NewJersey/110/2004	H10N7	cha-wild	NewJersey	Atlantic		NS
A/ruddyturnstone/NJ/238/2004	H10N7	cha-wild	NewJersey	Atlantic		NS
A/leastsandpiper/SouthCentralAlaska/1/2007	H4N8	cha-wild	Alaska	Pacific		NS
A/leastsandpiper/SouthCentralAlaska/7KW0434/2007	H4N8	cha-wild	Alaska	Pacific		NS
A/leastsandpiper/SouthCentralAlaska/2/2007	H4N8	cha-wild	Alaska	Pacific		NS
A/leastsandpiper/Alaska/7KW0411/2007	H4N8	cha-wild	Alaska	Pacific		NS
A/leastsandpiper/SouthCentralAlaska/3/2007	H4N8	cha-wild	Alaska	Pacific		NS
A/greenwingedteal/Ohio/464069/2006	H5N2	ans-wild	Ohio	Atlantic		NS
A/northernshoveler/Washington/44249-675/2006	H10N2	ans-wild	Washington	Pacific		NS
A/blue-wingedteal/Ohio/1339/2005	H4N6	ans-wild	Ohio	Mississippi		NS
A/green-wingedteal/California/K218/2005	H4N6	ans-wild	California	Pacific		NS
A/northernshoveler/Missouri/10OS4718/2010	H10N7	ans-wild	Missouri	Mississippi		NS
A/northernpintail/InteriorAlaska/6MP0792/2006	H2N3	ans-wild	Alaska	Pacific		NS
A/mallard/InteriorAlaska/6MP0038BR2/2006	H2N3	ans-wild	Alaska	Pacific		NS
A/mallard/InteriorAlaska/7MP0709/2007	H3N8	ans-wild	Alaska	Pacific		NS
A/northernshoveler/Alaska/7MP1113/2007	H4N6	ans-wild	Alaska	Pacific		NS
A/northernshoveler/InteriorAlaska/7MP1649/2007	H3N8	ans-wild	Alaska	Pacific		NS
A/northernshoveler/InteriorAlaska/7MP0953/2007	H3N8	ans-wild	Alaska	Pacific		NS
A/northernshoveler/InteriorAlaska/7MP0944/2007	H3N8	ans-wild	Alaska	Pacific		NS
A/northernpintail/InteriorAlaska/1/2007	H3N8	ans-wild	Alaska	Pacific		NS

A/ring-neckedduck/California/K90/2005	H6N8	ans-wild	California	Pacific	NS
A/mallard/Maryland/182/2006	H5N2	ans-wild	Maryland	Atlantic	NS
A/mallard/Quebec/11281/2006	H2N3	ans-wild	Quebec	Atlantic	NS
A/wildbird/Minnesota/460613-12/2006	H5N2	ans-wild	Minnesota	Mississippi	NS
A/greenwingedteal/Delaware/458672-5/2006	H5N2	ans-wild	Delaware	Atlantic	NS
A/mallard/California/19524-001/2005	H6N8	ans-wild	California	Pacific	NS
A/mallard/Alberta/330/2007	H4N6	ans-wild	Alberta	Central	NS
A/northernpintail/Alberta/265/2007	H4N6	ans-wild	Alberta	Central	NS
A/mallard/California/6469/2008	H6N2	ans-wild	California	Pacific	NS
A/pintailduck/Alberta/49/2003	H12N5	ans-wild	Alberta	Central	NS
A/mallard/Alberta/221/2006	H12N6	ans-wild	Alberta	Central	NS

Table A5.1 Sequences for the discrete traits analysis in Chapter 5.
Including the strain name and subtype, host, location, flyway and the segments being selected in the AIV dataset

	AK	AB	BC	CA	DE	IL	IA	JA	LA	MB	MD	MN	MS	MO	NE	NB	NJ	NC	NS	OH	PA	QC	SK	TX	WA	WI
Alaska(AK)		0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Alberta(AB)	0		0	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0
BritishColumbia(BC)	1	0		1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
California(CA)	1	0	1		0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Delaware(DE)	0	0	0	0		0	0	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
Illinois(IL)	0	1	0	0	0		1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0
Iowa(IA)	0	0	0	0	0	1		1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0
Jalisco(JA)	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Louisiana(LA)	0	0	0	0	0	0	0	1		0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
Manitoba(MB)	0	1	0	0	0	1	1	1	0		0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0
Maryland(MD)	0	0	0	0	1	0	0	1	0	0		0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
Minnesota(MN)	0	0	0	0	0	0	0	1	1	0	0		1	1	0	0	0	0	0	0	0	0	0	0	0	1
Mississippi(MS)	0	0	0	0	0	0	0	1	1	0	0	1		1	0	0	0	0	0	0	0	0	0	0	0	1
Missouri(MO)	0	0	0	0	0	0	0	1	1	0	0	1	1		0	0	0	0	0	0	0	0	0	0	0	1
Nebraska(NE)	0	1	0	0	0	1	1	1	0	1	0	0	0	0		0	0	0	0	0	0	0	1	1	0	0
NewBrunswick(NB)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0		1	1	1	1	1	1	0	0	0	0
NewJersey(NJ)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1		1	1	1	1	1	0	0	0	0
NorthCarolina(NC)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1		1	1	1	1	0	0	0	0
NovaScotia(NS)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1		1	1	1	0	0	0	0
Ohio(OH)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	1		1	1	0	0	0	0
Pennsylvania(PA)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	1	1		1	0	0	0	0
Quebec(QC)	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	1	1	1		0	0	0	0
Saskatchewan(SK)	0	1	0	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0		1	0	0
Texas(TX)	0	1	0	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1		0	0
Washington(WA)	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
Wisconsin(WI)	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	

Table A5.2 Reduced indicator matrix with between-flyway states being turned off.

The indicator matrix of non-reversible BSSVS model contains 325 transition pairs between states, which are all initially switched on (indicator=1). In the reduced matrix, the pair of states belonged to different flyways are switched on (indicator=0). There are 98 non-reversible transition pairs in the new indicator matrix

Chapter 6

Discussion

Chapter abstract

Avian influenza viruses present an epidemiological and economic threat on a global scale, being associated with substantial economic losses and having the potential for pandemics in humans and other host species. However, questions remain unanswered as to how AIV evolve in their natural reservoir of wild birds and in the non-natural host such as domestic birds. In this thesis, I have explored the evolution and dynamics of AIV by using the largest dataset of complete AIV genomes compiled to date. The complete structure of genetic diversity of AIV was described, the frequency and patterns of virus reassortment were quantified, and the associations between phylogeny and different traits including subtypes, host range and temporal–spatial distribution were further uncovered. In particular, I investigated the evolution of character traits along phylogenies. In this chapter, the main results of the thesis and the contribution to the understanding of viral phylogenetics are summarized. In addition, possible extensions of the current work that could be performed are suggested.

6.1 Chapter Summary

Compared to the limited number of HA, NA subtypes being observed in humans and other animals, a larger number of combinations among different HA and NA subtypes of influenza A viruses are detected in bird populations. Some HA-NA subtypes are rarely detected, indicating that some restrictions on possible combinations may exist. Thus, two important questions regarding to the general pattern of genetic interaction between different influenza A virus subtypes remain: whether certain subtypes are more likely to reassort than the other subtypes, and why this might be so. With the aim of exploring the rate of genetic exchange among viral subtypes during evolution, the evolution of internal segments of avian influenza A viruses were systematically analysed with a comprehensive avian dataset. A phylodynamic analysis was carried out upon the influenza A viruses circulating in avian populations (in both wild birds and domestic birds) in a 55 years time-span. In particular, I focused on investigating how AIV segments specific to the HA, NA and combined HA-NA subtypes evolve on the backbones of the six internal gene segments: I quantified the inter-subtype reassortment rates and tested possible ecological and evolutionary factors that are associated with such pattern of reassortment, including the relative proportion of each host species, the inferred evolutionary relaxed clock rates and the selective constraint acting on the genes. The results indicated that the overall reassortment rates for the external protein coding segments on internal protein coding segments were highly correlated with the internal segment genetic diversity (divergence time) and negatively correlated with dN/dS ratio and selective constraint. In addition, I found that different subtypes are characterized by reassortment rates that are strongly dependent on host type (wild and domestic birds). These results are consistent with the hypothesis that lack of immunological cross-protection at the subtype level in avian species allows frequent mixed infections in wild bird populations and provide novel insights into the evolutionary dynamics of avian influenza viruses. However, avian influenza surveillance efforts have been localized in places where outbreaks in poultry have occurred in Asia. Though surveillance activities in wild birds have increased since the emergence of HPAI H5N1, nucleotide sequence data from wild birds are still limited in comparison to those from North America. More sequences from this region need to

be generated to determine the role of host migration and other factors on influenza diffusion.

China plays a central role in maintaining a source population for influenza diversity globally (Claas *et al.*, 1998; Kilbourne, 2006; Lindstrom *et al.*, 2004). Two types of locations in China are considered as potential hotspots of influenza outbreaks and therefore of great importance. In Chapter 4, I described the pattern of spatial diffusion of diverse subtypes of avian influenza virus in bird populations and indicated the related risk factors across the diversity of economy, agriculture, ecological and production systems. To investigate the spatial diffusion pattern of avian influenza virus in China, the sampling locations were categorized into larger regions according to different meanings: one type is of high ecological complexity; another type of locations of great importance is highly developed in economy and trade, which also has been considered as a potential epicentre. I explored the key sources and destinations of the AIV special diffusion in China by performing a Bayesian stochastic search variable selection procedure to quantify the significant viral spatial transmission among regions classified by geographic meaning and economic meaning. The spatial diffusion of AIV isolated in China was explored and the determinants of spatial diffusion were tested using the six internal segments with an advanced discrete mapping methodology. I also evaluated the impacts of potential predictors on viral diffusion by integrating agriculture and ecology information with viral genetic data in the phylogeographic reconstruction. My results described the AIV diffusion pattern among geographic regions and economic zones in China. I found that AIV in China are mainly imported into the east coast areas; while they are mainly exported from the Southeast and central areas that agro-economically developed and also are likely exported from west regions along the migration routes. The results highlight the importance of the economic-agricultural impacts on the evolution and spatial dissemination of the influenza virus, and could be used to inform future strategies for the surveillance and control of influenza. In addition, policy makers in China may again be faced with a decision on whether or not to use vaccination in poultry to contain this disease so as to reduce the likelihood of exposure of humans to the H7N9 virus. Although there is no evidence so far that this virus will result in a human

pandemic, this outbreak reminds the importance for all countries to ensure the agriculture policy should be coherent with public health policy with the aim to reduce the risk of emergence of cross-host transmission from animal to human.

AIV of the H7 subtype constitutes a major threat to animal health in the whole world. H7N3 AIV is responsible for all lethal influenza outbreaks in poultry in the Americas over the past decade. There are two consequent cases of human infections of HPAI H7N3. A series of HPAI outbreaks which occurred in poultry in Mexico from June 2012 to August 2014 were caused by an H7N3 virus from an unknown source. In Chapter 5, the origin of this novel H7N3 outbreak influenza virus in Mexico since 2012 was explored using phylogenetic methodologies. In this study, the origin of this HPAI H7N3 AIV was explored by a combined analysis of the phylogenetic history of AIV with its host distribution and ecology, using joint Bayesian discrete phylogenetic models with sequence data from all gene segments. I discovered that the virus is a novel reassortant carried by wild waterfowl migrating from different flyways in North America during the time period studied. Importantly, I conclude that Mexico, and Central America in general, might be potential hotspots for AIV reassortment events – a possibility which to date has not attracted widespread attention. My results identify a significant threat of AIV arising from wild birds and indicate comprehensive surveillance in Mexico and Central America is highly desirable.

Early phylogenetic studies suggested that AIV sampled from wild birds in North America between 1998 to 2008 exhibit a strong spatially structured population, with relatively infrequent gene flow among localities, especially between those AIV that are spatially distant or belong to different flyways (Lam *et al.*, 2012; Munster *et al.*, 2007). However, the opposite was demonstrated recently in a study which emphasized that the long-term persistence of the influenza A virus gene pool in North American wild birds may be independent of migratory flyways, and the short-term evolutionary consequences of these ecological barriers may be rapidly erased by East-West virus migration (Bahl *et al.*, 2013). In Chapter 5, I confirmed the genetic interactions between flyways using a similar discrete trait model and extended these observations in the context of the Mexican H7N3 outbreak. Interestingly, we found

that gene flow from three flyways (Pacific, Central and Mississippi) generated the reassortants which acted as the predecessor of HPAI H7N3 in Mexico. It is possible that the reassortment events occurred in Mexico itself among passage migrants, which is an important source reason for the between-flyway gene flow to be documented. On the other hand, the under-sampling and lack of sequencing of influenza viruses from both wild and domestic birds in Central and South America means that these are potential regions in which surveillance could reveal previously unobserved influenza virus diversity. The extent to which North and South American avian influenza sequences cluster phylogenetically, and the role of migration in the spread of avian influenza, cannot be formally assessed without additional sequence data from Central and South America. Two recent studies (Gonzalez-Reiche & Perez, 2012; Scotch *et al.*, 2014) explored the evolution of AIV in South USA and South America using sequences which were not available at the time of my study. A study of AIV surveillances and phylogenetic studies in Latin America and the Caribbean, including Barbados, Guatemala, Argentina, Brazil, Mexico, and Peru (Gonzalez-Reiche & Perez, 2012), which found the AIV reports are sparse across the region and two distinct genetic lineages are co-circulating.

Results from phylogenetic analyses reflect bias in time and location where sampling has been conducted. In their study, sequences were isolated from 1994 to 2007, in which 22 H7N3 AIV isolated from Mexico in 2006 were described. However, only one sequence is available in pubic databases. Scotch et al (Scotch *et al.*, 2014) did a phylogeographic study using similar discrete trait models, they sequenced the HA gene of AIV strains of different subtypes isolated from south America countries (Arizona and New Mexico) from 2007 to 2009. By using discrete trait models and BSSVS methods (see Chapter 2 section 2.4), they found significant transmissions between states and identified significant routes both within and across flyways. This result is similar to my study in Chapter 5. They stated that the inter-flyway spread of influenza was probably the result of birds from four flyways co-mingling on breeding grounds in northern regions before moving south each fall, however, by jointly analysing 8 segments of North America AIV sequences, I found the inter-flyway reassortment events are more probably occurred in the narrow regions of Central

America which is the stopover and breeding region for the wild birds of the four flyways migration south in winter.

6.2 Future work

Newly developed phylogenetic methods for analysing viral evolution may facilitate future exploration of the evolutionary dynamics of avian influenza and other rapidly evolving RNA viruses. In this sectionsection, I will describe what is left unanswered from my thesis studies and provide suggestions for the possible way forward.

The discrete trait model is the key method being applied in Chapter 2 to 5. If the dates and locations of all phylogenetic nodes are known or posited, then each phylogeny branch represents a conditionally independent trajectory of viral movement, defined by a start location, end location, and duration. Recent Bayesian inference methods that consider phylogenetic diffusion of continuously distributed traits (latitude and longitude coordinates are used as spatial locations for viral samples) offer a unique opportunity to explore genotypic and phenotypic evolution in greater detail (Lemey *et al.*, 2010). A relaxed random walk (RRW) model allows a dispersal rate drawn independently on each phylogeny branch according to a discretized rate distribution. The heterogeneity in epidemic spread can be quantified by evaluating the variability of dispersal paths among phylogeny branches. Under a Bayesian statistical framework, phylogeographic diffusion in continuous space could be inferred using RRW while the time-scaled evolutionary history could be simultaneously reconstructed from molecular sequence data (Lemey *et al.*, 2010; Pybus *et al.*, 2012).

In Chapter 4, I used GLM model integrated in BEAST to evaluate the genetic and ecology predictors of viral spatial diffusion. By using GLM, we can also compare the discrete trait models in the phylogeographic sense with the widely used gravity model, which is originally inspired by Newton's Law of Gravitation:

$$F = G\frac{m1 \times m2}{R^2} \quad (6.1)$$

The magnitude of the attractive force F is equal to G (the gravitational constant, a number the size of which depends positively on the system of units used multiplied by the product of the mass (m1 and m2), and divided by the square of the distance R between the centers of the masses. Gravity models have been widely used to analyse patterns of trade in economics (Anderson & van Wincoop, 2003). In ecology, spatial transmission of directly transmitted infectious diseases is ultimately linked to movement by the host (donors and recipients) populations. The network of spatial spread may therefore be expected to be related to the transportation network within the hosts. Therefore, assuming a gravity transmission between host communities, a quantitative description of the topology of the epidemic network could be achieved by quantify the parameters of the gravity model (Xia *et al.*, 2004). Gravity models have a long history of use in describing and forecasting the movements of people as well as goods and services, making them a natural basis for disease transmission rates over distance. In agent-based micro-simulations, gravity models can be directly used to represent movement of individuals and hence disease. For example, a study examined the ability of synthetic networks considered a range of gravity models to test whether these models fit to movement data from the UK and the US. For UK data, epidemic behaviour on synthetic networks works well on the original data. In contrast, synthetic networks initially perform poor on US movement data, the networks were considerably improved by a assortative model based on node population size (Truscott & Ferguson, 2012). However, there are many limitations of the gravity law. For example, the gravity law has systematic predictive discrepancies; it is also unable to predict mobility in regions if the systematic traffic data, areas of major interest are missing in modelling of infectious diseases. A newly developed modelling framework overcomes the problems, specifically by replacing the actual distance with a network effective distance based upon host movement fluxes. And because the model relies only on the accurately estimated population densities, it can significantly improve the accuracy of predicting the transport patterns in a wide range of areas, even in those areas with poorly collected data (Simini *et al.*, 2012).

Sequences from effectively unlinked loci are rapidly becoming the norm in the era of next-generation sequencing. Increasing the number of loci improves

estimation of past population dynamics in terms of both bias and precision. Improvement in coalescent methods allows estimation of population trends from contemporary multi-locus DNA or RNA with increasing accuracy. The Extended Bayesian Skyline Plot (EBSP) model incorporates data from multiple unlinked genetic loci (Heled & Drummond, 2008). In addition, Bayesian Skygrid model generalizes the Gaussian Markov random field model for population dynamic inference. It provides a flexible demographic tree prior, allows the estimated trajectory to change at coalescent times (Gill *et al.*, 2013). For example, the past population dynamics for HIV-1 group M and non-pandemic group O were indicated by usingBayesian Skygrid. The results showed that the epidemic histories of group M and group O were similar until 1960, after which group M underwent an epidemiological transition and outpaced regional population growth (Faria *et al.*, 2014). In addition, a phytogeography study on South-Central Skunk Rabies Virus (SCSK) reconstructed the spatiotemporal dynamics of SCSK inferred from the Bayesian Skygrid model (Kuzmina *et al.*, 2013). In their study, the partitioned Skygrid model was selected for estimating TMRCA and substitution rates under the relaxed uncorrelated log-normal molecular clock. Their results indicated that different viral lineages circulate in their areas have limited evidence of geographic spread during decades. However, after a long period of stability the dispersal rate is increasing (Kuzmina *et al.*, 2013). In this thesis, I found similar pattern of diffusion among six internal segments in Chapter 3 and 4, though each of which have unique phylogenies. In Chapter 5, I used HPM model to link all eight segments of North America AIV to infer the overall evolution history of the novel HP outbreak strains in Mexico. I could alternatively use Skygrid to joint analyse different genes of AIV and try to describe a comprehensive spatio-temporal transmission networks.

Many wild bird species (mainly waterfowl) travel long distances between their breeding places and non-breeding areas in different time of year. For the North America flyway, the mean spring migration departure time is around March to April; the fall migration departure time is around September to October. As natural reservoirs for AI viruses, the regular seasonal migrations of these species can play an important role in the maintenance and spread of AIV. Migrating birds may exchange

viruses with other populations at staging, stopover or wintering sites (Gunnarsson *et al.*, 2012). In Chapter 5, the phylogeographic origin of HPAI H7N3 strains sampled in Mexico was described by using discrete trait model, and found the origin of the new virus was around winter season when wild birds in North America migrated to south. It is also possible to explore the phylodynamics of AIV by slicing the evolution into different timings of wild bird migration. To explore the temporal dynamics in evolutionary histories, an evolutionary approach called epochs can be adopted, which explicitly relaxes the time-homogeneity assumption by allowing the specification of different infinitesimal substitution rate matrices across different time intervals along the evolutionary history. A recently developed epoch model implementation in a Bayesian inference framework offers modelling flexibility in drawing inference about any discrete data type characterized as a continuous-time Markov chain, which is critical to accommodate discrete phylogeographic inference (Bielejec *et al.*, 2014). It has been shown that epoch model performs well in recovering evolutionary parameters from data generated according to different evolutionary scenarios: the selection dynamics on within-host HIV populations throughout infection and the seasonality of global influenza circulation. In the latter case, diffusion rates were estimated for the spring/summer epoch, and autumn/ winter epoch, respectively by applying BSSVS. The results suggested seasonal dynamics with spring/summer circulation to a large extent mirroring autumn/winter circulation. The epoch model captures key features of temporal heterogeneity that are difficult to test, thus is useful to detect changing selective dynamics in rapidly evolving viral populations as well as fluctuations in historical circulation dynamics (Bielejec *et al.*, 2014).

In addition, my study demonstrated the importance of undertaking surveillance in wild birds to characterise the influenza viruses carried by these birds. The genetic information obtained so far on the H7N9 outbreak virus sequences suggests that the H and N components of this virus were probably derived from wild birds, and also possibly from poultry (Lam *et al.*, 2013; Liu *et al.*, 2013). It is also evident from the genetic studies that the surveillance systems in place have not detected close relatives of the original host of these viruses and need to be strengthened. Bayesian

likelihood-based methods could be utilised to overcome missing data issues, so that the imputation of network properties can be carried out whilst simultaneously accounting for missing data attributes. Increased surveillance is needed to address the major determinants for AIV ecology, evolution, and transmission in the region. Previous phylogenetic studies also mentioned the importance of filling gaps in viral sampling in these regions. In the long run, influenza virus surveillance in animals plays a vital role for pandemic preparedness. And earlier detection of novel pandemic viruses once they have emerged in humans would provide longer lead times for vaccine development which is crucially important. The typical example is the 2009 pandemic virus, it was circulating in humans for about 4 months, and the predecessors were maintained in pig population for 10 years before being recognized (Poon *et al.*, 2010; Smith *et al.*, 2009b). A more recent example is the outbreaks of HPAI H5N8 in Europe in November 2014. Initial phylogenetic analysis indicated that wild migratory birds might be a source of virus introduction from Asia to Europe. However, as few AIV sequences in East Asia and Europe are available in flu database, the clear route of virus introduction remains unknown: Whether the novel HPAI is generated within Europe or it is introduced via bird migration? Or is it possible that the virus is circulating in wild birds or backyard poultries in Europe before infecting poultries in the outbreak areas? Therefore, it is crucial to continue to urge more data collection. Efforts have to be made to help countries to improve their surveillance and diagnose system of viruses in bird. It is expected that this will provide additional intelligence and warning of emerging problems so that appropriate action can be taken. The viruses isolated from surveillance can then be fully characterised and gene sequences can be uploaded to gene databases. Their contributions to the pool of data are essential for scientists trying to unravel the origin of novel viruses (Peiris *et al.*, 2012).

6.3 Final conclusion

In conclusion, I systematically analysed full genome AIV sequences obtained from different bird species around the world by using developed phylogenetic methods. In this study, I described the structure of genetic diversity of influenza A virus circulating in bird population and quantified the viral reassortment in the context of inter-subtypes, cross-host species and between-locations; I also uncovered the strong

associations between AIV evolution and virulence, host types, ecology and agro-economics impacts. I evaluated the linkages between bird flyway and possible routes of disease transmission and spread. These results may further enhance the understanding of the evolution pattern of AIV as well as the ecology and epidemiology impact of AIV.

Appendix

An electronic CD appendix is also provided containing the XMLs of the joint analysis in chapter 5; R code for classifying sequences and R code for the diffusion rate estimation.

Bibliography

Alexander, D. J. (2007). An overview of the epidemiology of avian influenza. *Vaccine* **25**, 5637-5644.

Anderson, J. E. & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *Am Econ Rev* **93**, 170-192.

Ayala, F. J. (1997). Vagaries of the molecular clock. *P Natl Acad Sci USA* **94**, 7776-7783.

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. & Suchard, M. A. (2012). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol* **61**, 170-173.

Baba, K., Shibata, R. & Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics* **46**, 657-664.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. & Alekseyenko, A. V. (2012). Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Molecular Biology and Evolution* **29**, 2157-2167.

Baele, G., Li, W. L., Drummond, A. J., Suchard, M. A. & Lemey, P. (2013). Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol* **30**, 239-243.

Bahl, J., Krauss, S., Kuhnert, D., Fourment, M., Raven, G., Pryor, S. P., Niles, L. J., Danner, A., Walker, D., Mendenhall, I. H., Su, Y. C., Dugan, V. G., Halpin, R. A., Stockwell, T. B., Webby, R. J., Wentworth, D. E., Drummond, A. J., Smith, G. J. & Webster, R. G. (2013). Influenza a virus migration and persistence in North American wild birds. *PLoS Pathog* **9**, e1003570.

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

Banks, J., Speidel, E. C., McCauley, J. W. & Alexander, D. J. (2000). Phylogenetic analysis of H7 haemagglutinin subtype influenza A viruses. *Arch Virol* **145**, 1047-1058.

Belser, J. A., Blixt, O., Chen, L. M., Pappas, C., Maines, T. R., Van Hoeven, N., Donis, R., Busch, J., McBride, R., Paulson, J. C., Katz, J. M. & Tumpey, T. M. (2008). Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc Natl Acad Sci U S A* **105**, 7558-7563.

Belser, J. A., Bridges, C. B., Katz, J. M. & Tumpey, T. M. (2009). Past, present, and possible future human infection with influenza virus A subtype H7. *Emerg Infect Dis* **15**, 859-865.

Belshe, R. B. (2005). The origins of pandemic influenza--lessons from the 1918 virus. *N Engl J Med* **353**, 2209-2211.

Berhane, Y., Hisanaga, T., Kehler, H., Neufeld, J., Manning, L., Argue, C., Handel, K., Hooper-McGrevy, K., Jonas, M., Robinson, J., Webster, R. G. & Pasick, J. (2009). Highly pathogenic avian influenza virus A (H7N3) in domestic poultry, Saskatchewan, Canada, 2007. *Emerg Infect Dis* **15**, 1492-1495.

Bielejec, F., Lemey, P., Baele, G., Rambaut, A. & Suchard, M. A. (2014). Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst Biol* **63**, 493-504.

Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**, 2910-2912.

Biswas, S. K. & Nayak, D. P. (1994). Mutational Analysis of the Conserved Motifs of Influenza-a Virus Polymerase Basic-Protein 1. *Journal of Virology* **68**, 1819-1826.

Blok, J. & Air, G. M. (1982). Variation in the membrane-insertion and "stalk" sequences in eight subtypes of influenza type A virus neuraminidase. *Biochemistry* **21**, 4001-4007.

Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R., de Jong, M. C. & van Boven, M. (2007). Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Comput Biol* **3**, e71.

Bourouiba, L. (2011). The Interaction of Migratory Birds and Domestic Poultry and Its Role in Sustaining Avian Influenza. *SIAM JAppl Math* **71(2)**, 487–516.

Brown, I. H. (2010). Summary of Avian Influenza Activity in Europe, Asia, and Africa, 2006-2009. *Avian Diseases* **54**, 187-193.

Campitelli, L., Mogavero, E., De Marco, M. A., Delogu, M., Puzelli, S., Frezza, F., Facchini, M., Chiapponi, C., Foni, E., Cordioli, P., Webby, R., Barigazzi, G., Webster, R. G. & Donatelli, I. (2004). Interspecies transmission of an H7N3 influenza virus from wild birds to intensively reared domestic poultry in Italy. *Virology* **323**, 24-36.

Capua, I. & Alexander, D. J. (2007). Avian influenza infections in birds--a moving target. *Influenza Other Respi Viruses* **1**, 11-18.

Capua, I. & Alexander, D. J. (2008). Ecology, Epidemiology and Human Health Implications of Avian Influenza Viruses: Why do We Need to Share Genetic Data? *Zoonoses and Public Health* **55**, 2-15.

Cardona, C., Yee, K. & Carpenter, T. (2009). Are live bird markets reservoirs of avian influenza? *Poultry Sci* **88**, 856-859.

Cauthen, A. N., Swayne, D. E., Schultz-Cherry, S., Perdue, M. L. & Suarez, D. L. (2000). Continued circulation in China of highly pathogenic avian influenza viruses encoding the hemagglutinin gene associated with the 1997 H5N1 outbreak in poultry and humans. *J Virol* **74**, 6592-6599.

Chen, H., Li, Y., Li, Z., Shi, J., Shinya, K., Deng, G., Qi, Q., Tian, G., Fan, S., Zhao, H., Sun, Y. & Kawaoka, Y. (2006a). Properties and dissemination of H5N1 viruses isolated during an influenza outbreak in migratory waterfowl in western China. *J Virol* **80**, 5976-5983.

Chen, H., Smith, G. J., 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., Wu, W. L., Lu, H. R., Chen, Y., Xia, N. S., Naipospos, T. S., Yuen, K. Y., Hassan, S. S., Bahri, S., Nguyen, T. D., Webster, R. G., Peiris, J. S. & Guan, Y. (2006b).

Establishment of multiple sublineages of H5N1 influenza virus in Asia: implications for pandemic control. *Proc Natl Acad Sci U S A* **103**, 2845-2850.

Chen, H., Smith, G. J., Zhang, S. Y., Qin, K., Wang, J., Li, K. S., Webster, R. G., Peiris, J. S. & Guan, Y. (2005). Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* **436**, 191-192.

Chen, J., Lee, K. H., Steinhauer, D. A., Stevens, D. J., Skehel, J. J. & Wiley, D. C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **95**, 409-417.

Chen, J., Wharton, S. A., Weissenhorn, W., Calder, L. J., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1995). A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in Escherichia coli into the low-pH-induced conformation. *Proc Natl Acad Sci U S A* **92**, 12205-12209.

Cheung, C. L., Vijaykrishna, D., Smith, G. J., Fan, X. H., Zhang, J. X., Bahl, J., Duan, L., Huang, K., Tai, H., Wang, J., Poon, L. L., Peiris, J. S., Chen, H. & Guan, Y. (2007). Establishment of influenza A virus (H6N1) in minor poultry species in southern China. *J Virol* **81**, 10402-10412.

Chitnis, N., Hyman, J. M. & Cushing, J. M. (2008). Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bull Math Biol* **70**, 1272-1296.

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

Cline, T. D., Karlsson, E. A., Freiden, P., Seufzer, B. J., Rehg, J. E., Webby, R. J. & Schultz-Cherry, S. (2011). Increased pathogenicity of a reassortant 2009 pandemic H1N1 influenza virus containing an H5N1 hemagglutinin. *J Virol* **85**, 12262-12270.

Collins, D. W. & Jukes, T. H. (1994). Rates of Transition and Transversion in Coding Sequences since the Human-Rodent Divergence. *Genomics* **20**, 386-396.

Cox, N. J. & Subbarao, K. (2000). Global epidemiology of influenza: past and present. *Annu Rev Med* **51**, 407-421.

Cui, L., Liu, D., Shi, W., Pan, J., Qi, X., Li, X., Guo, X., Zhou, M., Li, W., Li, J., Haywood, J., Xiao, H., Yu, X., Pu, X., Wu, Y., Yu, H., Zhao, K., Zhu, Y., Wu, B., Jin, T., Shi, Z., Tang, F., Zhu, F., Sun, Q., Wu, L., Yang, R., Yan, J., Lei, F., Zhu, B., Liu, W., Ma, J., Wang, H. & Gao, G. F. (2014). Dynamic reassortments and genetic heterogeneity of the human-infecting influenza A (H7N9) virus. *Nat Commun* **5**, 3142.

Cui, P., Hou, Y., Xing, Z., He, Y., Li, T., Guo, S., Luo, Z., Yan, B., Yin, Z. & Lei, F. (2011). Bird migration and risk for H5N1 transmission into Qinghai Lake, China. *Vector Borne Zoonotic Dis* **11**, 567-576.

Cybis, G. B., Sinsheimer, J. S., Lemey, P. & Suchard, M. A. (2013). Graph hierarchies for phylogeography. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120206.

Davison, S., Galligan, D., Eckert, T. E., Ziegler, A. F. & Eckroade, R. J. (1999). Economic analysis of an outbreak of avian influenza, 1997-1998. *J Am Vet Med Assoc* **214**, 1164-1167.

Deng, G., Tan, D., Shi, J., Cui, P., Jiang, Y., Liu, L., Tian, G., Kawaoka, Y., Li, C. & Chen, H. (2013). Complex reassortment of multiple subtypes of avian influenza viruses in domestic ducks at the Dongting Lake region of China. *J Virol*.

Dohna, H. Z., Li, J. L., Cardona, C. J., Miller, J. & Carpenter, T. E. (2009). Invasions by Eurasian Avian Influenza Virus H6 Genes and Replacement of Its North American Clade. *Emerging Infectious Diseases* **15**, 1040-1045.

Dong, G., Xu, C., Wang, C., Wu, B., Luo, J., Zhang, H., Nolte, D. L., Deliberto, T. J., Duan, M., Ji, G. & He, H. (2011). Reassortant H9N2 influenza viruses containing H5N1-like PB1 genes isolated from black-billed magpies in Southern China. *PLoS One* **6**, e25808.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307-1320.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185-1192.

Drummond, A. J. & Suchard, M. A. (2010). Bayesian random local clocks, or one rate to rule them all. *Bmc Biol* **8**.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973.

Duan, L., Campitelli, L., Fan, X. H., Leung, Y. H., Vijaykrishna, D., Zhang, J. X., Donatelli, I., Delogu, M., Li, K. S., Foni, E., Chiapponi, C., Wu, W. L., Kai, H., Webster, R. G., Shortridge, K. F., Peiris, J. S., Smith, G. J., Chen, H. & Guan, Y. (2007). Characterization of low-pathogenic H5 subtype influenza viruses from Eurasia: implications for the origin of highly pathogenic H5N1 viruses. *J Virol* **81**, 7529-7539.

Dugan, V. G., Chen, R., Spiro, D. J., Sengamalay, N., Zaborsky, J., Ghedin, E., Nolting, J., Swayne, D. E., Runstadler, J. A., Happ, G. M., Senne, D. A., Wang, R., Slemons, R. D., Holmes, E. C. & Taubenberger, J. K. (2008). The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog* **4**, e1000076.

Dugan, V. G., Dunham, E. J., Jin, G., Sheng, Z. M., Kaser, E., Nolting, J. M., Alexander, H. L., Jr., Slemons, R. D. & Taubenberger, J. K. (2011). Phylogenetic analysis of low pathogenicity H5N1 and H7N3 influenza A virus isolates recovered from sentinel, free flying, wild mallards at one study site during 2006. *Virology* **417**, 98-105.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Edo-Matas, D., Lemey, P., Tom, J. A., Serna-Bolea, C., van den Blink, A. E., van 't Wout, A. B., Schuitemaker, H. & Suchard, M. A. (2011). Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. *Mol Biol Evol* **28**, 1605-1616.

Edwards, C. J., Suchard, M. A., Lemey, P., Welch, J. J., Barnes, I., Fulton, T. L., Barnett, R., O'Connell, T. C., Coxon, P., Monaghan, N., Valdiosera, C. E., Lorenzen, E. D., Willerslev, E., Baryshnikov, G. F., Rambaut, A., Thomas, M. G., Bradley, D. G. & Shapiro, B. (2011). Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol* **21**, 1251-1258.

Edwards, W., Lindman, H. & Savage, L. J. (1963). Bayesian Statistical-Inference for Psychological-Research. *Psychol Rev* **70**, 193-242.

Efron, B., Halloran, E. & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *P Natl Acad Sci USA* **93**, 7085-7090.

Engelhardt, O. G. & Fodor, E. (2006). Functional association between viral and cellular transcription during influenza virus infection. *Rev Med Virol* **16**, 329-345.

Fang, L. Q., Li, X. L., Liu, K., Li, Y. J., Yao, H. W., Liang, S., Yang, Y., Feng, Z. J., Gray, G. C. & Cao, W. C. (2013). Mapping spread and risk of avian influenza A (H7N9) in China. *Sci Rep* **3**, 2722.

FAO (2007). Avian Influenza and Wild Birds: An introduction to applied field research and disease sampling techniques. In *FAO Animal Production and Health Manual*, p. 20: the Food and Agriculture Organization.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pepin, J., Posada, D., Peeters, M., Pybus, O. G. & Lemey, P. (2014). HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56-61.

Faria, N. R., Suchard, M. A., Rambaut, A. & Lemey, P. (2011). Toward a quantitative understanding of viral phylogeography. *Curr Opin Virol* **1**, 423-429.

Faria, N. R., Suchard, M. A., Rambaut, A., Streicker, D. G. & Lemey, P. (2013). Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120196.

Fauci, A. S. (2006). Pandemic influenza threat and preparedness. *Emerg Infect Dis* **12**, 73-77.

Feare, C. J. (2010). Role of wild birds in the spread of highly pathogenic avian influenza virus H5N1 and implications for global surveillance. *Avian Dis* **54**, 201-212.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-376.

Felsenstein, J. (1985). Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution* **39**, 783-791.

Fodor, E., Crow, M., Mingay, L. J., Deng, T., Sharps, J., Fechter, P. & Brownlee, G. G. (2002). A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J Virol* **76**, 8989-9001.

Fukami-Kobayashi, K. & Tateno, Y. (1991). Robustness of maximum likelihood tree estimation against different patterns of base substitutions. *J Mol Evol* **32**, 79-91.

Fuller, T. L., Gilbert, M., Martin, V., Cappelle, J., Hosseini, P., Njabo, K. Y., Abdel Aziz, S., Xiao, X., Daszak, P. & Smith, T. B. (2013). Predicting hotspots for influenza virus reassortment. *Emerg Infect Dis* **19**, 581-588.

Gaidet, N., Cappelle, J., Takekawa, J. Y., Iverson, S. A., Perry, W. M., Douglas, D. C., Prosser, D. J., Mundkur, T. & Newman, S. H. (2011). Potential Spread of Highly Pathogenic

Avian Influenza H5N1 by Wildfowl: Dispersal Ranges and Rates Determined from Large-scale Satellite Telemetry. *Ecohealth* **7**, S35-S35.

Gaidet, N., Cappelle, J., Takekawa, J. Y., Prosser, D. J., Iverson, S. A., Douglas, D. C., Perry, W. M., Mundkur, T. & Newman, S. H. (2010). Potential spread of highly pathogenic avian influenza H5N1 by wildfowl: dispersal ranges and rates determined from large-scale satellite telemetry. *J Appl Ecol* **47**, 1147-1157.

Gaidet, N., Cattoli, G., Hammoumi, S., Newman, S. H., Hagemeijer, W., Takekawa, J. Y., Cappelle, J., Dodman, T., Joannis, T., Gil, P., Monne, I., Fusaro, A., Capua, I., Manu, S., Micheloni, P., Ottosson, U., Mshelbwala, J. H., Lubroth, J., Domenech, J. & Monicat, F. (2008). Evidence of infection by H5N2 highly pathogenic avian influenza viruses in healthy wild waterfowl. *Plos Pathogens* **4**.

Ge, F. F., Zhou, J. P., Liu, J., Wang, J., Zhang, W. Y., Sheng, L. P., Xu, F., Ju, H. B., Sun, Q. Y. & Liu, P. H. (2009). Genetic Evolution of H9 Subtype Influenza Viruses from Live Poultry Markets in Shanghai, China. *Journal of Clinical Microbiology* **47**, 3294-3300.

Gerloff, N. A., Khan, S. U., Balish, A., Shanta, I. S., Simpson, N., Berman, L., Haider, N., Poh, M. K., Islam, A., Gurley, E., Hasnat, M. A., Dey, T., Shu, B., Emery, S., Lindstrom, S., Haque, A., Klimov, A., Villanueva, J., Rahman, M., Azziz-Baumgartner, E., Rahman, M. Z., Luby, S. P., Zeidner, N., Donis, R. O., Sturm-Ramirez, K. & Davis, C. T. (2014). Multiple reassortment events among highly pathogenic avian influenza A(H5N1) viruses detected in Bangladesh. *Virology* **450**, 297-307.

Ghedin, E., Sengamalay, N. A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D. J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y., St George, K., Taylor, J., Lipman, D. J., Fraser, C. M., Taubenberger, J. K. & Salzberg, S. L. (2005). Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162-1166.

Gilbert, M., Newman, S. H., Takekawa, J. Y., Loth, L., Biradar, C., Prosser, D. J., Balachandran, S., Subba Rao, M. V., Mundkur, T., Yan, B., Xing, Z., Hou, Y., Batbayar, N., Natsagdorj, T., Hogerwerf, L., Slingenbergh, J. & Xiao, X. (2010). Flying over an infected landscape: distribution of highly pathogenic avian influenza H5N1 risk in South Asia and satellite tracking of wild waterfowl. *Ecohealth* **7**, 448-458.

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B. & Suchard, M. A. (2013). Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci. *Molecular Biology and Evolution* **30**, 713-724.

Girard, Y. A., Runstadler, J. A., Aldehoff, F. & Boyce, W. (2012). Genetic structure of Pacific Flyway avian influenza viruses is shaped by geographic location, host species, and sampling period. *Virus Genes* **44**, 415-428.

Goldberg, E. E. & Igic, B. (2008). On phylogenetic tests of irreversible evolution. *Evolution* **62**, 2727-2741.

Goldman, N. & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**, 725-736.

Gonzalez-Reiche, A. S. & Perez, D. R. (2012). Where Do Avian Influenza Viruses Meet in the Americas? *Avian Diseases* **56**, 1025-1033.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.

Greenbaum, B. D., Li, O. T., Poon, L. L., Levine, A. J. & Rabadan, R. (2012). Viral reassortment as an information exchange between viral segments. *Proc Natl Acad Sci U S A* **109**, 3341-3346.

Gu, M., Huang, J. Q., Chen, Y. X., Chen, J., Wang, X. Q., Liu, X. W. & Liu, X. F. (2012). Genome Sequence of a Natural Reassortant H5N2 Avian Influenza Virus from Domestic Mallard Ducks in Eastern China. *Journal of Virology* **86**, 12463-12464.

Gu, M., Liu, W., Cao, Y., Peng, D., Wang, X., Wan, H., Zhao, G., Xu, Q., Zhang, W., Song, Q., Li, Y. & Liu, X. (2011). Novel reassortant highly pathogenic avian influenza (H5N5) viruses in domestic ducks, China. *Emerg Infect Dis* **17**, 1060-1063.

Guan, Y., Peiris, J. S., Lipatov, A. S., Ellis, T. M., Dyrting, K. C., Krauss, S., Zhang, L. J., Webster, R. G. & Shortridge, K. F. (2002). Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc Natl Acad Sci U S A* **99**, 8950-8955.

Guan, Y., Poon, L. L., Cheung, C. Y., Ellis, T. M., Lim, W., Lipatov, A. S., Chan, K. H., Sturm-Ramirez, K. M., Cheung, C. L., Leung, Y. H., Yuen, K. Y., Webster, R. G. & Peiris, J. S. (2004). H5N1 influenza: a protean pandemic threat. *Proc Natl Acad Sci U S A* **101**, 8156-8161.

Guan, Y., Shortridge, K. F., Krauss, S. & Webster, R. G. (1999). Molecular characterization of H9N2 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? *Proc Natl Acad Sci U S A* **96**, 9363-9367.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W., Ortin, J., Hart, D. J. & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**, 500-506.

Gunnarsson, G., Latorre-Margalef, N., Hobson, K. A., Van Wilgenburg, S. L., Elmberg, J., Olsen, B., Fouchier, R. A. & Waldenstrom, J. (2012). Disease dynamics and bird migration--linking mallards Anas platyrhynchos and subtype diversity of the influenza A virus in time and space. *PLoS One* **7**, e35679.

Hall, B. G. (2013). Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution* **30**, 1229-1235.

Hasegawa, M. & Kishino, H. (1989). Heterogeneity of Tempo and Mode of Mitochondrial-DNA Evolution among Mammalian Orders. *Jpn J Genet* **64**, 243-258.

Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160-174.

Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97-&.

Heled, J. & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol Biol* **8**, 289.

Hinshaw, V. S., Bean, W. J., Geraci, J., Fiorelli, P., Early, G. & Webster, R. G. (1986). Characterization of two influenza A viruses from a pilot whale. *J Virol* **58**, 655-656.

Hinshaw, V. S., Bean, W. J., Webster, R. G., Rehg, J. E., Fiorelli, P., Early, G., Geraci, J. R. & Staubin, D. J. (1984). Are Seals Frequently Infected with Avian Influenza-Viruses. *Journal of Virology* **51**, 863-865.

Hirst, M., Astell, C. R., Griffith, M., Coughlin, S. M., Moksa, M., Zeng, T., Smailus, D. E., Holt, R. A., Jones, S., Marra, M. A., Petric, M., Krajden, M., Lawrence, D., Mak, A., Chow, R., Skowronski, D. M., Tweed, S. A., Goh, S., Brunham, R. C., Robinson, J., Bowes, V., Sojonky, K., Byrne, S. K., Li, Y., Kobasa, D., Booth, T. & Paetzel, M. (2004). Novel avian influenza H7N3 strain outbreak, British Columbia. *Emerg Infect Dis* **10**, 2192-2195.

Holder, M. & Lewis, P. O. (2003). Phylogeny estimation: Traditional and Bayesian approaches. *Nat Rev Genet* **4**, 275-284.

Hollich, V., Milchert, L., Arvestad, L. & Sonnhammer, E. L. (2005). Assessment of protein distance measures and tree-building methods for phylogenetic tree reconstruction. *Mol Biol Evol* **22**, 2257-2264.

Holmes, E. C., Ghedin, E., Miller, N., Taylor, J., Bao, Y. M., St George, K., Grenfell, B. T., Salzberg, S. L., Fraser, C. M., Lipman, D. J. & Taubenberger, J. K. (2005). Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *Plos Biology* **3**, 1579-1589.

Holmes, E. C. & Grenfell, B. T. (2009). Discovering the Phylodynamics of RNA Viruses. *Plos Computational Biology* **5**.

Horimoto, T. & Kawaoka, Y. (2005). Influenza: lessons from past pandemics, warnings from current incidents. *Nat Rev Microbiol* **3**, 591-600.

Horimoto, T., Rivera, E., Pearson, J., Senne, D., Krauss, S., Kawaoka, Y. & Webster, R. G. (1995a). Origin and Molecular-Changes Associated with Emergence of a Highly Pathogenic H5n2 Influenza-Virus in Mexico. *Virology* **213**, 223-230.

Horimoto, T., Rivera, E., Pearson, J., Senne, D., Krauss, S., Kawaoka, Y. & Webster, R. G. (1995b). Origin and molecular changes associated with emergence of a highly pathogenic H5N2 influenza virus in Mexico. *Virology* **213**, 223-230.

Hossain, M. J., Hickman, D. & Perez, D. R. (2008). Evidence of expanded host range and mammalian-associated genetic changes in a duck H9N2 influenza virus following adaptation in quail and chickens. *PLoS One* **3**, e3170.

Hu, X., Liu, D., Wang, M., Yang, L., Zhu, Q., Li, L. & Gao, G. F. (2011). Clade 2.3.2 avian influenza virus (H5N1), Qinghai Lake region, China, 2009-2010. *Emerg Infect Dis* **17**, 560-562.

Huang, K., Bahl, J., Fan, X. H., Vijaykrishna, D., Cheung, C. L., Webby, R. J., Webster, R. G., Chen, H., Smith, G. J., Peiris, J. S. & Guan, Y. (2010). Establishment of an H6N2 influenza virus lineage in domestic ducks in southern China. *J Virol* **84**, 6978-6986.

Huang, K., Zhu, H., Fan, X., Wang, J., Cheung, C. L., Duan, L., Hong, W., Liu, Y., Li, L., Smith, D. K., Chen, H., Webster, R. G., Webby, R. J., Peiris, M. & Guan, Y. (2012). Establishment and lineage replacement of H6 influenza viruses in domestic ducks in southern China. *J Virol* **86**, 6075-6083.

Ito, T., Suzuki, Y., Mitnaul, L., Vines, A., Kida, H. & Kawaoka, Y. (1997). Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. *Virology* **227**, 493-499.

Jackson, S., Van Hoeven, N., Chen, L. M., Maines, T. R., Cox, N. J., Katz, J. M. & Donis, R. O. (2009). Reassortment between avian H5N1 and human H3N2 influenza viruses in ferrets: a public health risk assessment. *J Virol* **83**, 8131-8140.

Jiao, P. R., Wei, L. M., Yuan, R. Y., Gong, L., Cao, L., Song, Y. F., Luo, K. J., Ren, T. & Liao, M. (2012). Complete Genome Sequence of an H5N2 Avian Influenza Virus Isolated from a Parrot in Southern China. *Journal of Virology* **86**, 8890-8891.

Jukes, T. & Cantor, C. (1969). *Evolution of Protein Molecules*. New York: New York: Academic Press.

Kalthoff, D., Globig, A. & Beer, M. (2010). (Highly pathogenic) avian influenza as a zoonotic agent. *Vet Microbiol* **140**, 237-245.

Kandeel, A., Manoncourt, S., Abd el Kareem, E., Mohamed Ahmed, A. N., El-Refaie, S., Essmat, H., Tjaden, J., de Mattos, C. C., Earhart, K. C., Marfin, A. A. & El-Sayed, N. (2010). Zoonotic transmission of avian influenza virus (H5N1), Egypt, 2006-2009. *Emerg Infect Dis* **16**, 1101-1107.

Kapczynski, D. R., Pantin-Jackwood, M., Guzman, S. G., Ricardez, Y., Spackman, E., Bertran, K., Suarez, D. L. & Swayne, D. E. (2013). Characterization of the 2012 highly pathogenic avian influenza H7N3 virus isolated from poultry in an outbreak in Mexico: pathobiology and vaccine protection. *J Virol* **87**, 9086-9096.

Kash, J. C., Tumpey, T. M., Proll, S. C., Carter, V., Perwitasari, O., Thomas, M. J., Basler, C. F., Palese, P., Taubenberger, J. K., Garcia-Sastre, A., Swayne, D. E. & Katze, M. G. (2006). Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* **443**, 578-581.

Kass, R. E., Carlin, B. P., Gelman, A. & Neal, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *Am Stat* **52**, 93-100.

Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *J Am Stat Assoc* **90**, 773-795.

Kawaoka, Y., Naeve, C. W. & Webster, R. G. (1984). Is virulence of H5N2 influenza viruses in chickens associated with loss of carbohydrate from the hemagglutinin? *Virology* **139**, 303-316.

Kayali, G., Setterquist, S. F., Capuano, A. W., Myers, K. P., Gill, J. S. & Gray, G. C. (2008). Testing human sera for antibodies against avian influenza viruses: Horse RBC hemagglutination inhibition vs. microneutralization assays. *J Clin Virol* **43**, 73-78.

Kida, H., Yanagawa, R. & Matsuoka, Y. (1980). Duck influenza lacking evidence of disease signs and immune response. *Infect Immun* **30**, 547-553.

Kilbourne, E. D. (2006). Influenza pandemics of the 20th century. *Emerging Infectious Diseases* **12**, 9-14.

Kilpatrick, A. M., Chmura, A. A., Gibbons, D. W., Fleischer, R. C., Marra, P. P. & Daszak, P. (2006). Predicting the global spread of H5N1 avian influenza. *Proc Natl Acad Sci U S A* **103**, 19368-19373.

Kim, H. R., Park, C. K., Oem, J. K., Bae, Y. C., Choi, J. G., Lee, O. S. & Lee, Y. J. (2010). Characterization of H5N2 influenza viruses isolated in South Korea and their influence on the emergence of a novel H9N2 influenza virus. *Journal of General Virology* **91**, 1978-1983.

Kimura, M. (1980). A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide-Sequences. *Journal of Molecular Evolution* **16**, 111-120.

Kingman, J. F. (1982). The coalescent. *Stochastic Processes and their Applications* **13**, 235-248.

Kishino, H., Thorne, J. L. & Bruno, W. J. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* **18**, 352-361.

Kitakado, T., Kitada, S., Kishino, H. & Skaug, H. J. (2006). An integrated-likelihood method for estimating genetic differentiation between populations. *Genetics* **173**, 2073-2082.

Klenk, H. D. & Garten, W. (1994). Host cell proteases controlling virus pathogenicity. *Trends Microbiol* **2**, 39-43.

Kosakovsky Pond, S. L. & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-1222.

Krauss, S., Obert, C. A., Franks, J., Walker, D., Jones, K., Seiler, P., Niles, L., Pryor, S. P., Obenauer, J. C., Naeve, C. W., Widjaja, L., Webby, R. J. & Webster, R. G. (2007). Influenza in migratory birds and evidence of limited intercontinental virus exchange. *PLoS Pathog* **3**, e167.

Krauss, S. & Webster, R. G. (2012). Predicting the next influenza virus. *Science* **337**, 644.

Kuhner, M. K., Yamato, J. & Felsenstein, J. (1998). Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429-434.

Kuzmina, N. A., Lemey, P., Kuzmin, I. V., Mayes, B. C., Ellison, J. A., Orciari, L. A., Hightower, D., Taylor, S. T. & Rupprecht, C. E. (2013). The phylogeography and spatiotemporal spread of south-central skunk rabies virus. *PLoS One* **8**, e82348.

Labadie, K., Dos Santos Afonso, E., Rameix-Welti, M. A., van der Werf, S. & Naffakh, N. (2007). Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. *Virology* **362**, 271-282.

Lam, T. T., 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. & Holmes, E. C. (2012). Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among North American birds. *Ecol Lett* **15**, 24-33.

Lam, T. T., Wang, J., Shen, Y., Zhou, B., Duan, L., Cheung, C. L., Ma, C., Lycett, S. J., Leung, C. Y., Chen, X., Li, L., Hong, W., Chai, Y., Zhou, L., Liang, H., Ou, Z., Liu, Y., Farooqui, A., Kelvin, D. J., Poon, L. L., Smith, D. K., Pybus, O. G., Leung, G. M., Shu, Y., Webster, R. G., Webby, R. J., Peiris, J. S., Rambaut, A., Zhu, H. & Guan, Y. (2013). The genesis and source of the H7N9 influenza viruses causing human infections in China. *Nature* **502**, 241-244.

Lanave, C., Preparata, G., Saccone, C. & Serio, G. (1984). A New Method for Calculating Evolutionary Substitution Rates. *Journal of Molecular Evolution* **20**, 86-93.

Lebarbenchon, C., Feare, C. J., Renaud, F., Thomas, F. & Gauthier-Clerc, M. (2010). Persistence of highly pathogenic avian influenza viruses in natural ecosystems. *Emerg Infect Dis* **16**, 1057-1062.

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. & Lee, H. S. (2014). Novel reassortant influenza A(H5N8) viruses, South Korea, 2014. *Emerg Infect Dis* **20**, 1087-1089.

Lei, F. & Shi, W. (2011). Prospective of Genomics in Revealing Transmission, Reassortment and Evolution of Wildlife-Borne Avian Influenza A (H5N1) Viruses. *Curr Genomics* **12**, 466-474.

Lemey, P., Minin, V. N., Bielejec, F., Kosakovsky Pond, S. L. & Suchard, M. A. (2012). A counting renaissance: Combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. *Bioinformatics*.

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D. & Suchard, M. A. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog* **10**, e1003932.

Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput Biol* **5**, e1000520.

Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* **27**, 1877-1885.

Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. (2007). A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* **24**, 2669-2680.

Li, C., Yu, K., Tian, G., Yu, D., Liu, L., Jing, B., Ping, J. & Chen, H. (2005). Evolution of H9N2 influenza viruses from domestic poultry in Mainland China. *Virology* **340**, 70-83.

Li, K. S., Guan, Y., Wang, J., Smith, G. J., Xu, K. M., Duan, L., Rahardjo, A. P., Puthavathana, P., Buranathai, C., Nguyen, T. D., Estoepangestie, A. T., Chaisingh, A., Auewarakul, P., Long, H. T., Hanh, N. T., Webby, R. J., Poon, L. L., Chen, H., Shortridge, K. F., Yuen, K. Y., Webster, R. G. & Peiris, J. S. (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209-213.

Li, X., Liu, X., Xu, L. & Zhang, Z. (2009). Spatial transmission of avian influenza (type H5) in birds. *Integr Zool* **4**, 418-425.

Lindstrom, S. E., Cox, N. J. & Klimov, A. (2004). Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101-119.

Liu, D., Shi, W., Shi, Y., Wang, D., Xiao, H., Li, W., Bi, Y., Wu, Y., Li, X., Yan, J., Liu, W., Zhao, G., Yang, W., Wang, Y., Ma, J., Shu, Y., Lei, F. & Gao, G. F. (2013). Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet* **381**, 1926-1932.

Liu, M., He, S. Q., Walker, D., Zhou, N. N., Perez, D. R., Mo, B., Li, F., Huang, X. T., Webster, R. G. & Webby, R. J. (2003). The influenza virus gene pool in a poultry market in South Central China. *Virology* **305**, 267-275.

Liu, W. Y. (2012). *China Ariculture Yearbook*. Beijing: China Agriculture Press.

Liu, Z., Xu, B., Chen, Q. & Chen, Z. (2014). Complete Genome Sequence of a Mixed-Subtype (H5N1 and H6N6) Avian Influenza Virus Isolated from a Duck in Hunan Province, China. *Genome Announc* **2**.

Lopez-Martinez, I., Balish, A., Barrera-Badillo, G., Jones, J., Nunez-Garcia, T. E., Jang, Y., Aparicio-Antonio, R., Azziz-Baumgartner, E., Belser, J. A., Ramirez-Gonzalez, J. E., Pedersen, J. C., Ortiz-Alcantara, J., Gonzalez-Duran, E., Shu, B., Emery, S. L., Poh, M. K., Reyes-Teran, G., Vazquez-Perez, J. A., Avila-Rios, S., Uyeki, T., Lindstrom, S., Villanueva, J., Tokars, J., Ruiz-Matus, C., Gonzalez-Roldan, J. F., Schmitt, B., Klimov,

A., Cox, N., Kuri-Morales, P., Davis, C. T. & Diaz-Quinonez, J. A. (2013). Highly pathogenic avian influenza A(H7N3) virus in poultry workers, Mexico, 2012. *Emerg Infect Dis* **19**, 1531-1534.

Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *Plos Pathogens* **3**, 1470-1476.

Lu, L., Lycett, S. J. & Leigh Brown, A. J. (2014). Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evol Biol* **14**, 16.

Ludwig, S., Schultz, U., Mandler, J., Fitch, W. M. & Scholtissek, C. (1991). Phylogenetic relationship of the nonstructural (NS) genes of influenza A viruses. *Virology* **183**, 566-577.

Lycett, S. J., Baillie, G., Coulter, E., Bhatt, S., Kellam, P., McCauley, J. W., Wood, J. L., Brown, I. H., Pybus, O. G. & Leigh Brown, A. J. (2012). Estimating reassortment rates in co-circulating Eurasian swine influenza viruses. *J Gen Virol* **93**, 2326-2336.

Lycett, S. J., Ward, M. J., Lewis, F. I., Poon, A. F. Y., Kosakovsky Pond, S. L. & Brown, A. J. L. (2009). Detection of Mammalian Virulence Determinants in Highly Pathogenic Avian Influenza H5N1 Viruses: Multivariate Analysis of Published Data. *Journal of Virology* **83**, 9901-9910.

Marinova-Petkova A., M. M. F., SM Rabiul Alam, M Kamrul H., Sharmin A., Lisa J.E., David W., Laura M., Adam Rubrum, John Franks, Patrick S., Trushar J., Pamela M., Scott K., Webby, R. G.& Webster, R. G. (2014). Multiple introductions of highly pathogenic avian influenza H5N1 viruses into

Bangladesh. *Emerging Microbes & Infections* **3**, 1-14.

Martin, V., Pfeiffer, D. U., Zhou, X., Xiao, X., Prosser, D. J., Guo, F. & Gilbert, M. (2011). Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog* **7**, e1001308.

Maurer-Stroh, S., Lee, R. T., Gunalan, V. & Eisenhaber, F. (2013). The highly pathogenic H7N3 avian influenza strain from July 2012 in Mexico acquired an extended cleavage site through recombination with host 28S rRNA. *Virol J* **10**, 139.

McHardy, A. C. & Adams, B. (2009). The role of genomics in tracking the evolution of influenza A virus. *PLoS Pathog* **5**, e1000566.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J Chem Phys* **21**, 1087-1092.

Meyn, S. P. & Tweedie, R. L. (1993). *Markov chains and stochastic stability*. London ; New York: Springer-Verlag.

Minin, V. N., Bloomquist, E. W. & Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* **25**, 1459-1471.

Minin, V. N. & Suchard, M. A. (2008a). Counting labeled transitions in continuous-time Markov models of evolution. *J Math Biol* **56**, 391-412.

Minin, V. N. & Suchard, M. A. (2008b). Fast, accurate and simulation-free stochastic mapping. *Philos T R Soc B* **363**, 3985-3995.

Munster, V. J., Baas, C., Lexmond, P., Waldenstrom, J., Wallensten, A., Fransson, T., Rimmelzwaan, G. F., Beyer, W. E., Schutten, M., Olsen, B., Osterhaus, A. D. & Fouchier, R. A. (2007). Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog* **3**, e61.

Murhekar, M., Arima, Y., Horby, P., Vandemaele, K. A., Vong, S., Zijian, F., Lee, C. K. & Li, A. (2013). Avian influenza A(H7N9) and the closure of live bird markets. *Western Pac Surveill Response J* **4**, 4-7.

Naffakh, N., Massin, P., Escriou, N., Crescenzo-Chaigne, B. & van der Werf, S. (2000). Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *Journal of General Virology* **81**, 1283-1291.

Naffakh, N., Tomoiu, A., Rameix-Welti, M. A. & van der Werf, S. (2008). Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu Rev Microbiol* **62**, 403-424.

Nagarajan, N. & Kingsford, C. (2011). GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res* **39**, e34.

Neumann, G., Noda, T. & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931-939.

Newman, S. H., Hill, N. J., Spragens, K. A., Janies, D., Voronkin, I. O., Prosser, D. J., Yan, B., Lei, F., Batbayar, N., Natsagdorj, T., Bishop, C. M., Butler, P. J., Wikelski, M., Balachandran, S., Mundkur, T., Douglas, D. C. & Takekawa, J. Y. (2012). Eco-virological approach for assessing the role of wild birds in the spread of avian influenza H5N1 along the Central Asian Flyway. *PLoS One* **7**, e30636.

Newton, M. A. & Raftery, A. E. (1994). Approximate Bayesian-Inference with the Weighted Likelihood Bootstrap. *J R Stat Soc B* **56**, 3-48.

Nidom, C. A., Takano, R., Yamada, S., Sakai-Tagawa, Y., Daulay, S., Aswadi, D., Suzuki, T., Suzuki, Y., Shinya, K., Iwatsuki-Horimoto, K., Muramoto, Y. & Kawaoka, Y. (2010). Influenza A (H5N1) viruses from pigs, Indonesia. *Emerg Infect Dis* **16**, 1515-1523.

O'Brien, J. D., Minin, V. N. & Suchard, M. A. (2009). Learning to count: robust estimates for labeled distances between molecular sequences. *Mol Biol Evol* **26**, 801-814.

Octaviani, C. P., Ozawa, M., Yamada, S., Goto, H. & Kawaoka, Y. (2010). High level of genetic compatibility between swine-origin H1N1 and highly pathogenic avian H5N1 influenza viruses. *J Virol* **84**, 10918-10922.

Ogata, T., Yamazaki, Y., Okabe, N., Nakamura, Y., Tashiro, M., Nagata, N., Itamura, S., Yasui, Y., Nakashima, K., Doi, M., Izumi, Y., Fujieda, T., Yamato, S. & Kawada, Y. (2008). Human H5N2 avian influenza infection in Japan and the factors associated with high H5N2-neutralizing antibody titer. *J Epidemiol* **18**, 160-166.

Okada, M. (1974). [Pattern recognition of renograms with the "Learning Machine"]. *Iyodenshi To Seitai Kogaku* **12**, 60-68.

Olsen, B., Munster, V. J., Wallensten, A., Waldenstrom, J., Osterhaus, A. D. & Fouchier, R. A. (2006). Global patterns of influenza a virus in wild birds. *Science* **312**, 384-388.

Oner, A. F., Bay, A., Arslan, S., Akdeniz, H., Sahin, H. A., Cesur, Y., Epcacan, S., Yilmaz, N., Deger, I., Kizilyildiz, B., Karsen, H. & Ceyhan, M. (2006). Avian influenza A (H5N1) infection in eastern Turkey in 2006. *N Engl J Med* **355**, 2179-2185.

Pagel, M., Meade, A. & Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* **53**, 673-684.

Parker, J., Rambaut, A. & Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* **8**, 239-246.

Parrish, C. R. & Kawaoka, Y. (2005). The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. *Annu Rev Microbiol* **59**, 553-586.

Pasick, J., Berhane, Y. & Hooper-McGrevy, K. (2009). Avian influenza: the Canadian experience. *Rev Sci Tech* **28**, 349-358.

Pavade, G., Awada, L., Hamilton, K. & Swayne, D. E. (2011). The influence of economic indicators, poultry density and the performance of veterinary services on the control of high-pathogenicity avian influenza in poultry. *Rev Sci Tech* **30**, 661-671.

Pearson, J. E. (2003). International standards for the control of avian influenza. *Avian Dis* **47**, 972-975.

Peiris, J. S., Poon, L. L. & Guan, Y. (2012). Public health. Surveillance of animal influenza for pandemic preparedness. *Science* **335**, 1173-1174.

Peiris, M., Yuen, K. Y., Leung, C. W., Chan, K. H., Ip, P. L., Lai, R. W., Orr, W. K. & Shortridge, K. F. (1999). Human infection with influenza H9N2. *Lancet* **354**, 916-917.

Perez, D. R., Dugan, V. G., Chen, R., Spiro, D. J., Sengamalay, N., Zaborsky, J., Ghedin, E., Nolting, J., Swayne, D. E., Runstadler, J. A., Happ, G. M., Senne, D. A., Wang, R., Slemons, R. D., Holmes, E. C. & Taubenberger, J. K. (2008). The Evolutionary Genetics and Emergence of Avian Influenza Viruses in Wild Birds. *PLoS Pathogens* **4**, e1000076.

Perk, S., Banet-Noach, C., Golender, N., Simanov, L., Rozenblut, E., Nagar, S., Pokamunski, S., Pirak, M., Tendler, Y., Garcia, M. & Panshin, A. (2007). Molecular characterization of the glycoprotein genes of H5N1 influenza A viruses isolated in Israel and the Gaza Strip during 2006 outbreaks. *Virus Genes* **35**, 497-502.

Plague, F. (2010). Highly Pathogenic Avian Influenza. *OIE*, 1-18.

Poon, L. L., Mak, P. W., Li, O. T., Chan, K. H., Cheung, C. L., Ma, E. S., Yen, H. L., Vijaykrishna, D., Guan, Y. & Peiris, J. S. (2010). Rapid detection of reassortment of pandemic H1N1/2009 influenza virus. *Clin Chem* **56**, 1340-1344.

Prosser, D. J., Cui, P., Takekawa, J. Y., Tang, M., Hou, Y., Collins, B. M., Yan, B., Hill, N. J., Li, T., Li, Y., Lei, F., Guo, S., Xing, Z., He, Y., Zhou, Y., Douglas, D. C., Perry, W. M. & Newman, S. H. (2011a). Wild bird migration across the Qinghai-Tibetan plateau: a transmission route for highly pathogenic H5N1. *PLoS One* **6**, e17622.

Prosser, D. J., Cui, P., Takekawa, J. Y., Tang, M. J., Hou, Y. S., Collins, B. M., Yan, B. P., Hill, N. J., Li, T. X., Li, Y. D., Lei, F. M., Guo, S., Xing, Z., He, Y. B., Zhou, Y. C., Douglas, D. C., Perry, W. M. & Newman, S. H. (2011b). Wild Bird Migration across the Qinghai-Tibetan Plateau: A Transmission Route for Highly Pathogenic H5N1. *PLoS One* **6**.

Prosser, D. J., Wu, J., Ellis, E. C., Gale, F., Van Boeckel, T. P., Wint, W., Robinson, T., Xiao, X. & Gilbert, M. (2011c). Modelling the distribution of chickens, ducks, and geese in China. *Agric Ecosyst Environ* **141**, 381-389.

Pybus, O. G., Barnes, E., Taggart, R., Lemey, P., Markov, P. V., Rasachak, B., Syhavong, B., Phetsouvanah, R., Sheridan, I., Humphreys, I. S., Lu, L., Newton, P. N. & Klenerman, P. (2009). Genetic History of Hepatitis C Virus in East Asia. *Journal of Virology* **83**, 1071-1082.

Pybus, O. G., Fraser, C. & Rambaut, A. (2013). Evolutionary epidemiology: preparing for an age of genomic plenty Introduction. *Philos T R Soc B* **368**.

Pybus, O. G. & Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* **10**, 540-550.

Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W., Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P. & Delwart, E. L. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci U S A* **109**, 15066-15071.

Rabaa, M. A., Ty Hang, V. T., Wills, B., Farrar, J., Simmons, C. P. & Holmes, E. C. (2010). Phylogeography of recently emerged DENV-2 in southern Viet Nam. *PLoS Negl Trop Dis* **4**, e766.

Raftery, A., Newton, M., Satagopan, J. & Krivitsky, P. (2007). *Estimating the integrated likelihood via posterior simulation using the harmonic mean identity*: New York: Oxford University Press.

Raftery, M. A. N. A. E. (1994). Approximating Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society* **56**, 3-48.

Ragonnet-Cronin M, L. S., Hodcroft E, Hué S, Fearnhill E, Dunn D, Delpech V, Leigh Brown AJ (2013). Automated Analysis of Phylogenetic Clusters. *BMC Bioinformatics*.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K. & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615-619.

Reid, A. H., Fanning, T. G., Janczewski, T. A., McCall, S. & Taubenberger, J. K. (2002). Characterization of the 1918 "Spanish" influenza virus matrix gene segment. *J Virol* **76**, 10717-10723.

Reid, A. H., Fanning, T. G., Janczewski, T. A. & Taubenberger, J. K. (2000). Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc Natl Acad Sci U S A* **97**, 6785-6790.

Rogers, J. S. & Swofford, D. L. (1998). A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. *Syst Biol* **47**, 77-89.

Saitou, N. & Nei, M. (1987). The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* **4**, 406-425.

Schnebel, B., Dierschke, V., Rautenschlein, S., Ryll, M. & Neumann, U. (2007). Investigations on infection status with H5 and H7 avian influenza virus in short-distance and long-distance migrant birds in 2001. *Avian Dis* **51**, 432-433.

Scholtissek, C. (1995). Molecular evolution of influenza viruses. *Virus Genes* **11**, 209-215.

Scotch, M., Lam, T. T. Y., Pabilonia, K. L., Anderson, T., Baroch, J., Kohler, D. & DeLiberto, T. J. (2014). Diffusion of influenza viruses among migratory birds with a focus on the Southwest United States. *Infection Genetics and Evolution* **26**, 185-193.

Seifi, S., Asasi, K. & Mohammadi, A. (2010). Natural co-infection caused by avian influenza H9 subtype and infectious bronchitis viruses in broiler chicken farms. *Vet Arhiv* **80**, 269-281.

Sen, B. (2013). *China Statistical Yearbook*. Beijing: China Statistics Press.

Senne, D. A., Suarez, D. L., Pedersen, J. C. & Panigrahy, B. (2003). Molecular and biological characteristics of H5 and H7 avian influenza viruses in live-bird markets of the northeastern United States 1994-2001. *Avian Diseases* **47**, 898-904.

Shapiro, B., Rambaut, A. & Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23**, 7-9.

Sharp, G. B., Kawaoka, Y., Jones, D. J., Bean, W. J., Pryor, S. P., Hinshaw, V. & Webster, R. G. (1997). Coinfection of wild ducks by influenza A viruses: distribution patterns and biological significance. *J Virol* **71**, 6128-6135.

Sherrilyn Wainwright, C. T., Filip Claes, Moisés Vargas-Terán, Vincent Martin, Juan Lubroth (2012). Highly Pathogenic Avian Influenza in Mexico (H7N3) - A significant threat to poultry production not to be underestimated. . *EMPRES WATCH* **26**, 1-9.

Shi, W. F., Gibbs, M. J., Zhang, Y. Z., Zhang, Z., Zhao, X. M., Jin, X., Zhu, C. D., Yang, M. F., Yang, N. N., Cui, Y. J. & Ji, L. (2008). Genetic analysis of four porcine avian influenza viruses isolated from Shandong, China. *Arch Virol* **153**, 211-217.

Si, Y., de Boer, W. F. & Gong, P. (2013). Different environmental drivers of highly pathogenic avian influenza H5N1 outbreaks in poultry and wild birds. *PLoS One* **8**, e53362.

Simini, F., Gonzalez, M. C., Maritan, A. & Barabasi, A. L. (2012). A universal model for mobility and migration patterns. *Nature* **484**, 96-100.

Sims, L. D., Domenech, J., Benigno, C., Kahn, S., Kamata, A., Lubroth, J., Martin, V. & Roeder, P. (2005). Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet Rec* **157**, 159-164.

Smith, G. J. (2005). Migratory Bird Pathways and the Gulf of Mexico USGS Fact Sheet 2005-3069.

Smith, G. J., Bahl, J., Vijaykrishna, D., Zhang, J., Poon, L. L., Chen, H., Webster, R. G., Peiris, J. S. & Guan, Y. (2009a). Dating the emergence of pandemic influenza viruses. *Proc Natl Acad Sci U S A* **106**, 11709-11712.

Smith, G. J., Fan, X. H., Wang, J., Li, K. S., Qin, K., Zhang, J. X., Vijaykrishna, D., Cheung, C. L., Huang, K., Rayner, J. M., Peiris, J. S., Chen, H., Webster, R. G. & Guan, Y. (2006). Emergence and predominance of an H5N1 influenza variant in China. *Proc Natl Acad Sci U S A* **103**, 16936-16941.

Smith, G. J., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S., Guan, Y. & Rambaut, A. (2009b). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122-1125.

Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S. M., Guan, Y. & Rambaut, A. (2009c). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122-1125.

Smith, W., Andrewes, C. H. & Laidlaw, P. P. (1933). A virus obtained from influenza patients. *Lancet* **2**, 66-68.

St John, K., Warnow, T., Moret, B. M. E. & Vawter, L. (2003). Performance study of phylogenetic methods: (unweighted) quartet methods and neighbor-joining. *J Algorithm* **48**, 173-193.

Stamatakis, A. (2005). An Efficient Program for phylogenetic Inference Using Simulated Annealing. In *Parallel and Distributed Processing Symposium* p. 8.

Stamatakis, A., Ludwig, T. & Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456-463.

Steinhauer, D. A. (1999). Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* **258**, 1-20.

Stienekegrober, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H. D. & Garten, W. (1992). Influenza-Virus Hemagglutinin with Multibasic Cleavage Site Is Activated by Furin, a Subtilisin-Like Endoprotease. *Embo J* **11**, 2407-2414.

Streicker, D. G., Lemey, P., Velasco-Villa, A. & Rupprecht, C. E. (2012). Rates of Viral Evolution Are Linked to Host Geography in Bat Rabies. *Plos Pathogens* **8**.

Suarez, D. L. & Perdue, M. L. (1998). Multiple alignment comparison of the non-structural genes of influenza A viruses. *Virus Res* **54**, 59-69.

Suarez, D. L., Senne, D. A., Banks, J., Brown, I. H., Essen, S. C., Lee, C. W., Manvell, R. J., Mathieu-Benson, C., Moreno, V., Pedersen, J. C., Panigrahy, B., Rojas, H., Spackman, E. & Alexander, D. J. (2004). Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis* **10**, 693-699.

Suarez, D. L., Spackman, E. & Senne, D. A. (2003). Update on molecular epidemiology of H1, H5, and H7 influenza virus infections in poultry in North America. *Avian Diseases* **47**, 888-897.

Suchard, M. A., Kitchen, C. M. R., Sinsheimer, J. S. & Weiss, R. E. (2003). Hierarchical phylogenetic models for analyzing multipartite sequence data. *Systematic Biology* **52**, 649-664.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18**, 1001-1013.

Sun, Y., Qin, K., Wang, J., Pu, J., Tang, Q., Hu, Y., Bi, Y., Zhao, X., Yang, H., Shu, Y. & Liu, J. (2011). High genetic compatibility and increased pathogenicity of reassortants derived from avian H9N2 and pandemic H1N1/2009 influenza viruses. *Proc Natl Acad Sci U S A* **108**, 4164-4169.

Swayne, D. E. (2008). *Avian influenza*. Ames, Iowa: Blackwell Pub.

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. & Ji, W. (2010). Migration of waterfowl in the East Asian flyway and spatial relationship to HPAI H5N1 outbreaks. *Avian Dis* **54**, 466-476.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512-526.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Taubenberger, J. K. & Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440-451.

Tavaré, S. (1986a). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *American Mathematical Society* **17**, 57–86.

Tavaré, S. (1986b). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect Math Life Sci* **17**, 57–86.

Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo, D. 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. & Donis, R. O. (2012). A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci U S A* **109**, 4269-4274.

Tong, S. X., Zhu, X. Y., Li, Y., Shi, M., Zhang, J., Bourgeois, M., Yang, H., Chen, X. F., Recuenco, S., Gomez, J., Chen, L. M., Johnson, A., Tao, Y., Dreyfus, C., Yu, W. L., 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. & Donis, R. O. (2013). New World Bats Harbor Diverse Influenza A Viruses. *Plos Pathogens* **9**.

Treanor, J. J., Snyder, M. H., London, W. T. & Murphy, B. R. (1989a). The B-Allele of the Ns Gene of Avian Influenza-Viruses, but Not the a-Allele, Attenuates a Human Influenza-a Virus for Squirrel-Monkeys. *Virology* **171**, 1-9.

Treanor, J. J., Snyder, M. H., London, W. T. & Murphy, B. R. (1989b). The B allele of the NS gene of avian influenza viruses, but not the A allele, attenuates a human influenza A virus for squirrel monkeys. *Virology* **171**, 1-9.

Truscott, J. & Ferguson, N. M. (2012). Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput Biol* **8**, e1002699.

Tweed, S. A., Skowronski, D. M., David, S. T., Larder, A., Petric, M., Lees, W., Li, Y., Katz, J., Krajden, M., Tellier, R., Halpert, C., Hirst, M., Astell, C., Lawrence, D. & Mak, A. (2004). Human illness from avian influenza H7N3, British Columbia. *Emerg Infect Dis* **10**, 2196-2199.

Van Boeckel, T. P., Prosser, D., Franceschini, G., Biradar, C., Wint, W., Robinson, T. & Gilbert, M. (2011). Modelling the distribution of domestic ducks in Monsoon Asia. *Agric Ecosyst Environ* **141**, 373-380.

Vijaykrishna, D., Bahl, J., Riley, S., Duan, L., Zhang, J. X., Chen, H., Peiris, J. S., Smith, G. J. & Guan, Y. (2008). Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. *PLoS Pathog* **4**, e1000161.

Wainwright, S., Carlene Trevennec, Filip Claes, Moisés Vargas-Terán, Vincent Martin, Juan Lubroth (2012). Highly Pathogenic Avian Influenza in Mexico (H7N3) - A significant threat to poultry production not to be underestimated. . *EMPRES WATCH* **26**, 1-9.

Wan, X. F., Dong, L. B., Lan, Y., Long, L. P., Xu, C. L., Zou, S. M., Li, Z., Wen, L. Y., Cai, Z. P., Wang, W., Li, X. D., Yuan, F., Sui, H. T., Zhang, Y., Dong, J., Sun, S. H., Gao, Y., Wang,

M., Bai, T., Yang, L., Li, D. X., Yang, W. Z., Yu, H. J., Wang, S. W., Feng, Z. J., Wang, Y., Guo, Y. J., Webby, R. J. & Shu, Y. L. (2011). Indications that Live Poultry Markets Are a Major Source of Human H5N1 Influenza Virus Infection in China. *Journal of Virology* **85**, 13432-13438.

Wang, G. J., Deng, G. H., Shi, J. Z., Weiyu, L. A., Zhang, G. Q., Zhang, Q. Y., Liu, L. L., Jiang, Y. P., Li, C. J., Sriwilaijaroen, N., Hiramatsu, H., Suzuki, Y., Kawaoka, Y. & Chen, H. L. (2014). H6 Influenza Viruses Pose a Potential Threat to Human Health. *Journal of Virology* **88**, 3953-3964.

Wang, R., Soll, L., Dugan, V., Runstadler, J., Happ, G., Slemons, R. D. & Taubenberger, J. K. (2008). Examining the hemagglutinin subtype diversity among wild duck-origin influenza A viruses using ethanol-fixed cloacal swabs and a novel RT-PCR method. *Virology* **375**, 182-189.

Wang, S., Shi, W. M., Mweene, A., Wei, H. L., Bai, G. R. & Liu, J. H. (2005). Genetic analysis of the nonstructural (NS) genes of H9N2 chicken influenza viruses isolated in China during 1998-2002. *Virus Genes* **31**, 329-335.

Wasilenko, J. L., Lee, C. W., Sarmento, L., Spackman, E., Kapczynski, D. R., Suarez, D. L. & Pantin-Jackwood, M. J. (2008). NP, PB1, and PB2 viral genes contribute to altered replication of H5N1 avian influenza viruses in chickens. *J Virol* **82**, 4544-4553.

Watanabe, Y., Ibrahim, M. S., Suzuki, Y. & Ikuta, K. (2012). The changing nature of avian influenza A virus (H5N1). *Trends Microbiol* **20**, 11-20.

Webster, R. G. (1998). Influenza: an emerging disease. *Emerg Infect Dis* **4**, 436-441.

Webster, R. G. (2002). The importance of animal influenza for human disease. *Vaccine* **20 Suppl 2**, S16-20.

Webster, R. G. (2004). Wet markets--a continuing source of severe acute respiratory syndrome and influenza? *Lancet* **363**, 234-236.

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol Rev* **56**, 152-179.

WHO (2005). Avian Influenza: assessing the pandemic threat. Geneva: World Health Organization WHO/CDS/2005.29

WHO (2014). WHO RISK ASSESSMENT: Human infections with avian influenza A(H7N9) virus: World Health Organization.

Wilke, C. O., Nelson, M. I., Lemey, P., Tan, Y., Vincent, A., Lam, T. T.-Y., Detmer, S., Viboud, C., Suchard, M. A., Rambaut, A., Holmes, E. C. & Gramer, M. (2011). Spatial Dynamics of Human-Origin H1 Influenza A Virus in North American Swine. *PLoS Pathogens* **7**, e1002077.

Wise, H. M., Foeglein, A., Sun, J., Dalton, R. M., Patel, S., Howard, W., Anderson, E. C., Barclay, W. S. & Digard, P. (2009). A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *J Virol* **83**, 8021-8031.

Wolf, Y. I., Viboud, C., Holmes, E. C., Koonin, E. V. & Lipman, D. J. (2006). Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol Direct* **1**, 34.

Wood, G. W., McCauley, J. W., Bashiruddin, J. B. & Alexander, D. J. (1993). Deduced amino acid sequences at the haemagglutinin cleavage site of avian influenza A viruses of H5 and H7 subtypes. *Arch Virol* **130**, 209-217.

Worobey, M., Han, G. Z. & Rambaut, A. (2014). Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc Natl Acad Sci U S A* **111**, 8107-8112.

Xia, Y., Bjornstad, O. N. & Grenfell, B. T. (2004). Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am Nat* **164**, 267-281.

Xiao, X. M., Gilbert, M., Slingenbergh, J., Lei, F. & Boles, S. (2007). Remote sensing, ecological variables, and wild bird migration related to outbreaks of highly pathogenic H5N1 avian influenza. *J Wildlife Dis* **43**, S40-S46.

Xu, K. M., Smith, G. J. D., Bahl, J., Duan, L., Tai, H., Vijaykrishna, D., Wang, J., Zhang, J. X., Li, K. S., Fan, X. H., Webster, R. G., Chen, H., Peiris, J. S. M. & Guan, Y. (2007). The genesis and evolution of H9N2 influenza viruses in poultry from southern china, 2000 to 2005. *Journal of Virology* **81**, 10389-10401.

Xu, X., Subbarao, Cox, N. J. & Guo, Y. (1999). 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**, 15-19.

Yu, H., Wu, J. T., Cowling, B. J., Liao, Q., Fang, V. J., Zhou, S., Wu, P., Zhou, H., Lau, E. H., Guo, D., Ni, M. Y., Peng, Z., Feng, L., Jiang, H., Luo, H., Li, Q., Feng, Z., Wang, Y., Yang, W. & Leung, G. M. (2014a). Effect of closure of live poultry markets on poultry-to-person transmission of avian influenza A H7N9 virus: an ecological study. *Lancet* **383**, 541-548.

Yu, X., Jin, T., Cui, Y., Pu, X., Li, J., Xu, J., Liu, G., Jia, H., Liu, D., Song, S., Yu, Y., Xie, L., Huang, R., Ding, H., Kou, Y., Zhou, Y., Wang, Y., Xu, X., Yin, Y., Wang, J., Guo, C., Yang, X., Hu, L., Wu, X., Wang, H., Liu, J., Zhao, G., Zhou, J., Pan, J., Gao, G. F. & Yang, R. (2014b). Influenza H7N9 and H9N2 viruses: coexistence in poultry linked to human H7N9 infection and genome characteristics. *J Virol* **88**, 3423-3431.

Zhang, L., Guo, Z. W., Bridge, E. S., Li, Y. M. & Xiao, X. M. (2013). Distribution and dynamics of risk factors associated with highly pathogenic avian influenza H5N1. *Epidemiol Infect* **141**, 2444-2453.

Zhao, G., Gu, X. B., Lu, X. L., Pan, J. J., Duan, Z. Q., Zhao, K. K., Gu, M., Liu, Q. T., He, L., Chen, J., Ge, S. Q., Wang, Y. H., Chen, S. J., Wang, X. Q., Peng, D. X., Wan, H. Q. & Liu, X. F. (2012). Novel Reassortant Highly Pathogenic H5N2 Avian Influenza Viruses in Poultry in China. *PLoS One* **7**.

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. (2013). Characterization of three H5N5 and one H5N8 highly pathogenic avian influenza viruses in China. *Vet Microbiol* **163**, 351-357.

Zhu, Q., Yang, H., Chen, W., Cao, W., Zhong, G., Jiao, P., Deng, G., Yu, K., Yang, C., Bu, Z., Kawaoka, Y. & Chen, H. (2008). A naturally occurring deletion in its NS gene contributes to the attenuation of an H5N1 swine influenza virus in chickens. *J Virol* **82**, 220-228.

Zimmer, S. M. & Burke, D. S. (2009). Historical Perspective - Emergence of Influenza A (H1N1) Viruses. *New Engl J Med* **361**, 279-285.