Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biomedicina Comparata e Alimentazione

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE VETERINARIE

CICLO XXVI

EVOLUTIONARY DYNAMICS OF RNA VIRUSES WITH ZOONOTIC POTENTIAL

Direttore della Scuola: Ch.mo Prof. Gianfranco GABAI

Supervisore: Dott. Enrico Massimiliano NEGRISOLO

Co-supervisore: Dott. Giovanni CATTOLI

Dottoranda: Alice FUSARO

31 GENNAIO 2014

RIASSUNTO

I virus a RNA con potenziale zoonosico rappresentano una seria minaccia per la salute pubblica a livello mondiale. L'elevato tasso di mutazione, i rapidi tempi di replicazione e le ingenti dimensioni della popolazione sono tre peculiarità all'origine delle potenzialità di questi patogeni in termini di variabilità genetica e capacità di adattamento a diverse condizioni ambientali. Comprendere le caratteristiche genetiche e le dinamiche evolutive dei virus a RNA con potenziale zoonosico è fondamentale al fine di prevenire, controllare, curare e ridurre i danni che provocano alla salute degli animali e dell'uomo.

Questa tesi si occupa dello studio dell'epidemiologia e dell'evoluzione molecolare di due zoonosi di origine virale diffuse a livello mondiale: l'influenza aviaria e la rabbia. Un elevato numero di dati genetici, generati con l'utilizzo di tecniche di sequenziamento di prima (sequenziamento Sanger) e seconda (*Next Generation Sequencing*) generazione, sono stati analizzati mediante l'utilizzo di strumenti bioinformatici. Tali sequenze sono state ottenute da campioni raccolti nel corso di quattro diverse epidemie: un'epidemia di rabbia silvestre verificatasi nel nord-est Italia tra il 2008 e il 2011, focolai epidemici di influenza aviaria causati dal sottotipo H5N1 ad alta patogenicità descritti in Egitto tra il 2006 e il 2010, e due ondate epidemiche provocate da due diversi sottotipi influenzali aviari H7N1 e H7N3 che hanno colpito l'Italia settentrionale nei periodi 1999 - 2001 e 2002 - 2004.

Attraverso l'analisi filogenetica e filogeografica di sequenze virali rappresentative a livello spazio-temporale delle varie epidemie, è stato possibile identificare la co-circolazione di diversi gruppi genetici, determinare il flusso genico e studiare le dinamiche evolutive di virus sottoposti a pressioni selettive di varia natura, come ad esempio la vaccinazione. Inoltre, grazie all'applicazione di un approccio di tipo *deep sequencing*, questo studio ha permesso di comprendere meglio i meccanismi di trasmissione delle sottopopolazioni virali da un ospite all'altro, la variabilità intra-ospite della popolazione virale e l'evoluzione della patogenicità.

I risultati presentati in questa tesi permettono di ampliare le nostre conoscenze a) sull'impatto e l'efficacia delle misure di sorveglianza e controllo applicate nel corso delle epidemie studiate, b) sulle dinamiche evolutive e sulla diffusione spaziale di virus appartenenti a diversi gruppi genetici, c) sull'emergenza di mutazioni amminoacidiche potenzialmente correlate a un aumento della fitness virale, d) sulla trasmissione a livello inter-ospite di varianti virali e e) sull'acquisizione di determinanti di virulenza.

Infine, il presente studio evidenzia le enormi potenzialità della tecnologia di *Next Generation Sequencing* nel favorire la comprensione delle complicate dinamiche evolutive dei patogeni emergenti.

SUMMARY

RNA viruses with zoonotic potential represent a public health threat throughout the world. High mutation rates, short generation times and large population sizes are three factors responsible for their high genetic variability and enormous adaptive capacity to new environmental conditions. Understanding the genetic properties and the evolutionary dynamics of RNA viruses with zoonotic potential is crucial to prevent, control, treat and lessen the damage to animal and human health.

This thesis investigates the most essential aspects related to the evolution and epidemiology of two widespread zoonotic diseases caused by two RNA viruses: Avian Influenza and rabies. Through the application of bioinformatics tools, I analysed a large amount of sequence data, generated using both first and second generation sequencing technology, from viruses collected during four distinct epidemics: a fox-rabies virus epidemic occurred in north-eastern Italy between 2008 and 2011, highly pathogenic H5N1 avian influenza outbreaks reported in Egypt between 2006 and 2010, and two avian influenza epidemics caused by two distinct subtypes that took place in northern Italy from 1999 to 2001 and 2002 to 2004.

Through phylogenetic and Bayesian phylogeographic analyses of viral sequences sampled over multiple discrete spatio-temporal scales, the studies in this thesis reveal the co-circulation of multiple viral lineages, explore the viral gene flows and investigate the evolutionary dynamics of viruses under different selection pressures. In addition, to better understand the pattern of transmission of viral subpopulations from host to host, the intra-host variability and the evolution of viral pathogenicity, I employed an ultra-deep sequencing approach to assess the diversity of viral populations.

The data generated in this thesis provide important insights into the a) impact and efficacy of surveillance strategies and control measures implemented during an outbreak, b) differences in the evolutionary dynamics and spatial spread between distinct genetic groups, c) emergence of amino acid mutations that may increase viral fitness, d) inter-host transmission of viral variants and e) gain of virulence determinants.

Finally, this thesis shows the great opportunity offered by next generation sequencing technology for dramatic advancement in our understanding of the complicated evolutionary dynamics of these pathogens.

PUBLICATIONS INCLUDED IN THIS THESIS

Fusaro A, Monne I, Salomoni A, Angot A, Trolese M, Ferrè N, Mutinelli F, Holmes EC, Capua I, Lemey P, Cattoli G, De Benedictis P. The introduction of fox rabies into Italy (2008-2011) was due to two viral genetic groups with distinct phylogeographic patterns. Infect Genet Evol. 2013;17:202-9. doi:10.1016/j.meegid.2013.03.051.

Cattoli G, **Fusaro A**, Monne I, Coven F, Joannis T, El-Hamid HS, Hussein AA, Cornelius C, Amarin NM, Mancin M, Holmes EC, Capua I. Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. Vaccine. 2011;29(50):9368-75. doi: 10.1016/j.vaccine.2011.09.127.

Monne I^, **Fusaro A**^, Nelson MI, Bonfanti L, Mulatti P, Hughes J, Murcia PR, Valastro V, Moreno A, Holmes EC, Cattoli G. Emergence of a Highly Pathogenic Avian Influenza Virus from a Low Pathogenic Progenitor. J Virol. 2014. Accepted. ^These authors contributed equally to this work.

TABLE OF CONTENTS

PREFACE
- CONTRIBUTIONS TO THIS THESIS -

The studies contained in this thesis were mainly carried out at the Research & Innovation Department, Division of Biomedical Science, Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy. I performed all the analyses reported in chapters 3, 4, 5 and 6 and contributed to much of the writing. Contributions from other authors are described below:

Chapter 3. Paola De Benedictis, Isabella Monne, Giovanni Cattoli, Franco Mutinelli and Ilaria Capua of the Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy advised on the study and contributed to writing of the paper. Angela Salomoni, Angélique Angot, of the Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy undertook the viral genome sequencing, Matteo Trolese, Nicola Ferre' of the Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy provided all the maps and advised on the epidemiological aspects of the study. Edward C. Holmes from Sydney Emerging Infections and Biosecurity Institute, University of Sydney, Sydney, Australia contributed to writing the paper. Technical assistance for the sequence analysis was provided by Philippe Lemey of the department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium.

Chapter 4. Giovanni Cattoli, Isabella Monne, Ilaria Capua of Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy contributed to writing of the paper. Fethyie Coven (Bornova Veterinary Control and Research Institute, I˙zmir, Turkey), Tony Joannis (National Veterinary Research Institute, P.M.B. 01 VOM, Plateau State, Nigeria), Hatem S. Abd El-Hamid (Faculty Veterinary Medicine Damanhour University, Egypt), Aly Ahmed Hussein (Faculty Veterinary Medicine, Cairo University, Giza, Egypt), Claire Cornelius (US Naval Medical Research Unit No. 3, Cairo, Egypt), Nadim Mukhles Amarin (Boehringer Ingelheim, Regional Technical Manager, Dubai, United Arab Emirates), provided the viral samples. Marzia Mancin of Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy performed the statistical analysis. Edward C. Holmes from Sydney Emerging Infections and Biosecurity Institute, University of Sydney, Sydney, Australia provided invaluable technical insights and contributed to writing of the paper. Annalisa Salviato and Alessia Schivo of the

Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy undertook the viral genome sequencing

Chapter 5. Isabella Monne of Istituto Zooprofilattico Sperimentale delle Venezie, Padova contributed to writing of the paper and on NGS data analysis. Giovanni Cattoli (Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy) and Pablo R. Murcia (MRC-University of Glasgow Center for Virus Research, Glasgow, United Kingdom) advised on the study. Technical assistance for the sequence analysis was provided by Martha I. Nelson (Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA) and Joseph Hughes (MRC-University of Glasgow Center for Virus Research, Glasgow, United Kingdom). Lebana Bonfanti and Paolo Mulatti of Istituto Zooprofilattico Sperimentale delle Venezie, Padova advised on epidemiological aspects. Viral genome sequencing was conducted by Viviana Valastro and Alessia Schivo of Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy. Ana Moreno of Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, Brescia, Italy provided a subset of samples. Edward C. Holmes from Sydney Emerging Infections and Biosecurity Institute, University of Sydney, Sydney, Australia provided invaluable advices and contributed to writing of the paper.

Chapter 6. Isabella Monne, Giovanni Cattoli of Istituto Zooprofilattico Sperimentale delle Venezie, Padova and Pablo R. Murcia of MRC-University of Glasgow Center for Virus Research, Glasgow, United Kingdom advised on the study and provided invaluable direction. Technical assistance for the sequence analysis was provided by Enrico Massimiliano Negrisolo (Department of Comparative Biomedicine and Food Science, University of Padova, Padova, Italy) Joseph Hughes (MRC-University of Glasgow Center for Virus Research, Glasgow, United Kingdom), Richard Orton (3IBAHCM, University of Glasgow, Glasgow, United Kingdom) and Luca Tassoni (Istituto Zooprofilattico Sperimentale delle Venezie, Padova). Viviana Valastro, Alessia Schivo and Angela Salomoni of Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy undertook the viral genome sequencing. Lebana Bonfanti of Istituto Zooprofilattico Sperimentale delle Venezie, Padova advised on epidemiological aspects.

Chapter 1

Introduction

As an effect of increased globalization, animal diseases, in particular those transmissible to man, have an important global economic and social impact (Hill et al., 2012). RNA viruses with zoonotic potential represent a public health threat throughout the world. Due to their high mutation rate, RNA viruses are particularly well-suited to highlight the relevance of evolutionary dynamics occurring simultaneously on multiple spatial and temporal scales. Understanding their evolution is crucial to prevent, control, treat and lessen the damage to animal and human health. At an individual level, RNA viruses likely evolve in response to a hostile environment presented by the immune system or drug treatment, while, at population level, they evolve to evade immune control, thus optimizing their level of virulence. Through the 7 chapters of this thesis, I have studied the evolutionary dynamics of two representative RNA viruses with high potential for crossing the species: influenza viruses and lyssaviruses. This thesis focuses on four different case studies of outbreaks which have caused severe health, ecological and socio-economic consequences: a fox-rabies virus epidemic occurred between 2008 and 2011 (chapter 3) in north-eastern Italy, highly pathogenic H5N1 avian influenza outbreaks reported in Egypt between 2006 and 2010 (chapter 4), and two avian influenza epidemics caused by two distinct subtypes (H7N1 and H7N3) that took place in northern Italy between 1999 and 2001 (chapter 5 and 6) and between 2002 and 2004 (chapter 6). In each study, classical or high-throughput sequencing strategies combined with recent developed bioinformatic tools provided new information on the inter- and intra-host evolutionary process, as well as on the spatial movements and evolution pathogenicity of RNA viruses during the course of an epidemic.

Chapter 2 reports the information about the epidemiology and molecular biology of the two RNA viruses, influenza A and rabies, analysed in this thesis. In addition, it reviews the existing knowledge on the RNA virus evolution, and provided information about the technologies and bioinformatics tools applied in the research chapters of this thesis.

Chapters 4, 5 and 6 provide a detailed comparison of the evolutionary dynamics between a) viruses circulating under different selection pressures (chapter 4), b) strains with different pathogenic properties (chapter 5) or c) different epidemics which affected the same geographic area (chapter 6). As a whole, these studies contribute to better understand how evolution of RNA viruses can be shaped by different epidemiological conditions.

The intra-host variability of viral populations is investigated in chapters 5 and 6, providing important insights on the emergence of viruses with a virulent phenotype and on the transmission of subpopulations from host to host.

Molecular phylogeography, which offers a framework in which specific hypotheses regarding pathogen gene flow and dispersal within an ecological context can be compared, has been investigated throughout chapters 3 and 5.

Thanks to these studies, I aimed to shed new light on the origin, evolutionary dynamics, acquisition of virulence determinants and spreading of several epidemic events, providing a better understanding of the complexity of RNA virus evolution.

Chapter 2

Background

2.1. POPULATION DYNAMICS OF RNA VIRUS

Viruses are obligate intracellular parasites which require a host cell for their growth and replication. The replication of the RNA virus genomes is unique, considering that the host cell does not contain a RNA dependent RNA polymerase. To overcome this constraint, the majority of RNA viruses encode their own RNA-dependent RNA replicases and reverse transcriptases that are either packaged with the virus genome or synthesised shortly after infection. As a consequence of the lack of a proofreading activity of their RNA polymerases, RNA viruses posses an extremely high mutation rate, which leads to frequent errors during viral RNA synthesis. This characteristic, in addition to short generation times and large populations, is responsible for the high genetic variability and the enormous adaptive capacity of RNA viruses to new environmental conditions (Moya et al., 2000). Because of their high mutation rates, these pathogens are moving targets for therapeutic intervention and frequently develop resistance to vaccines and antiviral drugs, limiting the efficacy of the attempts to control outbreaks. Moreover, their diversity allows them to jump species boundaries and establish productive infections in new hosts.

The average mutation rates of RNA viruses fall in the range of 10^{-5} to 10^{-2} nucleotide substitutions per site, per year (sub/site/year), with most of the viruses exhibiting rates within one order of magnitude of 1×10^{-3} sub/site/year (Duffy et al., 2008). Mutation rate must be distinguished from the mutation frequency, which depends on the replicative ability of each virus in competition with the rest of the population. Positive selection favours the presence of

mutations with an adaptive value, such as amino acid changes which facilitate immune escape or mutations that are selected at inter-host transmission (Duffy et al., 2008).

Point mutations, as well as recombination and reassortment, contribute to modulate the genetic diversity of RNA viruses. Recombination is the process by which two molecules result in a new mutant genome after the exchange of genetic information, while the genome segment reassortment takes place in viruses with a segmented genome and consists of the encapsidation in the same viral particle of genomic segments belonging to different parental viruses (Manrubia and Lazaro, 2006). As a consequence of all the mechanisms generating genetic diversity, RNA viruses exist as a cloud of diverse variants or quasispecies that interact cooperatively and collectively determine the biological behaviour of a viral population (Domingo et al., 2012).

When the environmental conditions are relatively constant, the virus population can be maintained in a sort of evolutionary stasis, which represents a balance between mutation and selection and keeps a consensus sequence (the sequence that corresponds to the most represented nucleotide at each genomic position) almost unvaried through time (Manrubia and Lazaro, 2006). However, in nature, viral populations encounter a range of environments. After infection of a new host or recognition by the immune system, a virus can experience an acceleration of fixation of beneficial mutations to adapt to the new conditions. For this reason it might be advantageous for the virus to maintain an assortment of pre-adapted variants, allowing a more rapid fitness gain (Lauring et al., 2013).

The term fitness is used to indicate how well an organism fits into its environment, which reflects its ability to survive and reproduce (Lauring and Andino, 2010). According to the theory of "survival of the fattest", natural selection pushes a population to the "flat" region of the fitness landscape (figure 1). In this landscape, the "ground level" is the range of genotypes in the sequence space, while the altitude is the fitness of each genotype. Environment and selective pressure determine the contours of the landscape. At low mutation rates, the variants occupy a tall peak, where there is little genotypic diversity and maximal fitness. When the mutation rate is high, as for RNA viruses, the populations are pushed away from the fitness peaks and variants are spread out over their corresponding peak (figure 2.1). The population on the flattest peak accepts mutations, with minimal consequences on fitness, and is ready to adapt to rapid environmental change, such as activation of immune response, antiviral

treatments, host-switch (Lauring and Andino, 2010; Domingo et al., 2012; Lauring et al, 2013; Tejero et al., 2011).

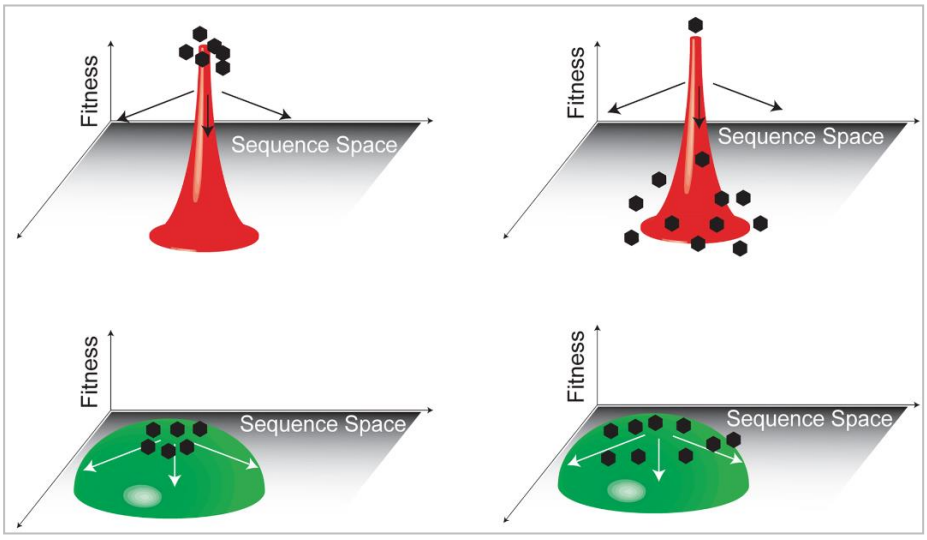


Figure 2.1. High mutation rates and survival of the flattest. In a fitness landscape, the ground level represents the sequence space and the vertical axis gives the fitness value for each sequence. When the mutation rate is low, the population will be genotypically stable and cluster on the top of the fitness peak. When the mutation rate is high, the population on the flatter peak (green) remains near its fitness optimum and has a higher fitness than the population on the steeper peak (red). Adapted from Lauring and Andino, 2010.

In some fitness landscapes, the existence of a maximum mutation rate, called the error threshold, has been predicted. Beyond this threshold, the quasispecies enters into an error catastrophe, losing its genetic information: consequently, the virus population extinguishes. RNA viruses have evolved the perfect error rate: a compromise between too many mutations which result into an error catastrophe and too few mutations which result into a viral population unable to cope with selective pressure (Duffy et al., 2008; Lauring and Andino, 2010; Tejero et al., 2011).

Population size is another relevant factor of evolution. Large population size permits competition among a large number of genomes, thus increasing the probability of fixation of beneficial mutations, which results in the increase of the fittest sequence. In this way, large population sizes preserve the invariance of phenotype, despite the mutational perturbation, ensuring their survival in a given environment (Lauring et al., 2013).

Evolution is always a result of changes in variant frequencies whereby some variants are lost over the time, while other variants sometimes increase their frequency to 100% and become fixed in the population (figure 2.2). The rate at which this occurs is called fixation rate. Differently, the substitution rate or the rate of molecular evolution represents the rate at which individuals accrue genetic differences to each other over the long-term evolution (Vandamme et al., 2009).

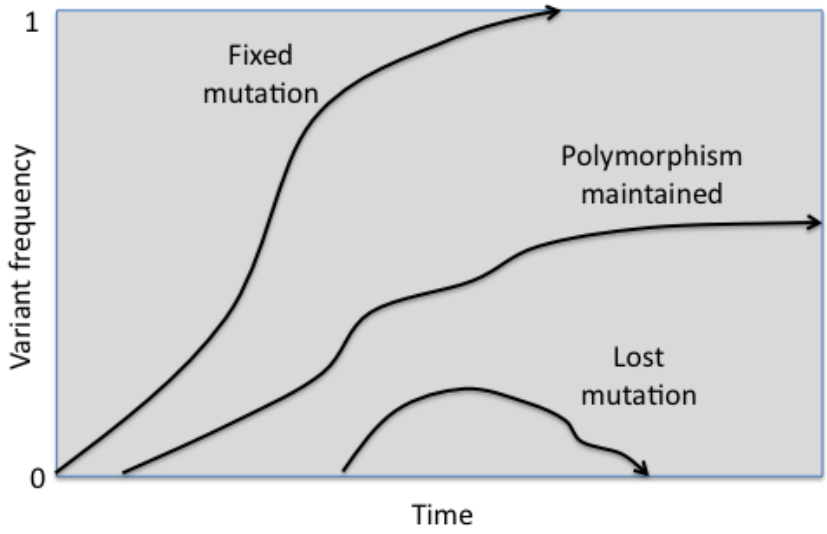


Figure 2.2. Loss or fixation of mutations in a population.

2.2. AVIAN INFLUENZA

Influenza viruses have a segmented, negative-sense, single-stranded RNA genome and belong to the Family *Orthomyxoviridae*, which consists of five genera: Influenzavirus A, Influenzavirus B, Influenzavirus C, Isavirus, and Thogotovirus (Capua and Alexander, 2008). Influenza A viruses are further divided into subtypes based on the antigenic relationships of the haemagglutinin (HA) and neuraminidase (NA) surface glycoproteins. To date, eighteen HA (H1–H18) and eleven NA subtypes (N1–N11) have been identified (Tong et al 2013). Aquatic birds are considered the major influenza A viruses reservoir in nature, harbouring diverse combinations of most of the HA (H1-H16) and NA (N1-N9) subtypes.

2.2.1 Epidemiology of avian influenza virus

2.2.1.1. Avian influenza virus in birds

Influenza viruses infect a great variety of birds (Alexander 2000; Alexander 2001; EFSA 2005; Hinshaw et al. 1981; Lvov 1978), including free-living birds, captive caged birds and domestic ducks, chickens, turkeys and other domestic poultry. In particular, the wild bird population, especially anseriforms (Order Anseriformes, including ducks, geese and swans) and charadriiforms (Order Charadriiformes, including shorebirds, gulls, terns) hosts an enormous pool of influenza viruses (Webster et al., 1992; Olsen et al., 2006; Hurt et al., 2006).

Until the widespread emergence of the highly pathogenic H5N1 in 2002, influenza viruses were thought to cause little morbidity and mortality in wild birds. However, from 2002, H5N1 has caused illness and killed hundreds of birds from several different orders including ducks (Vandergrift et al., 2010).

Generally, the primary introduction of LPAI viruses into a poultry population is the result of wild bird activity, usually waterfowl, but gulls and shorebirds have also been implicated. This may not necessarily involve direct contact, as infected waterfowl may carry the viruses into an area and these may then be introduced to poultry from other routes. Surface water used for drinking may also be contaminated with influenza viruses and thus serve as a source of infection.

The greatest threat of secondary spread of AI viruses is by mechanical transfer of infective organic material. Birds or other animals not susceptible to infection themselves may become contaminated and spread the virus mechanically in infective faeces or exudates from the waterfowl. However, for domestic poultry the main source of secondary spread appears to be through humans. In several specific accounts, strong evidence has implicated the movements of caretakers, farm owners, staff, trucks and drivers moving birds or delivering food as being responsible for the spread of the virus both onto and through a farm (Glass et al 1981; Homme et al 1970; Capua and Alexander 2008). However, the mechanism by which influenza viruses pass from one bird to another is extremely complex and depends on several causes, such as the strain of the virus, the species of bird and other environmental factors, such as biosecurity levels, concentration of poultry in the vicinity of the initial outbreaks, etc.

2.2.1.2. Avian influenza virus in mammals

Influenza A viruses have been shown to infect also mammalian species, including humans (figure 2.3). The availability of structurally distinct sialic acid linked receptors in the sites of human and avian influenza infection is one of the main factors which influence susceptibility to infection. Avian influenza (AI) viruses exhibit a strong preference for using sialic acid (SA)-α2,3-Gal-terminated saccharides as receptors, which are abundant in avian cells. Human influenza viruses, in contrast, preferentially bind to SA-α2,6-Gal-terminated saccharides receptors, well-represented on the human respiratory tract. This different binding preference is believed to be one of the major factors that prevents crossing the species barrier. However, the fact that AI viruses do occasionally infect people and other mammals indicates that this barrier is not insurmountable (Capua and Alexander, 2008).

Since 1996 the zoonotic potential of AI viruses has become a serious and growing public health concern (Capua and Alexander, 2008). With some exceptions, almost all the reported symptomatic cases of AI virus infection in humans have been caused by highly pathogenic avian influenza (HPAI) viruses belonging to the H5 or H7 subtypes, directly transmitted from infected birds to humans. The first human epidemic of avian influenza, popularly known as bird flu, caused by HPAI H5N1 virus circulating in the domestic poultry, was reported in Hong Kong in 1997. This strain infected 18 people, causing the death of 6 of them (Chan et al., 2002). In February 2003, two more human cases of avian influenza belonging again to the HPAI H5N1 subtype, were reported in Hong Kong (Peiris et al., 2004). In that year, this HPAI virus circulating in poultry in South East China began to spread westwards among wild and domestic birds throughout Asia, reaching Europe and Africa in 2005 and 2006, respectively. Since then, the infection in humans has started to occur with greater frequency in many countries, being responsible for a high mortality rate. As of 22 November 2013, the World Health Organization has reported a total of 641 confirmed cases of HPAI H5N1 infections in humans, with 380 human deaths (WHO, 2013).

H7 is another avian influenza subtype transmitted from birds to humans on several occasions. In 1996, a LPAI H7N7 virus was passed from ducks to a woman, resulting in a case of conjunctivitis. During a large outbreak of HPAI H7N7 virus in the Netherlands in 2003, 89 human infections, mostly conjunctivitis cases, and one death, were reported. In 2004, an outbreak of HPAI H7N3 virus in Canada resulted in conjunctivitis and in a mild influenza-like illness in two men. Human infections with H7 subtype were also reported in the

UK in 2006 and 2007 during two distinct outbreaks caused by LPAI H7N3 and LPAI H7N2 viruses, respectively (de Wit and Fouchier, 2008). In 2012, two human cases linked to the Mexican HPAI H7N3 outbreak were reported. Both patients showed only mild symptoms, such as conjunctivitis (CDC, 2012) and fully recovered. In August 2013, the H7N7 HPAI outbreaks reported to have occurred in Italy had caused conjunctivitis in three workers who took part in depopulation measures (Promed-mail, 2013). Viruses of the H7 subtype display an unusual tissue tropism compared to other subtypes, since cases of conjunctivitis had occurred in many outbreaks of LPAI and HPAI H7 viruses, but have rarely been reported with other subtypes.

Also the avian influenza virus of H9 subtype has been responsible of influenza like symptoms and mild upper respiratory tract infections in adults and children from Hong Kong (1997, 1999, 2003 and 2009) and China (1998 and 1999) (Butt et al., 2010).

On the whole, these findings indicate that transmission of avian influenza viruses from birds to humans is not common but may cause potential severe disease consequences.

Pigs play a crucial role in influenza ecology and epidemiology, primarily because of their dual susceptibility to human and avian viruses. They possess both SA-α2,3-Gal-terminated saccharides and SA-α2,6-Gal-terminated saccharides in cells lining the trachea and are therefore considered a potential "mixing vessel" for influenza viruses, from which reassortants may emerge. The introduction of AI viruses to pigs is not an uncommon occurrence. Nonetheless, the only subtypes to have truly established in the pig populations are H1N1, H3N2 and H1N2, although genotype analysis of isolates of these subtypes suggests that they can be the result of the reassortment of viruses from different progenitor host species (pig, human and avian) (Capua and Alexander 2008). Serological and virological evidence of infections of pigs with viruses of the H4, H5, H7 and H9 subtypes (Karasin et al., 2000, Karasin et al., 2004; Loeffen et al., 2003, 2004; Peiris et al., 2001; Rui-Hua et al 2011) has been largely documented.

The introduction of AI viruses in other mammals has been demonstrated in several instances (figure 2.3). Evidence of infections with avian influenza viruses of the subtypes H1N1, H2N2 and H3N2 has been reported in horses (Tumová, 1980). H7N7, H4N5, H13N2, H13N9, H1N3 and H3N8 avian influenza subtypes have been isolated in marine mammals (Webster et al., 2012; Anthony et al., 2012). Mustelids, dogs and cats have also been demonstrated to be naturally and experimentally susceptible to AI infection (Okazaki et al.,

1983; Berg et al 1990; Songserm et al 2006; Park et al 2012; Sun et al, 2012; Driskell et al 2012).

Fortunately, most of the zoonotic transmission from avian species to mammals are dead-end infections and are not further transmitted within the new species due to several barriers. However, on rare occasions influenza A viruses can break the species barrier and establish an entirely new virus lineage in a mammalian species, as exemplified by the human pandemic of 1918 (Taubenberger and Kash, 2011), the Eurasian classical swine influenza virus lineage (Pensaert et al., 1981), H3N8 influenza viruses in horses (Horimoto et al., 2005).

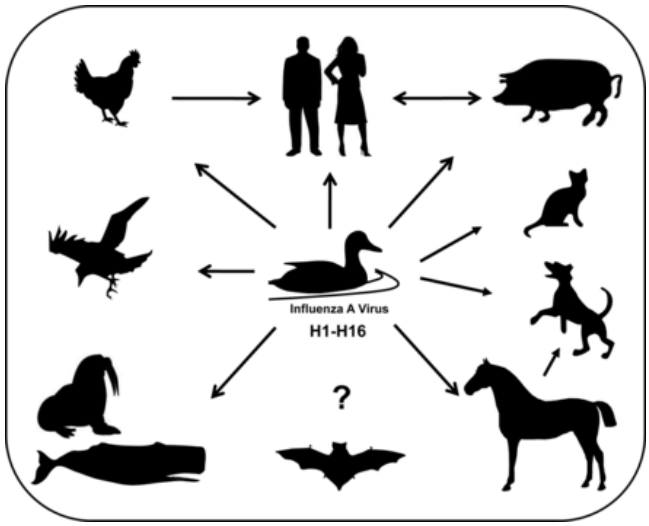


Figure 2.3. Host range of influenza A virus. Avian influenza virus can be transmitted from wild birds, the natural reservoir of the virus, to a wide range of other hosts. From Manz et al., 2013.

2.2.2. Molecular biology of influenza A virus

2.2.2.1 Genes and proteins

The influenza A virus genome (~13.5 Kb) consists of eight separate segments, named 1-8 from the longest to the shortest, which encode for about 15 viral proteins (figure 2.4).

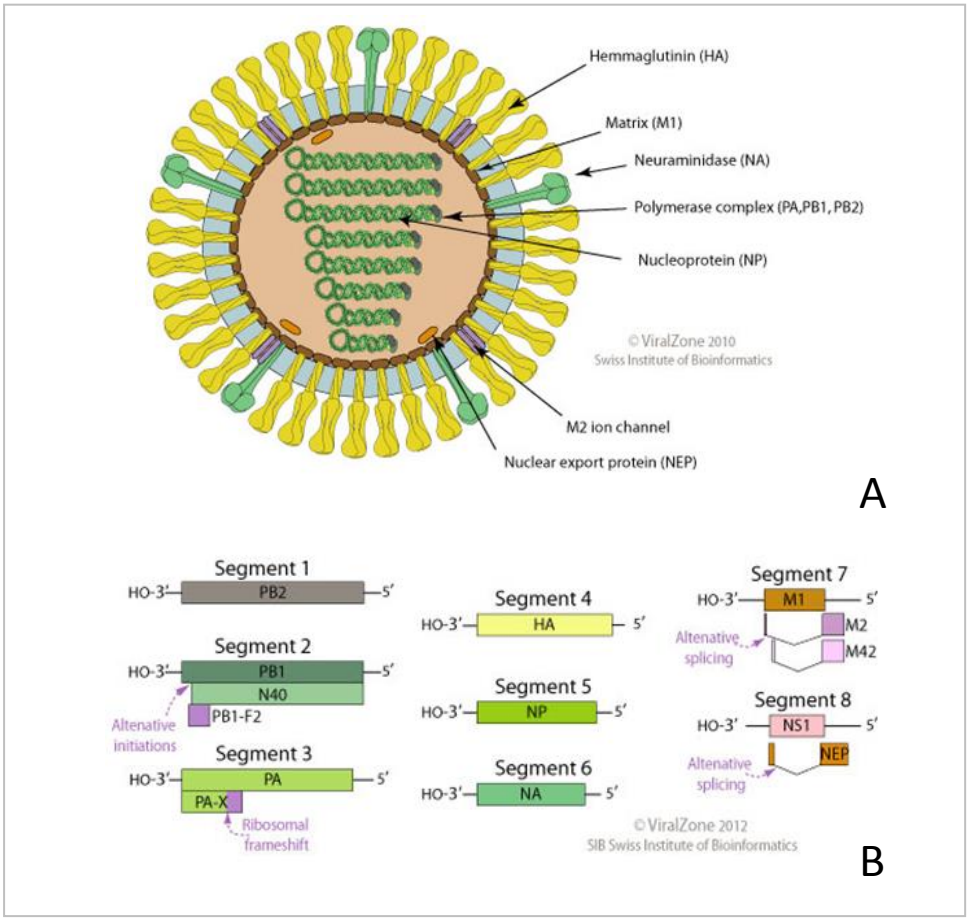


Figure 2.4. The virion structure of the influenza A virus and its genome. A) Two surface glycoproteins, HA and NA, and the M2 ion-channel protein are embedded in the viral envelope. The ribonucleoprotein complex comprises a viral RNA segment associated with the nucleoprotein and three polymerase proteins (PA, PB1, and PB2). The matrix protein is associated with both ribonucleoprotein and the viral envelope. B) Graphical representation of the influenza virus proteins codified by each segment. From ViralZone, SIB Swiss Institute of Bioinformatic (http://viralzone.expasy.org/).

Segment 1: PB2 polymerase. The basic polymerase protein 2 (PB2) is known to function during initiation of viral mRNA transcription as the protein, which recognizes and binds the 5' cap structures of host cell mRNAs to use as viral mRNA transcription primers (Webster et al., 1992).

Segment 2: PB1 polymerase, PB1-F2 and N40. The basic polymerase protein 1 (PB1) is the main polymerase protein responsible for the elongation of the primed nascent viral mRNA and template RNA and for the vRNA synthesis (Webster et al., 1992).

Segment 2 also encodes for the PB1-F2 protein in an alternate reading frame, which appears to be involved in cell apoptosis (Zamarin et al 2006).

A third major polypeptide is synthesized from PB1 mRNA via differential AUG codon usage. PB1 codon 40 directs translation of an N-terminally truncated version of the polypeptide (N40) that lacks transcriptase function but nevertheless interacts with PB2 and the polymerase complex in the cellular environment (Wise et al., 2009).

Segment 3: PA polymerase and PA-X. The acid polymerase protein (PA) (a) has endonuclease and protease activities, (b) is involved in viral RNA or complementary RNA promoter binding and (c) interacts with the PB1 subunit (Das et al., 2010).

It has recently been demonstrated that segment 3 of the virus contains a second open reading frame (X-ORF), accessed via ribosomal frameshifting. The frameshift product, termed PA-X, comprises the endonuclease domain of the viral PA protein with a C-terminal domain encoded by the X-ORF and functions to repress cellular gene expression. Loss of PA-X expression leads to changes in the kinetics of the global host response, which includes increases in inflammatory, apoptotic, and T lymphocyte-signaling pathways (Jagger et al., 2012).

In addition, novel PA-related proteins in influenza A virus-infected cells have been recently identified. These newly proteins are translated from the 11th and 13th in-frame AUG codons in the PA mRNA and are, therefore, N-terminally truncated forms of PA, which we named PA-N155 and PA-N182, respectively. The function of these proteins needs to be investigated (Muramoto et al., 2013).

Segment 4: Hemagglutinin. The hemagglutinin (HA) is a homotrimeric integral membrane glycoprotein and the major surface antigen of the influenza virus. It is responsible for binding to host sialic acid receptors for viral entry into cells and for fusion of viral and endosomal membranes to release viral nucleocapsids into the cell cytoplasm (Skehel and

Wiley, 2000). The HA is cleaved into two subunits, HA1 and HA2, connected by disulfide linkages. This cleavage is mediated by host trypsin-like proteases and is required for the virus-cell fusion. The HA consists of two distinct domains: (a) a globular head formed by the HA1 that contains a conserved receptor binding site and five antigenic regions A-E and (b) a fibrous stem formed by both HA1 and HA2 (Wiley and Skehel, 1987; Weis et al., 1988).

Segment 5: Nucleoprotein. The nucleoprotein (NP) is the core of the ribonucleoprotein (RNP) complex binding the vRNA with the polymerase complex. It has also been suggested that NP controls the switching of the RNA polymerase from transcription to replication (Shapiro and Krug 1988).

Segment 6: Neuraminidase. Segment 6 encodes for another integral membrane glycoprotein, the neuraminidase (NA), which is instrumental for cleaving the sialic acid receptor to facilitate viral release from infected cells (Palese et al 1974).

Segment 7: M1 and M2. Segment 7 contains two overlapping reading frames encoding for two proteins, matrix protein 1 (M1) and matrix protein 2, (M2) which are translated from spliced mRNAs (Lamb et al 1981). The M1 protein forms a layer to separate the RNPs from the viral envelope and is involved in numerous functions related to the assembly and disassembly of viral particles, the transport of the RNPs to the nucleus, the nuclear export of viral proteins and viral morphology (Martin and Helenius, 1991).

M2 is a tetramer integral membrane protein situated within the viral membrane as a proton ion channel that regulates the pH level. Its activity is essential for viral infection of host cells (Pinto et al., 1992).

Segment 8: NS1 and NS2. The smallest viral segment 8 also contains two overlapping reading frames, which require mRNA splicing, that encode for two proteins: nonstructural protein 1 and 2 (NS1 and NS2). NS1 is the main non structural protein, involved in host cell processes to promote viral protein synthesis, as well as in sequestering cellular mRNAs in the nucleus so that their caps may be used as primers for viral mRNA synthesis (Qian et al 1994). NS1 also sequesters dsRNA to prevent induction of the host interferon response (Kochs et al., 2007).

NS2 is frequently referred to as the “nuclear export protein” (NEP) for its role in transporting newly synthesized RNP from the nucleus to the cytoplasm (O'Neill et al, 1998).

2.2.2.2 Replication cycle

Infection of host cells by influenza is initiated by the binding of sialic acid (N-acetyl-neuraminic acid) receptors on host cells by the globular head of the viral HA, followed by viral entry by endocytosis (Carroll et al., 1981; Tumpey et al., 2007). Upon cellular entry of the virus as an endosome, the acid environment induces a pH-dependent conformational change in the HA, which causes the viral and endosomal membranes to fuse and release viral RNPs into the cytoplasm (figure 2.5). This fusion occurs thanks to the cleavage of the HA0 precursor molecule into the HA1 and HA2 subunits by host cell proteases (Skehel et al., 1982; Bullogh et al., 1994). Following viral entry, H ions flood through the M2 ion channel, causing the separation of the M1 from the RNPs to allow the transport of the RNPs from the cytoplasm into the nucleus, mediated by the nuclear localization signal (NLS) located at the extreme N-terminus of NP. All the vRNA syntheses, including viral transcription, replication and the assembling of vRNPs occur in nucleus. The viral RNA polymerase complex uses the negative sense vRNA as a template to synthesize both viral mRNA and cRNA, a full length copy of the vRNA which serves as a template for the synthesis of additional vRNAs (Herz et al., 1981). Following transcription, the mRNAs are transported to the cytoplasm where they are translated by the host ribosome. Newly synthesized viral proteins NP, PB2, PB1 and PA are transported back to the nucleus to bind to vRNA to form RNPs for new viral particles. The HA, NA and M2 are transported via the Golgi apparatus to the cell membrane, where they aggregate with the RNPs and M1 proteins to form new viral particles and to be released from the cell upon NA cleavage of host sialic acid receptors (Baigent and McCauley, 2003) (figure 2.5).

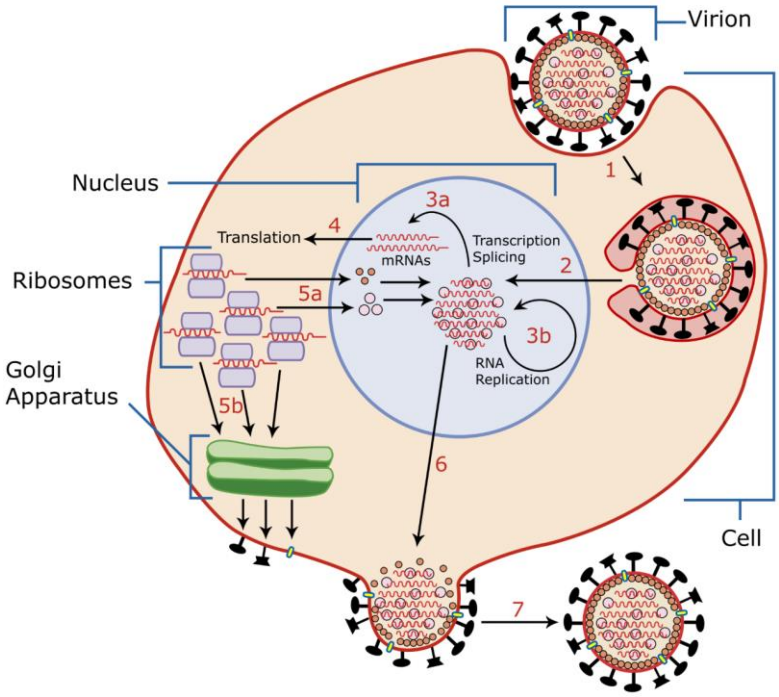


Figure 2.5. Influenza A virus replication. 1) A virion attaches to the host cell membrane using HA and enters the cytoplasm by receptor-mediated endocytosis, thereby forming an endosome. 2) HA2 promotes fusion of the virus envelope and the endosome membranes. M2 acts as an ion channel making the inside of the virion more acidic. As a result, the major envelope protein M1 dissociates from the nucleocapsid and vRNAs are translocated into the nucleus. 3) In the nucleus, the viral polymerase complexes transcribe (3a) and replicate (3b) the vRNAs. 4) Newly synthesized mRNAs migrate to cytoplasm where they are translated. 5) Post-translational processing of HA, NA, and M2 includes transportation via Golgi apparatus to the cell membrane (5b). NP, PB2, PB1 and PA move to the nucleus (5a) and bind the newly synthesized copies of vRNAs. 6) The nucleocapsids migrate into the cytoplasm and eventually interact via protein M1 with a region of the cell membrane where HA, NA and M2 have been inserted. 7) NA destroys the sialic acid moiety of cellular receptors, releasing the progeny virions from infected cell. By User:YK Times GFDL (http://www.gnu.org/copyleft/fdl.html), CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/), via Wikimedia Commons

2.2.2.3 Evolution

Influenza A viruses are dynamic and are continuously evolving. They can change in two different ways: drift and shift. While influenza viruses are changing by genetic drift all the time, genetic shift happens only occasionally.

Genetic drift refers to small gradual changes, which occur through point mutations. These may happen unpredictably and result in minor changes to influenza genome. When they occur in the two genes that codify for the main surface glycoproteins (HA and NA), new virus strains that may not be recognized by antibodies to earlier influenza strains may be produced. The antigenic drift in human influenza viruses has been well described. This is the main reason why people can be infected with influenza viruses more than once, and why global surveillance is critical in order to monitor the evolution of human influenza virus stains. This allows the selection of the most suitable strains to be included in the annual production of influenza vaccine. There is scientific evidence indicating that evolution and antigenic drift of human influenza (H1 and H3 subtypes) viruses are driven by multiple mutations within major antigenic sites located in the receptor binding subdomain of the HA protein (RBD) (Hensely et al., 2009; Shih et al., 2007). In avian influenza viruses, the occurrence of the antigenic drift is not as well documented and the mechanisms remain to be clarified, particularly with reference to the use of vaccines as control measures. Antigenic diversity has been described for H5N2 viruses causing a low-pathogenic AI epidemic in Mexico (Lee et al., 2004) and for H5N1 HPAI viruses in Asia and in Egypt (Balish et al., 2010; Beato et al., 2013; Cattoli et al., 2011; Wu et al., 2008). However, the driving mechanisms and the impact on control measures have not yet been established.

Antigenic shift refers to an abrupt, major change to produce a novel influenza A virus subtype. Due to their segmented nature, influenza genomes can be readily mixed in host cells infected with more than one virus. As a consequence, new strains can suddenly be produced by reassortment. For example, when a cell is infected with influenza viruses from different species, reassortment can result in progeny viruses containing genes from strains that normally infect birds, and genes from strains which normally infect humans. This leads to the creation of new strains which may cause influenza pandemics in humans, as observed in the case of the 1957 and 1968 pandemics (Clancy et al., 2008).

In the last century, three subtypes of the influenza A virus have been established in humans: 1918 - H1N1; 1957 - H2N2; 1968 - H3N2; 2009 - H1N1 (Morens et al., 2009). In

1957, the H2N2 pandemic virus acquired the HA, NA and PB1 segments through reassortment with an avian H2N2 influenza virus, while retaining five segments (PB2, PA, NP, M and NS) of the H1N1 virus of human origin. Similarly, the 1968 H3N2 pandemic virus acquired avian HA and PB1 segments, while retaining six human H2N2 segments (PB2, PA, NP, NA, M and NS) (Scholtissek et al., 1978). Differently, the 2009 pandemic H1N1 virus was derived by reassortment of two pre-existing swine IAV (Garten et al., 2009).

Today H5, H7 and H9 avian influenza virus subtypes top the World Health Organization's list for the greatest pandemic potential. Indeed, the spread of HPAI and LPAI viruses of the H5, H7 or H9 subtypes amongst birds and sporadic infection in humans (as described in chapter 2.2.1.2.) - where adaptive mutations or reassortment events between human and avian influenza strains may occur - continues to pose a threat to public health (Lin et al., 2000).

2.2.3 Avian influenza pathotypes

So far, only some strains of viruses of the H5 and H7 subtypes have been shown to cause the highly pathogenic form of the disease in susceptible species.

The hemagglutinin glycoprotein is produced by all influenza A viruses as a precursor, HA0, which requires a post-translational cleavage by host proteases to be functional (Vey et al., 1992). The HA0 precursor proteins of avian influenza viruses of low virulence for poultry have a single arginine at the cleavage site and another basic amino acid at position -3 or -4 from the cleavage site. These viruses are limited to cleavage by extracellular host proteases, such as trypsin-like enzymes and thus restricted to replication at sites in the host where such enzymes are present, i.e. the respiratory and intestinal tracts. H5 and H7 HPAI viruses possess multiple basic amino acids (arginine and lysine) at their HA0 cleavage sites (Wood et al., 1993; Senne et al., 1996; Vey et al., 1992) and are cleavable by one or more intracellular ubiquitous proteases (Stieneke-Gröber et al., 1992). H5 and H7 HPAI viruses are able to replicate throughout the bird, damaging vital organs and tissues and causing disease and death.

It appears that HPAI viruses arose by mutation after the introduction of LPAI H5 or H7 viruses into poultry from the wild bird reservoir. Several mechanisms seem to be responsible for this mutation: a) insertion/substitution of basic amino acids at the HA cleavage site, which most likely occurred due to a transcription error by the polymerase complex, or b) non-homologous recombination which results in the insertion of a foreign nucleotide sequence

(Perdue et al., 1998). The factors that bring about mutation from LPAI to HPAI are not known. In some instances, mutation seems to have taken place rapidly (at the index case site) after introduction from wild birds; in others, the LPAI virus progenitor had been circulating in poultry for months before mutating. Therefore, it is impossible to predict if and when this mutation will occur (Alexander, 2000).

Besides mutations in the HA cleavage site, other amino acid substitutions located in different gene segments can be responsible of the viral pathogenicity and can enhance various aspects of the viral life cycle, including virus binding and entry, genome transcription and translation, virion assembly and release, and evasion of innate immune responses. Mutations in the viral polymerase complex, such as the PB2 627K or D701N mutations, in the PB1-F2 protein, such as N66S and in the NS1, as amino acid 92E or mutations at the carboxyterminal portion, have been recognized as important contributors to viral pathogenicity (Tscherne and García-Sastre, 2011).

In addition to the H5 and H7 subtypes, also some H10 viruses fall within the definition of HPAI (Alexander, 2008). However, these viruses do not possess a multi-basic cleavage site and do not cause systemic infection. They are known to have tropism for the kidney, and the impaired function of this organ is what determines death of the bird when the virus is administered intravenously (Swayne and Alexander, 1994).

2.3. RABIES VIRUS

Rabies virus (RABV), a member of the genus Lyssavirus in the family Rhabdoviridae, is a neurotropic virus that causes fatal encephalitis in warm-blooded animals.

At present, the Lyssavirus genus includes 12 species: Rabies virus (RABV), Lagos bat virus (LBV), Mokola virus (MOKV), Duvenhage virus (DUVV), European bat lyssavirus type 1 (EBLV-1), European bat lyssavirus type 2 (EBLV-2), Australian bat lyssavirus (ABLV), Aravan virus (ARAV), Khujand virus (KHUV), Irkut virus (IRKV), West Caucasian bat virus (WCBV), and Shimoni bat virus (SHIBV) (Kuzmin and Tordo, 2012).

In addition, new lyssaviruses are routinely described, such as Bokeloh bat lyssavirus (BBLV; Freuling et al., 2011), Ikoma lyssavirus (IKOV; Marston et al., 2012) and Lleida bat lyssavirus (LLEBV; Ceballos et al., 2013), which still await classification assessment. Among

the currently recognized species, RABV is the most broadly distributed and causes over 99% human rabies cases in the world (Kuzmin and Tordo, 2012).

2.3.1. Epidemiology of rabies virus

Rabies is an enzootic disease widespread throughout the world and is a serious problem in developing countries. According to WHO, almost 55,000 people die because of rabies every year (WHO, 2005). Rabies viruses are adapted to various animal species (acting as reservoir) on which they depend for their existence. Viral infection of a species that does not normally maintain that virus (spill-over infection) occurs frequently, but only rarely results in spread within the new host (Nadin-Davis & Real, 2011). Although rabies is a zoonotic disease, human infections and deaths are an unfortunate consequence of biologic processes of virus maintenance in which humans play no significant role. The route of infection is usually, but not always, by bite. Indeed, in many cases, the infected animal is exceptionally aggressive, it attacks without provocation and exhibits otherwise uncharacteristic behaviour.

On a global basis, the domestic dog remains the most important reservoir of rabies. Although rabies has been eliminated in most European countries, canine rabies remains largely uncontrolled throughout most of Asia and Africa (Niezgoda et al., 2002), where dogs remain the main reservoir and transmitters of rabies to humans (Arai, 2005). Majority of wild reservoirs belongs to the family Carnivora and include foxes in the Artic (*Alopex lagopus*), Canada, central and western Europe and the Middle East (*Vulpes vulpes*), and scattered foci elsewhere throughout North America (e.g., *Urocyon cinereoargenteus*) (de Mattos et al., 2001; King, 1998). Other important terrestrial mammal hosts of RABV include common racoons (*Procyon lotor*) and shunks (*Menphitis menphitis*) in North America, and the racoon dogs (*Nyctereus procyonoides*) in Europe. Furthermore, insectivorous bats in North America and Australia and haematophagic bats in Latin America also act as reservoirs for RABV. Indeed, most human deaths in the USA are due to bat-associated RABV (Messenger et al., 3003).

2.3.2. Molecular biology of rabies virus

Rabies virus has a single stranded, negative sense RNA genome of about 12 Kb, which encodes five genes - nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and a viral RNA polymerase (L) - preceded at the 3’ end by a non coding leader

sequence of about 50 nucleotides (LeRNA). The genes are separated by short sequences that represent the intergenic regions of the genome (figure 2.6). With regards to the relative conservation of these genes, the N gene is the most conserved, followed by the L, M, G and P genes (Wu *et al*., 2007).

2.3.2.1. Genes and proteins

Nucleoprotein. The nucleoprotein (N) is a phosphorylated protein consisting of 450 amino acids, which protect the RNA genome from ribonuclease digestion (figure 2.6). It efficiently and specifically encapsidates the viral RNA to form the ribonucleoprotein (RNP) complex, which provides the template for RNA transcription and replication by the viral polymerase complex, which includes the P and L proteins (Tordo and Kouknetzoff, 1993).

Phosphoprotein. The phosphoprotein (P) contains 297 amino acids and is expressed in a variety of phosphorylated forms. It plays a pivotal role as a cofactor in transcription and replication of the viral genome (Jackson & Wunner, 2002, Gupta *et al*., 2000), mediating the connection between the nucleoprotein-RNA complex and the RNA polymerase (Chenik *et al*., 1994). It also binds to free nucleoprotein, thereby preventing the nucleoprotein aggregation in the cytoplasm (Liu *et al*., 2004).

Matrix protein. The matrix protein (M) is the smallest (202 amino acids) and most abundant protein in the virion (Jackson and Wunner, 2002). It is a multifunctional protein: it plays a role in packaging the RNP into new virus particles, adheres to the inner wall of the envelope, interacts with the RNP and wraps the host cell membrane around the newly formed virus particles to form the viral envelope (Kiezle and Alcamo, 2007) (figure 2.6).

Glycoprotein. The glycoprotein (G) is 505 amino acids protein. It is a trimeric type I transmembrane glycoprotein that mediates both receptor recognition and low pH-induced membrane fusion and is responsible for inducing production of and binding with protective, virus-neutralizing antibodies (Mailard and Gaudin, 2002) (figure 2.6).

RNA polymerase. The RNA polymerase or large protein (L) is the largest protein in the virion, which binds to the RNA complex via the phosphoprotein to form a complex responsible of replication and transcription of viral RNA (Chenik *et al*., 1994). It has an important role in the start of infection by initiating the primary transcription of genomic RNA once the nucleocapsid is released into the cytoplasm of the infected cell (Jackson & Wunner, 2002).

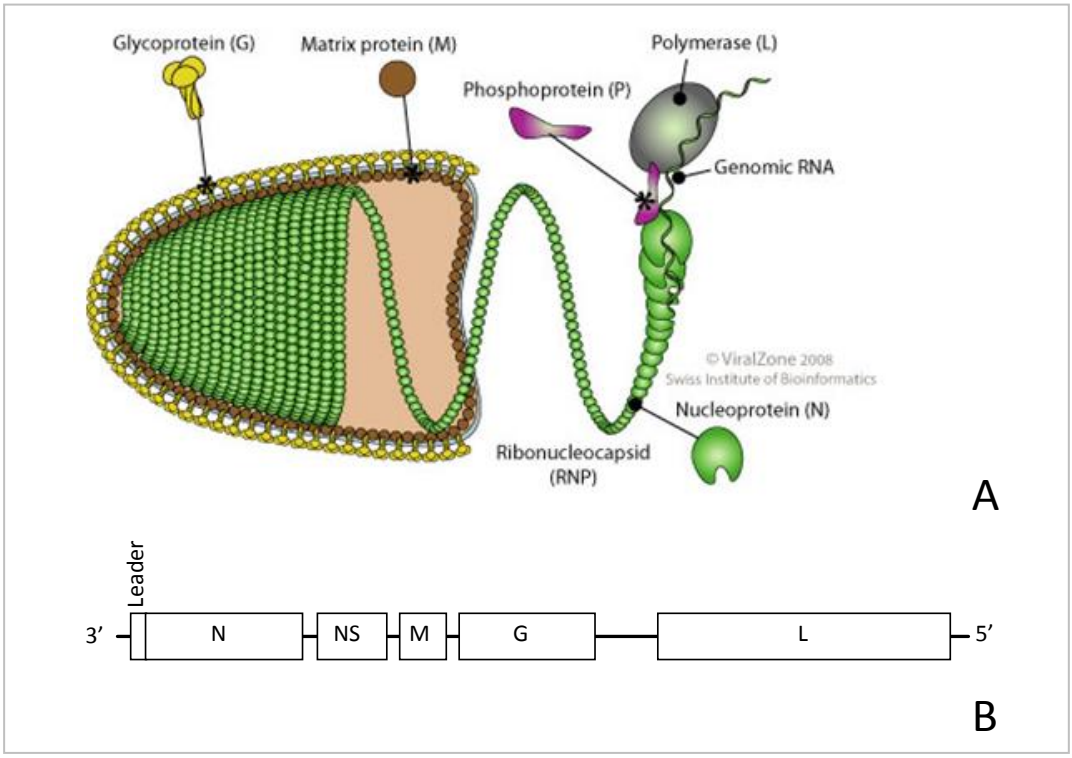


Figure 2.6. **The virion structure of the rabies virus and its genome.** A) The negative-stranded single-stranded RNA genome (12 kilobases) is encapsidated in the nucleoprotein (green), forming the ribonucleoprotein (RNP). Two other viral proteins are associated with the RNP, the viral polymerase (grey) and the phosphoprotein (purple). This complex forms the capsid, which is associated to the membrane with two additional proteins: the matrix protein (brown) and the glycoprotein (yellow). From ViralZone, SIB Swiss Institute of Bioinformatic (http://viralzone.expasy.org/). B) A schematic diagram showing the organization of the Rabies genome.

2.3.2.2. Replication cycle

Rabies virus infection is initiated by attachment of the virus to a receptor on the host cell surface, which is mediated by the rabies virus glycoprotein. However, it is unclear which host cell molecule mediates viral entry into the cell, and recent evidence indicates that several different receptors can be used (Schennel et al., 2009). After binding to its cellular receptor, rabies virus enters into the host cell by endocytosis. A change in the pH inside the vesicle allows the fusion of the viral and endosomal membrane and the release of the viral genome into the cytoplasm (Whitt et al., 1991). Because the entry site of rabies virus is most likely a motor neuron at the neuromuscular junction, which does not provide the biochemical environment required for protein synthesis, the virion is subsequently transported in a

retrograde direction through the axon of the infected neuron (Kelly and Strick, 2000). Two different mechanisms have been suggested for rabies virus transportation through the axon: transportation of the only virus capsid or of the whole rabies virion within the vesicle (Schennel et al., 2009). In this stage the virus keeps a low profile and does not replicate until it has reached the cell body, so that it can evade the immune system and reach the brain.

Once the virus particles arrive at the cell body of the neuron, the production of the virion components begins. Transcription of the viral genomic RNA occurs in the cytoplasm of the infected cell once the RNP core is released from the endosome. The transcription process is carried out on the RNP complex by the RNA-dependent RNA polymerase, which initiates transcription at the 3' end of the genomic RNA, where it first synthesizes a small 55 nucleotide RNA, the leRNA (figure 2.6). This is followed by the transcription of the five individual mRNAs from each gene (N, P, M, G and L), which encode the viral proteins. The polymerase complex stops at the signal sequence (U/ACUUUUUU), ignores the intergenic region of 2–24 nucleotides and restarts transcription at the transcription start signal sequence (UUGURRNGA). Successful re-initiation of transcription at each intergenic junction does not always occur. An estimated 20-30% of the polymerase complexes that reach the gene junction dissociate from the nucleocapsid (Emerson and Yu, 1982). As a result, this "stop-start" mode of transcription leads to relatively high levels of the 3' terminal N mRNA, but progressively less P, M, G and L transcripts (Emerson and Yu, 1982).

Besides regulating the transcription of rabies virus genes, the leRNA appears to contain the encapsidation signal for the nucleoprotein. It has been speculated that when sufficient nucleoprotein is available, encapsidation of the leRNA sequence results in the synthesis of a full-length anti-genome rather the the production of individual viral mRNAs, therefore switching transcription to replication (Liu et al., 2004).

The last phase of the life cycle is the assembly of the viral components, budding and release of the rabies virus virions, which can start a new round of infection. Budding of rabies virus takes place at the plasma membrane, but it is unknown how the capsid is transported to the site of budding, although it has been demonstrated that both rabies virus matrix protein and glycoprotein play an important part in budding (Mebatsion et al., 1996; Mebatsion et al., 1999).

2.3.2.3. Evolution

The molecular diversity of rabies viruses has been increasingly reported both at a national and international level. Seven clades - namely Indian, Asian, Cosmopolitan, Africa-2, Africa-3, Artic related and American indigenous clade - have been identified. With the exception of the American indigenous clade, most of RABV clades are associated totally or in part to dogs or other *Canidae* species, such as foxes, skunks, mongoose, jackals and wolves (figure 2.7).

Indian clade. The Indian clade is distributed only within southern India and Sri Lanka. In the phylogenetic tree of all RABVs associated with terrestrial mammals, it has the most basal position, suggesting that it may have been the progenitor for all RABVs lineages harboured by terrestrial mammals (Bourhy *et al*., 2008). It is maintained in nature by dogs (Nanayakkara *et al*., 2003).

Asian clade. The Asian clade is a heterogeneous lineage, for which the dog is the principal reservoir (Bourhy *et al*., 2008). It includes isolates from China to Indonesia and South-eastern Asia (Gong *et al*., 2010, Nishizono *et al*., 2002, Susetya et al., 2008).

Cosmopolitan clade. The Cosmopolitan clade includes isolates circulating among terrestrial animals in Europe, the Middle-East, Iran, Kazakhstan, Russia, America and the genetic group previously referred to as Africa 1 (Bourhy *et al*., 1993, Kissi *et al*., 1995). This clade has been worldwide distributed mainly throughout human-assisted movement of pets from Europe during colonialism (Bourhy *et al*., 2008, Smith *et al*., 1992).

Africa 2 clade. Viruses from the Africa 2 clade are distributed in West and Central Africa, with the dog acting as host reservoir (Bourhy *et al*., 1993,Kissi *et al*., 1995).

Africa 3 clade The Africa 3 clade comprises viruses well adapted to the carnivores of the family Herpestidae (mainly the yellow mongoose), which is the main vector of rabies in Sub-Saharan Africa (Bourhy *et al*., 1993,Kissi *et al*., 1995).

Arctic-related clade. Arctic-related or Arctic-like includes viruses collected from a large area across the Northern hemisphere, ranging from central to eastern Asia as well as Greenland and North America (Mansfield *et al*., 2006, Nadin-Davis *et al*., 2007, Nadin-Davis *et al*., 1993). Phylogenetic analysis suggests that Indian isolates fall into a basal position, thus suggesting that the entire clade evolved from Indian dog viruses, with subsequent differentiation into genetic sub-clades according to the geographical location (Kuzmin *et al*., 2008, Nadin-Davis *et al*., 2007).

American indigenous clade The majority of the viruses of the American indigenous clade circulate in chiropteran hosts together with a limited number of strains from terrestrial carnivores (Nadin-Davis & Real, 2011). It has been associated to different insectivorous bats in North America and to vampire bats (*Desmodus rotundus*) in Latin America (Nadin-Davis & Real, 2011).

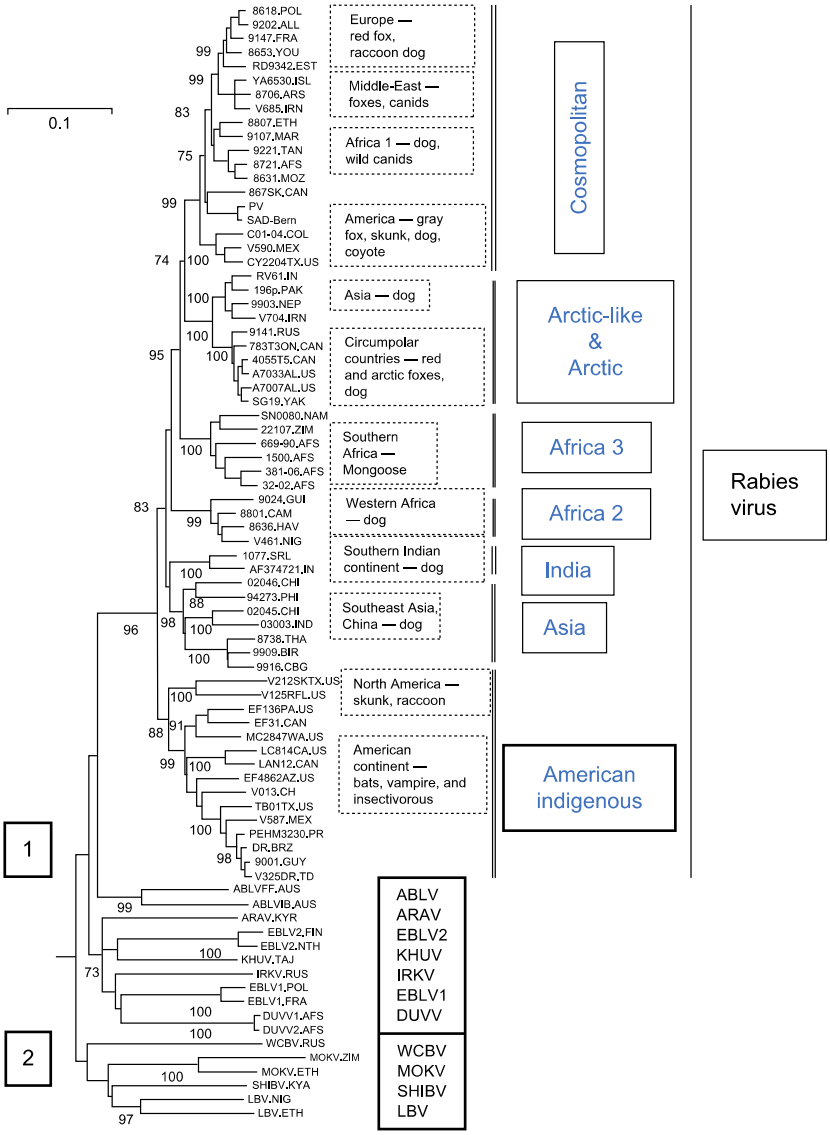


Figure 2.7. Neighbor-joining phylogenetic tree of the N gene sequences of 80 representative RABVs implemented in MEGA 4 software. Numbers at the nodes represent bootstrap values >70. Species assignments of all clades are shown in the boxes at the top right of the figure. The seven RABVs clades (cosmopolitan, Arctic, Africa 2, Afrca 3, India, Asia and American indigenous) are marked in blue. To the left of each clade name, the countries affected and the main reservoir species are indicated. (from Nadin-Davis & Real, 2011).

2.4. SEQUENCING TECHNOLOGIES

2.4.1. Sanger sequencing

Sequencing allows the nucleotide sequence of a portion of DNA to be determined. Sanger sequencing, also called "first generation sequencing technology", is one of the sequencing methods I applied in this thesis. It was developed by Frederick Sanger and colleagues in 1977. Since then, many improvements have been made to the method (Sanger et al, 1992), the most notable being the introduction of fluorescent labeled dideoxy nucleotides with different colours.

The Sanger method is a mixed-mode process involving synthesis of a complementary DNA template using natural 2′-deoxynucleotides (dNTPs) and termination of synthesis using 2′,3′-dideoxynucleotides (ddNTPs) by DNA polymerase (Sanger et al., 1977). For automated Sanger sequencing the four ddNTPs are labeled with different fluorescent tags. The ratio of dNTPs/dye-labeled ddNTPs in the sequencing reaction determines the frequency of chain termination, and hence the distribution of lengths of terminated chains. The dye-labeled DNA fragments are then separated by their size using high-resolution gel electrophoresis. Shorter DNA strands migrate through the gel material more quickly, and come out from the bottom of the capillary first, while longer strands become tangled in the gel material and take longer to emerge out from the bottom. As the strands emerge from the bottom of the capillary they pass through a laser beam that excites the fluorescent dye attached to the dideoxynucleotide at the end of each strand. This causes the dye to glow at a specific wavelength. The determination of the color is the underlying method for assigning a base call, and the order of the fluorescent fragments reveals the DNA sequence (Sambrook and Russel, 2001).

Although nowadays the first generation sequencing technologies are highly reproducible and accessible tools, their application to large projects, such as whole genome sequencing or deep sequencing analysis, is expensive, time consuming and often require prior knowledge of the target genome for specific template application. New methods of sequencing have now been developed (second generation or, more commonly, next generation sequencing), which have entirely revolutionized our ability to sequence, as described in the next paragraphs.

2.4.2. Next Generation Sequencing

Recent advances in nucleic acid sequencing technologies, referred to as *Next Generation Sequencing* (NGS), have produced a true revolution in our ability to sequence. NGS platforms provide unprecedented throughput, generating hundreds of gigabases of data in a single experiment, with a dramatically reduction of the price per information unit (nucleotide) in comparison to first generation sequencing. Moreover, these technologies allow unbiased sequencing without prior knowledge of the complete DNA content in a sample, whilst retaining the flexibility to allow for targeted sequencing.

2.4.2.1. Technologies – Illumina's sequencing

A number of different platforms are currently available, each utilizing different sequencing chemistries and detection strategies. Current NGS methods use a three-step sequencing process:

- Library preparation. During this phase, DNA or cDNA fragments of appropriate lengths are prepared by either breaking long molecules or by using PCR amplification. Adapter sequences are joined to the DNA molecules (by ligation or amplification) and can include a barcode sequence which allows multiplexing of several samples in an experiment.
- Individual DNA molecules in each library are clonally amplified.
- Clonal DNA is sequenced by massive parallel sequencing.

The NGS sequencing platforms use different detection principles including pyrosequencing (454 Life Sciences, acquired by Roche, available since 2005), Illumina's sequencing by synthesis (previously Solexa, available since 2007), SOLiD ligation based sequencing (Life Technologies, available since 2006), and, more recently, the Ion Torrent semiconductor sequencing technology.

Since the Illumina sequencing is the technology applied in this thesis the remaining part of this paragraph will focus on its technical features.

The sequencing technology developed by Illumina, is, like other high throughput sequencing methods, based on sequencing of short DNA molecules (at the time of writing the maximum size is 500 bp). After preparing the DNA fragments of an appropriate length, adaptor sequences are ligated to the fragment ends. The single-stranded fragments are then annealed to a glass surface already covered with small pieces of DNA complementary to the adaptor sequences. These oligos serve to capture the template DNA and as primers for

subsequent amplification. Amplification occurs on the slide by a process termed “bridge amplification”, in which each single stranded molecule binds at both ends to the oligo primers on the slide. Successive rounds of PCR result in the generation of small clusters consisting of many single stranded copies of the same fragment (figure 2.8).

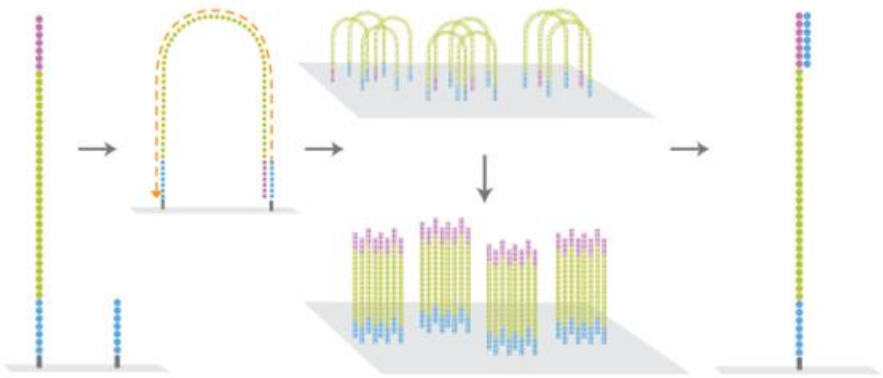


Figure 2.8. Bridge amplification. Free DNA end binds to complementary primer to form a bridge. Strands are copied. A cluster of fragments is formed after many iterations.

When the clusters of fragments have been created, the process of sequencing-by-synthesis begins. Illumina uses fluorescently labelled reversible terminators, such that each single base incorporation on each molecule temporarily terminates the reaction. After incorporation of nucleotides, the unused ones are washed away and a laser is used to scan the surface which makes the fluorescent dyes emit light of different colors, one color for each type of nucleotide. A high-resolution digital image is used to determine which nucleotide is incorporated in each cluster of fragments. The terminator is then removed, allowing the template molecule to be extended again in the next round of sequencing (figure 2.9) (Radford et al., 2012; Pareek et al., 2011)

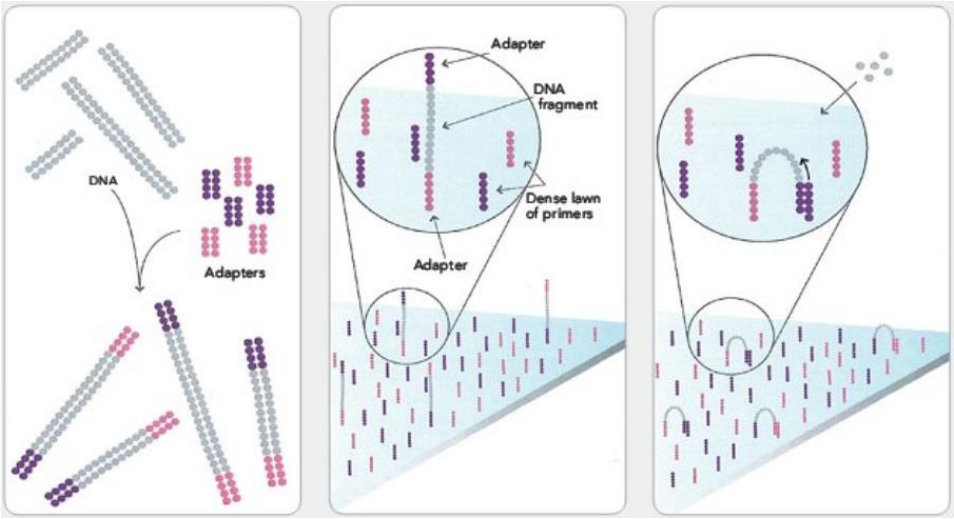


Figure 2.9. Sequencing by synthesis. a) All four labeled reversible terminator, DNA polymerase enzyme and primers are added to the flowcell. b) A laser excites the fluorescent dyes and a camera captures which colors the clusters emit. c) The reversible terminators are removed and the cycle is repeated.

2.4.2.2. Application of NGS to RNA viruses

Characterization of intra-host variability. At the level of the quasispecies, NGS technologies can detect low frequency mutations providing a snapshot of the entire viral population. The technology allows a) the comparison of genetically diverse populations from different replication sites within a host (Stack et al., 2013; Wright et al., 2011); b) the study of the evolution of intra-sample sequence diversity and the modification of the frequency of variants during serial transmission in different hosts (Morelli et al., 2013; Wilker et al., 2013); c) the understanding of the evolutionary dynamics of viral populations under selection pressures, with implications for immune response, antiviral drug resistance and pathogenicity (Capobianchi et al., 2013).

Complete viral genome reconstruction. The advent of NGS has also led to the cost-efficient sequencing of complete viral genomes, even in the case of unknown or poorly characterized viruses, starting either from cultured-enriched viral preparations, or directly from clinical samples through metagenomic sequencing of random library or amplicon genome sequencing (Capobianchi et al., 2013).

Metagenomics. Metagenomics is the determination of the sequence content of a complete microbial community without the need of previous knowledge of the sequences.

Metagenomic applications of NGS include a) the discovery of novel viruses, as, for example, the identification in 2011 of the new Orthobunyavirus, subsequently named Schmallenberg virus, identified in an epidemiological cluster of diseased cattle in Germany (Hoffmann et al., 2011); b) the characterization of the viriome in the environment, in animals or in humans (Capobianchi et al., 2013; Radford et al., 2012).

2.4.2.3. NGS challenges

While the advantages of NGS are numerous, several challenges remain to be addressed. Analysis of NGS data is complicated, due to short read lengths and to error introduction. This includes the errors associated with the sequencing chemistry and the library construction, as well as point mutations and insertions/deletions that may arise during reverse transcription and PCR amplification. For this reason, NGS data need to be "cleaned". This includes filtering to remove low-quality reads, alignment, error correction of reads, variant calling and frequency estimation. Discriminating true biological variants from those deriving from experimental noise is an important issue when trying to identify low frequency variants in a population, for example in viral quasispecies or metagenomic analyses (Beerenwinkel et al., 2011). Each different NGS available platform has its own distinct characteristics in terms of read and error profiles, with the Illumina platform (the one used in this thesis) often regarded as having the lowest error rate (Loman et al., 2012; Quail et al., 2012).

Currently, a multitude of software has been developed to address different aspects of NGS analyses (Beerenwinkel et al., 2011). However, the available algorithms can present some limitations (Finotello et al., 2012) and in-house resolution of bioinformatics problems are often needed to investigate novel datasets and specific hypotheses. In this context, researchers are frequently faced with the need to acquire computer skills and bioinformatics expertise.

A further issue is the scale of genetic data produced by NGS technologies, which presents a physical constraint in terms of data storage and analysis. Large datasets typically require high performance computational clusters, which imply a considerable investment and require sufficient information technology support.

2.5. PHYLOGENETIC AND EVOLUTIONARY ANALYSIS OF RNA VIRUSES

The field of viral evolutionary analysis has greatly benefited from the increasing availability of viral genome sequences, the growth in computer processing power and from the development of sophisticated statistical methods. Application of these methods can help to a) determine the relationship among groups of organisms, b) infer the origin and the onset of a newly emergent epidemic, c) resolve the order of transmission events during an outbreak, d) follow the viral movements between different geographic regions or different tissues of a single infected host. This section will outline the methodologies used to explore the evolutionary biology of viral genome, including methods for phylogenetic inference, estimation of selective pressure and evolutionary analysis.

2.5.1. Phylogenetic inference

A phylogenetic tree describes the relationships among genes or among organisms. One of the first steps in the analysis of aligned nucleotide sequences is the computation of the matrix of genetic distances between all pair of sequences, which provides a measure of the similarity between sequences. The proportion of different sites between two aligned sequences is called observed distance or p-distance. However, p-distance generally underestimates the true genetic distance (d). Indeed, when two sequences are very divergent it is likely that, at a certain position, two or more consecutive mutations have occurred. As a consequence, a model of substitution, which corrects for multiple hits, is necessary. The simplest one-parameter Jukes-Cantor model assumes that all nucleotides (A, C, T, G) occur in equal proportions and that the probabilities of each nucleotide substituting for another are equal among all six nucleotide pairings (A↔C, A↔T, A↔G, C↔T, C↔G, T↔G) (Jukes and Cantor 1969). However, the Jukes-Cantor model is considered an over-simplification of the process of nucleotide substitution and more complex models generally provide a statistically better fit to observed patterns of sequence evolution. The studies presented in this thesis employ the more complex general time reversible (GTR) model (Yang, 1994a). This model permits variation a) in the frequency of the four nucleotide bases and b) in the rate of substitution between nucleotide pairs, allowing the rate substitution between all the six nucleotide pairings to vary. In addition, rate of nucleotide substitution may vary at different sites across the genome. For example, in protein coding sequences third codon positions

mutate usually faster than first and second positions. The Γ-distribution accommodates for varying degree of rate heterogeneity.

The methods for constructing phylogenetic trees from molecular data can be grouped according to the kind of data they use, discrete character states or a distance matrix and according to the algorithm, clustering algorithm and optimality criterion, as illustrated in table 2.1.

Distance methods. Distance methods start by calculating the evolutionary distance between each pair of operational taxonomic units (OUT), usually employing a model of nucleotide substitution to produce a pairwise distance matrix and then infer the phylogenetic relationship from this matrix by a variety of methods, including UPGMA, minimum evolution (ME) neighbour-joining (NJ). Among distance methods, the NJ is currently the most commonly used method to construct a distance tree. Starting with a star-like tree, the principle of this method is to find pairs of OTUs that minimize the total branch length at each stage of clustering of OTUs (Saitou and Nei, 1987).

Maximum parsimony (MP). MP is based on the assumption that the most likely tree is the one that requires the fewest number of changes to explain the data in the alignment. Parsimony operates by selecting the tree or trees that minimise the number of evolutionary steps required to explained the data. Parsimony, or minimum change, is the criterion for choosing the best tree. An algorithm is used to determine the minimum number of steps necessary for any given tree to be consistent with the data. That number is the score for the tree, and the tree or trees with the lowest scores are the most parsimonious trees (Nei and Kumar, 2000).

Maximum likelihood methods (ML). The main idea behind the ML method is to determine the tree topology, branch lengths and parameters of the model of substitution that maximize the probability of observing the sequence data. ML methods outperform other methods in term of accuracy (i.e. finding the correct tree), according to simulation studies (Kuhner and Felsenstein 1995; Huelsenbeck 1995). A main drawback of the ML method is its computational intensity, especially with a large number of sequences. Rather than evaluating every possible tree, heuristic searches use hill-climbing methods to limit the tree space that needs to be explored only to the one adjacent to the temporary tree (Holder and Lewis, 2003). If a tree with a higher likelihood score is found, the search is repeated, and then again, until no further score improvements are made. However, it is possible that this tree represents a

local peak rather than the "summit" of the tree space. To avoid missing the global optimal tree, which may be located beyond the explored tree space, perturbations in tree topology are generated by branch-swapping algorithms, such as nearest-neighbor interchange (NNI), subtree pruning and regrafting (SPR) and tree bisection-regrafting (TBR). These algorithms remove entire sections of the tree and reposition them in other parts of the tree to test any potential improvements in likelihood scores.

Bayesian methods. Bayesian analysis of phylogenies generates a posterior distribution for a parameter, composed of a phylogenetic tree and of a model of evolution, based on the prior for that parameter and the likelihood of the data, generated by a multiple alignment. Similarly to what happens for maximum likelihood, the user postulates a model of evolution and the program searches for the best trees that are consistent with both the model and the data. However, while ML seeks the tree that maximizes the probability of observing the data given that tree, Bayesian analysis seeks the tree that maximizes the probability of the tree given the data and the model of evolution. However, it is usually not possible to calculate the posterior probabilities of all the trees analytically. The Markov chain Monte Carlo (MCMC) method, as implemented by the MrBayes program can be used for this purpose (Ronquist and Huelsenbeck, 2003).

Table 2.1. Classification of phylogenetic analysis methods

	Optimality search criterion	Clustering
Character state	Maximum Parsimony Maximum likelihood Bayesian inference	
Distance matrix	Fitch-Margoliash	UPGMA Neighbour-joining

Maximum likelihood and Bayesian methods are considered the preferable method for tree-building, particularly when evolutionary rates differ among lineages, and therefore are employed in the studies contained in this thesis.

Confidence levels for each node of the phylogeny are statistically assessed using the bootstrapping method (Felsenstein, 1985). The bootstrapping method assesses the reliability of a phylogenetic structure by determining the proportion of "pseudoreplicate" datasets

supporting each node. These datasets are generated by randomly sampling the original character matrix (the columns of the alignment) to create new matrices of the same size as the original. These proportions (expressed as percentages) can be used as a measure of the reliability of individual branches in the optimal tree, with percentages >70 considered statistically significant.

The basis for phylogenetic analysis is the identification of discrete clades of genetically related viral isolates that cluster together on a tree and share a unique common ancestor. Clades (or genetic clusters) are topologically separated by long branches and are considered statistically well supported when bootstrap values exceed 70 (Felstein, 1985).

2.5.2. Estimating selection pressure

Molecular adaptation by natural selection is one of the most important processes in biology. Mutation generates multiple genetic variants that can be fixed or eliminated from the population. This depends on a) the degree to which the mutation increases or decreases an individual's ability to survive and reproduce in the current environment (fitness) and b) the size of the population. The effect of selection is to increase the frequency of beneficial mutations (those that increase an individual's fitness) until they become fixed in the population (positive selection), or to decrease the frequency of deleterious mutations (those that decrease an individual's fitness) until it is eliminated (negative selection). Selection does not affect the frequency of mutations that have no appreciable fitness effect, the neutral mutations.

Mutations in protein coding sequences can be classified as either synonymous (silent) or non-synonymous (replacement). If we assume that selection acts less strongly on silent mutations, comparison of synonymous (α) and non-synonimous (β) changes should reflect the action of natural selection. The ratio $\omega=\beta/\alpha$ (also referred to as dN/dS) has become a standard measure of selective pressure (Nielsen & Yang, 1998). $\omega\approx1$ signifies neutral evolution, $\omega>1$ negative selection and $\omega>1$ positive selection.

In this thesis, my main concern is the study of site-by site selection. This estimate has been made using different statistical methods available on the Datamonkey online version of the Hy-Phy package.

Single likelihood ancestor counting (SLAC). SLAC is a quick counting method, which estimates the number of nonsynonymous and synonymous changes which have occurred at

each codon position throughout the evolutionary history of the sample. This approach involves reconstructing the ancestral sequences using likelihood-based methods and nucleotide and codon substitution parameters. SLAC is computationally fast, is adapted to process large dataset (>50 sequences) and does not make any assumption regarding the distribution of rate across sites. However, SLAC as other counting methods may lack power, especially for data sets comprising a small number of sequences or low divergence sequences, as the power of the test is limited by the total number of inferred substitutions at a site. In addition, counting the number of changes between ancestral states may underestimate the true number of substitutions, and hence, the number of changes inferred using this approach may not accurately reflect the rate at which a site is evolving (Kosakovsky Pond and Frost, 2005).

Random effects likelihood (REL). The REL method originally proposed by Nielsen and Yang (1998) assumes a distribution of substitution rates across sites and infers the rate at which individual sites evolve. This inference can be based on the maximum likelihood estimates of the rate parameters (empirical Bayes) or on the posterior distribution of rate parameters (Hierarchical Bayes approach). The latter takes into account the rate distribution parameters, which can be subject to error, but, on the other hand, can be affected by an inadequate prior assumptions of the distribution of rate parameters. Consequently, for small or low divergence datasets, the errors in estimation of the rate distribution may be large, such that empirical Bayes approaches may give misleading results, while a hierarchical Bayes approach may be very sensitive to prior assumptions (Kosakovsky Pond and Frost, 2005).

Fixed effects likelihood (FEL). The FEL method estimates the substitution rates directly at each site. Like counting methods, FEL makes no assumption regarding the distribution of rates across site, making the estimation of α and β potentially more accurate. However, this method may require a substantial number of sequences (20-30) to gain power. In addition, parameters such as nucleotide substitution, codon frequencies and branch lengths are treated as known rather than subjected to error. However, this assumption has been shown to not lead to false positives (Kosakovsky Pond and Frost, 2005).

Mixed Effects Model of Evolution (MEME). The MEME is a generalization of FEL that is capable of identifying instances of both episodic and pervasive positive selection at the level of an individual site. Differently from FEL, which assumes that the same dN/dS (ω) ratio applies to all branches, MEME models variable ω across lineage at an individual site. MEME is nearly always preferable to FEL because it matches the performance of FEL when

there is no lineage-to-lineage variation in dN/dS, and significantly improves upon it when such variation is present (Murrel et al., 2012).

Fast, Unconstrained Bayesian AppRoximation (FUBAR). The FUBAR is an approximate hierarchical Bayesian method using a Markov chain Monte Carlo (MCMC) routine, which ensures robustness against model misspecification by averaging over a large number of predefined site classes. This leaves the distribution of selection parameters essentially unconstrained, and also allows sites experiencing positive and purifying selection to be identified orders of magnitude faster than by existing methods (Murrel et al., 2013).

2.5.3. Bayesian evolutionary analysis by sampling trees

The BEAST software package is a Bayesian statistical framework implementing Markov chain Monte Carlo (MCMC) algorithms for phylogenetic inference, divergence time dating, coalescent analysis, phylogeography and other molecular evolutionary analyses.

The BEAST software package explicitly models the rate of molecular evolution for each branch of the phylogenetic tree, either assuming a uniform rate across all branches (based on a "strict" molecular clock) or allowing rates to vary among lineages (a "relaxed" molecular clock). In addition, BEAST implements various models for the changes in the population size over time (e.g. constant, exponential, logarithmic, expansion, Bayesian skyline or skyride), different nucleotide substitution models (including HKY and GTR), models that allow different rates for the 1st, 2nd and 3rd codon positions, amino acid substitution models (Blosum62, CPREV, Dayhoff, JTT, MTREV and WAG) and so on (Drummond et al., 2013, Drummond and Rambaut, 2007).

The BEAST package also analyzes the demographic history of a population of sequences. Under coalescent theory, the demographic history of a viral population can be viewed as a series of coalescent events with sequences traced back in time along parallel lineages to points of convergence that represent a most shared common ancestor (MRCA) (Drummond et al., 2013, Drummond and Rambaut, 2007).

Phylogeography inference is another important tool to trace the patterns of virus dispersal available in BEAST. Reconstructing the spatial process provides fundamental understanding of the evolutionary dynamics underlying epidemics. BEAST enables the reconstruction of timed viral dispersal patterns while accommodating phylogenetic uncertainty. Standard Markov model inference is extended with a stochastic search variable selection (BSSVS)

procedure that allows to construct a Bayes factor test that identifyes the most parsimonious description of the phylogeographic diffusion process (Lemey et al., 2009).

2.5.4. BaTS – Bayesian Tip-association Significance testing

BaTS is a Bayesian MCMC approach to quantify the association of phenotypic characters (e.g. geographic locations, physical characteristics, epidemiological risk group, cell tropism, and so on) with the shared ancestry, as represented by a viral phylogenetic tree. Assigning a discrete character to each taxa of a phylogenetic tree, BaTS determines if closely related taxa are more likely to share the same character than expected by chance alone. If the trait represents geographic location, a strong association (the character is tightly correlated with phylogeny) reflects low lineage migration, on the contrary a weak association (the character is fully interspersed in the phylogny) represents high rates of gene flow. However, in many cases the phylogenetic distribution of the characters may be intermediate and their correlation with phylogeny less clear. BaTS allows to quantify and statistically test the strength of the association against the distribution of characters expected by chance. Starting from the posterior sample of trees, calculated by BEAST, BaTS estimates the parsimony score (PS) statistic, which represents the number of state changes in the phylogeny and the association index (AI) statistic. This takes into account the shape of the phylogeny by measuring the imbalance of internal phylogenetic nodes and the monophyletic clade (MC) statistic for each trait value, which quantify the phylogeny-trait associations, based on the dimension of the monophyletic clades whose tips all share the same trait value. This method incorporates statistical errors arising from phylogenetic uncertainty, returning the posterior distributions (95% confidence intervals) of the statistics and providing the significance estimation for these (Parker et al., 2007).

2.6. AIMS OF THE THESIS

RNA viruses are characterized by large population size, high replication rate and short generation time, which, collectively, are responsible for extremely high genetic variability. As a consequence, each viral population consists of dynamic mutant spectra (termed quasispecies) rather than of a defined genome sequence. Genetic diversity allows the viruses

to evolve in an ever-changing environment with shifting selection pressure. The advantage of maintaining multiple subpopulations is that, when the virus is shifted to a new environmental condition, a variant, which will be more fit in the new environment, may already be present in the population.

The objective of my research is to provide a better understanding of the mechanisms underlying the evolutionary dynamics of RNA viruses. This would be beneficial to improve our capacity to identify effective target surveillance and potentially to predict evolutionary trajectory for appropriate selection of control strategies.

To this end I applied first (Sanger sequencing) and second generation sequencing technologies and recently developed bioinformatics tools to explore the evolution of viral pathogenicity (chapter 5), the inter-host transmission of viral variants and their fixation in the viral population (chapter 5 and 6), the pattern of viral gene flows (chapters 3 and 5) and the evolution of viruses under different selection pressures (chapter 4).

Chapter 3

The introduction of fox rabies into Italy (2008-2011) was due to two viral genetic groups with distinct phylogeographic patterns

3.1. ABSTRACT

Fox rabies re-emerged in north-eastern Italy at the end of 2008 and circulated until early 2011. As with previous rabies epidemics, the Italian cases were linked to the epidemiological situation in adjacent regions. To obtain a comprehensive picture of the dynamics of the recent Italian epidemic, we performed a detailed evolutionary analysis of RABVs circulating in north-eastern Italy. Sequences were obtained for the hyper-variable region of the nucleoprotein gene, the complete glycoprotein gene, and the intergenic region G-L from 113 selected fox rabies cases. We identified two viral genetic groups, here referred to as Italy-1 and Italy-2. Phylogenetic and phylogeographic analyses revealed that both groups had been circulating in the Western Balkans and Slovenia in previous years and were only later introduced into Italy (into the Friuli Venezia Giulia region-FVG), occupying different areas of the Italian territories. Notably, viruses belonging to the Italy-1 group remained confined to the region of introduction and their spread was minimised by the implementation of oral fox vaccination campaigns. In contrast, Italy-2 viruses spread westward over a territory of 100 km from their first identification in FVG, likely crossing the northern territories where

surveillance was inadequate. A genetic sub-group (Italy-2A), characterised by a unique amino acid mutation (D106A) in the N gene, was also observed to occupy a distinct geographic cluster. This molecular epidemiological analysis of the 2008-2011 fox rabies epidemic will contribute to future control programmes both at national and regional levels. In particular, our findings highlight the weaknesses of the national surveillance strategy in the period preceding rabies re-emergence, and of control plans implemented immediately after rabies notification, and underline the need of a coordinated approach at the regional level for both the surveillance and control of wildlife rabies.

3.2 INTRODUCTION

Terrestrial rabies in Europe is caused by rabies virus (RABV; Rhabdoviridae) of the Cosmopolitan clade (Bourhy et al., 1999). Two main animal reservoir species are responsible for disease maintenance in the European continent – the red fox (Vulpes vulpes) and the racoon dog (Nyctereutes procyonoides) – with the latter mainly distributed in North-eastern Europe (Bourhy et al., 1999). The ancestry of this viral lineage in Europe dates back to the 18th century, before it spread widely on a global scale by means of human-mediated dispersal and locally by natural wildlife migration (Bourhy et al., 1999,Badrane and Tordo, 2001,Smith et al., 1992). A phylogenetic analysis of European RABV revealed a number of distinct groups, each associated with a specific geographical area, as well as the westward and southward spread of the virus across Europe during the last century(Bourhy et al., 1999). A recent study tracing the evolution of RABV in the Balkans supported the phylogeographic pattern previously suggested for RABV in Europe (McElhinney et al., 2011). In particular, this study confirmed that the red fox is responsible for the maintenance of several sub-lineages of viruses in this region, either those documented previously – namely the Eastern European (EE) and Western European (WE) lineages (Bourhy et al., 1999) – or newly described, including the so-called “Serbian fox” group (McElhinney et al., 2011).

The north-eastern territories of Italy were affected by rabies in the 1970s and 1980s, and more recently during the period 1991 to 1995. Epidemics of rabies in foxes were linked to infections in Austria and the nearby territories of former Yugoslavia. For this reason, the risk of rabies re-introduction into Italy from the bordering areas has long been recognised.

Vaccination campaigns using oral rabies vaccines have been conducted targeting the fox population in these areas in 1989 and between 1992 and 2004 (Mutinelli et al., 2004). The last case of rabies was diagnosed in a fox in the province of Trieste on the border with Slovenia in December 1995, and Italy has been classified as free from terrestrial rabies since 1997.

In late 2008, fox rabies re-emerged in North-eastern Italy, with a total of 287 rabies cases confirmed in animals by February 2011. Despite the implementation of four oral rabies vaccination (ORV) campaigns in the affected territories in 2009, the infection spread westwards reaching the province of Belluno (Veneto region) in November 2009 (De Benedictis et al., 2009,Nouvellet et al., 2013). Due to the epidemic recrudescence, an emergency ORV campaign was undertaken in January 2010, followed by three emergency campaigns in the same year (Mulatti et al., 2011). The peak of the epidemic was recorded in the first half of 2010 (199/287 – from January to June), with 174 cases notified in the Belluno province. Infection has predominantly affected the red fox population (242/287) with occasional spills-over into other wild species and domestic animals (45/287) (figure 3.1). Emergency vaccinations help alleviate the epidemic, with the last confirmed case in a fox in February 2011. Two ORV campaigns in both 2011 and 2012 were also carried out (Nouvellet et al., 2013).

As with previous rabies epidemics, the recent Italian outbreaks have been linked to the epidemiological situation in the adjacent Balkan region (De Benedictis et al., 2008). To help inform future strategies for rabies control, we aimed to obtain a comprehensive picture of the epidemiological and evolutionary dynamics of these outbreaks. To this end we undertook a detailed evolutionary analysis of recent RABVs circulating in north-eastern Italy.

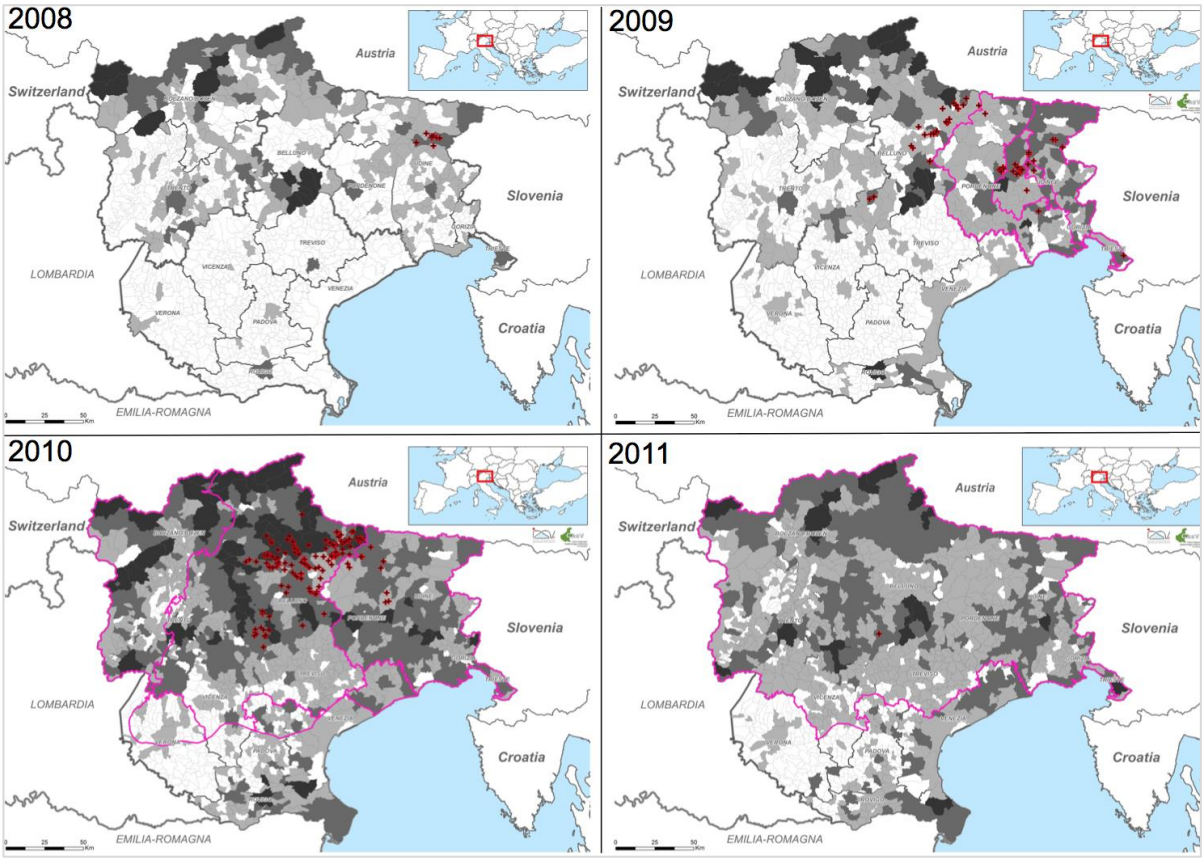


Figure 3.1. Annual distribution of red foxes collected through passive surveillance and assessment of vaccination efficacy in the Italian territories during the 2008-2011 epidemic (Nouvellet et al., 2013). Different grey gradations indicate different average levels of red fox sampling: bright grey, deep grey and black correspond to 1-4, 5-15 or more than 16 collected foxes, respectively. Black and red daggers: positive rabies cases detected in red foxes. Purple lines: areas subject to ORV.

3.3. MATERIALS AND METHODS

3.3.1. Sample collection and criteria for selection of viruses

Most of samples analysed here were submitted for testing at Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe) within the framework of enhanced passive surveillance, comprising dead animals found by hunters and forest rangers. In a few cases animals were killed due to the observation of severe neurological signs. However, to monitor the impact of the vaccination campaigns on fox populations, foxes were also sampled following these vaccination campaigns, and were tested for presence/absence of viral infection to assess both the vaccination immunity and the rate of bait uptake (Nouvellet et al., 2013).

To investigate the genetic diversity of RABV circulating in north-eastern Italy, a fragment of approximately 600 nucleotides (nt) of the N gene was sequenced from all rabies positive brain specimens (n=287). Only red fox samples were used for further analyses, as the others likely represent cases of transient spill-over (n=242/287). Of these, the identical sequences of samples collected from the same municipality and in the same month have been removed, leaving a data set of 128 sequences. Finally, only good quality sequences (i.e. with no degenerate bases) with a minimum length of 480 nt were retained for sequence analyses (n = 113 sequences). Additional sequencing of the entire G gene (1572 nt) and the intergenic region G-L (520 nt) was performed for these 113 selected fox brain specimens. For all samples, the date of sampling and geographical location were available. Relevant epidemiological information for all the RABV isolates analysed in this study is provided in the Annex A (Annex A - TableA1).

3.3.2. RNA extraction, RT-PCR and sequencing

Viral RNA was extracted from brain specimens using the Nucleospin RNA II kit, according to the manufacturer's instructions (Macherey–Nagel GmbH & Co., Düren, Germany). Briefly, one hundred microliters of sample suspension were used for the extraction, and RNA was eluted in a final volume of 60 µl and stored at –80 °C.

One-Step RT-PCR amplification was performed using the Qiagen OneStep RT-PCR kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. Primers used for the amplification of partial N gene sequences and complete G gene and G-L intergenic region sequences are listed in the Annex A (Annex A - Table A2). PCR products were analyzed for purity and size by electrophoresis in 2% agarose gel after staining with GelRedTM Nucleic Acid Gel Stain (Biotium, Hayward, CA). Amplicons were subsequently purified with ExoSAP-IT (USB Corporation, Cleveland, OH) and sequenced in both directions using the Big Dye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). The products of the sequencing reactions were cleaned-up using the Performa DTR Ultra 96-well kit (Edge BioSystems, Gaithersburg, MD) and analyzed on a 16-capillary ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

3.3.3. Phylogenetic analysis

Sequences were manually aligned using SeAl (http://tree.bio.ed.ac.uk/software/seal/) and compared with publicly available sequences obtained using remote BLAST searches in GenBank. Maximum likelihood (ML) trees were estimated using the best-fit general time-reversible (GTR) model of nucleotide substitution with gamma-distributed rate variation among sites, and a heuristic SPR branch-swapping search available in PhyML version 3.0 (Guindon and Gascuel, 2003). Parameter values for the GTR substitution matrix, base composition, gamma distribution of the rate variation among sites (with four rate categories, Γ4) were estimated directly from the data using MODELTEST(Posada and Crandall, 1998). A bootstrap resampling process (1,000 replications) using the neighbor-joining (NJ) method was used to assess the robustness of individual nodes of the phylogeny using PAUP* 4.0 package (Wilgenbusch and Swofford, 2003) and incorporating the substitution model obtained by MODELTEST as described above. The ML tree topology was also compared to that obtained using the Bayesian phylogenetic method available in the MrBayes v.3.1.2 package (Ronquist and Huelsenbeck, 2003), again using the model of nucleotide substitution estimated by MODELTEST, with posterior probabilities values used to depict support for individual groupings. Amino acid substitutions along the branches of the phylogeny were identified using the parsimony algorithm available in MacClade program (Maddison and Maddison, 1989).

3.3.4. Estimating a time-scale for the RABV evolutionary history

The Time to the Most Recent Common Ancestor (TMRCA) of the Italian RABV sequence data was estimated using the concatenated N and G genes of the Italian viruses using the BEAST program v1.6.1 (Drummond and Rambaut, 2007), which employs a Bayesian Markov chain Monte Carlo (MCMC) approach. For each analysis, a flexible Bayesian skyline (BSP) coalescent tree prior (Drummond et al., 2005) and the HKY85 model of nucleotide substitution were used. Prior to the BEAST analysis we plotted genetic distances estimated from the ML tree (above) against day of sampling using the Path-O-Gen program (kindly provided by Andrew Rambaut, University of Edinburgh) to investigate the temporal signal in the data. Due to the low sequence divergence, the short sampling time span, and the insufficient phylogenetic signal, this analysis revealed no clear accumulation of sequence divergence over the sampling time span. For this reason, we specified informative

normal prior distributions on the evolutionary rate in BEAST and assuming a strict molecular clock. For this, we employed a mean substitution rate of 3.30 x 10-4 (95% HPD, 2.02 x 10-4 - 4.82 x 10-4) nucleotide substitutions per site per year (sub/site/year) for the G gene, 2.89 x 10-4 (95% HPD, 1.34 x 10-4 - 4.60 x 10-4) sub/site/year for the N gene, and 3.40 x 10-4 (95% HPD, 2.57 x 10-4 - 4.28 x 10-4) sub/site/year for the concatenated N and G-gene sequences, according to the results obtained by McElhinney et al. (2011). Similar rates have been obtained in other studies of RABV (reviewed in (Nadin-Davis and Real, 2011)) and hence are likely to be robust. In all cases, uncertainty in the data is reflected in the 95% highest probability density (HPD) values and, in each case, chain lengths were run for sufficient time to achieve convergence as assessed using the Tracer v1.5 program (Drummond and Rambaut, 2007). Finally, maximum clade credibility (MCC) phylogenetic trees were summarized from the posterior distribution of trees generated by BEAST using the program TreeAnnotator v1.5.3 (Drummond and Rambaut, 2007) after the removal of an appropriate burn-in (10% of the samples). The MCC trees were visualized using the program FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/), which allowed us to depict the divergence time for individual nodes in the trees.

3.3.5. Phylogeography of RABV in Italy

To determine the spatial dissemination of the RABVs in Italy and the Balkans, the sequences of the N gene were grouped into six geographic regions, namely; (i) Italy-A (yellow area in figure 3.3, corresponding to the South-central part of the Friuli Venezia Giulia region, FVG), (ii) Italy-B (green area, including the northern part of FVG, the mountainous areas in Belluno province, north-western area of the autonomous province of Trento and Bolzano), (iii) Italy-C (blue area, covering the piedmont areas in the Veneto region and the south-eastern part of the autonomous province of Trento), (iv) Bosnia and Herzegovina (BiH), (v) Croatia, and (vi) Slovenia. Phylogenetically divergent sequences from other geographic regions were excluded. This resulted in a final data set of 160 viruses that could be used for in-depth spatial analysis. Given the limited number of G gene sequences closely related to the Italian viruses available on GenBank (23 sequences from Slovenia and 1 sequence from BiH), this phylogeographic analysis was performed on the N gene data set and the concatenated N and G-gene sequences only.

Posterior distributions under the Bayesian phylogeographic model (Lemey et al., 2009) were estimated using a MCMC method implemented in BEAST. This model incorporated the date of sampling (year) and used a strict molecular clock, BSP prior, and the SRD06 model of nucleotide substitution. This method estimates a reversible diffusion rate for each potential diffusion pathway among the predefined locations while simultaneously estimating evolutionary and coalescent parameters, thereby allowing quantification of the uncertainty in ancestral state reconstructions given the available sample (i.e. ancestral geographic locations). We used a Bayesian stochastic search variable selection (BSSVS) to identify the links between these locations among the posterior sets of trees that explain the most likely migration patterns among the rabies viruses under study (Lemey et al., 2009). The BSSVS enables us to estimate Bayes factors (BF) that depict the support for individual transition rates between pairs of location states. We calculated the BFs using SPREAD (Bielejec et al., 2011) and consider rates yielding a BF> 5 (Annex A - Table A3) as well supported diffusion rates between specific localities. All parameters reached convergence, as visually assessed using Tracer v1.5. A MCC phylogenetic tree was summarized from the posterior distribution of trees generated by BEAST using the program TreeAnnotator v1.6.1 (Drummond and Rambaut, 2007) after the removal of 10% of the samples as burn-in, and visualized using FigTree v1.3.1. To summarize the posterior distribution of ancestral location states, nodes in the MCC tree were annotated with the modal location state for each node. The expected number of location state transitions conditional on the observed data was obtained using Markov jump counts (Minin and Suchard, 2008a,Minin and Suchard, 2008b) using BEAGLE (Suchard and Rambaut, 2009).

3.3.6. Nucleotide sequence accession numbers

The nucleotide sequences obtained in this study are available from GenBank under accession numbers KC197856 to KC197968 for the entire G gene and the intergenic region G-L, and EPI223068 to EPI223123 for the N gene.

3.4. RESULTS

To investigate the genetic diversity of RABVs circulating in North-eastern Italy, the G and N genes of 113 Italian strains under study were aligned and compared with all the related sequences of viruses from Europe available on GenBank (120 for the N gene and 62 for the G gene). Bayesian phylogenetic trees inferred for the two genes confirmed the topologies of the ML phylogenies (figure 3.2 and Annex A - Figure A1). As expected, all RABVs circulating in north-eastern Italy from 2008 to 2011 cluster within the Cosmopolitan lineage. In particular, they belong to the 'Western European sublineage', which includes viruses from France, Austria, Belgium, Germany, Slovenia, Croatia, BiH and Montenegro (Bourhy et al., 1999,McElhinney et al., 2011,Rihtaric et al., 2011,Lojkic et al., 2012) and show the greatest similarity with strains from Western Balkans (Croatia and BiH) and Slovenia.

Analysis of the N and G phylogenies of the Italian fox samples identifies two clearly distinct genetic groups of RABVs with relatively strong nodal support, designated Italy-1 and Italy-2 (Figure 3.2, 3.3 and Annex A - Figure A1). These groups exhibit 31 nucleotide differences in the G gene and the intergenic region G-L, with a single amino acid signature in the G protein - 102R for Italy-1 and 121T for Italy-2 (antigenic site II). The Italy-1 group includes all the samples collected in the eastern and central part of FVG (area A in figure 3.3) from October 2008 to April 2010, with the exception of one virus (RS09-25/fox/dec08/Udine) that clusters within the Italy-2 group. This second group includes all the sequences of fox samples collected between December 2008 and February 2011 in the Belluno province (Veneto region), the autonomous provinces of Trento and Bolzano and the western part of Udine province in FVG (areas B and C in figure 3.3, figure 3.2 and Annex A - Figure A1). Interestingly, RS09-25/fox/dec08/Udine, the oldest isolate of the Italy-2 group and the only sequence from area A of FVG belonging to this genetic group, was collected in December 2008, at the beginning of the epidemic, and in the same municipality (Resia, Lat: 46°21'21.59"N; Lon: 13°17'40.16"E) that notified the first rabies case in October 2008. In addition, the Italy-2 group contains a distinct cluster of viruses in the N gene phylogeny, termed the Italy-2A subgroup, which represents viruses collected in 2010 from a limited area including the far western province of Belluno, and the autonomous provinces of Trento and Bolzano (grey dots in figure 3.3). Viruses clustering in the Italy-2A subgroup contain a unique amino acid mutation (D106A) in the N gene.

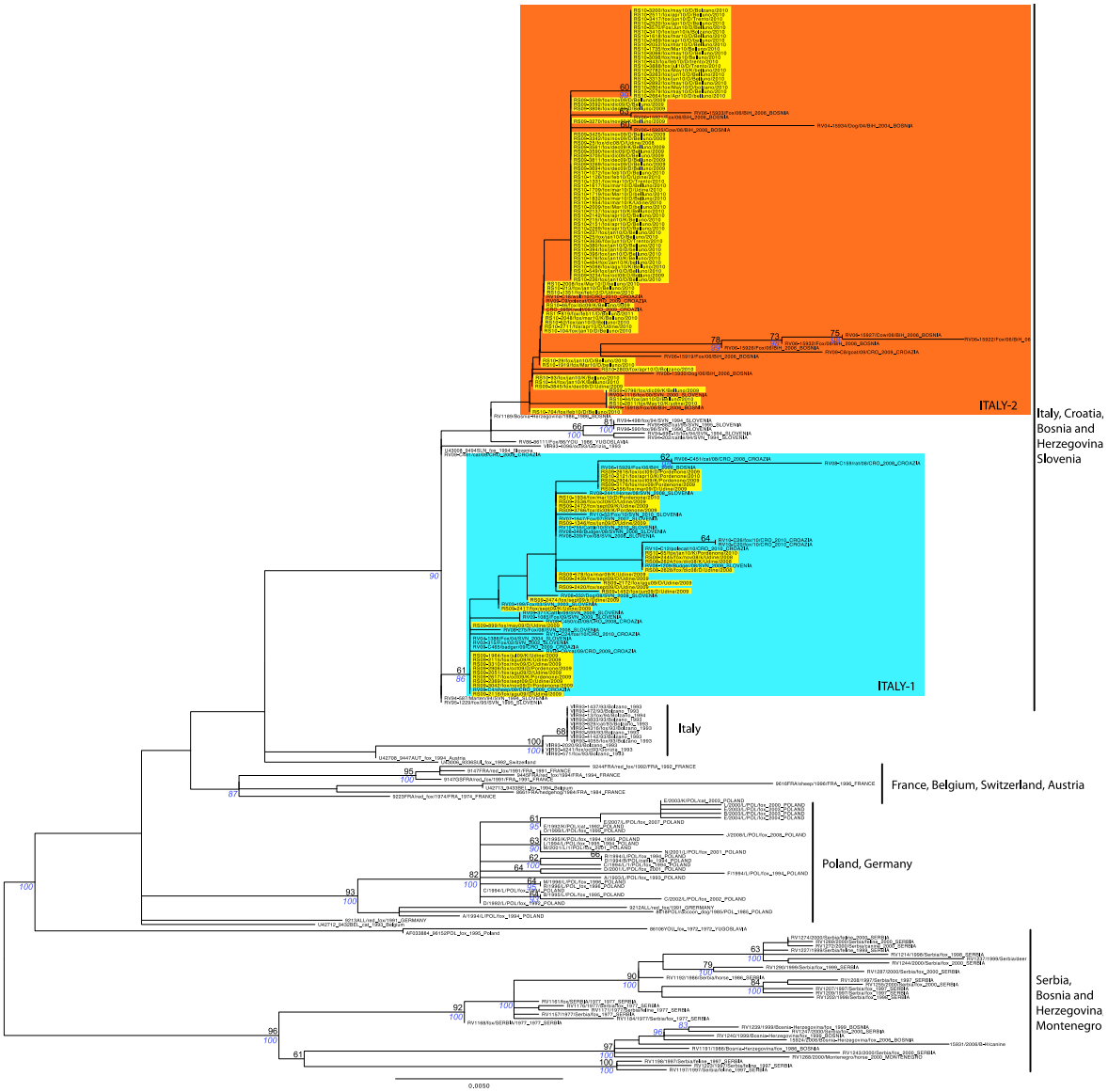


Figure 3.2. ML phylogenetic tree of the N gene of RABVs collected in Italy 2008-2011, combined with those from neighbouring countries. The sequence identification includes: sample registration number, species of origin, month and year of collection, province where the carcass was collected. The numbers at branch points represent bootstrap values (>60, in black) or posterior probabilities (>80, in blue). The Italian samples are highlighted in yellow. Two different genetic groups are indicated in different colours: blue – Italy-1 and orange – Italy-2.

Our phylogenetic analysis also provided strong clues as to the origin of the Italian viruses. Specifically, the N gene phylogeny (figure 3.2) showed that viruses belonging to Italy-1 group clustered with viruses from Croatia (2008 to 2010), BiH (2006) and Slovenia (2002 to

2010), while Italy-2 samples were closely related to viruses collected in BiH (2004-2006), Croatia (2009 to 2010) and Slovenia (2000). Hence, these two distinct RABV lineages have clearly been circulating in the Balkans prior to their introduction into Italy.

We estimated the TMRCA for Italy-1 and Italy-2 groups using a Bayesian coalescent method and previously established rates of nucleotide substitution. For the concatenated N and G genes data sets the 95% HPDs for the TMRCA were January 2007-September 2008 for Italy-1 (mean of March 2008) and May 2008-December 2008 for Italy-2 (mean of October 2008). These dates overlap with the period of circulation of these two clusters in the neighbouring countries to the East. We also estimated TMRCAs of the two Italian groups using the data set including sequences from Balkans (Croatia, BiH and Slovenia). Given that only the year of collection was available for these viruses, the estimation of the tMRCAs was less precise and ranged from 2004 to 2007 (95%HPD) for Italy-1 and from 2006 to 2008 (95% HPD) for Italy-2. Although we did not observe a perfect overlap between these two sets of estimates, both are compatible with the introduction of Italy-2 after the Italy-1 group. The uncertainty in these date estimates, particularly for Italy-1, likely reflects the limited number of sequences available for study and their low levels of genetic variation.

To examine the extent and pattern of geographical structure in these RABV data in more detail, we inferred a Bayesian (MCC) phylogeny for the N gene (figure 3.4) and the concatenated N and G-gene sequences (Annex A - Figure A2), in which each RABV was assigned to a different geographic area. The MCC trees confirmed the existence of two major groups co-circulating in Italy. Both included samples collected from Italy, BiH, Croatia and Slovenia. More specifically, the MCC trees annotated with most probable nodal locations confirmed the existence of two distinct introductions of Italy-1 and Italy-2 groups in Friuli Venezia Giulia (corresponding to area A in figure 3.3), with viral migration of the Italy-2 group toward western areas of northern Italy (from area A to area B and from area B to area C in figure 3.3, Annex A - Figure A2).

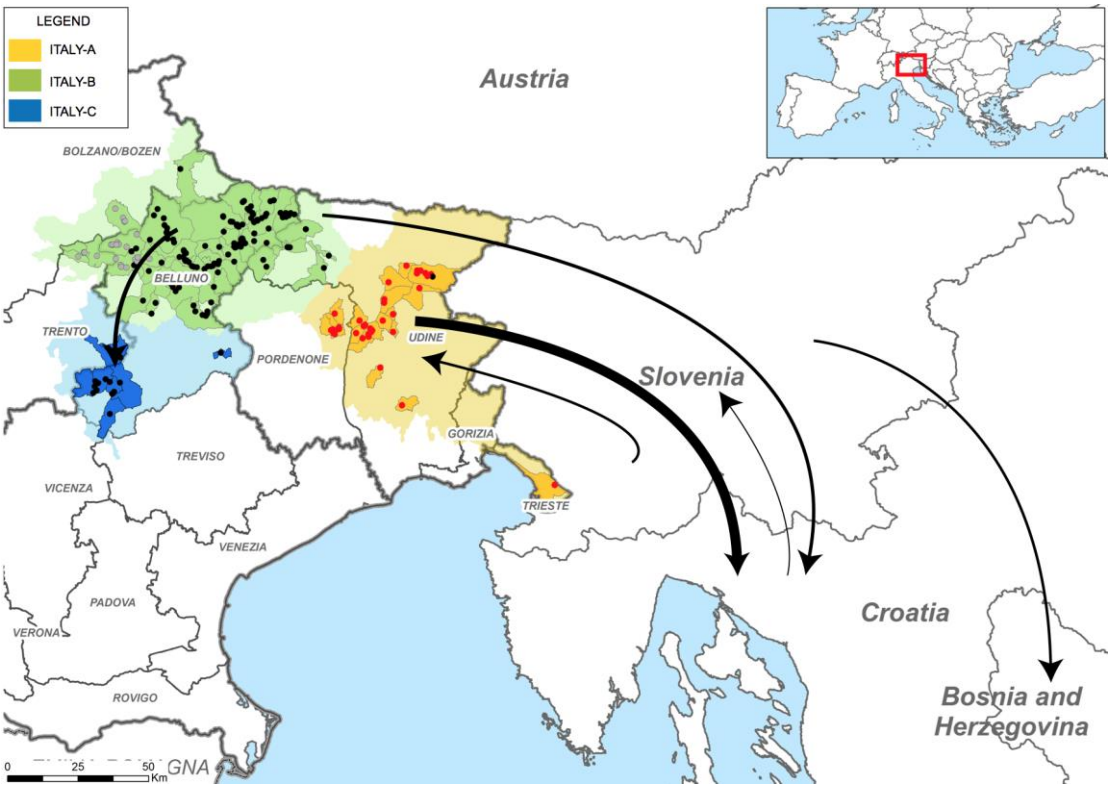


Figure 3.3. The Italian territories that experienced the fox rabies epidemic from 2008 to 2011. Geographical areas are coloured as follows: yellow: southern and central part of FVG (Italy A); green: northern part of FVG, the mountainous areas in the Belluno province, north-western area of the autonomous provinces of Trento and Bolzano (Italy B); blue: piedmont areas in the Veneto region and the south-eastern part of the autonomous province of Trento (Italy C). Red, black and grey circles indicate the RABVs detected in foxes belonging to Italy-1, Italy-2 and -2A genetic groups, respectively. Municipalities where RABV positive samples were identified are highlighted with darker colours. Arrows indicate the well-supported (Bayes factor > 5) non-zero rates (Annex A - Table A3). The thickness of the arrows is proportional to the median Markov jumps obtained for those specific transitions.

Our phylogeographic study suggests that the ancestor strains for group Italy-1 were likely introduced from Slovenia (according to the N MCC tree, figure 3.3) or Croatia (according to the N-G MCC tree, Annex A - Figure A2), while group Italy-2 appears to have been introduced from BiH, with a plausible westwards spread, although back migration events also seem to occur, for example to Croatia within the Italy-2 clade (figure 3.3 and Annex A - Figure A2). Although the direction of viral spread is not always clear, both MCC trees provide evidence for migration between Croatia and Italy (area A and B).

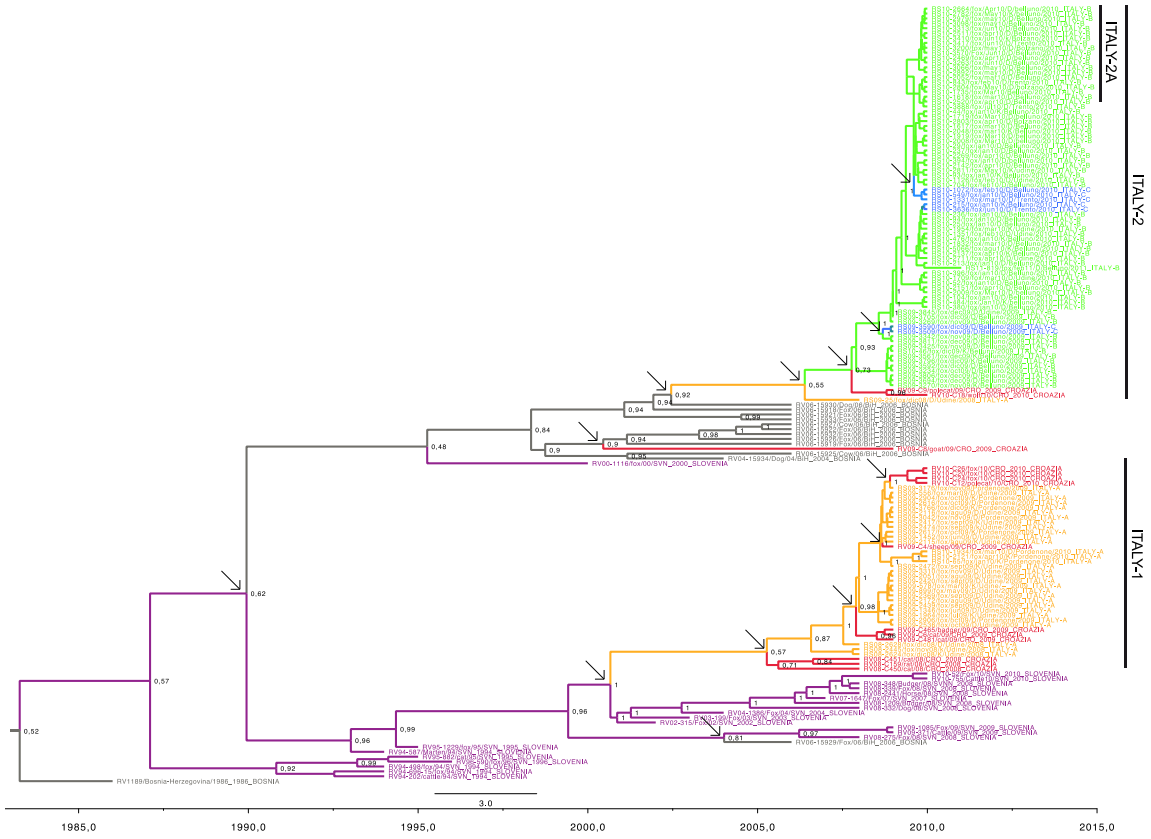


Figure 3.4. MCC tree inferred for the N gene sequences from Italy, BiH, Croatia and Slovenia. Sequences are coloured according to the geographical origin. Yellow: Italy A; green: Italy B; blue: Italy C; grey: BiH; red: Croatia; purple: Slovenia. The numbers at branch points represent the probability of spatial states. The arrows indicate the state probabilities with values higher than 0.55.

By specifying a prior on the number of non-zero rates, we performed Bayes factor (BF) tests to evaluate the support of individual diffusion rates (Annex A - Table A3). Three out of six well-supported rates (BF>5) obtained for the N gene are concordant with the ones obtained from the analyses of the concatenated N and G sequences (and all rates with a BF>5 are displayed in figure 3.3). In particular, two well-supported migration links confirm the spread from east-to-west (area B to area C, Slovenia to area A) and from Italy to the western Balkans (area A to Croatia and area B to Croatia). The lack of support for movement from area A to area B is probably due to the identification of only one virus belonging to the Italy-2 group in the area A (figure 3.3), but this strain is immediately basal to the Italy-2 clade.

3.5. DISCUSSION

Our phylogenetic analysis of the N and G genes of RABV revealed congruent evolutionary patterns, both indicating that the rabies viruses responsible for the recent Italian epidemic belong to the WE sub-lineage of the Cosmopolitan lineage. In addition, both the N and G phylogenies suggest that this sub-lineage could be further subdivided into at least two genetic groups: one which mainly spread westwards to Belluno province (Veneto region), while the other circulated in FVG. These two phylogenetically and geographically distinct viral populations were likely introduced into Italy a few months apart and occupied two distinct geographic areas. Unfortunately, our estimates for the time-scale of RABV evolution in Italy do not provide the accuracy necessary to clearly pinpoint the time of introduction of those two genetic groups, as reflected in the 95% HPD values, and may be biased to some extent by our sampling strategy. The Italy-1 group might have circulated undetected from the beginning of 2007 to September 2008 (mean TMRCA March 2008), when the virus was officially notified, while the Italy-2 group might have been identified shortly after its introduction (mean TMRCA October 2008). However, the inherent uncertainty in these estimates means that we cannot completely exclude that the two viral groups were introduced at the same time between May and September 2008. More sequence data from the period prior to rabies re-emergence in 2008 may be useful in clarifying this issue. Indeed, only 123 red foxes (1.56/100 km2) were submitted from FVG within the framework of rabies passive surveillance from 1st October 2007 to 30th September 2008. As expected, such surveillance activities were intensified only after the first rabies detection in October 2008 (figure 3.1). The reduced passive surveillance in Italy in the period prior to the first rabies notification may be explained by several factors. These include the difficulty in retaining the high awareness of field staff for the risk for rabies re-introduction at border areas as these had experienced a rabies-free status for over ten years, and the nature of epidemiological data available from neighbouring countries, namely Austria and Switzerland which are officially rabies-free, and Slovenia, where all but one of the rabies cases detected in 2008 were collected from areas bordering Croatia (WHO, Rabies Bulletin Europe, 2008. Available at http://www.who-rabies-bulletin.org/journal/Archive/Bulletin_2008_1.pdf and http://www.who-rabies-bulletin.org/journal/Archive/Bulletin_2008_2.pdf). Interestingly, a few weeks after the detection of RABV in Italy, a single fox rabies case was reported from a Slovenian

municipality bordering FVG not subject to ORV (De Benedictis et al., 2009). This notification likely occurred secondary to the increased surveillance in this area previously considered not at risk.

Notably, viruses from both the Italy-1 and Italy-2 groups were found at the beginning of the epidemic in the same municipality. However, from its first detection in December 2008, viruses of the Italy-2 group (black dots in figure 3.3) reappeared 10 months later and more than 100 km further west. One possible explanation is that the Italy-2 group might have been introduced from the western Balkans in an area further north than where the Italy-1 group was circulating and crossed undetected into the northern territories of FVG, spreading westwards to the Belluno province (Veneto region). The absence of RABV in the northern bordering countries, Austria and Switzerland (see above) means that southward spread is unlikely, although it cannot be entirely excluded, and nor can unofficial fox translocation. During 2009, official data collected at the regional level indicated that the disease was spreading through the FVG region, mainly southwards. The absence of rabies cases in the northern part of FVG may be explained by an insufficient surveillance in that area (figure 3.1). Oral rabies vaccination (ORV) campaigns were implemented at the regional level, covering limited areas of the FVG region during 2009 (approximately 1,600 kms2) and not all the territory at the same time (see figure 3.1), and hence failing to comply with the European guidelines for ORV of foxes (EU European Commission, 2002). Northern territories were left unvaccinated until the end of 2009 (De Benedictis et al., 2009), taking into account previous experiences of fox rabies control (Mutinelli et al., 2004) and assuming that foxes would experience difficulties to settle into a mountain area in winter. Such an assumption was later refuted by spatial analyses of field observations, which confirmed rabies hotspots in foxes in mountainous areas during winter (altitude > 1,000 m. above sea level) (Capello et al., 2010,Mulatti et al., 2012). Thus, ORV undertaken in 2009 was able to control the disease only partially and, in those areas where Italy-1 viruses were circulating, Italy-1 viruses remained confined to FVG with no westward spread. In contrast, infection with Italy-2 might have been able to overcome the ORV campaigns undertaken in 2009, perhaps assisted by heterogeneous and insufficient surveillance, settling in the large area encompassing Belluno province. Interestingly, viruses of the Italy-2A sub-group were mostly collected from a single valley (Fiorentina valley) and to a lesser extent from the adjacent Fassa and Badia valleys

during the peak of the epidemic (February to July 2010), suggesting that these viruses are geographically segregated.

Our phylogeographic analysis also provides information on the spatial dispersal patterns of the two Italian RABV groups. Each group likely has an ancestor that originates in the Western Balkans (Former Republic of Yugoslavia), thereby supporting the notion that the WE lineage circulates in a region encompassing Montenegro, BiH, Croatia, Slovenia and Italy. However, it is notable in the MCC tree that viruses of the Italy-2 group are separated from BiH viruses by a relatively long branch of approximately two years. Whether, during this time, RABV circulated in the fox population within the Balkan area is unclear. However, we can speculate that infected foxes likely moved through Croatia and Slovenia, where rabies cases were reported. Accordingly, transient descendants along this branch in these countries have yet to be sampled. Indeed, to date there are few Italy-2 sequences from Croatia and Slovenia. The lack of sequence data from the Balkans could be also a possible explanation for the uncertainty regarding the exact origin of Italy-1.

No viruses of the Eastern European (EE) lineage were identified in Italy. Although there is recent evidence for the circulation of EE viruses in central areas of BiH (McElhinney et al., 2011) and eastern parts of Slovenia and Croatia (Rihtaric et al., 2011,Lojkic et al., 2012), the extent of their circulation in Western Balkans has not been determined. As speculated by McElhinney et al. (2011), the presence of natural barriers or different hunting pressures may have played a role in the segregation of WE viruses in Western Balkans. However, why the EE variant has not spread westwards to Italy is unclear and merits further investigation. Indeed, a more complete ecological and genetic study of the fox population in the pre-Alpine and Alpine areas of Italy and in neighbouring countries, as well as of the possible effects of concomitant infectious diseases in the reservoir species, such as a distemper epidemic in foxes (Nouvellet et al., 2013,Monne et al., 2011), would clearly add important information on the risk for rabies re-emergence and spread.

In conclusion, we have undertaken a comprehensive molecular epidemiological analysis of the fox rabies epidemic that occurred in Italy from 2008 to 2011, and which will contribute to future control programmes at both national and regional levels. Although informative, our phylogenetic and phylogeographic analyses highlight the need for improvements in the genetic and epidemiological data collected from rabies affected countries. Availability of these data is of paramount importance to understand epidemic dynamics and to identify gaps

in both surveillance and control. Our findings also underline the need for vigilance and early warning systems to react to transboundary diseases, such as fox rabies. Indeed, the lack of adequate and homogeneous surveillance, particularly at border areas, has clearly increased the risk of introduction of wildlife rabies. A coordinated regional approach, such as the early warning system implemented between the Italian and Slovenian authorities, has been crucial in synergising the efforts for rabies control in these countries.

ACKNOWLEDGMENTS

This study was funded in part by the European Commission through FP7, project PREDEMICS (Grant agreement no 278433), and by the Italian Ministry of Health through the RC IZSVE 19/10. Paolo Mulatti and Monica Lorenzetto, are gratefully acknowledged for their technical support. Authors wish to thank two anonymous reviewers for their constructive comments.

Chapter 4

Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry

4.1. ABSTRACT

Highly Pathogenic Avian Influenza (HPAI) H5N1 (clade 2.2) was introduced into Egypt in early 2006. Despite the control measures taken, including mass vaccination of poultry, the virus rapidly spread among commercial and backyard flocks. Since the initial outbreaks, the virus in Egypt has evolved into a third order clade (clade 2.2.1) and diverged into antigenically and genetically distinct subclades. To better understand the dynamics of HPAI H5N1 evolution in countries that differ in vaccination policy, we undertook an in-depth analysis of those virus strains circulating in Egypt between 2006 and 2010, and compared countries where vaccination was adopted (Egypt and Indonesia) to those where it was not (Nigeria, Turkey and Thailand). This study incorporated 751 sequences (Egypt n=309, Indonesia n=149, Nigeria n=106, Turkey n=87, Thailand n=100) of the complete haemagglutinin (HA) open reading frame, the major antigenic determinant of influenza A virus. Our analysis revealed that two main Egyptian subclades (termed A and B) have co-circulated in domestic poultry since late 2007 and exhibit different profiles of positively selected codons and rates of nucleotide substitution. The mean evolutionary rate of subclade A H5N1 viruses was 4.07 x 10^{-3} nucleotide substitutions per site, per year (HPD 95%, 3.23-

4.91), whereas subclade B possessed a markedly higher substitution rate (8.87 x 10^{-3}; 95% HPD 7.0-10.72 x 10^{-3}) and a stronger signature of positive selection. Although the direct association between H5N1 vaccination and virus evolution is difficult to establish, we found evidence for a difference in the evolutionary dynamics of H5N1 viruses among countries where vaccination was or was not adopted. In particular, both evolutionary rates and the numbers of positively selected sites were higher in virus populations circulating in countries applying avian influenza vaccination for H5N1, compared to viruses circulating in countries which had never used vaccination. We therefore urge a greater consideration of the potential consequences of inadequate vaccination on viral evolution.

4.2. INTRODUCTION

Since its emergence in 1996 in Guangdong, China, highly pathogenic avian influenza (HPAI) viruses of the H5N1 subtype have spread widely across Asia, Europe, and Africa, infecting a range of domestic and wild avian species and sporadically spilling over into humans and other mammals (Chen et al., 2005; Xu et al., 1996). The spread of HPAI H5N1 virus has resulted in the implementation of a vaccination policy in several countries including PR China, Hong Kong, Vietnam, Indonesia (Eagles et al., 1997) and Egypt (Peyre et al., 2009a). While vaccination in some countries seems to have successfully contributed to the reduction of the incidence of the disease, as is the case for Hong Kong in 2002 (Ellis et al., 2006) and Vietnam (Soares Magalhaes et al., 2010), in other countries, such as Indonesia and Egypt, outbreaks are still occurring and the virus has become endemic in poultry (Peyre et al., 2009a; Domenech et al., 2009).

In Egypt, H5N1 HPAI virus was first notified in poultry in February 2006, with the first human case occurring shortly afterwards (March 2006). Since then, 151 human infections and 52 deaths have been reported in Egypt by the WHO (WHO, 2011). Egypt experienced the introduction of only one H5N1 clade (clade 2.2) virus, which spread rapidly among commercial and backyard flocks in most of the Governorates (WHO, 2011; Abdel-Moneim et al., 2009). Since the initial outbreaks, the virus in Egypt has evolved into a distinct third order clade (clade 2.2.1) and diverged further into antigenically and genetically distinct subclades (Abdel-Moneim et al., 2009; WHO/OIE/FAO, 2009; Balish et al., 2010). In particular, a new

variant of HPAI H5N1 viruses exhibiting significant antigenic diversity from previous Egyptian isolates appears to have spread widely in Egypt since late 2007 (Balish et al., 2010; Terregino et al., 2010; Grund et al., 2011; Rauw et al., 2011).

To control the disease in poultry and reduce the risk to human health, the large-scale vaccination of commercial and backyard sectors has been used officially since March 2006 and May 2007, respectively. However, in July 2009 vaccination in household poultry was suspended. Despite all the control measures undertaken, poultry outbreaks are regularly reported and human cases still occur (Peyre et al., 2009b; WHO, 2011). Previous studies have suggested that the limited impact of vaccination in controlling the infection is probably due to the improper use of vaccines rather than to the efficacy of the vaccine itself (Terregino et al., 2010; Kim et al., 2010). Indeed, in Egypt vaccination coverage was poor in the household sector (estimated range 1% to 30%) and about 50-60% in the commercial farms. In addition, the vaccination monitoring system seems inefficient and the training of the field technicians inadequate (Peyre et al., 2009b; EMPRES/FAO-GLEWS, 2010).

The prolonged circulation of influenza viruses in a partially immunized population might also have a direct impact on the evolution of influenza virus, and particularly on the occurrence of antigenic drift (Shih et al.,2007; Smith et al., 2004). However, information on the impact of poultry vaccination on the evolutionary dynamics of avian influenza viruses in natural situations is limited. One previous study investigated the occurrence of antigenic drift in low pathogenic avian influenza isolates (H5N2) circulating in Mexico (Lee et al., 2004). However, detailed information on viral evolution was not available and the observed antigenic drift could not be directly associated to vaccination. To better understand whether or not HPAI H5N1 evolution is different in countries where poultry vaccination is applied, we investigated the evolutionary and population dynamics of the viruses circulating in Egypt between 2006 and 2010. We then compared the evolutionary rates and the selection profiles of H5N1 viruses isolated from countries where vaccination was adopted (Egypt and Indonesia) or not adopted (Nigeria, Turkey and Thailand).

4.3. MATERIALS AND METHODS

4.3.1. Viruses and genetic sequences included in this study

We analyzed the HA gene cds of 309 H5N1 viruses isolated in Egypt from February 2006 to July 2010, 149 H5N1 viruses collected in Indonesia from October 2003 to March 2007, 87 Turkish isolates of H5N1 isolated between December 2005 and March 2008, 106 H5N1 viruses circulating in Nigeria from January 2006 to July 2008, and 100 H5N1 isolates collected in Thailand between 2004 to 2008. In total, 126 HA sequences were newly determined in this study; the remaining sequences have been retrieved from GenBank (www.ncbi.nlm.nih.gov). All complete HA open reading frame sequences from the above mentioned countries available or newly determined at the time of writing were included in the study.

4.3.2. Nucleotide sequencing

The HA gene segment of 39 H5N1 viruses isolated from poultry in Egypt in 2010 were sequenced here. In addition, we sequenced the HAs of 87 viruses collected in Turkey from 2005 to 2008. For these isolates, viral RNA was extracted from the infective allantoic fluid of SPF fowls' eggs using the Nucleospin RNA II Kit (Machery-Nagel, Duren, Germany) and was reverse transcribed with the SuperScript III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA - USA). PCR amplifications were performed by using specific primers (primer sequences available on request). The complete coding sequences were generated using the Big Dye Terminator v3.1 cycle sequencing kit (Applied Biosystem, Foster City, CA - USA). The products of the sequencing reactions were cleaned-up using PERFORMA DTR Ultra 96-Well kit (Edge BioSystems, Gaithersburg, MD - USA) and sequenced in a 16-capillary ABI PRISM 3130xl Genetic Analyzer (Applied Biosystem, Foster City, CA - USA). Sequence data were assembled and edited with SeqScape software v2.5 (Applied Biosystem).

4.3.3. Nucleotide sequence accession numbers

The nucleotide sequences obtained in this study are made available in the GISAID database under accession numbers: EPI287349 to EPI287384 (HA of Egyptian viruses), and EPI293896, EPI293904, EPI293912, EPI293920, EPI293928, EPI293936, EPI293944, EPI293952, EPI293960, EPI305540 to EPI305616 (HA of Turkish viruses).

4.3.4. Evolutionary Analysis

Rates of nucleotide substitution per site, per year and the time to the Most Recent Common Ancestor (tMRCA) for five different data sets comprising the HA gene of H5N1 viruses from Egypt, Indonesia, Nigeria, Turkey and Thailand were estimated using the BEAST program, version 1.6.1 (Drummond and Rambaut, 2007), which employs a Bayesian Markov chain Monte Carlo (MCMC) approach. All data sets were analysed with the codon-based SRD06 nucleotide substitution model (Shapiro et al., 2006). For each analysis the Bayesian skyride (Minin et al., 2008c) coalescent prior was utilized. Two molecular clock models – the strict (i.e. constant) and uncorrelated lognormal (UCLN) relaxed molecular clocks – were compared by analyzing values of the coefficient of variation (CoV) in the Tracer program (Drummond and Rambaut, 2007). For the HA gene of viruses from Egypt, Indonesia and Thailand the lower 95% HPDs of CoV values were >0, indicative of rate among viral lineages. In contrast, the HA gene of Turkish and Nigerian viruses showed CoV values approximately of 0 and hence indicative of strictly clock-like evolution. For consistency, the comparison of parameter estimates was performed using the uncorrelated lognormal relaxed molecular clock and the Bayesian skyride coalescent prior as these are the most general models and likely better reflect the complex population dynamics of influenza A virus (Rambaut et al., 2008). However, the results of all model comparisons – strict versus relaxed molecular clock – are given in Annex B - Table B1. In all cases, statistical uncertainty in the data is reflected in the 95% highest probability density (HPD) values for each parameter estimate, and in each case chain lengths were run for sufficient time to achieve coverage as assessed using the Tracer program (up to 100 million generations in the case of the Egyptian HA sequences). In addition, an independent estimate of nucleotide substitution rate was obtained using a root-to tip regression method available in the Path-O-Gen program (kindly provided by Andrew Rambaut, University of Edinburgh; http://tree.bio.ed.ac.uk/software/pathogen/), in which genetic distances estimated from a maximum likelihood (ML) tree (see below) are plotted against day of sampling (table 4.1).

We used the Bayesian skyride plot to infer the population dynamics of Egyptian viruses in terms of changing levels of relative genetic diversity [$N_e t$] through time, in which N_e represents the effective population size and t the generation time (Minin et al., 2008). Finally, a Maximum Clade Credibility (MCC) phylogenetic tree was estimated for the Egyptian HA

sequences using the posterior distribution of trees obtained from the BEAST analysis described above and after the removal of an appropriate burn-in (10% of the samples). To increase phylogenetic rigor, the Bayesian tree topology was compared with those of neighbor-joining (NJ) and ML phylogenetic trees inferred using the PAUP* 4.0 (Wilgenbusch and Swofford, 2003) and PhyML (Guindon and Gascuel, 2003) programs, respectively. For both these later trees the best-fit general time-reversible (GTR)+I+Γ_4 model of nucleotide substitution was employed, and a bootstrap analysis comprising 1000 replicates was also undertaken. Parameter values for the GTR substitution matrix, base composition, gamma distribution of among-site rate variation (with four rate categories, Γ_4), and proportion of invariant sites (I) were estimated directly from the data using MODELTEST (Posada and Crandall, 1998) (available on request). All trees were visualized using the program FigTree v1.3.1 (Rambaut, 2008).

4.3.5. Statistical analysis

A comparison of the evolutionary rates of viruses collected from five distinct countries was performed using the probability density function of posterior simulated data for each population mean rate obtained using BEAST. Assuming a normal distribution of mean substitution rates with a standard deviation derived from the 95% HPD values, we calculated the probability that the distributions of the evolutionary rates from viruses sampled in vaccinated countries was higher than those of viruses sampled from unvaccinated countries (Pr). A probability of >0.90 was evaluated relevant (R program version 2.10.1).

4.3.6. Analysis of selection pressures

Gene and site-specific selection pressures for the HA gene segment of H5N1 viruses from Egypt, Indonesia, Nigeria, Turkey and Thailand were measured as the ratio of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitutions per site (ratio d_N/d_S). In all cases these estimates were made using the Single Likelihood Ancestor Counting (SLAC) and Fixed Effects Likelihood (FEL) methods available at the Datamonkey online version of the Hy-Phy package (Kosakovsky Pond and Frost, 2005). All analyses used input NJ phylogenetic trees estimated using the GTR model of nucleotide substitution.

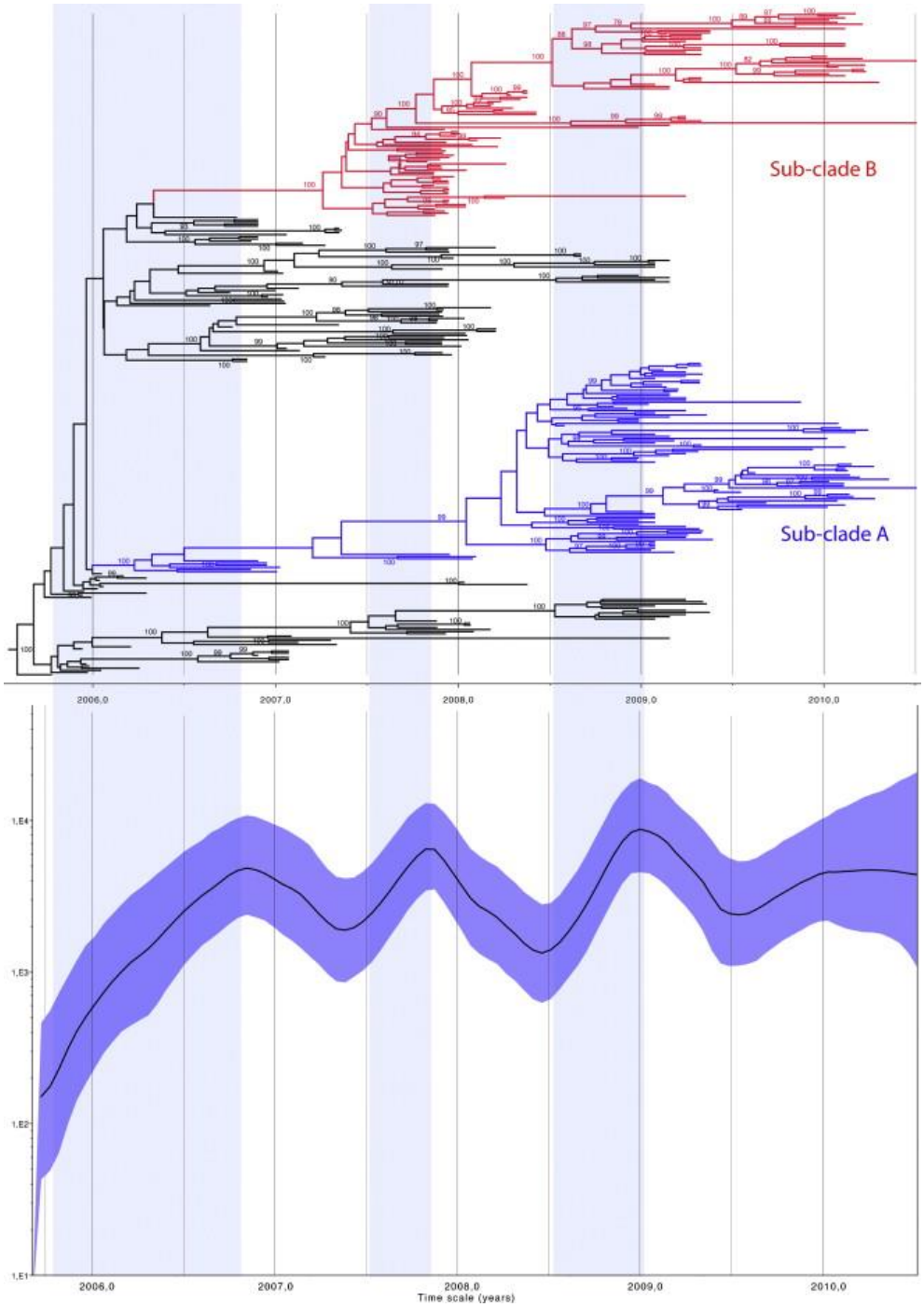


Figure 4.1. Phylogenetic tree of the HA gene segment of clade 2.2.1 Egyptian viruses scaled to real time (x-axis). Posterior probability values higher than 90 are indicated at nodes. Subclade A viruses are marked in blue, while viruses belonging to subclade B are marked in red. Bayesian skyline plot of the HA gene showing changes in genetic diversity of all H5N1 viruses collected in Egypt between 2006 and 2010 . A measure of genetic diversity (Net) is given on the y-axis with 95% HPD values shown in blue. Vertical blue shaded lines highlight peaks of drastic increase in genetic diversity.

4.4. RESULTS

4.4.1. Phylogenetic and molecular analysis

All the 309 HPAI H5N1 viruses sampled from February 2006 to July 2010 in Egypt fell within clade 2.2.1 (WHO/OIE/FAO, 2009) (Figure 4.1). The maximum clade credibility (MCC) tree inferred for the HA gene segment showed that most of the viruses collected at the beginning of the outbreak (i.e. 2006 – early 2007) were characterized by a poorly resolved branching structure. In contrast, the majority of viruses sampled from more recent infections (2007 – 2010) fell within two main genetic groups, denoted here as groups A and B to maintain the subdivision nomenclature adopted in a previous publication (Abdel-Moneim et al., 2009). These two clusters have been subsequently identified by Balish et al (2010) using a smaller data set, as groups C and E, respectively. Of note, these groups meet the three specific clade definition criteria proposed by the WHO/OIE/FAO H5N1 Evolution Working Group (WHO/OIE/FAO, 2008) and could be defined as additional fourth order clades. Indeed, all phylogenetic trees (MCC, ML, NJ) of the HA gene of the Egyptian H5N1 viruses estimated here show that subclades A and B comprise two well-supported monophyletic groups (with bootstrap values of 90 and 100, respectively, and posterior probabilities of 100, at both clade-defining nodes). Subclade A shows an average percentage pairwise within-clade nucleotide distance of 1% and a between-clade distance of 1.8% - 7.4%, while for subclade B the average within and between-clade percentage pairwise nucleotide distance were 1.5% and 2.9% - 7.9%, respectively. We focus our analysis on these two subclades as they clearly represent the two dominant variants co-circulating in Egypt since 2007.

Among those viruses belonging to subclade A (95 viruses), 25% were isolated from chickens, 23% from ducks, 6% from geese, 1% from turkeys and 45% from humans, while 94% of the analysed viruses from subclade B (92 viruses) were collected from chickens, 3% from ducks, 2% from turkey, 1% from geese and only one isolate was obtained from a human case. Compared to the Egyptian H5N1 viruses circulating in this country in 2006 (i.e. the progenitor viruses), the HA gene of isolates belonging to subclade B showed eight amino acid differences, while three amino acid substitutions were acquired by subclade A. In particular, two mutations at positions 74 and 165 (H3 numbering) generated two potential additional glycosylation sites in the HA gene of subclade B and A viruses, respectively. A total of 11 amino acid differences in HA were observed between subclades A and B. Interestingly, most

of these fell within the receptor binding domain, in the H3-corresponding A or B antigenic sites (positions 120, 129, 140, 141, 151, 184, 226, H3 numbering) confirming previous observations (Nobusawa et al., 1991; Wilson et al., 1981; Cattoli et al., 2011).

4.4.2. Evolutionary rates and population dynamics

We inferred the evolutionary rates and population dynamics of the HA gene segment using a Bayesian coalescent approach (Drummond and Rambaut, 2007). The mean rate of nucleotide substitution of all the Egyptian viruses collected from February 2006 to July 2010 was 5.36 x 10^{-3} substitutions/site/year (95% HPD from 4.81 x 10^{-3} to 5.96 x 10^{-3} subs/site/year) and the mean time of the most recent common ancestor (TMRCA) was estimated to be September 2005 (95% HPD from July 2005 to November 2005). Analysis of virus population dynamics revealed three distinct increases in genetic diversity of clade 2.2.1 in poultry in Egypt (Figure 4.1). The first, very rapid, increase at the beginning of the epidemic in 2006 corresponding with the initial invasion of H5N1. A second, peak was identified at the end of 2007 followed by a gradual decrease during the first half of 2008, while the third increase in genetic diversity occurred approximately from June to December 2008.

To determine the contribution of each subclade to the viral population behaviour we analyzed the evolutionary dynamics of subclades A and B separately. Interestingly, the evolutionary rates estimated for the two subclades were different with no overlap in HPD values (table 4.1): a mean rate of 4.07 x 10^{-3} substitutions/site/year (95% HPD from 3.23 x 10^{-3} to 4.91 x 10^{-3}) for subclade A, and a rate of 8.87 x 10^{-3} substitutions/site/year (95% HPD from 7.00 x 10^{-3} to 10.72 x 10^{-3}) for subclade B, making the latter clade the fastest evolving group of viruses. The estimated TMRCAs for these groups were January 2006 (95% HPD, July 2005 – June 2006) for subclade A and September 2007 (95% HPD, July – November 2007) for subclade B, such that subclade B was detected soon after its appearance in the country.

The MCC tree also showed a long branch within subclade A. The absence of this long branch in the ML phylogenetic tree and the lack of sequences of samples collected during 2007 suggest that this branch could be an artefact of the Bayesian analysis. To explore its possible influence on the evolutionary dynamics of subclade A, the analysis was repeated

using subclade A viruses collected from January 2008. Essentially identical substitution rates were observed using this subpopulation (results available on request).

4.4.3. Selection pressures

We investigated the selection profiles of the HA protein of all Egyptian viruses, as well as those from subclade A and subclade B. Overall, the total number of nonsynonymous to synonymous substitutions (d_N/d_S) for each data set was less than 1 (table 4.1), indicating that purifying selection has been the main pressure shaping the evolution of Egyptian H5N1 viruses, although mean d_N/d_S was markedly higher in subclade B (0.46) compared to subclade A (0.29). Nevertheless, we identified several individual codons likely under positive selection pressure in the HA gene of all the Egyptian viruses (nine residues), of subclade A (four residues), and of subclade B (five residues) (table 4.1). In addition, six, three, and two positively selected sites identified in the HA gene segment of all Egyptian viruses, subclade A and subclade B, respectively, were located at antigenic sites previously identified in H5, H1 and H3 subtypes (table 4.1) (Bush et al., 1999; Kaverin et al., 2007; Stevens et al., 2006; Li et al., 2009).

4.4.4. Comparison of evolutionary rates and selection profiles between viruses collected from vaccinated and unvaccinated poultry populations

To investigate the possible effect of poultry vaccination on viral evolution we compared the evolutionary rates of the HA gene of H5N1 viruses isolated both in countries where vaccination was (Egypt and Indonesia) and was not adopted (Turkey, Nigeria, Thailand). Importantly, each of these data contained a large number of complete HA cds sequences (>80), either from GenBank (H5N1 sequences from Indonesia, Egypt, Thailand and Nigeria) or obtained in this study (87 sequences of H5N1 from Turkey and 39 from Egypt). Except for Thailand, the H5N1 isolates identified in each country belonged to a single genetic clade (table 4.1). Specifically, viruses from Egypt, Nigeria and Turkey clustered into the major clade 2.2, while Indonesian isolates grouped within clade 2.1. Viruses from Thailand fell within two distinct clades, namely 1 and 2.3.4. However, only few outbreaks in northeast Thailand were caused by 2.3.4 viruses (Suwannakarn et al., 2009). In the present analysis we included the genetic variant prevalently circulating in this country (clade 1) as this allows for more comparable results.

Table 4.1. Evolutionary profiles of H5N1 HPAI viruses included in this study.

COUNTRY	N. SEQ	COLLECTION DATE	GENETIC CLADE	EVOLUTIONARY RATE - sub/site/year ($x10^{-3}$)		Mean d_N/d_S	POSITIVELY SELECTED SITES§		
				BEAST Mean (95% HPD)	Path-O-Gen (correlation coefficient; R^2)		N	aa pos.§	Antig. Site $
Egypt	309	February 2006 – July 2010	2.2.1	5.36 (4.81 - 5.96)	4.07 (0.72-0.52)	0.31	9	61 136 137 157 170* 171* 204 205 254	- A B A B B - B -
Subclade A	95	January 2007 – July 2010	2.2.1	4.07 (3.23-4.91)	2.14 (0.67; 0.45)	0.29	4	102 136 157 205	- A A B
Subclade B	92	December 2007 – July 2010	2.2.1	8.87 (7.00-10.72)	6.15 (0.89; 0.79)	0.46	5	87 131 157 201 291*	- - A B -
Indonesia	149	October 2003 – March 2007	2.1	6.13 (5.14-7.06)	6.12 (0.78; 0.6)	0.21	6	104 145 156 172* 211* 529	E A A B - -
Nigeria	106	February 2006 - July 2008	2.2	5.20 (4.11-6.27)	4.67 (0.87-0.75)	0.23	2	152 489	A -
Turkey	87	December 2005 – March 2008	2.2	4.04 (2.16-6.04)	5.3 (0.91; 0.84)	0.24	0	-	-
Thailand	100	2004 - 2008	1	2.52 (1.93-3.19)	2.14 (0.7-0.5)	0.29	3	8 156 167	- A -

§ H5 numbering; * potential additional glycosylation sites. $ Bush et al., 1999; Kaverin et al., 2007; Stevens et al., 2006; Li et al., 2009.

The two countries applying vaccination, Egypt and Indonesia, shared a number of features. Specifically, they experienced a single introduction of H5N1 (EMPRES/FAO-GLEWS , 2010; Lam et al., 2008; Takano et al., 2009); after its appearance, the virus rapidly became endemic and evolved creating a new third-order clades or subclades: 2.1.1, 2.1.2, 2.1.3 in Indonesia; clade 2.2.1 and subclades A and B in Egypt. In contrast to what occurred in Egypt, where both subclades co-circulated widely, in Indonesia clade 2.1.3 viruses have widely spread in recent years, whereas clade 2.1.1 and 2.1.2 viruses have progressively become rarer since 2005. Furthermore, in both countries sporadic zoonotic transmissions to humans have occurred, as reported in literature (Takano et al., 2009; Peyre et al., 2009a).

In the countries where vaccination was not enforced (i.e. Nigeria, Thailand and Turkey) we noted multiple introductions of H5N1 viruses belonging to the same clade, such as in the case of Nigeria (Fusaro et al., 2010) or Turkey (this study, data not shown), or to different clades, such as in Thailand (Suwannakarn et al., 2009). Thailand and Turkey experienced seven and three waves of H5N1, respectively (Suwannakarn et al., 2009; OIE, 2011). In Nigeria, the virus circulated extensively from the beginning of 2006 to the end of 2007, when outbreaks appeared to have been successfully restrained without the need of vaccination. In July 2008 new cases were reported (Fusaro et al., 2009), but since then no cases of H5N1 have been noted.

Interestingly, the substitution rate of Egyptian subclade B viruses (95% HPD 7.00 x 10^{-3} to 10.72 x 10^{-3}) and H5N1 viruses collected in Indonesia (95% HPD 5.14 x 10^{-3} to 7.06 x 10^{-3}) were higher than the rates of the HA gene of H5N1 viruses from Nigeria (95% HPD 4.11 – 6.27 x 10^{-3}), Turkey (95% HPD 2.16 – 6.04 x 10^{-3}) and Thailand (95% HPD 1.93 – 3.19 x 10^{-3}) (table 4.1). Indeed, the probability that the evolutionary rates of viruses collected from the unvaccinated countries (Nigeria, Turkey and Thailand) were lower than those of viruses circulating in countries where vaccination was applied (Egyptian subclade B and Indonesia) was Pr>0.90 (Annex B - Table B2). An equivalent analysis using root-to-tip regression against sampling time (program Path-O-Gen) resulted in a similar distinction between the vaccinated and unvaccinated countries, even though most of the values obtained were lower than those achieved with BEAST (table 4.1).

Finally, there was also a marked difference in the number of positively selected sites between the HA gene of viruses collected from vaccinated poultry populations and those from unvaccinated ones. Specifically, zero (Turkey), two (Nigeria) or three (Thailand) codons

predicted to be under positive selection were identified in the HA gene of viruses from unvaccinated countries (table 4.1), compared to nine and six positively selected sites from Egypt and Indonesia, respectively. Although our analysis of the number of positive selected sites is clearly affected by sample size, the high number of codons predicted to be under positive selection (five) in the Egyptian subclade B viruses alone supports the conclusion that there is an elevated level of positive selection in vaccinated countries.

4.5. DISCUSSION

Sequence analysis of the HA gene segment of H5N1 viruses isolated in Egypt revealed that two main distinct subclades (here named A and B) fitting the specific clade definition criteria described by WHO/OIE/FAO H5N1 Evolution Working Group (WHO/OIE/FAO, 2008) have co-circulated in domestic poultry since late 2007. Although viruses belonging to both subclades were collected from the same country and during the same time period, they exhibited different selection profiles and rates of nucleotide substitution, suggesting that they are subject to differing evolutionary pressures. This could reflect the circulation of viruses belonging to each of these subclades in distinct host environments such as separate poultry sectors (e.g. backyard versus commercial flocks) and which in turn leads to differences in selection pressure. The idea of independent maintenance of different viruses in distinct compartments was proposed in a previous Egyptian study (Arafa et al., 2010). Indeed, we note that subclade B was mainly isolated from chickens, the poultry species most represented in the integrated and commercial sectors (FAO, 2008; Peyre, 2009b), while subclade A had a wider host range (i.e distinct avian species). However, additional epidemiological information is clearly required to test this hypothesis.

Interestingly, subclade A which seemingly circulates in multiple species experienced a lower evolutionary rate than subclade B, which is almost exclusively detected in chickens, indicating that, in the case of the HA gene, multiple inter-species transmission of these viruses has not resulted in an elevated rate of viral evolution. It is therefore possible that the higher population density of chickens, which are the main target species for vaccination, associated with subclade B has facilitated the action of natural selection in favouring antigenic variants in HA which has in turn increased the substitution rate. This appears to be

also consistent with the evolutionary rates of subclade A which is comparable to the lower rate of evolution observed for viruses circulating in countries not applying vaccination (table 4.1).

Analysis of virus population dynamics of the entire data set of the Egyptian H5N1 viruses showed a rise in genetic diversity at the beginning of 2006, immediately after the introduction of the H5N1 viruses in the country, at the end of 2007, corresponding to the emergence of subclade B, and during the second half of 2008. The second and third peaks in relative genetic diversity occurred approximately a year after the beginning of vaccination in commercial farms (March 2006) and backyard birds (May 2007), respectively. Interestingly, a previous study estimated the emergence of antigenically distinct viruses in Egypt as also occurring 18 months after the beginning of vaccination in poultry (Balish et al., 2010). Although China was not included in the present study, H5N1 genetic diversity has been documented in this country (Chen et al., 2006) and the emergence of vaccine escape variant viruses was reported one year after the implementation of vaccination in poultry (Smith et al., 2006). As such, the increase of the viral genetic diversity we observe may reflect the emergence and the subsequent selection of mutants escaping vaccine pressure, particularly where vaccination was not completely or properly applied as it was the case of Egypt, China and Indonesia (Peyre et al., 2009a; Smith et al., 2006; Elly, 2008).

We also noted evidence of a difference in the evolutionary dynamics of H5N1 among countries where vaccination was adopted or not adopted. In particular, rates of nucleotide substitution were higher for viruses circulating in poultry population where vaccination was applied, and most strikingly in subclade B from Egypt followed by Indonesia. In addition, we noted that virus HA sequences analysed from vaccinating countries seemingly had greater numbers of positively selected sites compared to the HA sequences of viruses from countries which had never used vaccination, and a markedly higher mean d_N/d_S value in the case of Egyptian subclade B. However, the direct association between H5N1 vaccination and virus evolution is clearly difficult to establish since other factors, difficult to evaluate under field conditions, might contribute to the apparent differences in evolutionary dynamics observed here. For example, poultry density, a variety of ecological factors or environmental changes may contribute to the different rates of nucleotide substitution and strength of natural selection in those viruses circulating in poultry (Suarez, 2010; Vandegrift et al., 2010). In

addition, the endemicity of H5N1 viruses, the number of distinct introductions of H5N1 (e.g. in Nigeria and Turkey), may also have influenced the pattern of viral evolution.

When properly planned and adopted, vaccination is a powerful tool for the control and eradication of avian influenza in poultry, as demonstrated by past experiences in Italy, Vietnam and Hong Kong, where it reduced economic losses and the risks for zoonotic transmission (Capua and Marangon, 2007; Ellis et al., 2006; Domenech et al., 2009). However, if not properly applied and not coupled with careful surveillance, robust vaccine strategies and strict bio-security precautions, vaccination may contribute to the rapid evolution and antigenic change of H5N1 viruses, creating opportunities for the viruses to escape from vaccine protection. Clearly, more sophisticated and effective animal health control and prevention measures are urgently required, such as a sustainable and efficacious vaccination strategies for poultry.

ACKNOWLEDGEMENTS

We acknowledge the Ministry of Agriculture of Egypt for its continuous support and for assistance with the avian influenza control program in the country. The laboratories that have shared and deposited the genetic sequences included in the present study in public database (GenBank) are gratefully acknowledged. In particular, National Laboratory for Veterinary Quality Control on Poultry Production, Animal Health Research Institute (Dokki, Giza, Egypt); Chulalongkorn University (Pathumwan, Bangkok, Thailand); Yuko Uchida National Institute of Animal Health (Tsukuba, Japan); Molecular Biology and Virology Unit, Faculty of Veterinary Science, Mahidol University (Phuttamonthon District, Nakornpathom, Thailand); Division of Virology, Institute of Medical Science,University of Tokyo (Shirokanedai, Minato-ku, Tokyo; Japan), Southeast Poultry Research Laboratory, Athens, GA 30605, USA); WHO Collaborating Center for Surveillance, Epidemiology and Control of Influenza, Influenza Branch, Centers for Disease Control and Prevention (Atlanta,GA 30333, USA), Department of Microbiology, The University of Hong Kong (Hong Kong SAR, China); WHO Collaborating Centre for Reference and Research on Influenza Centre (Parkville, Australia). Ms Francesca Ellero is gratefully acknowledged for the proof reading of the article. Part of this work was financially supported by the Italian Ministry of Health through the RC IZS VE 14/09.

Chapter 5

Emergence of a Highly Pathogenic Avian Influenza Virus from a Low Pathogenic Progenitor

5.1. ABSTRACT

Avian influenza (AI) viruses of the H7 subtype have the potential to evolve into highly pathogenic (HP) viruses that represent a major economic problem for the poultry industry and a threat to global health. However, the emergence of HPAI viruses from low pathogenic (LPAI) progenitor viruses is currently poorly understood. To investigate the origin and evolution of one of the most important avian influenza epidemics described in Europe, we investigated the evolutionary and spatial dynamics of the entire genome of 109 H7N1 (46 LPAI and 63 HPAI) viruses collected during Italian H7N1 outbreaks between March 1999 and February 2001. Phylogenetic analysis revealed that the LPAI and HPAI epidemics shared a single ancestor, that the HPAI strains evolved from the LPAI viruses in the absence of reassortment, and that there was a parallel emergence of mutations among HPAI and later LPAI lineages. Notably, an ultra-deep sequencing analysis demonstrated that some of the amino acid changes characterizing the HPAI virus cluster were already present with low frequency within several individual viral populations from the beginning of the LPAI H7N1 epidemic. A Bayesian phylogeographic analysis revealed stronger spatial structure during the LPAI outbreak, reflecting the more rapid spread of the virus following the emergence of

HPAI. The data generated in this study provide the most complete evolutionary and phylogeographic analysis of epidemiologically intertwined high and low pathogenic viruses undertaken to date, and highlight the importance of implementing prompt eradication measures against LPAI to prevent the appearance of viruses with fitness advantages and unpredictable pathogenic properties.

5.2. INTRODUCTION

Influenza A viruses (IAVs) evolve rapidly due to the high error rate of the RNA polymerase, strong immune-driven natural selection, and the capacity for genome segments to be exchanged via reassortment (Webster et al., 1992). The rapidity of evolution, in combination with the increasing availability of genome-scale sequence data and new computational tools for phylogenetic analysis, facilitate the inference of the evolutionary history and spatial spread of IAVs, in turn providing a better understanding of the patterns and processes that underpin epidemics (Rambaut et al., 2008; Lemey et al., 2009; Fusaro et al., 2010; Cattoli et al., 2011; Bataille et al., 2012; Baillie et al., 2012; Hughes et al., 2012). Eighteen antigenically distinct hemagglutinin (HA) subtypes of influenza viruses have been identified to date (Fouchier et al., 2005; Tong et al., 2013). In avian species, infections with the H5 and H7 subtypes are of greatest concern because of their potential to evolve into the highly pathogenic form of the virus that can devastate poultry populations and occasionally be transmitted to humans. Why the H5 and H7 subtypes are more prone to evolve into highly pathogenic forms than other subtypes remains poorly understood. Highly pathogenic avian influenza (HPAI) viruses can evolve directly from low pathogenic precursors following introduction into domestic poultry (Rohm et al., 1999; Banks et al., 2000; Banks et al., 2001). The emergence of HPAI viruses results from the insertion/substitution of basic amino acids at the HA cleavage site (Horimoto et al., 1995; Garcia et al., 1996; Spackman et al., 2003), or from non-homologous recombination resulting in the insertion of a foreign nucleotide sequence (Suarez et al., 2004; Pasik et al., 2005; Maurer-Stroh et al, 2013). HPAI viruses replicate systemically in birds, damaging vital organs and tissues, and frequently causing death (Capua and Alexander, 2008)). In addition to the HA, the virus polymerase gene complex (PB1, PB2, PA) and nucleoprotein (NP) also modulate the pathogenicity of AI

viruses, as demonstrated for H5N1 HPAI viruses (Hulse-Post et al., 2007; Cauthen et al., 2007; Wasilenko et al., 2008). Importantly, the insertion of multiple basic amino acids at the cleavage site is not always sufficient to convert a LPAI virus phenotype into a HPAI one as demonstrated in experimental studies (Stech et al., 2009) and sporadically in natural infections (Albonik et al., 2009; Londt et al., 2007; Gaidet et al., 2008).

Although there are numerous phylogenetic studies of the emergence, spread and evolution of HPAI viruses (Lemey et al., 2009; Fusaro et al., 2010; Cattoli et al., 2011; Bataille et al., 2012; Lam et al., 2008; Vijaykrishna et al., 2008; Cattoli et al., 2009), whole-genome sequence data from HPAI and the LPAI progenitor viruses collected during the same epidemic is limited. In most cases the number of LPAI isolates collected prior to the emergence of the HPAI strain was insufficient to determine the genesis of the virulent phenotype , including whether this occurs via a stepwise evolutionary process (Banks et al., 2001; Horimoto et al., 1995; Garcia et al., 1996; Suarez et al., 2004; Pasik et al., 2005; Abbas et al., 2010). Importantly, improved 'deep' sequencing technologies allow for greater insight into the emergence of new variants of influenza type A viruses (Croville et al., 2012; Selleri et al., 2012), potentially including HPAI strains.

During 1999 – 2001 in northern Italy, a LPAI H7N1 avian influenza virus evolved into a HPAI form that caused the death of over 16 million poultry and substantial economic losses to industry (Marangon et al., 2005). From the end of March to December 1999, 199 outbreaks of LPAI were detected in northern Italy, which is home to more than 65% of Italian poultry production (Capua et al., 2000; Capua et al., 2003). Despite control measures, a HPAI virus of the same subtype was isolated in December 1999 and rapidly spread to additional production facilities before being eradicated officially in April 2000. Following depopulation and restocking of the HPAI infected areas, LPAI re-emerged in August 2000 and 78 outbreaks were diagnosed and eradicated (Mulatti et al., 2010). To reduce the economic impact of the H7N1 infection in poultry (Sartore et al., 2010), a vaccination programme was initiated from November 2000 to May 2002.

The Italian HPAI H7N1 outbreak has been well characterized epidemiologically, including identifying risk factors and within-flock viral transmission dynamics (Mannelli et al., 2006; Mulatti et al., 2007; Busani et al., 2009; Dorigatti et al., 2010; Bos et al., 2010; Mulatti et al., 2010a). However, little genetic data have been published from this epidemic (Banks et al., 2000). Herein, we describe the analysis of 109 whole genomes from H7N1

viruses collected between March 1999 and February 2001. In particular, we compared the evolutionary patterns and dynamics of the HPAI and LPAI viruses, including rates of nucleotide substitution, selection pressures, and the emergence of amino acid mutations that may facilitate evolution from low to high pathogenicity. In addition, we investigated the patterns of spatial spread of the LPAI and HPAI H7N1 strains among the affected Italian provinces. To complement the phylogenetic analysis and to help identify viral subpopulations harbouring known molecular markers for viral pathogenicity, we utilized deep sequencing on representative clinical samples collected during the 1999-2001 H7N1 epidemic.

5.3. MATERIALS AND METHODS

5.3.1. Sanger sequencing

We generated the complete genome sequences of 72 H7N1 avian influenza A viruses collected from poultry in northern Italy from March 1999 to February 2001, comprising 22 LPAI and 50 HPAI viruses. In addition, 24 HA segments were sequenced from LPAI viruses from which whole genome sequences could not be obtained. Publicly available sequence data for this epidemic were downloaded from the Influenza Virus Resource at GenBank and also included in the analysis: the complete genome of 37 H7N1 viruses and the HA gene of 11 viruses. Epidemiological information (collection date, city and province of collection, and pathotype) for all the viruses included in this study (n = 144) is available in Annex C - Table C1. All sequence data generated here has been submitted to GenBank and assigned accession numbers KF492991-KF493575.

Viral RNA was extracted from the infected allantoic fluid of specific-pathogen-free fowls' eggs using the Nucleospin RNA II kit (Macherey-Nagel, Duren, Germany) and reverse transcribed with the SuperScript III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA). PCR amplifications were performed with the PfuTurbo DNA Polymerase (Stratagene, La Jolla, CA) by using specific primers (sequences are available on request). The complete coding sequences were generated using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). The products of the sequencing reactions were cleaned-up using the PERFORMA DTR Ultra 96-Well kit (Edge BioSystems, Gaithersburg, MD) and sequenced in a 16-capillary ABI PRISM 3130xl genetic analyzer (Applied

Biosystems, Foster City, CA). Sequence data were assembled and edited with SeqScape software v2.5 (Applied Biosystems). Sequences from all the eight gene segments were aligned and compared with the most related sequences available in GenBank. In addition, the phylogenetic relationship between the Italian H7N1 viruses and all publicly available Eurasian H7 sequences (n=443) was evaluated for the HA segment.

5.3.2. Library preparation, Illumina sequencing and data analysis

To assess virus population diversity, next-generation sequencing (NGS) was performed on all available H7N1 influenza A clinical samples: six LPAI and seven HPAI viruses. Full sample details are described in table 5.1. Viral RNA was extracted directly from the infected clinical samples such as trachea, lung and intestine using the Nucleospin RNA II kit (Macherey-Nagel, Duren, Germany). Complete influenza A genomes were prepared from the RNA using the SuperScript III One-Step RT-PCR system with Platinum®Taq High Fidelity (Invitrogen, Carlsbad, CA) following Zhou et al. (2009) protocol. All amplified products were visualized on a 0.7% agarose gel stained with GelRedTM. High-throughput sequencing of the amplified genome was performed using an Illumina MiSeq at IGA Technology Services srl (UD), Italy. The amplified genomes were fragmented down to 600-800 bp and tagged with sequencing adapters with indexes in a single step, using the Nextera DNA Sample Preparation Kit from Illumina. Libraries were run on the Caliper GX (Perkin Elmer) for quality control and QBit fluorimeter (Invitrogen) was used to determine the molar concentration of the libraries. Finally the indexed libraries were pooled in equimolar concentrations and sequenced in multiplex for 150bp paired-end on Illumina MiSeq, according to the manufacturer's instructions.

Raw sequence reads were inspected using FASTQC to assess the quality of data coming from the high throughput sequencing pipelines. Fastq files were cleaned with PRINSEQ and Trim Galore to remove adaptors and primers and exclude reads with a Phred quality score below 20. Only reads longer than 80 bp were considered and mapped to the reference sequences. Reads were aligned to both the A/Italy/4749/99 (HPAI) and A/Italy/3675/99 (LPAI) reference sequences using Stampy (Lunter and Goodson, 2011), an alignment tool particularly suited to map reads that contain sequence variation relative to the reference. This alignment protocol also enabled us to identify possible key mutations associated with the highly pathogenic virus in the low pathogenic virus samples and vice versa. Unmapped reads

were removed from the resulting BAM files using samtools version 0.1.14 (Li et al., 2009). Depth of coverage of the alignment was visualized using BEDTools (Quinlan and Hall, 2010) and R. The BAM alignment files were parsed using a custom perl script to identify the frequency of polymorphisms at each site relative to the reference used for the alignment. A binomial test, using the error probability calculated from the base qualities at a site and following the approach of Morelli et al. (2013), was used to determine whether the frequency of each mutation was above the artifactual value expected from Illumina sequencing. All low-frequency polymorphisms that were likely to be miscalled bases during sequencing were removed from downstream analyses.

5.3.3. Phylogenetic analysis

Nucleotide sequence alignments were manually constructed for each gene segment and for the concatenated whole genome using the Se-Al program (Rambaut 2002). To infer the evolutionary relationships for each gene segment, we employed the maximum likelihood (ML) method available in the PhyML program, incorporating a GTR model of nucleotide substitution with a gamma distribution of among site rate variation (with four rate categories, G4), and a SPR branch-swapping search procedure (Guindon and Gascuel, 2003). A bootstrap resampling process (1,000 replications), using the neighbour-joining (NJ) method and incorporating the ML substitution model defined above, was employed to assess the robustness of individual nodes of the phylogeny using PAUP* (Wilgenbusch and Swofford, 2003). Parameter values for the GTR substitution matrix, base composition, gamma distribution of among-site rate variation, and proportion of invariant sites (I) were estimated directly from the data using MODELTEST (Posada and Crandall, 1998).

Table 5.1. Identification number, epidemiological data and total number of sequences obtained for each viral population derived from clinical specimens using the Illumina deep sequencing platform.

ID Num	Sample name and pathotype	Type of sample	Date of collection	Location	Total sequences*	Mapped sequences
4618/99	A/turkey/Italy/4618/1999, HPAI	Trachea	14/12/1999	VR	423574	418238
4828/99	A/turkey/Italy/4828/1999, HPAI	Lung, Trachea	23/12/1999	VR	460708	460491
4827/99	A/turkey/Italy/4827/1999, HPAI	Lung, Trachea	23/12/1999	VR	431304	431085
4749/99	A/turkey/Italy/4749/1999, HPAI	Intestine	21/12/1999	VR	544596	541929
4708/99	A/turkey/Italy/4708/1999, HPAI	Trachea	21/12/1999	VR	501086	496408
4911/99	A/chicken/Italy/4911/1999, HPAI	Intestine	27/12/1999	RO	386926	378252
4756/99	A/turkey/Italy/4756/1999, HPAI	Intestine	22/12/1999	na^	463704	413277
2732/99	A/turkey/Italy/2732/1999, LPAI	Lung, Trachea	03/08/99	VR	533706	533293
3675/99	A/turkey/Italy/3675/1999, LPAI	Trachea	04/10/99	VR	591410	591173
4295/99	A/turkey/Italy/4295/1999, LPAI	Lung	22/11/1999	VR	520056	516798
3283/99	A/turkey/Italy/3283/1999, LPAI	Lung, Trachea	10/09/99	VR	534136	529407
4829/99	A/turkey/Italy/4829/1999, LPAI	Lung, Trachea	23/12/1999	VR	542062	541952
1744/99	A/turkey/Italy/1744/1999, LPAI	Lung, Trachea	26/04/99	VR	195496	194543

*After removal of adapter and low quality sequences

^na: not available

5.3.4. Analysis of selection pressures

Gene- and site-specific selection pressures for all segments of the Italian H7N1 viruses were measured as the ratio of nonsynonymous (dN) to synonymous (dS) nucleotide substitutions per site. In all cases, dN/dS ratios were estimated using the Single-Likelihood Ancestor Counting (SLAC), Fixed-Effects Likelihood (FEL) and Internal Fixed-Effects Likelihood (IFEL) methods available at the Datamonkey online version of the Hy-Phy package (Pond and Frost, 2005; Delport et al., 2010). In addition, we employed the Mixed Effects Model of Evolution (MEME), which is capable of identifying instances of both episodic and positive selection at individual sites (Murrel et al., 2012). All analyses utilized the GTR model of nucleotide substitution and employed input NJ phylogenetic trees.

5.3.5. Evolutionary dynamics and spatial analysis

Rates of nucleotide substitution per site per year (subs/site/year) and the time to the most recent common ancestor (tMRCA) of the sampled data were estimated using the Bayesian Markov chain Monte Carlo (MCMC) approach available in the BEAST program, version 1.6.2 (Drummond and Rambaut, 2007). For each analysis, we employed a flexible Bayesian skyline coalescent tree prior (10 piece-wise constant groups) as this is typically the best descriptor of the complex population dynamics of influenza A virus (Drummond et al., 2005). In addition, we utilized a HKY85 + G4 model of nucleotide substitution with two data partitions reflecting codon positions (1st + 2nd positions, 3rd position), and with base frequencies unlinked across all codon positions (i.e., the SRD06 substitution model). Two molecular clock models—strict (constant) and relaxed (uncorrelated lognormal) —were compared by analysing values of the coefficient of variation (CoV) in Tracer (Drummond and Rambaut, 2007), in which CoV values of >0 are evidence of non-clock-like evolutionary behaviour. The molecular clock model utilized for each data set is indicated in table 5.3. In all cases, statistical uncertainty is reflected in values of the 95% highest posterior density (HPD) for each parameter estimate, and in each case chain lengths were run for at least 50 million iterations to achieve convergence as assessed using Tracer v1.5 program (Drummond and Rambaut, 2007). Finally, Maximum Clade Credibility (MCC) phylogenetic trees were summarized from the posterior distribution of trees using TreeAnnotator v1.6.1 (Drummond and Rambaut, 2007) after the removal of 10% burnin. The topologies of the MCC trees,

visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/), were similar to those inferred using ML methods.

To infer the spatial spread of H7N1 we analysed two data sets separately, one containing the HA gene from 139 viruses and another including the eight gene segments concatenated together for 103 isolates, after the removal of all gaps (sequence concatenation was appropriate as we found no evidence for reassortment – see Results). Sequences of viruses sampled in regions for which fewer than three samples where available were excluded. We grouped the H7N1 sequences into eight discrete geographical locations each representing one province or adjacent provinces such that each group had at least three sequences: VR (Verona), BS (Brescia), MN (Mantova), FE-RA-RO-BO (Ferrara, Ravenna, Rovigo, Bologna), VE-PD-VI (Venezia, Padova, Vicenza), PN-UD (Pordenone and Udine), BG-CO (Bergamo and Como), MI-LO (Milano and Lodi) (see Annex C - Table C1) (figure 5.1). All these areas are located in northern Italy, and most are situated in the Veneto and Lombardia regions.

To determine the overall degree of geographical structure among the Italian H7N1 viruses, we employed the Association Index (AI) and Parsimony Score (PS) phylogeny-trait association statistics available in the Bayesian Tip-association Significance testing (BaTS) program (Parker et al., 2008). In this case, traits were defined as the eight geographical regions described above. This method uses the posterior distribution of trees obtained from the BEAST analysis and accounts for phylogenetic uncertainty in the data. The BaTS program also allowed us to assess the level of clustering in individual geographic locations using the Monophyletic Clade (MC) size statistic (Parker et al., 2008). In all cases, 1000 random permutations of tip locations were undertaken to create a null distribution for each statistic. This statistic determines the association between sampling location and phylogeny by estimating the size of the largest cluster of sequences from each sampling location. Again, the LPAI and HPAI viruses were analysed separately.

To infer phylogeographic histories in more detail we utilized a Bayesian phylogeographic approach (Lemey et al., 2009), employing the MCMC method implemented in BEAST. For this analysis we utilized the date (exact day) of viral sampling, a Bayesian skyline prior and the SRD06 model of nucleotide substitution, as described above. For the HA data set, a strict molecular clock was used, while for the concatenated genes a relaxed (uncorrelated lognormal) molecular clock model was preferred. Analyses were undertaken for the HA

segment of the LPAI and HPAI viruses independently. The MCMC chain was run for 150 million iterations. All parameters reached convergence, as assessed using the Tracer v1.5 program (Drummond and Rambaut, 2007). The initial 10% of the chain was removed as burn-in and MCC trees were again summarized using TreeAnnotator v1.6.1. We also produced posterior summaries of the root and HPAI internal node geographical states as described by Lemey et al. (2009).

Finally, to assess the role of poultry density in determining patterns of viral diffusion in northern Italy, we parameterized a discrete phylogeographic diffusion model using data derived from the Italian National Institute for Statistics (ISTAT) 5th Census of Agriculture (2000). Specifically, we collected information on the poultry population size for three provinces in the Lombardia region (Mantova, Brescia, Bergamo) and four in the Veneto region (Venezia, Verona, Padova, Vicenza) from which at least three sequences were available. From these data we determined (a) the poultry population size of the province of origin, (b) the poultry population size of the province of destination, and (c) the product of the poultry population sizes from the province of origin and the province of destination, all of which were considered as predictors in the phylogeographic analysis, in addition to geographic distance. Each of these predictors was incorporated into an asymmetric transition matrix that allows for separate directional rates between each pair of locations. We used a Bayes Factor (BF) comparison (Suchard et al., 2001) of the relative marginal likelihoods to select the most appropriate model for the data, compared to a null model of equal migration rates.

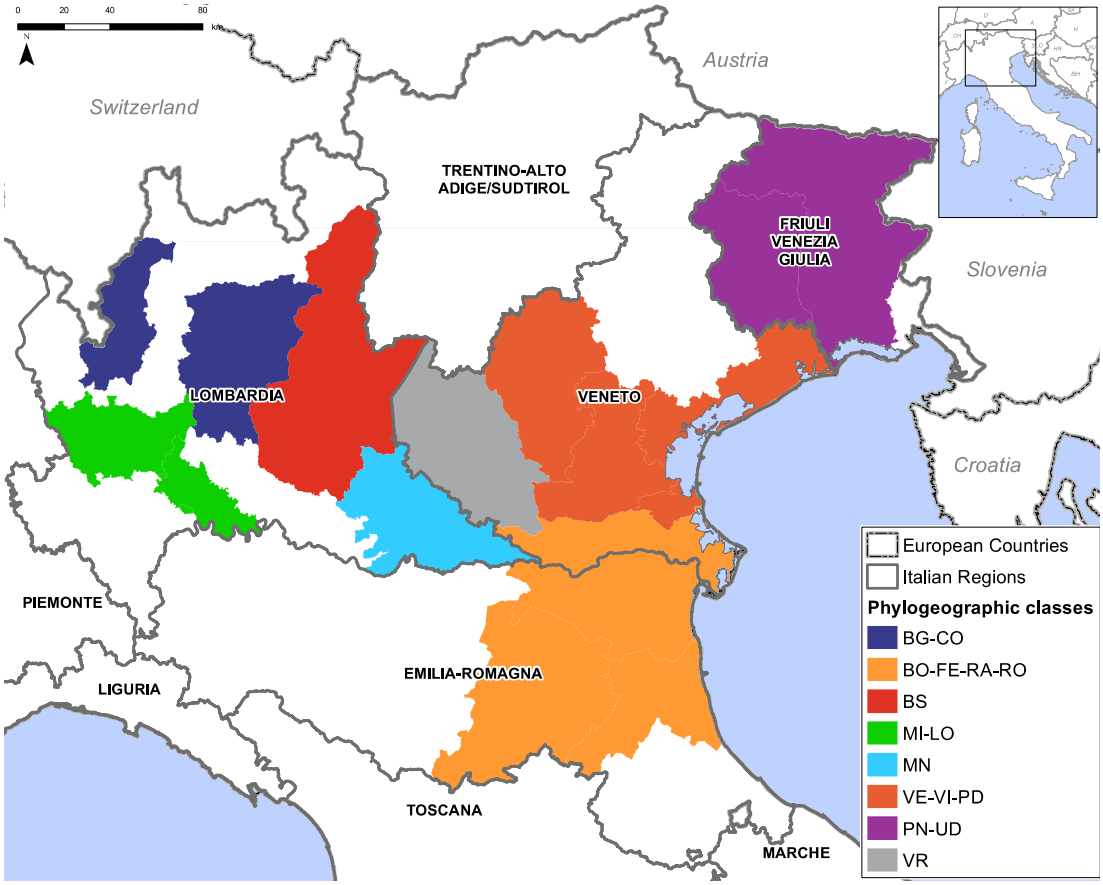


Figure 5.1. Map of the discrete geographical locations used in the spatial analysis. Locations are coloured as follows: grey for Verona, red for Brescia, light blue for Mantova, yellow for Ferrara, Ravenna, Rovigo and Bologna, orange for Venezia, Padova and Vicenza, violet for Pordenone and Udine, blue for Bergamo and Como, green for Milano and Lodi.

5.4. RESULTS

5.4.1. Phylogenetic analysis of LPAI and HPAI H7N1 viruses

We inferred the evolutionary relationships of HA sequences of 144 avian influenza H7N1 viruses collected during the 1999-2001 Italian epidemic (96 sequences generated in this study and 48 obtained from GenBank). These data included 57 LPAI viruses sampled from March to January 2000, 69 HPAI isolates detected between December 1999 and April 2000, and 18 LPAI samples isolated from August 2000 to February 2001. As background, 443 Eurasian H7 sequences available in GenBank were included in the analysis. The topology of the HA phylogeny indicates that the LPAI and HPAI Italian H7N1 viruses clustered together within a monophyletic group that was distinct from all other European H7 viruses (Annex C - Figure

C1), suggesting a single introduction of the virus into Italy (figure 5.2). Within this clade, we identified two main phylogenetic clusters, one that includes all Italian H7N1 HPAI viruses (yellow shading), and another that contains all Italian H7N1 LPAI samples collected from both time periods: late September 1999 – January 2000 and August 2000 – February 2001 (i.e. before and after the HPAI outbreak). However, the two clusters are not monophyletic as the HPAI viruses evolved from a subset of the LPAI viruses.

To explore the genome-scale evolution of the viruses collected during this epidemic, ML trees were inferred for each of the eight gene segments (Annex C - Figure C2 and C3) for a total of 109 H7N1 isolates: 63 HPAI isolates, and 46 LPAI viruses sampled before (n = 28) and after (n = 18) the HPAI outbreak, and a subset of the most closely related sequences publicly available on GenBank. Phylogenies inferred for each of the seven other genome segments show an identical pattern as observed in the HA tree: the H7N1 viruses represent a single viral introduction into Italy, and a distinct clade of Italian HPAI viruses (bootstrap values > 70%) has evolved from a subset of the precursor LPAI viruses. Importantly, there is no evidence for reassortment in the evolutionary history of all the H7N1 during this Italian epidemic. In addition to the HPAI viruses, a second genetically distinct cluster is present within the Italian H7N1 LPAI viruses, which we term 'LPAI-1' (shaded in grey on figure 5.2). The LPAI-I cluster is defined by high bootstrap values (>70%) and long branches in the HA, NA, PB2, PA, NS phylogenies (figure 5.2, Annex C - Figure C2 and C3). This clade, which is characterized by a total of 16 amino acid signature mutations (table 5.2), includes samples collected from the first wave of the LPAI epidemic (late September 1999 to January 2000) and from the second wave (August 2000 to February 2001), suggesting that this variant likely circulated undetected in Italy during the HPAI outbreak and re-emerged in the commercial flocks almost four months after the eradication of the HPAI strain. The sequence A/chicken/Italy/3283/1999 (marked with a black arrow in the phylogenetic trees), isolated in early September 1999, is the most closely related to this cluster on the HA, NA, NS, PB2, PB1 trees, and possesses five out 16 amino acid substitutions specific to LPAI-I isolates (table 5.2).

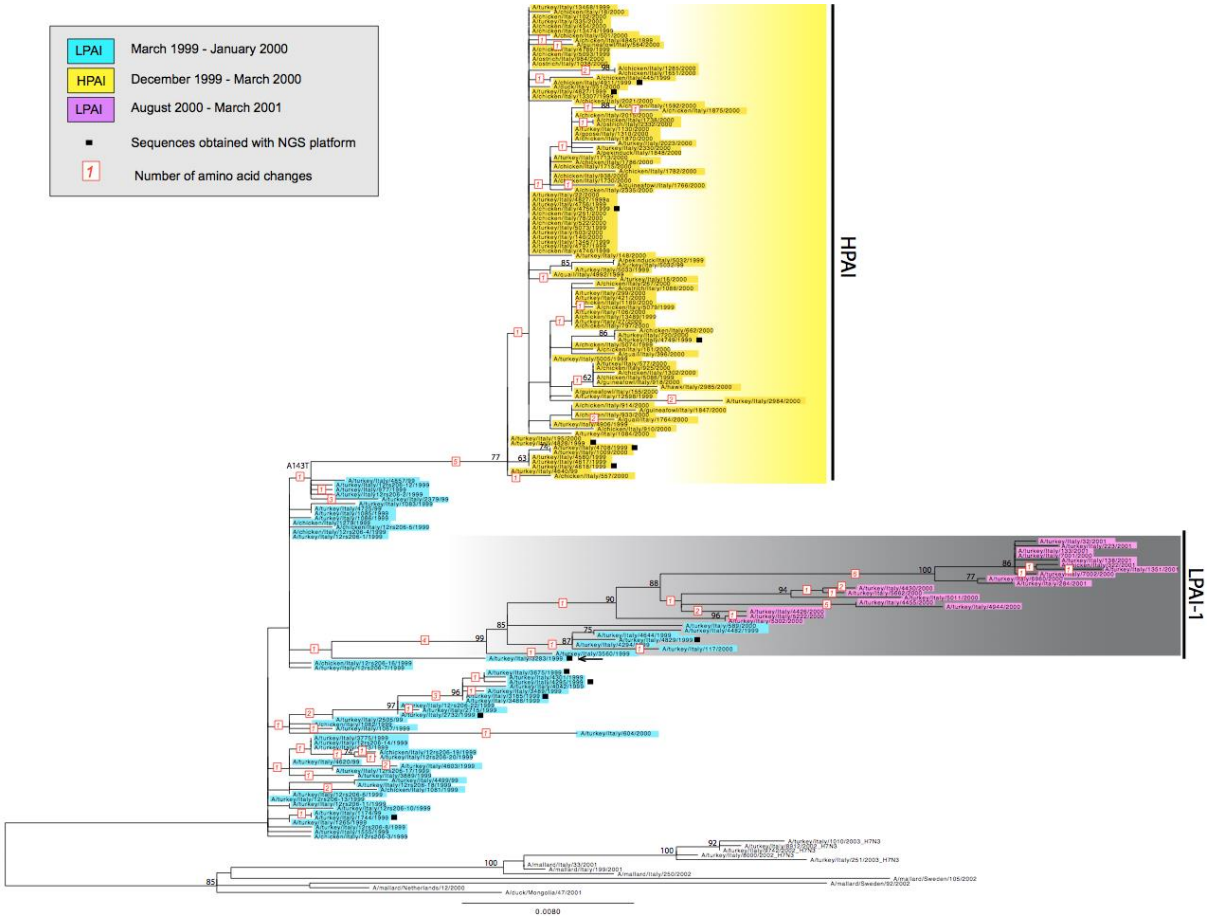


Figure 5.2. ML tree of the HA gene segment of H7N1 avian influenza virus. Viruses are coloured as follows: blue for LPAI viruses collected from March 1999 to January 2000, pink for LPAI viruses sampled between August 2000 and February 2001, and yellow for HPAI viruses (December 1999-March 2000). The numbers at nodes represent bootstrap values (>70%), while branch lengths are scaled according to the numbers of nucleotide substitutions per site. The number highlighted in red represents the number of amino acid changes identified along the main branches of the tree. Sequences obtained using the NGS platform are marked with a black square. The closest relative of the LPAI-I group is identified with a black arrow. The tree is mid-point rooted for clarity only.

Table 5.2. List of the amino acid signature mutations of HPAI and LPAI-I viruses

HPAI VIRUSES						LPAI-I VIRUSES				
Gene	**Position**	**Amino acids**			**Comments**	**Position**	**Amino acids**			**Comments**
		LPAI	**LPAI-I**	**HPAI**			**LPAI**	**LPAI-I**	**HPAI**	
PB2	398	T	T[1]	**I**	[1]2 viruses possess the aa I	527	L	**M**	L	
PB1	154	G	G	**D**		211	R	**K**	R	A/tk/It/3283/00 possesses the aa K
	216	S	S	**G**						
	745	K	K	**E**[2]	[2]1 virus possesses the aa G					
PA	61	I	I	**T**		190	S	**P**	S/Y	
	115	N	N	**K**						
	252	E	E	**K**[3]	[3]5 viruses possess the aa E	257	I	**V**	I	
HA	130	T	T	**A**		86	Q	**R**[9]	Q	[9]2 viruses possess the aa Q
						113	E	**G**	E	
	146	A/S	A/S	**T**[5]	[5]2 viruses possess the aa A or S	143	A/T	**V**	T[10]	[10]3 viruses which posses the aa A or K
	228	A[6]	A	**E**	[6]1 virus possesses the aa E	148	R	**K**	R	
	454	A	A[7]	**T**	[7]2 viruses possess the aa T	169	A/T	**T**[11]	A	A/tk/It/3283/00 possesses the aa T; [11]1 virus possesses the aa I
	554	K	K	**R**						
NP	349	T	T	**A**						
	376	S	S	**N**						
NA	173	R	R	**K**		163	V	**M/L**	V	
						314	I	**L**	I	A/tk/It/3283/00 possesses the aa L
						418	M	**I**	M[12]	[12]2 viruses possess the aa I
M1	166	V	V	**A**						
NS1	136	V	V	**I**		47	G	**C**	G	
	139	D	D	**N**		85	P	**T**	P	
	225	R[8]	-	**STOP**	[8]1 virus possesses a stop codon at position 221	221	Y/H[13]	**STOP**	Y	A/tk/It/3283/00 possesses a stop codon; [13]1 virus possesses a stop codon
NS2						64	T[14]	**K**	T	A/tk/It/3283/00 possesses the aa K; [4]1 virus possesses the aa K

5.4.2. Amino acid mutations of LPAI-1 and HPAI H7N1 viruses

The evolution from the low to the highly pathogenic H7N1 appears to have involved more than the multiple basic amino acids insertion at the HA cleavage site. Across the viral genome the HPAI isolates are characterized by 19 unique amino acid substitutions: five in the HA, one in the neuraminidase (NA), one in the polymerase PB2, three in the polymerase PB1, three in the polymerase PA, two in the nucleoprotein (NP), one in the matrix protein 1 (M1) and three in the non-structural protein 1 (NS1), including a nonsense mutation (table 5.2). This latter mutation results in a C-terminal truncation of six amino acids that promotes the transport of NS1 into the nucleoli, where it may favour viral transcription and translation or may alter the cell cycle for viral needs (Keiner et al., 2010). Of the 69 HPAI viruses analysed here, 66 possess a substitution at position 143 of the HA gene (A143T) that introduces a potential additional glycosylation site at position 141, located close to the 130 loop of the receptor binding domain (Skehel and Wiley, 1981) and within the equivalent antigenic site A in H3 viruses (Wiley et al., 1981). This mutation also was identified in a HPAI H7N7 virus isolated from a fatal human case in the Netherlands in 2003, where it was demonstrated to increase viral replication efficiency in MDCK cells and to alter viral attachment to human lung tissue (de Wit et al., 2010). Of note, this additional glycosylation site was also found among the five low-pathogenicity precursor viruses that are the most closely related to the HPAI viruses on the HA tree that are likely to be their evolutionary precursors.

Similar to the HPAI viruses, the LPAI-1 isolates have a truncation in the NS1 protein and an additional glycosylation site (AGS) in the HA molecule. However, in contrast to the HPAI viruses, the truncation in the NS1 protein is of ten rather than six amino acids in length (total length of 220 amino acids) and the potential AGS is located at position 169 (within the corresponding H3 and H5 antigenic site B of the HA molecule) instead of 143 (Kaverin et al., 2007; Lee et al., 2007). The NS1 truncation may have the same effect as the six amino acid deletion observed in the HPAI viruses, while the new glycosylation site in the HA protein has been demonstrated to promote influenza virus replication in cell culture (Wangner et al., 2000) and may also have an antigenic effect. Notably, the LPAI-1 viruses also possess an amino acid substitution at position 143 (A143V) of the receptor-binding domain of the HA. Unlike the HPAI strain, this mutation does not introduce any glycosylation site, but it could affect the receptor binding conformation. Of note, in the HA phylogeny all the viruses collected in south Verona and Padova from December 2000 to February 2001 following the

introduction of the vaccination programme (Capua et al., 2003) are separated from samples detected from the same area in the period immediately before (October 2000) by a long branch and at least six amino acid substitutions, two of them located in the H3-corresponding antigenic sites A (G151E) and B (G195V) (Bush et al., 1999).

5.4.3. Time-scaled evolutionary history of LPAI and HPAI H7N1 viruses

Rates of nucleotide substitution and the times to the most recent common ancestor (tMRCAs) were estimated using a Bayesian coalescent approach (Drummond and Rambaut, 2007) for each genome segment separately and for the eight concatenated segments. Evolutionary rates also were calculated for the entire H7N1 viral population and for the HPAI and LPAI viruses separately, although no significant difference in rates was observed between the LPAI and HPAI viruses for any of the genome segments (table 5.3). For the entire viral population, nucleotide substitution rates were high for the genes encoding the surface glycoproteins: mean rate of 10.15 x 10-3 substitutions per site per year (sub/site/year) (95% HPD, 8.5 - 11.9 x 10-3) for the HA gene, and 9.86 x 10-3 sub/site/year (95% HPD, 8.01 - 11.79 x 10-3) for the NA gene (table 5.3). These rates are similar to those calculated for the HA and NA from viruses collected during the 2003 HPAI H7N7 outbreak in the Netherlands (Bataille et al., 2011). The estimated rates for the other genome segments are lower, ranging from 5.42 x 10-3 sub/site/year (95% HPD, 3.81 - 7.15 x 10-3 sub/site/year) for the NP gene to 7.72 x 10-3 sub/site/year (95% HPD, 5.76 - 9.66 x 10-3) for the NS gene segment. Interestingly, the evolutionary rate for the branch leading from the low pathogenic progenitor strain to the HPAI lineage was similar – at 4.4-7.8 x 10-3 subs/site/year for the concatenated genome (95% HPD) – to the rest of the phylogeny, with no evidence of a major rate elevation.

Table 5.3. Estimated rates of nucleotide substitution and tMRCA for LPAI and HPAI H7N1 viruses collected during the Italian epidemic.

Gene	Genetic group	Models	Substitution rates (subs/site/year)		tMRCA	
			Mean ($x10^{-3}$)	95% HPD ($x10^{-3}$)	Mean	95% HPD
HA	H7N1	SRD06-SC-Skyline	10.15	8.5-11.9	Feb 1999	Jan 1999-Mar 1999
	H7N1-LPAI	SRD06-SC-Skyline	9.75	7.8-11.9	Feb 1999	Jan 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	9.89	6.54-13.22	Nov 1999	Oct 1999-Dec 1999
NA	H7N1	SRD06-SC-Skyline	9.86	8.01-11.79	Feb 1999	Dec 1998-Mar1999
	H7N1-LPAI	SRD06-SC-Skyline	10.06	7.89-12.37	Feb 1999	Jan 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	10.12	5.41-15.39	Nov 1999	Oct 1999-Dec 1999
PA	H7N1	SRD06-SC-Skyline	5.84	4.79-6.91	Feb 1999	Feb 1999-Mar 1999
	H7N1-LPAI	SRD06-SC-Skyline	5.49	4.20-6.77	Mar 1999	Feb 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	6.64	4.08-9.69	Nov 1999	Sep 1999-Dec 1999
MA	H7N1	SRD06-SC-Skyline	7.21	5.02-9.51	Feb 1999	Dec 1998-Mar1999
	H7N1-LPAI	SRD06-SC-Skyline	4.87	2.77-7.26	Feb 1999	Dec 1998-Mar1999
	H7N1-HPAI	SRD06-SC-Skyline	8.87	4.69-13.37	Nov 1999	Sep 1999-Dec 1999
PB1	H7N1	SRD06-SC-Skyline	5.57	4.54-6.68	Feb 1999	Gen 1999-Mar1999
	H7N1-LPAI	SRD06-SC-Skyline	5.70	4.31-7.15	Feb 1999	Jan 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	5.48	3.60-7.39	Oct 1999	Aug 1999-Nov 1999
PB2	H7N1	SRD06-SC-Skyline	6.81	5.54-8.04	Feb 1999	Dec 1998-Mar 1999
	H7N1-LPAI	SRD06-SC-Skyline	6.69	5.12-8.23	Jan 1999	Dec 1998-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	7.22	5.0-9.64	Oct1999	Aug 1999-Nov 1999
NP	H7N1	SRD06-SC-Skyline	5.42	3.81-7.15	Feb 1999	Dec 1999-Mar1999
	H7N1-LPAI	SRD06-SC-Skyline	5.11	3.00-7.39	Feb 1999	Dec 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	4.91	2.61-7.73	Nov 1999	Oct 1999-Dec 1999
NS	H7N1	SRD06-SC-Skyline	7.72	5.76-9.91	Jan 1999	Oct 1998-Mar 1999
	H7N1-LPAI	SRD06-SC-Skyline	7.19	4.85-9.66	Jan 1999	Nov 1998-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	6.44	2.90-10.31	Jun 1999	Jan 1999-Dec 1999
CG	H7N1	SRD06-SC-Skyline	6.67	5.97-7.34	Mar 1999	Feb 1999-Mar 1999
	H7N1-LPAI	SRD06-SC-Skyline	6.82	5.81-7.87	Feb 1999	Feb 1999-Mar 1999
	H7N1-HPAI	SRD06-SC-Skyline	6.41	5.25-7.62	Oct 1999	Sept 1999-Nov 1999

SC=strict molecular clock; UL=uncorrelated lognormal molecular clock; Skyline= Bayesian skyline coalescent prior; HPD=Highest Posterior Density

To explore the evolutionary origins of the H7N1 viruses in Italy we estimated the tMRCA of (a) the entire H7N1 population, (b) the HPAI viruses, and (c) LPAI viruses for (i) each genome segment and (ii) the eight segments concatenated. The tMRCAs for the entire Italian H7N1 viral population ranged from December 1998 to March 1999 (95% HPD value) for all the gene segments (with the exception of NS), suggesting that the detection of the outbreak in March 1999 occurred no more than three months after the entry of the virus into Italy and possibly simultaneously with the appearance of the virus. Similarly, the tMRCAs for the HPAI viruses ranged from August 1999 to December 1999 (95% HPD), 0-5 months before the first detection of the HPAI strain, for all eight segments (table 5.3). These dates overlap with the period of circulation of the LPAI viruses and are therefore consistent with the emergence of HPAI from the LPAI progenitor.

5.4.4. Selection pressures in the H7N1 genes

An analysis of selection pressures on the all virus genes revealed that the vast majority of codons were subject to purifying selection (mean dN/dS ratios ranged from 0.10 for the PB2 and 0.83 for the M2; table 5.4). Only the PB1-F2 protein, which is only 90 amino acids long, showed evidence of diversifying selection, although this is likely to be an artefact due to its encoding in an overlapping reading frame (dN/dS = 2.65; table 5.4) (Holmes et al., 2006). Using the SLAC, FEL, IFEL and MEME methods, we identified several sites (p-value<0.05) in the HA, NA, PB2 and NS2 with evidence of putative positive selection (table 5.4). Interestingly, positions 143 and 146 adjacent to the receptor binding site in the HA gene are highly polymorphic, which may affect virus interactions with the cell surface. Three (T, K and A) and two (A and V) distinct amino acids were observed at position 143 in the HPAI and LPAI viruses, respectively. At position 146 three distinct amino acids substitutions (A, T, S) were observed, with the majority of the HPAI and LPAI viruses retaining the amino acid T and A, respectively.

Table 5.4. Amino acid sites under putative positive selection detected using different analytical models and mean d_N/d_S ratio for each gene.

Gene	Mean d_N/d_S	Positively selected sites									
		SLAC	**p-value**	**FEL**	**p-value**	**IFEL**	**p-value**	**REL**	**Post. prob**	**MEME**	**p-value**
HA	0.37	-	-	143	0.04	77 468	0.03 0.04	NA	NA	146	0.05
HA HPAI	0.41	-	-	-	-	264 468	0.04 0.05	-	-	-	-
HA LPAI	0.34	-	-	-	-	-	-	8 86 143 146 169 195	0.97 0.87 0.97 0.97 0.97 0.97	-	-
NA	0.29	-	-	-	-	-	-	NA	NA	383	0.02
PB2	0.10	-	-	195 398	0.07 0.07	195 398 640	0.07 0.07 0.08	NA	NA	195 398 527 702 752	0.0009 0.04 0.05 0.04 0.02
PB1	0.14	-	-	-	-	-	-	NA	NA	-	-
PA	0.28	-	-	-	-	-	-	NA	NA	-	-
NP	0.14	-	-	-	-	-	-	NA	NA	-	-
M1	0.29	-	-	-	-	-	-	-	-	-	-
M2	0.83	-	-	-	-	-	-	-	-	-	-
NS1	0.50	-	-	-	-	-	-	-	-	-	-
NS2	0.31	-	-	-	-	-	-	46 49 68 116	0.99 0.99 0.99 0.99	46	0.04
PB1-F2	2.65	-	-	-	-	-	-	-	-	-	-

Post. Prob = posterior probability

5.4.5. Spatial movement of H7N1 in Northern Italy

To investigate the spatial diffusion of H7N1 avian influenza viruses within northern Italy, we inferred Bayesian phylogenies for the HA and the eight concatenated gene segments, considering eight discrete regions in northern Italy that are well sampled in our data (figure 5.1, 5.3, Annex C – Figure C4). To quantitatively identify geographical structure in these data, we measured the extent to which isolates clustered by geographical location using a phylogenetic trait-association test (Annex C - Table C2). Significant population subdivision was observed among the LPAI viruses in all locations ($P<0.01$, Annex C - Table C2) except Mantova ($P=0.11$, Annex C - Table C2). In contrast, HPAI viruses exhibited no statistically significant geographic structuring by province, indicative of greater spatial mixing ($P>0.13$) as expected during a rapidly spreading outbreak.

The Bayesian MCC tree inferred for the HA segment (figure 5.3) identifies the Mantova and Verona provinces as the most important in the emergence and spatial dissemination of the H7N1 virus within northern Italy. Verona and Mantova have the highest root state posterior probabilities (0.56 and 0.67, respectively, figure 5.3). These relatively low posterior probabilities, such that spatial origins are difficult to identify rigorously, are again suggestive of the rapid movement of viruses between geographical localities. The original LPAI H7N1 outbreak is most likely to have begun in Mantova (posterior probability = 0.67), with subsequent spread to Verona and Brescia, and from Verona to the Padova and Rovigo provinces (figures 5.1, 5.3), although the relatively low posterior probability values are indicative of rapid spread with little phylogenetic resolution. Although the HPAI virus appears to have originated first in Verona (posterior probability = 0.56), our phylogeographic analysis suggests that the poultry population in Verona has not played a major role in the spatial diffusion of this strain to other provinces. Instead, poultry from Mantova province most likely represent the key source for the spread of the HPAI virus to the six other provinces (figures 5.1, 5.3). In addition, the second LPAI outbreak during August 2000 – February 2001 also is likely to have begun in Verona, with subsequent spread to Padova but not to other regions. Separate analyses of the LPAI and HPAI viruses gave similar results to those obtained from the complete data set (data not shown).

To quantitatively estimate the importance of poultry densities and geographical distances in the spatial dynamics of the virus, we encoded four potential predictors of viral dissemination between locations and fitted these models individually to the sequence data: (a) the poultry population in the province of origin, (b) the poultry population in the province of destination, (c) the product of the poultry population in the provinces of origin and destination, and (d) the geographic distances between the different provinces. Bayes factor comparisons indicate that the spatial dissemination of the H7N1 viruses in Italian poultry are best described by the size of the poultry population in the region of destination (BF = 2.01 for the HA gene; BF = 3.99 for the complete genome; Annex C - Table C3), indicating viral migration from provinces of relatively lower poultry population size (i.e. Mantova, Bergamo, Venezia) to provinces with larger poultry populations (i.e. Verona) (Annex C - Table C3). Considering geographical distance as a predictor of viral movements also improved the marginal likelihood (BF = 1.41 for the HA gene; BF = 1.32 for the complete genome; Annex

C - Table C3), indicating higher rates of viral dissemination among geographically closer provinces (Annex C - Table C3).

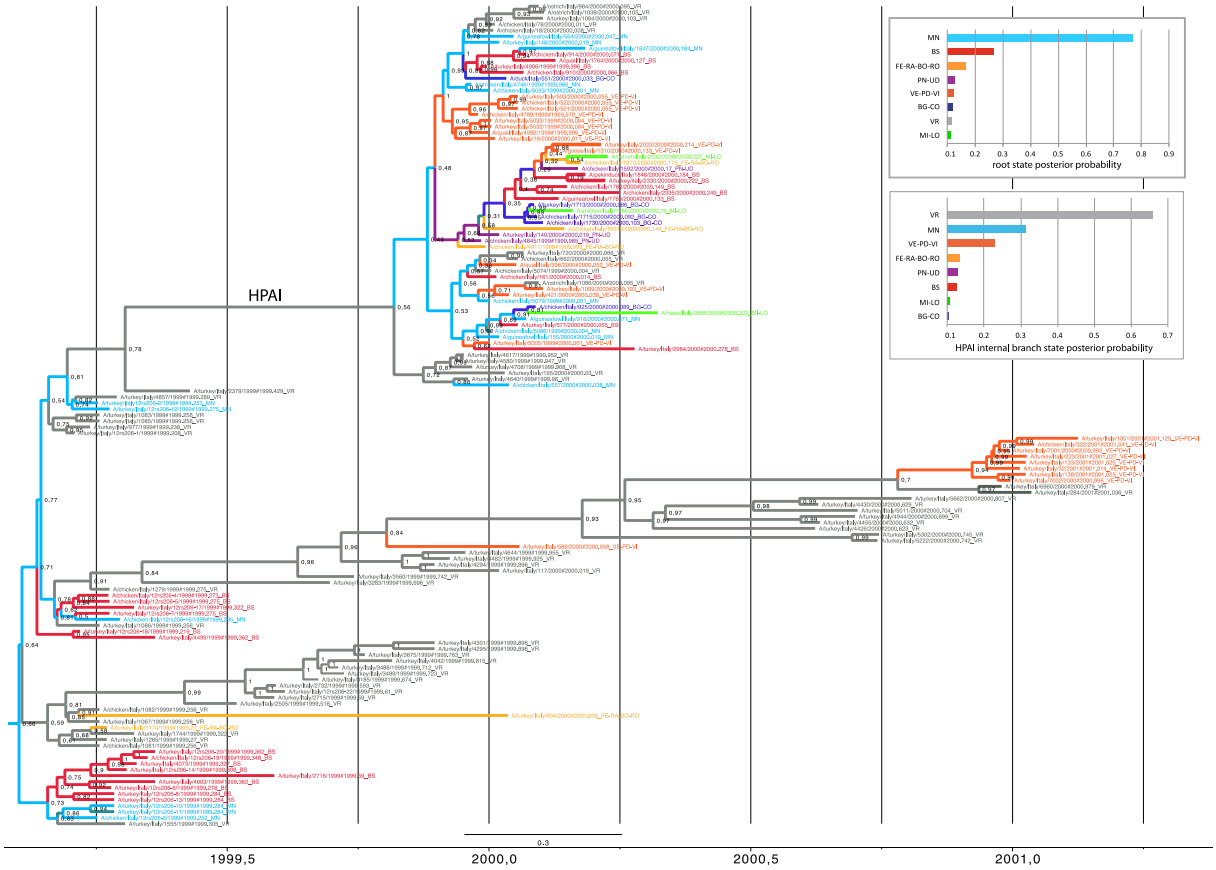


Figure 5.3. MCC tree inferred for HA gene sequences of the Italian H7N1 viruses. Sequences are coloured according to the province of origin. Light blue: Mantova; grey: Verona; red: Brescia; yellow: Ferrara-Ravenna-Bologna-Rovigo; orange: Venezia-Padova-Vicenza; blue: Bergamo-Como; green: Milano-Lodi; violet: Pordenone-Udine. The numbers at branch points represent the probability of spatial states. The location state posterior probability distributions for the root and for the internal branch of the HPAI group are shown in the two graphs to the upper right of the phylogenetic tree.

5.4.6. Intra-host genetic variation using NGS

We used an ultra-deep sequencing approach to determine whether mutations associated with the HPAI phenotype were present at low levels in LPAI viral populations. Complete genomes were obtained for all 13 clinical samples included in the study, with a mean of 472,508 reads per sample (table 5.1). The depth of coverage for the polymerase genes (PA, PB1 and PB2 segments) was lower (approximately between 1 and 1000X per site) than the

other five segments (approximately between 1000 and 100,000X per site), indicating that there might be a fragment length bias, where the longest fragments received less coverage (a coverage map is available for each sample in Annex C - Figures C5 and C6). We also obtained a smaller number of sequences (195,496) for sample 1744 and the coverage for the 5' end of the PB1 segment was not sufficient to study population variation (table 5.1 and Annex C - Figures C5).

To assess the variation within each viral population and to identify potential HPAI progenitor viruses, the reads obtained from each sample were aligned to both the low and highly pathogenic influenza reference sequences. In particular, LPAI samples possessing genetic signatures typical of highly pathogenic viruses described previously (table 5.2) and HPAI samples containing signature mutations of LPAI viruses were collated (table 5.5). We identified one LPAI sample (4829-99) collected in Verona province immediately after the first identification of the HPAI virus that contained a single read with an HA cleavage site associated with HPAI (table 5.5). This sample contained a complete insertion of multiple basic amino acids at the HA cleavage site as well as a minority of variants (0.58%) with the mutation A166V in the M1 gene that is also typical of the HPAI virus (table 5.5). Although at extremely low frequency, the occurrence of multiple HPAI-associated mutations suggests that it is unlikely to represent a sequencing artifact. However, none of the NGS sequenced samples collected from birds infected with the low pathogenic virus possessed together all the amino acid signatures typical of the HPAI virus, and there was no evidence for the stepwise evolution of HPAI mutations at the cleavage site. A number of mutations were consistently found above the normal Illumina error level, such as A166V found in six samples. Similar viral subpopulations were also present within some of the individuals. For example, samples 3675/99, 4295/99 and 3283/99 all have minority variants with mutations at site 745 of PB1, 349 of NP and 166 of M1, with a 10-fold increase of variants with K at site 745 in sample 4295/99 collected in November 1999 relative to 3283/99 collected in September 1999. The HPAI samples also contained a number of low-frequency mutations typical of the LPAI viruses. Insertion of the multiple basic amino acids in the HA cleavage site was detected in all the HPAI samples (table 5.5).

Table 5.5. Percentage of reads possessing mutations typical of the HPAI and LPAI viruses, respectively, in the LPAI and HPAI samples after removal of mutations with frequencies below Illumina artifacts (p-value<0.05).

Gene	Position	Amino acid		% of reads with the HPAI signatures in LPAI samples						% of reads with the LPAI signatures in HPAI samples							
		HPAI	LPAI	2732-99	3675-99	4295-99	3283-99	4829-99	1744-99	4618-99	4828-99	4827-99	4749-99	4708-99	4911-99	4756-99	4618-99
PB2	398	I	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PB1	154	D	G	0	0	0	0	0	n.a.*	0	0	0	0	0	0	0	0
PB1	216	G	S	0	0	0	0	0	n.a.*	0	0	0	0	0	0	0	0
PB1	745	E	K	0	0.34	2.8	0.23	0	0	0	1.12	0.68	0	0	0	0	0
PA	61	T	I	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PA	115	K	N	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PA	252	K	E	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HA	130	A	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HA	146	T	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HA	228	E	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HA	454	T	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HA	554	R	K	0	0	0	0	0	0	0	0	1.14	0	0	0	0.87	0
HA	339	Ins^	-	0	0	0	0	1 read	0	0	0	0	0	0	0	0	0
NP	349	A	T	0.29	0.37	0.39	0.25	0	0	0	0	0	0	0	0	0	0
NP	376	N	S	0	0.32	0	0	0	0	0.23	0.34	0.23	0.24	0.22	0.28	0.23	0.34
NA	173	K	R	0	0	0	0	0	0	0	0.48	0	0.76	0	0	0	0.48
M1	166	A	V	0.51	0.4	0.47	0.57	0.58	0.46	0	0	0	0	0	0	0	0
NS1	136	I	V	0	0.29	0	0	0	0.20	0	0	0.17	0	0	0	0	0
NS1	139	N	D	0	0	0	0	0	0.14	0	0	0	0	0	0	0	0
NS1	225	*	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0

^Abbreviation: *Ins*, Insertion; refers to the presence of a complete insertion of basic amino acids (-SRVR) at the HA cleavage site which is typical of HPAI viruses.

*Minimum read depth to call variants was not reached at these position (≤8)

5.5. DISCUSSION

The continual re-emergence of HPAI H7 avian influenza viruses presents a major concern for the poultry industry, as recently demonstrated by the 2012 H7N3 outbreak in Mexico and the 2013 Italian H7N7 outbreak (Maurer-Stroh et al., 2013; ProMED-mail, 2012; CDC, 2010; ProMED-mail, 2013). However, efforts to control these outbreaks are undermined by a poor understanding of the genesis and evolution of HPAI viruses. Here we employed newly developed deep-sequencing and Bayesian phylogenetic approaches to provide the most complete analysis of the evolution of high and low pathogenic viruses from a single epidemic undertaken to date. We showed that all the Italian LPAI and HPAI H7N1 isolates form a cluster that is distinct from all other Eurasian H7 isolates currently available (443 HA sequences), suggesting that the two epidemics resulted from a single viral introduction into Italy that occurred between December 1998 and March 1999. This inference is supported by the observation that the HPAI viruses have a common ancestor that dates from the period of circulation of the LPAI viruses (August-December 1999), and by the existence of minority genetic variants harbouring signature mutations of the HP viruses in the ultra-deep sequencing of the LP samples.

Although the original source of the Italian LPAI H7N1 outbreak has never been identified, some of the early isolates in our study lack the molecular signatures that frequently appear during the adaptation of viruses from aquatic bird hosts to domestic poultry (Gianecchini et al., 2010; Li et al., 2010), including the stalk deletion in the NA protein and the additional glycosylation sites in the HA protein. Hence, it is possible that wild birds have been the source of this virus. However, the availability of background AIV sequences was too sparse to identify any closely related viruses in either domestic or wild birds.

Critically, the HPAI viruses form a distinct group within the Italian H7N1 viral population on the phylogenies inferred for all eight segments. This HPAI lineage rapidly diverged from the LPAI viruses, with 19 amino acid changes observed in all eight segments that were acquired over a time period of only nine months. Despite their co-circulation in several Italian provinces, the phylogenies inferred for each segment show no evidence of reassortment between the HPAI and LPAI viruses, which continued to evolve as independent lineages. Among the low pathogenic AI populations, NGS technology identified minor variants with key mutations associated with the highly pathogenic viruses, including the insertion in the HA

cleavage site. Although we employed a conservative approach to exclude sequencing errors, excluding PCR errors is more problematic. However, that identical mutations are identified in multiple samples (table 5.5) reduces the likelihood that they are PCR errors.

These data confirm that the LPAI and HPAI epidemics shared a single common ancestor and suggest that some of the amino acid changes that characterize the HP virus cluster were already present with low frequency within several host viral populations from the beginning of the H7N1 epidemic, although the linkage between these amino acid sites could not be ascertained. Despite extensive sequencing, only complete insertions of the HA cleavage site were detected within the host viral populations of both the HP infected farms and a single LP sample collected early during the HPAI epidemic. Previous studies have suggested that the insertion can be acquired through several possible evolutionary mechanisms, including (a) the duplication of purine triplets due to a transcription fault in the polymerase complex (Horimoto et al., 1995; Garcia et al., 1996); (b) a stepwise accumulation of amino acid substitutions (Horimoto et al., 1995; Spackman et al., 2003); (c) non-homologous recombination resulting in the insertion of a foreign nucleotide sequence (Suarez et al., 2004; Pasik et al., 2005). The short length of the insertion (four amino acids) impedes efforts to distinguish between processes, as does the lack of deep sequencing information from the index case of the HPAI epidemic wave. However, it was notable that our deep sequencing data provided no evidence for the stepwise process of accumulation of mutations following polymerase complex errors. Hence, a disjunct evolutionary event, such as that caused by non-homologous recombination, is a more likely causative mechanism for the generation of HPAI viruses.

Experimental studies also are required to determine if any of the mutations that were observed across the HPAI genome had important compensatory effects or contributed to the virulence of these viruses. Of note, some of these substitutions have been previously demonstrated to modulate avian influenza virus pathogenicity. In particular, the NS1 protein has a deletion of six amino acids at the C terminus nucleolar localization signal (NoLS) (Baillie et al., 2012), and two mutations (D139N and V136I) are located within or closely adjacent to the nuclear export signal (NES) (Li et al., 1998). A recent study demonstrated that these mutations observed in the HPAI strain enhance cytoplasmic accumulation of NS1 and reduces IFN-b transcription, increasing viral pathogenicity (Skehel and Wiley, 2000). In addition, the C-terminal truncation promotes the transport of NS1 into the nucleoli, where it may favour viral transcription and translation or may alter the cell cycle to facilitate viral

replication (Keiner et al., 2010). Interestingly, some of the amino acid signatures identified in the Italian HPAI viruses have been previously described in other H7 HPAI strains, such as the mutation A146T in the HA protein, which has been detected in some HPAI H7N3 collected during a 2002 outbreak in Chile (Suarez et al., 2004); the truncation at the NS1 protein, identified in most of the HPAI H7N3 viruses from Pakistan (Abbas et al., 2010), and the additional glycosylation site at position 141-143 of the HA protein, found in both Pakistani (Aamir et al 2009; Lebarbenchon and Stallknecht, 2011) and Dutch (Bataille et al., 2011) isolates. The latter is noteworthy because it has also been suggested to be under positive selection (Bataille et al., 2011).

In September 1999, a few months before the identification of the HPAI variant, the LPAI-I lineage evolved from the LPAI population, forming a distinct phylogenetic group in six of the eight phylogenies that is characterized by 16 amino acid changes. The closest relative of the LPAI-I lineages appears to be the isolate A/chicken/Italy/3283/1999, which contains five of the 16 amino acid signatures of the LPAI-I cluster. Isolate A/chicken/Italy/3283/1999 was detected only a few days prior to the first LPAI-I virus (10 September 1999 and 27 September 1999, respectively), suggesting that these eleven mutations (PB2, L527M; PA, S190P, I257V; HA, Q86R, E113G, A143V, R148L; NA, V163M/L, M418I; NS1, G47C, P85T) were acquired over a short period of time. It is unclear whether this rapid accumulation of mutations reflects a strongly selectively driven event, as the virus responds to environmental changes, host switching and/or control measures, or was driven by a neutral population bottleneck and a subsequent founder effect. Although our Bayesian Skyline analysis does not show any major reduction in population size (data not shown), the available epidemiological data does not provide sufficient resolution to trace the cause of this rapid evolution.

It is also notable that the time of the most recent common ancestor obtained for the HPAI variant dates to August 1999 to December 1999, which is coincident with the time-scale of the emergence of the LPAI-I group. Whether the parallel emergence of these lineages, which clearly differ in virulence, is due to common evolutionary or ecological pressure clearly merits further investigation. Our phylogeographic analysis indicates that both lineages most likely emerged in Verona province, which has one of the highest densities of commercial poultry in Italy. Notably, these two viral lineages exhibited a degree of parallel evolution, including a glycosylation site in the HA protein and a truncation of the NS1 protein, suggesting that both viral populations may have been subject to similar selection pressures

despite their very different pathogenic effects. The presence of a potential additional glycosylation site in the HA gene and the deletion of the C-terminal of the NS1 protein observed in the HPAI group have been previously demonstrated to confer replication and selective advantages upon avian influenza viruses in the poultry hosts (Abbas et al., 2010; Keiner et al., 2010; Suarez and Perdue, 1998; Suwannakhon et al., 2008; Iqbal et al., 2012).

Our phylogeographic analysis also revealed striking spatial patterns of viral emergence and dissemination. The LPAI viruses were significantly more clustered by geographic region than the HPAI viruses. Although this could indicate a slower spatial diffusion of LPAI, particularly in the provinces of Verona, Venezia-Padova-Vicenza and Brescia, the less frequent spatial clustering may also be in part related to the clinical manifestation of disease. In particular, LPAI may manifest with limited or unnoticeable signs of disease and such that it might be not promptly identified. Moreover, in contrast to HPAI, LP strains were not subject to surveillance and control in 1999-2000 (CEC, 1992), and the notification of LP viruses was not mandatory at that time; therefore cases of LPAI may have been unreported in some instances. In contrast, HPAI viruses exhibited little structuring by province, reflecting their rapid spread among all the affected provinces, even those separated by large distances. Although the risk of infection has been found to be associated with proximity to an infected farm (Mannelli et al., 2006; Busani et al., 2009; Mulatti et al., 2010a), the movement of contaminated people, trucks, equipment and products can rapidly drive viral transmission between distant flocks during a HPAI outbreak (Mannelli et al., 2006; Dorigatti et al., 2010). Nevertheless, it is possible that we have missed the direct link between infected farms located in distant areas as several outbreaks were notified only on the basis of results of serological, epidemiological and clinical investigations (such that viruses and related genetic data were not available for these cases).

Although Verona province (located in the Veneto region) appears to be most important in the initial emergence of the HPAI and LPAI-I variants and the spread of LPAI viruses, the avian population in Mantova province (located in the Lombardia region) played the most important role in the spatial spread of the HPAI viruses to additional provinces. Indeed, limited dissemination of the virus toward other provinces was observed from Verona. These findings are in agreement with previous epidemiological studies (Mannelli et al., 2006; Mulatti et al., 2007; Mulatti et al., 2010a) and are possibly explained by differences among the provinces in the control strategies employed at the time. In Veneto, a ban of restocking

and pre-emptive culling were started 20 days earlier and on a greater number of farms than in Lombardia, reducing virus transmission within and outward the Veneto area. Hence, although the Mantova poultry population is markedly smaller than that of Verona, and is not as likely to be the geographic origin of the HPAI virus, Mantova has likely acted as a bridge between the poultry population of Verona and the large poultry production of Brescia province, thus becoming essential for the westward spread of both the LPAI and HPAI viruses. This is supported in our phylogeographic analyses, which indicate that viral migration occurred from provinces of relatively lower poultry population size to those with the largest poultry populations.

Finally, although the precise role of the set of mutations identified in the H7 HPAI viruses needs to be elucidated, the identification of minority variants with HP key mutations in the LP samples that became prevalent in the HP viral population highlights the importance of implementing prompt and effective eradication measures to prevent the appearance of viruses with fitness advantages and unpredictable pathogenic properties. Furthermore, this study demonstrated the utility of deep sequencing to detect viral subpopulations carrying biologically relevant mutations. In particular, awareness of the existence of viral variants associated with differences in virulence, antiviral resistance, or pandemic potential may boost the laboratory community's responsiveness to influenza epidemics. Monitoring of such variants may be especially useful in the surveillance of low pathogenic H7N9 avian influenza viruses in China, with the aim of predicting the emergence of variants resistant to antiviral drugs or with an increased capability of human-to-human transmission.

ACKNOWLEDGEMENTS

This work was financially supported by the European projects PREDEMICS (Grant Agreement No. 278433) and Epi-SEQ (research project supported under the 2nd joint call for transnational research projects by EMIDA ERA-NET [FP7 project nr 219235]). Alice Fusaro acknowledges the receipt of a fellowship from the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems in 2013.

The authors would like to acknowledge Richard Orton (IBAHCM, University of Glasgow, Glasgow, United Kingdom) for his support with statistical analysis of data, Philippe Lemey

(Rega Institute, Leuven, Belgium) for his useful suggestions in the application of phylogeographic analysis, and Alessia Schivo for her technical support.

Chapter 6

Parallel evolution of two distinct avian influenza epidemics

6.1. ABSTRACT

Avian influenza (AI) viruses of the H7 subtype continue to be an important concern both for the poultry industry and for global public health issues, as the recent H7N3 outbreak in Mexico and the H7N9 in China have demonstrated. Between 1999 and 2004 two avian influenza (AI) epidemics caused by different subtypes (H7N1, 1999-2001 and H7N3, 2002-2004) affected the poultry population in the northern part of Italy. Notably, while the low pathogenic (LP) H7N1 virus became highly pathogenic (HP) during the epidemic, the H7N3 did not. To better understand the mechanisms behind the evolution of these subtypes, we have focused our investigation on the evolutionary dynamics, intra-host variability and on patterns of subpopulation transmission and fixation of viruses collected during these two epidemics.

Analysis of the evolutionary dynamics revealed several similarities between the H7N1 and H7N3 strains, including i) 28 identical amino acid mutations acquired independently by both subtypes in the course of the two epidemics, ii) analogous rates of nucleotide substitutions, iii) similar pattern of variability in the amino acid diversity of the HA proteins and iv) a parallel emergence of genetic groups with similar genetic properties.

Deep sequencing analysis of 30 clinical samples identified a number of low frequency variants transmitted from outbreak to outbreak with occasional fixation in the viral population. In addition, the different intra-host genetic variability among the analysed

samples suggests that bottleneck events and the subsequent founder effect may have contributed in shaping the virus evolution during these two natural occurring epidemics.

These results provide novel insights into the evolution of H7 avian influenza viruses, which may help the international veterinary community to combat ongoing and future avian influenza epidemics.

6.2. INTRODUCTION

Avian H7 influenza viruses circulating in wild birds occasionally infect poultry species. Once introduced into domestic poultry, this subtype can acquire genetic properties able to confer high virulence, thus causing a serious impact on the poultry industry and on international trade. Outbreaks caused by low pathogenic (LP) and highly pathogenic (HP) avian influenza (AI) of the H7 subtype have been reported in poultry throughout the worlds: Europe with the HP H7N7 epidemic in the Netherlands (Bataille et al., 2011) and the LP/HP H7N1 and LP H7N3 epidemics in Italy (Capua and Marangon 2007); the Americas with the HP H7N3 in British Columbia (Pasick et al., 2005) and Chile (Suarez et al., 2004) and the recent HP H7N3 epidemic in Mexico (Maurer et al., 2013) and Asia with the ongoing LP H7N9 epidemic in China (Chen et al., 2013).

H7 avian influenza subtype, together with the H5 and H9, top the World Health Organization's list for the greatest pandemic potential, posing a continue threat to public health (Horimoto and Kawaoka, 2001). Since 1996, human infections with H7 influenza viruses have been reported in Canada, Italy, Mexico, the Netherlands, the United Kingdom, the United States of America, and more recently in China and Taiwan (Peiris, 2009; Chen et al., 2013). Before the occurrence of the H7N9 outbreaks, human infections mainly resulted in conjunctivitis and mild respiratory symptoms. However, since February 2013 H7N9 viruses have been responsible for 217 cases and 57 deaths in China and Taiwan (ProMed-mail, 2014).

Due to their high mutation rates, rapid replication kinetics and large population size, influenza viruses exist in the host as a diverse collection of genetically linked variants, termed quasispecies. The rate at which within-host genetic diversity is generated and maintained upon transmission and the degree of fixation of advantageous mutations in the viral population are parameters determining the virus survival in new environments. However, to

date the dynamics governing the intra- and inter-host evolution of avian influenza epidemics have been only partially elucidated.

Between 1999 and 2004 the densely populated poultry areas in northern Italy (Lombardia and Veneto regions) experienced two distinct epidemics of avian influenza of the H7 subtype. The first epidemic (1999-2001) was caused by a LPAI H7N1 strain, which, after having circulated in the industrial poultry population for approximately nine months (from the end of March to December 1999) causing 199 outbreaks, mutated into a HP form. The HPAI epidemic provoked the death of over 16 million poultry and substantial economic losses to industry before its eradication, officially in April 2000. Four months later, the LPAI H7N1 re-emerged, affecting 78 flocks. To reduce the economic impact of this second wave of LPAI viruses, a DIVA vaccination campaign was initiated in November 2000 (Capua and Marangon, 2001; Mulatti et al., 2010).

The second epidemic, caused by an H7N3 LPAI strain introduced from the wild bird reservoir into the domestic poultry, started in October 2002 (Campitelli et al., 2004). To contain the rapid spread of the infection, from January 2003 a DIVA vaccination program with an AI eterologous vaccine (a 1999 Italian H7N1 strain) was carried out in layers, capons and meat turkeys. The virus managed to circulate for one year (until October 2003) and to infect a total of 388 poultry holdings. In September 2004, one year after the last case, the LPAI H7N3 subtype re-emerged causing 28 new outbreaks, 27 of which in vaccinated meat turkey flocks (Capua and Marangon, 2007).

Using newly developed deep-sequencing and Bayesian phylogenetic approaches, the previous study (chapter 5) compared the evolutionary dynamics of the H7N1 HPAI viruses with those of low pathogenicity collected during the 1999-2001 epidemic and identified, in the LPAI samples, viral subpopulations which harboured known molecular markers for viral pathogenicity, thus providing evidence that HPAI viruses originated from their LPAI progenitors.

Starting from the results obtained in chapter 5, we compared the patterns of the evolutionary dynamics between the 1999-2001 H7N1 and the 2002-2004 H7N3 strains. These two epidemics, which had affected domestic species (mainly turkeys and chickens) reared in the same geographic area for similar periods of time, provide a formidable dataset to explore the role played by the environment in shaping the evolutionary dynamics of avian influenza viruses during natural occurring epidemics. In addition, using a deep sequencing approach we

analysed the intra-host genetic diversity of the HA gene segment and the inter-host transmission of viral variants among viruses with different pathogenic properties or belonging to different genetic lineages, providing a better understanding of the pattern of minority variant transmission during two distinct epidemics circulating in similar epidemiological conditions but with completely different virulence outcomes.

6.3. MATERIAL AND METHODS

6.3.1. Viruses included in this study

In this study, we generated the complete genome sequences of 35 H7N3 avian influenza A viruses collected from poultry in northern Italy from October 2002 to December 2004. In addition, we sequenced the partial genomes of five samples from which whole genome sequences could not be obtained. Sequences of 37 H7N3 viruses from the 2002-2004 epidemic publicly available in the Influenza Virus Resource at GenBank were also included in the analysis. These data were compared to the HA sequences of 144 samples and to the complete genome of 109 isolates collected during the 1999-2001 LP/HP H7N1 epidemic, sequenced and analysed in my previous study (chapter 5). Epidemiological information (collection date and province of collection) for all the H7N3 viruses included in this study is available in Annex D – Table D1.

6.3.2. Sanger sequencing

Viral RNA was extracted from the infected allantoic fluid of specific-pathogen-free fowls' eggs using the Nucleospin RNA II kit (Macherey-Nagel, Duren, Germany) and reverse transcribed with the SuperScript III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA). PCR amplifications were performed by using specific primers (sequences are available on request). Amplicons were subsequently purified with ExoSAP-IT (USB Corporation, Cleveland, OH) and sequenced using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). The products of the sequencing reactions were cleaned-up using the PERFORMA DTR Ultra 96-Well kit (Edge BioSystems, Gaithersburg, MD) and analysed on a 16-capillary ABI PRISM 3130xl genetic analyzer (Applied Biosystems, Foster City, CA).

6.3.3. Phylogenetic analysis

Sequences of the eight gene segments of the H7N1 and H7N3 viruses were aligned and compared with the most related sequences available in GenBank.

Nucleotide sequence alignments were manually constructed for each gene segment using the Se-Al program (Rambaut, 2002). The likelihood mapping analysis available in the PUZZLE TREE program (Schmidt et al., 2002) was adopted to visualize the phylogenetic content of the eight datasets. In particular, we investigated the phylogenetic signals for a) all codon positions, b) I codon position, c) II codon position, d) III codon position, e) I+II codon positions of the alignments. Since the results obtained for the three codon positions showed the highest percentage of resolved quartets, the following analyses were performed using all the alignment positions.

Maximum likelihood (ML) phylogenetic trees were constructed for all the eight gene segments using the best-fit general time-reversible (GTR) model of nucleotide substitution with gamma-distributed rate variation among sites (with four rate categories, Γ4) and a heuristic SPR branch-swapping search (Guindon and Gascuel, 2003) available in the PhyML program version 3.0. To assess the robustness of individual nodes of the phylogeny, one hundred bootstrap replicates were performed. Topologies of phylogenetic trees were confirmed using the ML method available in the RAXML program (Stamatakis, 2006a) (data not shown), incorporating GTR model of nucleotide substitution with CAT model of rate heterogeneity among sites (data not shown) (Stamatakis, 2006b). Phylogenetic trees were visualized with the program FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). Fixed amino acid changes along all the branches of the HA phylogeny were identified using the program Mesquite (Maddison and Maddison, 2011).

Within group mean p-distance was calculated using the program MEGA 5 (Tamura et al., 2011)

6.3.4. Analysis of selection pressures

Gene- and site-specific selection pressures for all segments of the Italian H7N1 and H7N3 viruses were measured as the ratio of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitutions per site. In all cases, d_N/d_S ratios were estimated using the fixed-effects likelihood (FEL) and Fast Unconstrained Bayesian AppRoximation (FUBAR) methods (Pond and Frost, 2005; Murrel et al., 2013) available at the Datamonkey online version of the Hy-

Phy package (Delport et al., 2010). All analyses utilized the GTR model of nucleotide substitution and employed input NJ phylogenetic trees.

6.3.5. Evolutionary dynamics

For each gene segment of the H7N3 viruses, rates of nucleotide substitution per site per year (subs/site/year) of the sampled data were estimated using the Bayesian Markov chain Monte Carlo (MCMC) approach available in the BEAST program, version 1.7.5 (Drummond and Rambaut, 2007). For each analysis, we employed a relaxed (uncorrelated lognormal) molecular clock, a flexible Bayesian skyline coalescent tree prior (10 piece-wise constant groups), a HKY85 + Γ_4 model of nucleotide substitution and the SRD06 codon position model with two data partitions of codon positions (1^{st} + 2^{nd} positions, 3^{rd} position) with base frequencies unlinked across all codon positions. In all cases, statistical uncertainty is reflected in values of the 95% highest probability density (HPD) for each parameter estimate. For each analysis chain lengths were run for sufficient time to achieve convergence as assessed using Tracer v1.5 program (Drummond and Rambaut, 2007).

Maximum clade credibility (MCC) tree of the HA gene was summarized from the posterior distribution of trees using the program TreeAnnotator v1.7.5 (Drummond and Rambaut, 2007) after the removal of 10% burnin. The MCC trees were visualized using the program FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

6.3.6. Library preparation, Illumina sequencing and data analysis

To assess virus population diversity, next-generation sequencing (NGS) was performed on seventeen clinical samples: eight LPAI H7N1 collected from September 1999 to February 2001 and nine H7N3 detected between October 2002 and January 2003. These newly generated sequences were analysed together with the NGS data generated in the previous chapter, for a total of 30 samples. Full sample details are described in Table 6.1. Unfortunately, no clinical samples of H7N3 viruses of the 2002-2004 epidemic collected after January 2003 were available.

Viral RNA was extracted directly from the infected clinical samples, i.e. trachea, lung, intestine, pool of organs or tracheal swabs, using the Nucleospin RNA II kit (Macherey-Nagel, Duren, Germany). Following the protocol by Zhou et al. (2009), amplification of the complete influenza A genomes was performed using the SuperScript III One-Step RT-PCR

system with Platinum®Taq High Fidelity (Invitrogen, Carlsbad, CA) and two primers complementary to the 3'-end and 5'-end conserved regions of each segment, allowing the simultaneous amplification of the eight genes. High-throughput sequencing was performed on an Illumina MiSeq at IGA Technology Services srl (UD), Italy, as detailed in chapter 5. Briefly, the Nextera DNA Sample Preparation Kit from Illumina was used to fragment and tag with sequencing adapters and indexes the amplified genomes in a single step. The molar concentration of the libraries was determined with the QBit fluorimeter (Invitrogen). Finally the indexed libraries were pooled in equimolar concentrations and sequenced in multiplex for 250bp paired-end on Illumina MiSeq, according to the manufacturer's instructions.

Raw sequence reads were inspected using FASTQC to assess the quality of data coming from the high throughput sequencing pipelines. Fastq files were cleaned with PRINSEQ and Trim Galore to remove low quality bases at the 5' and 3'-end of each read and to exclude reads with a Phred quality score below 30 and shorter than 80 bp. Reads of the H7N1 and H7N3 samples were aligned to A/Italy/3675/99 (H7N1) and A/turkey/Italy/8535/2002 (H7N3) reference sequences using Stampy (Lunter and Goodson, 2011). The number of assembled reads ranged from 194355 for the sample 5302 (H7N1) to 868287 for 7773 (H7N3) (table 6.1). The BAM alignment files were parsed using a custom perl script to identify the frequency of polymorphisms at each site relative to the reference used for the alignment. A binomial test, using the error probability calculated from the base qualities at a site and following the approach of Morelli et al. (2013), was used to determine whether the frequency of each mutation was above the artifactual value expected from Illumina sequencing. All the low-frequency polymorphisms likely to be miscalled bases during sequencing were removed from downstream analyses. Viral population variability (the amount of "disorder" in the population) was assessed by calculating Shannon entropy, as described by Morelli et al. (2013), for every site of the genome reaching a coverage of at least 500 reads. This coverage allows to include 60% of the total sites. Statistical analysis was carried out with GraphPad Prism v 6 software. Differences were tested for statistical significance using one way analysis of variance (ANOVA) followed by Tukey's multiple comparison test.

Table 6.1. Identification number, epidemiological data and total number of sequences obtained for each sample sequenced using the Illumina platform. T=trachea, L=lung; I=intestine

Sample	Collection date	Type of sample	Province	Specie	Subtype, pathotype	Total sequences^	Mapped sequences	HA mean coverage
4618/99*	14/12/99	T	VR	Turkey	H7N1, HP	423574	418238	2188,4
4749/99*	21/12/99	I	VR	Turkey	H7N1, HP	544596	541929	1184,5
4708/99*	21/12/99	T	VR	Turkey	H7N1, HP	501086	496408	381,5
4756/99*	22/12/99	I	na§	Chicken	H7N1, HP	463704	413277	1566,8
4827/99*	23/12/99	L,T	VR	Turkey	H7N1, HP	431304	431085	367,6
4828/99*	23/12/99	L,T	VR	Turkey	H7N1, HP	460708	460491	649,2
4911/99*	27/12/99	I	RO	Chicken	H7N1, HP	386926	378252	433,7
1744/99*	26/04/99	L,T	VR	Turkey	H7N1, LP	195496	194543	184,3
2732/99*	03/08/99	L,T	VR	Turkey	H7N1, LP	533706	533293	1245,7
3283/99*	10/09/99	L,T	VR	Turkey	H7N1, LP	534136	529407	862,5
3560/99	27/09/99	na	VR	Turkey	H7N1, LP	381841	275489	462,9
3675/99*	04/10/99	T	VR	Turkey	H7N1, LP	591410	591173	1515,6
4295/99*	22/11/99	L	VR	Turkey	H7N1, LP	520056	516798	827,2
4829/99*	23/12/99	L,T	VR	Turkey	H7N1, LP	542062	541952	1674,2
4981/00	14/09/00	T	VR	Turkey	H7N1, LP	311307	302359	1563,2
5222/00	28/09/00	na	VR	Turkey	H7N1, LP	614881	600952	797,6
5302/00	02/10/00	na	VR	Turkey	H7N1, LP	200835	194355	619,4
7001/00	28/12/00	L,T	PD	Turkey	H7N1, LP	262028	255091	4408,3
133/01	10/01/01	I	PD	Turkey	H7N1, LP	570972	511782	1778,8
138-1/01	10/01/01	L,T	PD	Turkey	H7N1, LP	564130	501098	1413,1
1351-1/01	15/02/01	L,T	PD	Turkey	H7N1, LP	539351	507008	8940,7
7773/02	31/10/02	T	VR	Turkey	H7N3, LP	882236	868287	4854,7
8093/02	12/11/02	T	VR	Chicken	H7N3, LP	531167	519352	5019,8
8303/02	13/11/02	T	VI	Turkey	H7N3, LP	586850	523301	228,4
9289/02	06/12/02	T	VR	Turkey	H7N3, LP	341836	320556	51483,2
9369/02	11/12/02	T	VR	Turkey	H7N3, LP	467093	450876	1373,4
9565/02	19/12/02	T	VR	Guinea Fowl	H7N3, LP	479903	466527	1428,4
9611/02	20/12/02	T	PD	Turkey	H7N3, LP	572001	543221	2608,8
4/03	03/01/03	T	VR	Turkey	H7N3, LP	754330	711224	4973,8
17/03	07/01/03	pool	PD	Turkey	H7N3, LP	537736	519973	1515,6

* data obtained in chapeter 5

^ After removal of adapter and low quality sequences

§ na: not available

6.4. RESULTS

6.4.1. Parallel evolution of the H7N1 and H7N3 viruses

Our maximum likelihood (ML) phylogenetic analyses of the complete genome of 181 (109 H7N1 and 72 H7N3) H7 viruses representative of the 1999-2001 and 2002-2004 Italian epidemics indicate that the two subtypes fall within two well-supported monophyletic clades, defined by both high bootstrap values (>80%) and long branch length, as exemplified by the HA phylogeny (figure 6.1 - H7N1 clade in blue and H7N3 clade in green). This finding indicates that these clades (H7N1 and H7N3) are likely to represent two separate introductions of the virus into the northern Italian regions. As observed in the previous chapter, the H7N1 HPAI viruses (marked in yellow in figure 6.1) form a separate cluster within the H7N1 clade in all the eight phylogenies and are characterized by 19 unique amino acid changes acquired across the entire genome.

The phylogenetic tree inferred for the HA gene for a total of 144 H7N1 and 77 H7N3 viruses identifies in both clades two main groups of LPAI viruses, namely H7N1 LPAI-I and H7N3 LPAI-I (marked in grey in figure 6.1). Both clusters had emerged seven months after the circulation of their respective progenitor viruses and had persisted until the end of the epidemic. The H7N1 LPAI-I group is characterized by a total of 16 amino acid signatures (chapter 5) and is defined by high bootstrap values (>70%) and long branches in the HA, NA, PB2, PA, NS phylogenies, while the H7N3 LPAI-I viruses form a distinct cluster with high bootstrap supports (>87%) only in the HA and PA phylogenies and showed only three amino acid signatures in the HA gene (A169T, Q219H and G195V).

Interestingly, these LPAI-I groups are characterized by several common features: a) both contain viruses collected seven months after the first outbreak, b) all the isolates of these two clusters contain an additional glycosylation site at position 167-169 of the HA gene, c) both emerged and circulated mainly in the Verona province.

The parallel evolution of the H7N1 and H7N3 epidemics can be observed in many other aspects, which go beyond the emergence of the LPAI-I groups. We identified a total of 28 identical amino acid substitutions across the viral genomes, which were independently acquired by both subtypes during their evolution. Most of these mutations (12 out 28) are positioned within the HA protein (mutations highlighted in yellow in figure 6.1; table 6.2) and, in particular, three are located within antigenic sites, six are positively selected sites and

two create new additional glycosylation sites.

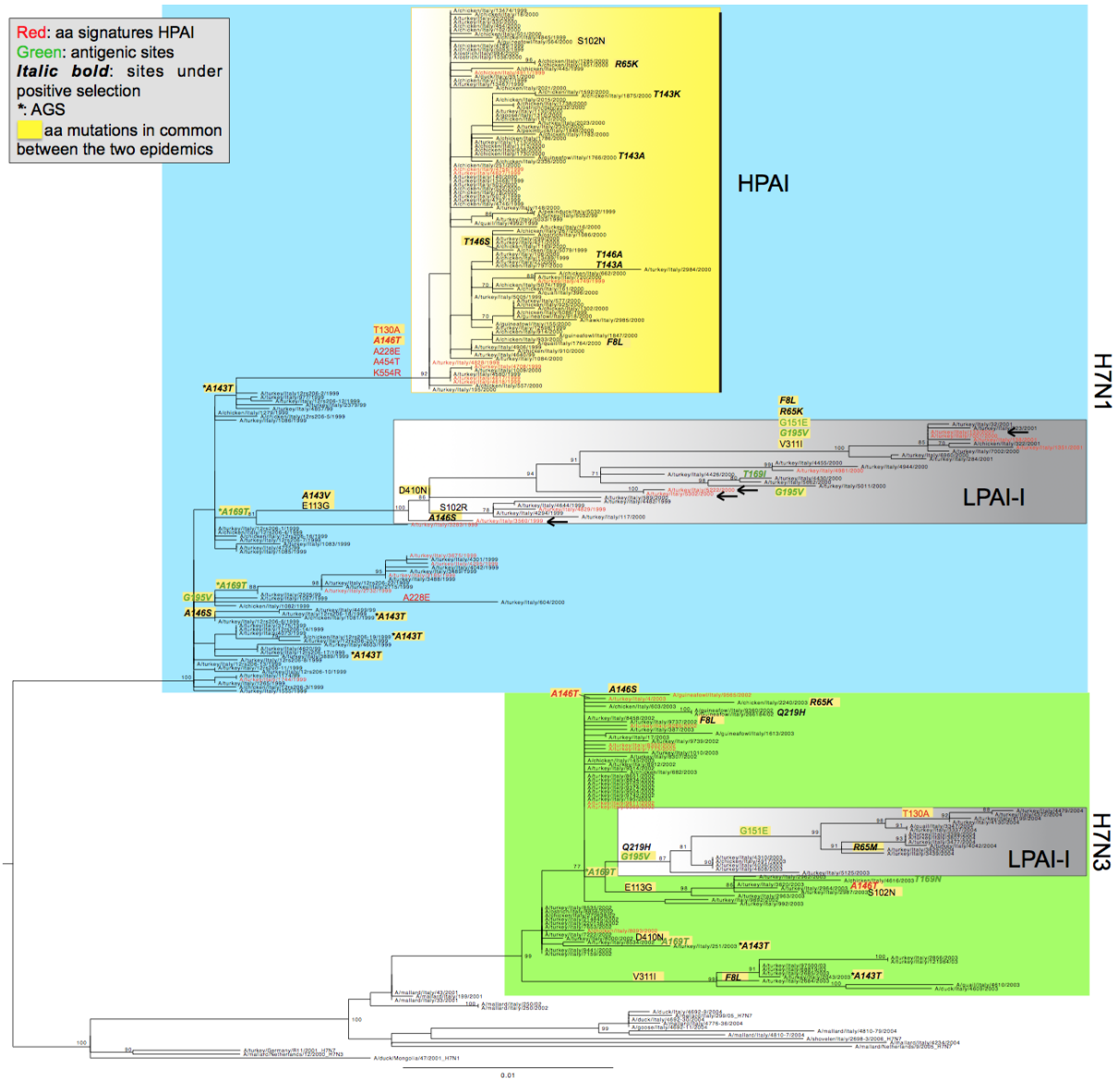


Figure 6.1. ML phylogenetic tree of the HA gene segment of Italian H7 avian influenza viruses analysed in this study. Viruses are coloured as follows: blue for H7N1 viruses collected during the 1999-2001 Italian epidemic, green for the H7N3 viruses collected during the 2002-2004 Italian epidemic. The HP H7N1 viruses are marked in yellow, while grey represents the two LPAI-I groups identified during the H7N1 and H7N3 epidemics. The numbers at nodes represent bootstrap values (>70%), while branch lengths are scaled according to the numbers of nucleotide substitutions per site. Sequences obtained using the NGS platform are marked in red. The tree is mid-point rooted for clarity only

Table 6.2. List of positions of amino acid mutations of the HA protein a) acquired independently by H7N1 and H7N3 viruses (parallel evolution); b) located in sites under positive selection; c) characteristic of the HPAI H7N1 viruses; d) positioned in antigenic sites; e) which generate new potential additional glycosylation site (AGS)

Site (H7 numbering)[§]	Parallel evolution	Positive selection	Typical of HPAI	Antigenic site	AGS
8	√	√			
65	√	√			
102	√				
113	√				
130	√		√		
143	√	√			√
146	√	√	√		
151	√			√*	
169	√	√		√^	√
195	√	√		√[$]	
219		√			
228			√		
311	√				
410	√				
454			√		
554			√		
TOT	**12**	**7**	**5**	**3**	**2**

§ numbering starting from the ATG of the HA gene

* Stevens et al., 2006; Bush et al., 1999 (position 142 in H3 numbering)

^ Kaverin et al., 2007; Li et al., 2009 (position 160 in H3 numbering)

$ Bush et al., 1999 (position 186 in H3 numbering)

The other 16 parallel substitutions are distributed across the internal proteins. In particular, three are situated in the PA (G213S, A212E, R221Q), five in the PB2 (R62K, D195N, R508K, M570I and D740N), four in the PB1 (A14V, T39A, I667V, K586R), two in the NS1 (M/I79V, S228P) and two in the M1 protein (R134K, G228R).

Most of these parallel mutations (23/28) occurred along the major branches of the phylogenies, in at least two different clustered sequences. Of note are the A169T, G195V, G151E substitutions, which were acquired during the evolution of the HA gene of both H7N1 LPA-I and H7N3 LPAI-I groups (figure 6.1). Differently, 5 parallel mutations - S102N and A146S in the HA gene (figure 6.1), D740N in the PB2, G213S in the PA and M/I79V in the NS1 - are peculiar of single isolates and were not transmitted, at least at the consensus level, to other viruses included in this study. In addition, 9 of the 28 parallel mutations (F8L, R65K, A143T, A146T, A169T, G195V in the HA, R221Q in the PA and D195N, D740N in the PB2)

were fixed along multiple branches of the same phylogeny, suggesting that these positions were subjected to a positive selection. This has been confirmed also by our analysis of selection pressure on the entire genome of all the H7N1 and H7N3 viruses. Using the FUBAR and FEL methods, we respectively identified in the HA gene six (posterior probabilities >0.9) and four (p-value<0.05) sites with evidence of putative positive selection (table 6.3). As expected, five of these sites (positions 65, 143, 169 and 195) coincide with the positions of parallel mutations detected in multiple branches of the HA phylogeny. Interestingly, two of them (169 and 195) are located in the H3-corresponding B antigenic site (Kaverin et al., 2007; Li et al., 2009; Bush et al., 1999). Moreover, mutation at position 169 introduces a potential additional glycosylation site at positions 167-169, which may have a strong antigenic effect.

Table 6.3. Amino acid sites under putative positive selection detected using different models (FUBAR and FEL) and mean dN/dS ratio for each gene.

Gene	**dN/dS**	**FUBAR**		**FEL**	
		PSS	**Posterior Prob**	**PSS**	**p-value**
HA	0,34	65	0,989	65	0,02
		143	0,997	143	0,007
		146	0,964		
		169	0,975	169	0,035
		195	0,938		
		219	0,977	219	0,047
PB2	0,135	/	/	/	/
PB1	0,098	/	/	/	/
PA	0,14	/	/	/	/
NP	0,07	/	/	/	/
M1	0,223	/	/	/	/
M2	0,738	/	/	/	/
NS1	0,243	/	/	/	/
NS2	0,292	/	/	/	/
PB1-F2	4,73	23	0,966		
		27	0,969		
		30	0,979		
		37	0,931		
		54	0,973		
		66	0,972		
		69	0,905		
		82	0,994	82	0,033
		89	0,933		

Analyses of selection pressure on the remaining gene segments did not identify further positively selected sites, except in the PB1-F2 protein, the only one showing evidence of diversifying selection (dN/dS = 4,73), although this is likely to be an artefact due to its encoding in an overlapping reading frame (Holmes et al., 2006).

To further explore the pattern of amino acid sequence variability of the HA protein during the H7N1 and H7N3 epidemics and the possible effect of bottleneck events, we grouped the viruses according to the month of collection and calculated the p-distances within each group containing at least three sequences (figure 6.2).

Surprisingly, the LPAI viruses belonging to the two subtypes show a similar pattern of genetic variability, with a rapid increase in the genetic diversity during the first ten months of each epidemic (corresponding to the first epidemic wave for both the LPAI H7N1 and H7N3), followed by a drastic reduction of the amino acid variability, that reached the lowest value in both epidemics about one year after the last outbreak of the first wave (March 1999- December 1999 for the H7N1, October 2002- October 2003 for the H7N3). For the H7N3 epidemic, this date corresponds to the re-occurrence, after almost a year with no reported cases, of the H7N3 strain in September 2004, when a single genetic group (H7N3 LPAI-I) was circulating, thus explaining this low variability. Similarly, all the LPAI H7N1 viruses that re-emerged in August 2000, approximately eight months after the last H7N1 LPAI outbreak, belong to the H7N1-LPAI-I group. However, viruses collected in August and September 2000 showed an intermediate variability, which progressively decreased reaching the lowest value in December 2000, twelve months after the last LPAI outbreak of the first wave (figure 6.2). This finding suggests that during both epidemics, bottleneck events, responsible for the selection and persistence of the LPAI-I lineages, may have occurred.

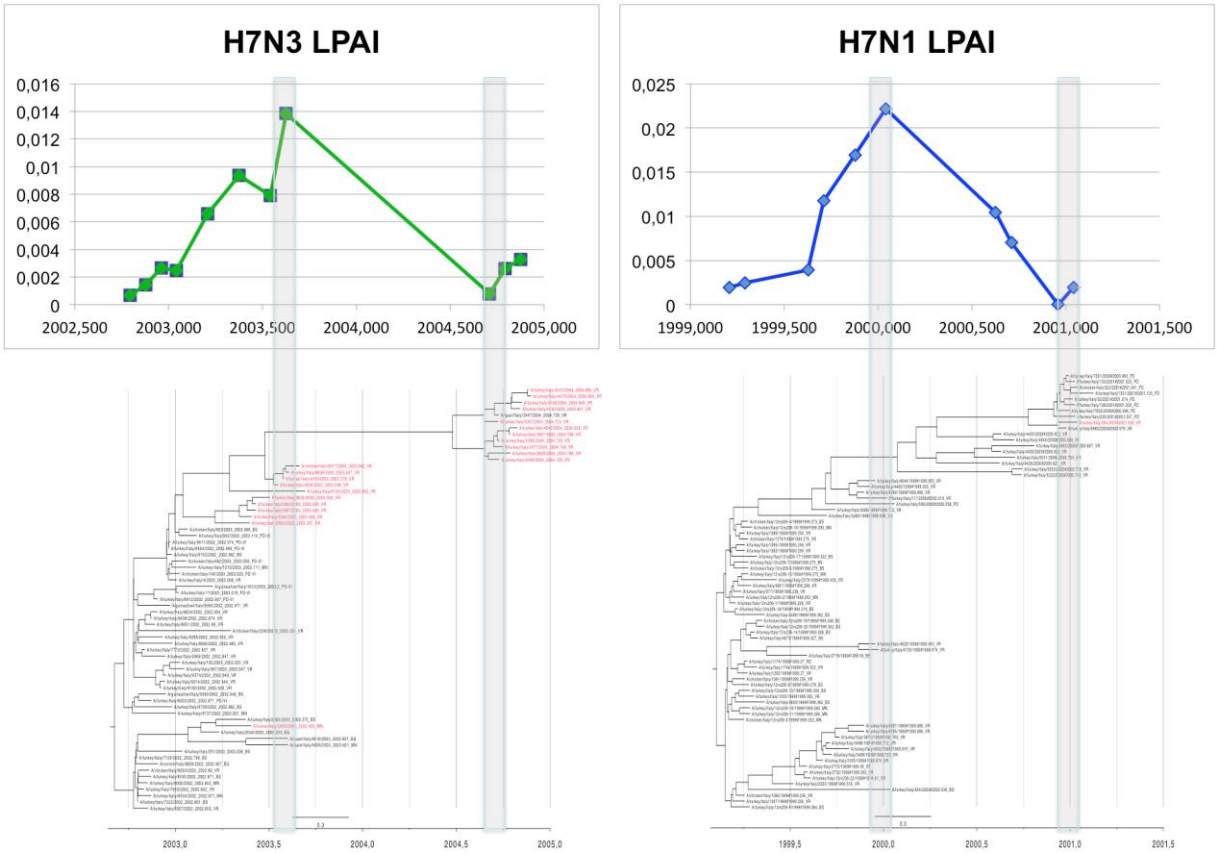


Figure 6.2. p-distance among the amino acid sequences of the HA gene of the H7N3 (left) and H7N1 (right) viruses collected during the same month. Maximum clade credibility trees estimated using a Bayesian coalescent approach of the H7N3 (left) and the H7N1 (right) viruses. Viruses collected from vaccinated birds are marked in red.

To further confirm the similar evolutionary dynamics of the H7 viruses collected during the 1999-2001 and 2002-2004 epidemics, we also compared the evolutionary rates of the eight gene segments of the H7N1 (calculated in chapter 5) and H7N3 (estimated in this study) viruses, both inferred using a Bayesian coalescent approach (Drummond and Rambaut, 2007). Despite the different vaccination strategies implemented during the two epidemics and the distinct pathogenic properties acquired by the two strains, the rate of nucleotide substitutions appears to be very similar between the two subtypes for most of the eight gene segments (7 out 8). Both viruses show high nucleotide substitution rate for the HA glycoprotein: mean rate of 10.15 substitutions per site per year (sub/site/year) (95% HPD, 8.5 – 11.9 x 10^{-3}) for the H7N1 viruses, and 8.46 x 10^{-3} sub/site/year (95% HPD, 6.46 – 10.65 x 10^{-3}) for the H7N3 viruses (table 6.4). While, different evolutionary rates have been observed for the NA genes

belonging to the two different subtypes (N1 and N3), with the N1 subtype evolving faster (mean rate of 9.86 sub/site/year, 95% HPD, 6.46 – 10.65 x 10^{-3} sub/site/year) than the N3 gene (mean rate of 3.78 sub/site/year, 95% HPD, 2.56 – 5.23 x 10^{-3} sub/site/year) (table 6.4).

Table 6.4. Estimated rates of nucleotide substitution for the H7N1 and H7N3 viruses collected during the two Italian epidemics.

Gene	Genetic group	Evolutionary rates (sub/site/y)		Comments
		Mean (x10^{-3})	95% HPD (x10^{-3})	
HA	H7N1	10.15	8.5-11.9	Chapter 5
	H7N3	8.46	6.46-10.65	This study
NA	H7N1	9.86	8.01-11.79	Chapter 5
	H7N3	3.78	2.56-5.23	This study
PA	H7N1	5.84	4.79-6.91	Chapter 5
	H7N3	4.65	3.17-6.19	This study
M	H7N1	7.21	5.02-9.51	Chapter 5
	H7N3	4.61	2.94-6.4	This study
PB1	H7N1	5.57	4.54-6.68	Chapter 5
	H7N3	4.56	3.24-5.99	This study
PB2	H7N1	6.81	5.54-8.04	Chapter 5
	H7N3	5.08	2.78-7.76	This study
NP	H7N1	5.42	3.81-7.15	Chapter 5
	H7N3	5.65	3.71-7.62	This study
NS	H7N1	7.72	5.76-9.91	Chapter 5
	H7N3	4.59	3.01-6.35	This study

6.4.2. Subpopulation transmission and bottlenecks

At the consensus level, we have demonstrated that H7N1 and H7N3 viruses accumulated several amino acid changes during the course of the two epidemics. To better characterize the dynamics of transmission and fixation of amino acid substitutions in the viral populations, we used a deep sequencing approach, which allowed to identify minority variants at each genomic site.

Complete genomes obtained for 8 H7N1 and 9 H7N3 clinical samples, were analysed together with the NGS data obtained in the previous study (chapter 5) for 7 HPAI and 6 LPAI H7N1 viruses. Here we focused on the HA, the gene that at the consensus level showed the highest amino acid variability and the highest number of mutations acquired during the course of the two epidemics. The mean depth of coverage for this gene ranges from 184 for the sample 1744/99 (H7N1) to 51483 for the sample 9289/02 (H7N3) (table 6.1).

To identify nucleotide variants possibly transmitted from farm to farm and occasionally fixed, which resulted in a consensus level polymorphism (mutation that appeared in more

than 50% of the reads), we divided our samples into four groups: H7N1 HP (7 samples), H7N1 LP-A (viruses collected before the HPAI outbreak, 7 samples), H7N1 LP-B (viruses collected after the HPAI outbreak, 7 samples) and H7N3 (9 samples) group. In order to reduce the likelihood to include amplification or sequencing errors in our analyses, we picked out i) all the minority variants above the Illumina error level that were consistently found in all the analysed samples of each group and ii) all the minority variants above the Illumina error level which fixed in the viral population becoming the prevalent population in at least one sample within the group.

We found a total of 109 low frequency nucleotide substitutions that might have been transmitted from farm to farm, of which 96 become fixed in the viral population (table 6.5). The H7N1 LP groups (H7N1 LP-A and B) are the ones showing the highest number of these minority variants (35 in the H7N1 LP-A and 52 in the H7N1 LP-B), while both the H7N1 HP and the H7N3 groups show only 11 host to host transmitted mutations.

Interestingly, 32 out of 109 are non-synonymous mutations: 8 in the H7N1 HP group (73% of the nucleotide transmitted variants), 13 (37%) in the H7N1 LP-A group, 8 (15%) in the H7N1 LP-B group and 3 (27%) in the H7N3 group (figure 6.3). Some of these mutations appear to have progressively increased their frequency during the viral evolution to finally fix in the viral population. The case of the mutation A454T identified in the group H7N1 LP-B is undoubtedly exemplificative. In the samples collected in September 2000 (4981/00, 5222/00) only 0.1-0.2% of the variants showed this amino acid change. From October 2000 to January 2001 the frequency of the amino acid T progressively increased from 7% (October) to 8.4% (December 2000) and 13.1% (January 2001) and was then fixed (99.1%) in the sample (1351/01) collected in February 2001 (figure 6.3). Differently, other consensus mutations seem to disappeared, i.e. T77A, D127N and A146S in the H7N1 LP-B group, although they continue to be present at a low frequency in some samples. Finally, other amino acid changes appear to have been constantly transmitted from outbreak to outbreak, without becoming fixed in the viral population, as in the case of V34E in the H7N1 HP group or of T296A, detected in all the H7N1 LP-A and H7N1 HP samples (figure 6.3, table 6.4). Most (7/8) of the transmitted variants identified in the H7N1 HP group belong to this last category, probably because of the short time-span collection of the samples of this group. Surprisingly, one of these low frequency substitutions is a nonsense mutation (Q304*), suggesting that viral subpopulations codifying for an HA protein of only 303 amino acids can be efficiently

transmitted from host to host.

Table 6.5. List of variants transmitted from farm to farm in the four groups (H7N1 HP; H7N1 LP-A; H7N1 LP-B; H7N3). Transmitted variants – nucleotide (nt) or amino acids (aa) mutations observed as minority variants in all the samples of the group. Fixed variant - nucleotide (nt) or amino acids (aa) mutations observed as minority variants in at least one sample of the group that become the prevalent viral population in at least one virus of the group.

Group	nt transmitted variants		aa transmitted variants		nt fixed variants		aa fixed variants		
	N	position	N	position	N	position	N	position	Comments*
H7N1 HP	8	101A 234A 886G 910T 924A 980A 1090G 1550G	7	V34E T296A Q304* S308R T327K W364G D517G	3	114T 1403A 1524G	1	L468Q	
H7N1 LP-A	1	886G	1	T296A	34	82G 135T 207T 233C 257A 285T 304C 338A 338G 436T 492C 505A 505G 516A 517T 549C 570G 677T 837G 947G 972G 994T 994G 1004C 1004T 1125C 1165A 1165G 1167G 1211C 1216A 1216G 1335C 1335T	12	I78T S102R E113G A146S A169T M173L P226L V332F I335T G393R L408S D410N	P P, PS P,PS,AGS,AS P
H7N1 LP-B	4	909G 1029G 1078G 1206G	2	F303L W364G	48	150A 186G 194G 223C 229A 229G 257A	6	 T77A	

						285T 305G 312C 333A 333G 366G 379A 379G 436T 436G 452G 516A 584T 584G 621A 621G 624C 648T 759C 759T 867C 867T 931G 975A 975G 999C 1066C 1110A 1110G 1191G 1212A 1212G 1220A 1220T 1288G 1335C 1348A 1348G 1413A 1413G 1422T 1557C 1614G 1627G 1664G		D127N A146S G195V N411I A454T 	 P, PS P, PS, AS HP
H7N3	0		0		11	337A 337G 388A 419C 420C 436A 436T 436G 505G 656A 1664G	3	E113K A146T A146S	 P, PS, HP P, PS
TOT	13		10		96		22		

* P=parallel mutations; PS=positive selection; HP=typical of HPAI viruses; AS=antigenic site; AGS=additional glycosylation site

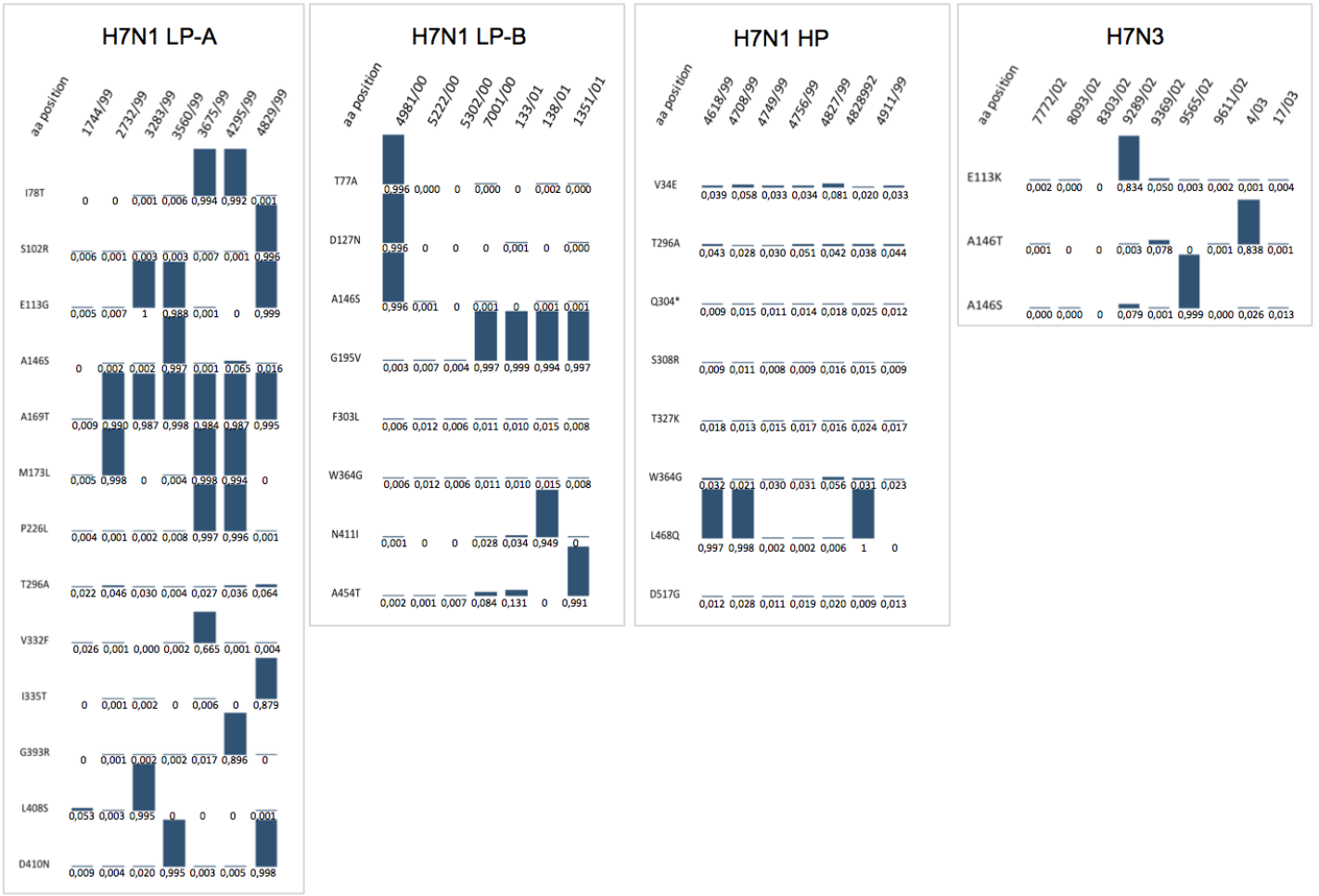
Of note, 8 of the 32 non-synonymous low frequency variants correspond to the previously (paragraph 6.4.1) identified i) parallel mutations (E113G, A146S, A169T, D410N, G195V), ii) substitutions in antigenic sites (169, 195) or iii) in positively selected sites (146, 169, 195) or iv) amino acid changes generating new potential AGS (A169T) (table 6.5).

To better understand the transmission dynamics of minority variants during a H7 epidemic and to evaluate whether bottleneck events may have played a role in the emergence of the H7N1 HPAI and the H7N1 LPAI-I groups, we looked in the progenitor H7N1 LP-A viruses for the presence of low frequency mutations identified in the H7N1 HP and H7N1 LP-B viruses. Unfortunately, the clinical samples available for the H7N3 viruses cover only the first two months of the epidemic and cannot be used to follow the dynamics of variant transmission throughout the entire epidemic wave.

All the low frequency polymorphisms detected in the H7N1 HP group and most (48 out 52) of the ones identified in the H7N1 LP-B samples (table 6.4) were already present in the H7N1 LP-A viruses as minority variants or as consensus-level nucleotides. In particular, 11 of these mutations appear to have been transmitted as minority variants from the progenitor to the H7N1 HP or H7N1 LP-B group, remaining at low frequency in all the analysed samples (figure 6.4); other 17 polymorphisms (2 in the H7N1 HP and 15 in the H7N1 LP-B group) present as consensus-level amino acid in the progenitor viruses had gradually been lost in the new genetic groups (figure 6.4). Conversely, 7 mutations, present at low frequency in some of the progenitor viruses, reached fixation in the H7N1 LP-B group (figure 6.4).

Notably, none of the identified transmitted variants are common to both the H7N1 HP and H7N1 LP-B groups, thus suggesting that they may have evolved from two distinct progenitor viral populations within the H7N1 LP-A group.

Figure 6.3. Graphical representation of changes in frequency (0 to 1) of amino acid mutations identified as minority variants in all the samples of a group (H7N1 LP-A, H7N1 LP-B, H7N1 HP, H7N3) or where at least one sample in a group reached the level of the consensus.



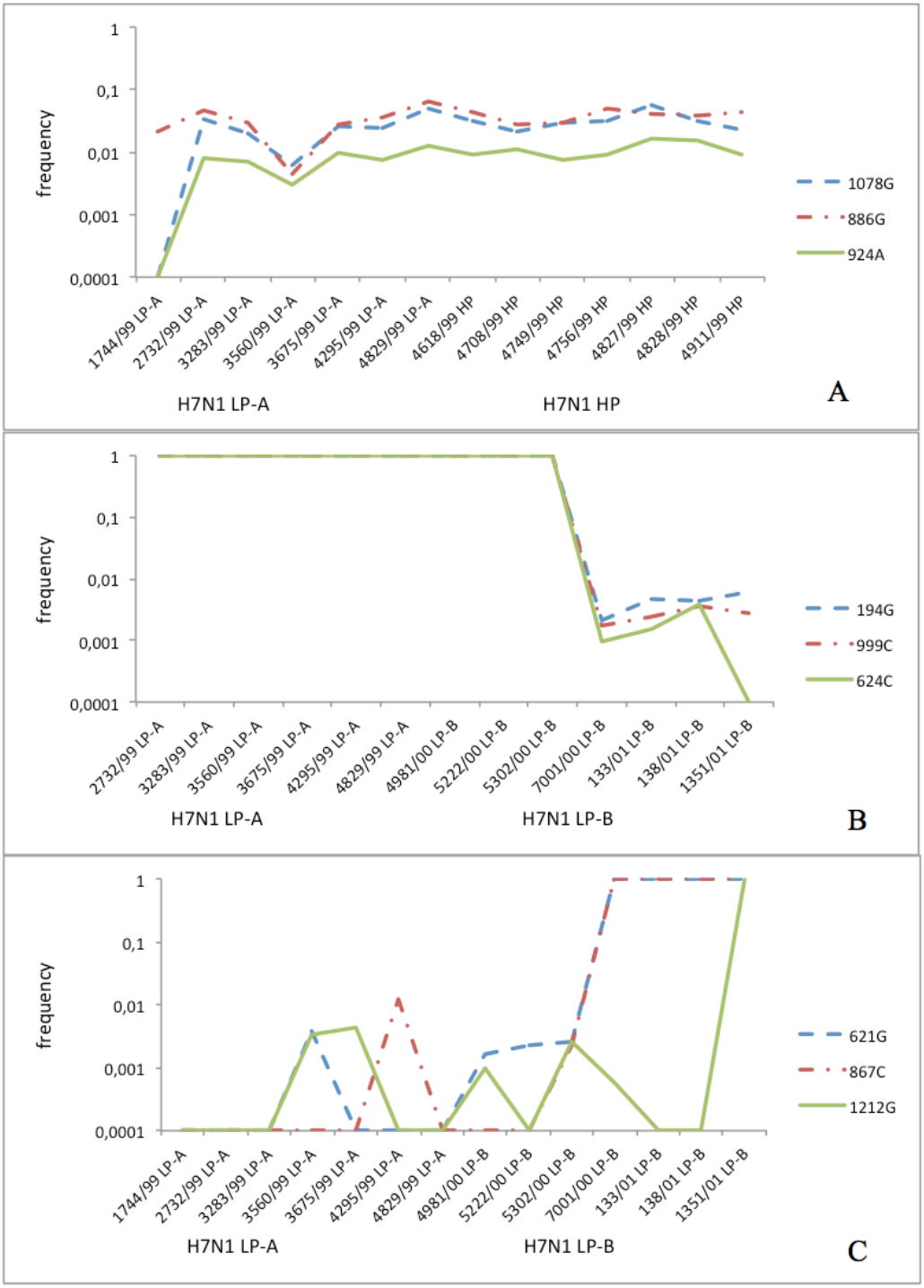


Figure 6.4. Changes in frequency of mutations found in the progenitor viruses (H7N1 LP-A) and transmitted to the descendant H7N1 HP (A) or H7N1 LP-B (B, C) viruses. Data shown is for 9 representative sites. Results are divided according to patterns. A) Minority variants transmitted from outbreak to outbreak (from H7N1 LP-A to H7N1 HP group); C) mutations present in the H7N1 LP-A viruses and gradually lost in the H7N1 LP-B samples; D) low frequency mutations of the H7N1 LP-A viruses that reach fixation in the H7N1 LP-B descendants.

6.4.3. Intra-host genetic variability

We measured the diversity of viral populations of the HA gene of each virus using the Shannon Entropy. Diversity depends on a) the number and b) the variability (presence of multiple nucleotides equitably represented) of polymorphic sites. Low entropy could be detected after a bottleneck that cause the loss of variability, while high entropy could be observed in samples founded by a heterogeneous viral population.

Figure 6.5 shows mean entropy of the HA gene of all the analysed samples considering only positions (N=1017) with a coverage higher than 500. Due to the low coverage, samples 4708/99, 4827/99, 4911/99 (H7N1 HP), 1744/99, 3560/99 (H7N1 LP-A) and 8303/02 (H7N3) have not been included in the analysis. The entropy values range from 0.004 for the sample 8093/02 (H7N3) to 0.01 for the sample 4/03 (H7N3), suggesting that the complexity of viral populations fluctuate considerably among the H7N3 viruses, with the samples collected at the beginning of the epidemic showing the lowest values. This may be due to the population bottleneck that may have accompanied the transmission from wild birds to poultry, after the introduction of the virus in commercial flocks. Other viruses displaying a low variability are 5222/00, 5302/00 (H7N1 LP-B) and 133/01 (H7N1 LP-B), with 5222/00 showing statistically significant lower values compared to most of the other samples. The HA consensus sequences of the 5222/00 and 5302/00 viruses are phylogenetically positioned as outgroups of different genetic groups within the H7N1 LPAI-I lineage (indicated with a black arrow in figure 6.1), while the 133/01 belongs to a subgroup, within the H7N1 LPAI-I lineage, separated by a long branch and five mutations from the previously collected viruses (indicated with a black arrow in figure 6.1). Because of the low coverage of sample 3560/99 in most of the sites of the HA gene, we decided to not include it in our analysis. However, it is noticeable that this virus, which is phylogenetically positioned as the outgroup of the entire H7N1 LPAI-I group, shows a mean entropy (calculated on 745 sites with a coverage >500) of 0.004, similar to the one obtained for sample 5222/00.

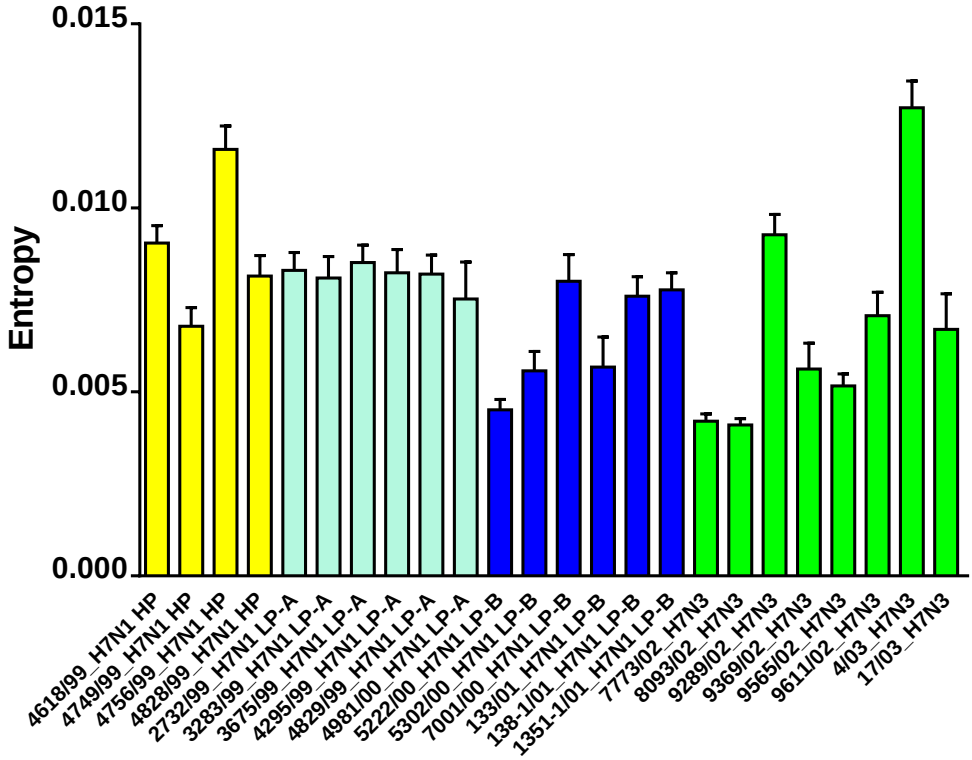


Figure 6.5. Mean entropy of the HA gene of 24 samples considering only validated sites with a coverage higher than 500 reads. Different groups are marked with distinct colours: yellow for the H7N1 HP viruses, light blue for the H7N1 LP-A viruses, blue for the H7N1 LP-B viruses and green for the H7N3 viruses. Error bars represent the standard error of the mean

6.5. DISCUSSION

Avian influenza viruses of the H7 subtype are of particular interest, since they are able to cause highly pathogenic influenza in poultry, as recently demonstrated by the 2012 H7N3 outbreak in Mexico and the 2013 H7N7 outbreak in Italy (Maurer et al., 2013; ProMED-mail (83), 2013), and to seriously affect human health. To date, the ongoing H7N9 epidemic in China has been responsible for 217 human cases (ProMed-mail, 2014), raising concern for the pandemic potential of this strain.

When and how genetic changes associated with increased virulence, host switching or antigenic changes occur is still not clear. By comparing evolutionary dynamics, intra-host variability and patterns of subpopulation transmission and fixation between two H7 epidemics occurring for similar periods of time and in similar epidemiological conditions, we provided a

better understanding of the evolution of H7 viruses during a poultry epidemic.

Analysis of the evolutionary dynamics revealed several similarities between the 1999-2001 H7N1 and 2002-2004 H7N3 epidemics, as above detailed.

1) We identified a total of 28 identical amino acid mutations which had been independently acquired by both subtypes in the course of the two epidemics. One of these (the additional glycosylation site (AGS) in the HA protein) corresponds to the parallel substitution described in chapter 5 between the H7N1 HPAI and H7N1 LPAI-I lineage. 2) The HA phylogeny showed that only two LPAI genetic groups (H7N1 LPAI-I and H7N3 LPAI-I) persisted until the end of the two epidemics. Both these clusters were detected seven months after the first outbreak, possessed a potential AGS in the HA protein and seemed to be circulating mainly in the Verona province. 3) We also demonstrated that both strains followed a similar pattern in the variability of the amino acid diversity among the HA proteins collected in the same month during the two LPAI epidemics. 4) Except for the NA gene, which belongs to different subtypes, the remaining seven segments displayed similar rates of nucleotide substitutions.

Vaccination is the key factor which may have differently affected the evolution of the two strains. Indeed, during the H7N1 and H7N3 epidemics vaccination was started respectively about 4 (November 2000) and 23 (January 2003) months before disease eradication. The similar evolutionary dynamics observed between the two strains suggest that vaccination had only a marginal effect in shaping viral evolution. This finding is very different from what described in chapter 4 for the H5N1, where vaccination has been suggested as influencing both evolutionary rates and selection profiles of the viral surface glycoprotein. We have speculated on two possible explanations: 1) robust vaccination strategies coupled with strict biosecurity precautions adopted during the H7N3 epidemic did not contribute to the rapid evolution and antigenic changes of the HA gene; 2) LP H7N1 viruses may have circulated in a immunized poultry population. Indeed, LP strains were not subject to surveillance and control in 1999-2000 (CEC, 1992), and the notification of LP viruses was not mandatory at that time, therefore multiple infections of the same flock could not be completely excluded.

Although, we observed several examples that suggest a parallel evolutionary trajectory between the two H7 strains, our findings suggest that circulation of distinct viruses under similar epidemiological conditions may led to the evolution of viruses with different genetic variations responsible of distinct virulence properties.

Today, thanks to the advent of the NGS technology, we can identify low frequency variants with key amino acid substitutions before they may become fixed in the viral population, as suggested in chapter 5, where minority variants with HP amino acid signatures were identified in the H7N1 LP samples. To better understand viral evolution also at a subpopulation level, we investigated the transmission dynamics of viral variants through the deep sequencing analysis of 30 clinical samples collected during the H7N1 (21 viruses) and H7N3 (9 viruses) Italian epidemics.

Our results indicate that low frequency variants can be transmitted from outbreak to outbreak throughout an entire epidemic: some of them remain as minority variants, while others fix in the viral population, or else reduce their frequency and disappear.

Detection of several consensus mutations, as the low frequency variants in the previously collected viruses, suggests that the emergence of mutations can be, in most cases, predicted. The number of the minority variants which had fixed in the viral population was higher in the H7N1 LPAI than in the H7N1 HPAI and H7N3 viruses. This difference may be due to the time span of collection of the samples within each group - respectively, 8 and 5 months for the H7N1 LP-A and B and 14 and 69 days for the H7N1 HP and H7N3. Although, the shortest collection period can reduce the probability to detect fixed variants, viruses sampled over a short-time period should share a higher number of low frequency polymorphisms. While, this was observed for the H7N1 HP samples, where the highest number of minority mutations (8, table 6.4) in common between all the analysed samples had been identified, no low frequency variants shared by all the H7N3 viruses were detected. This may be due to the low intra-host variability of the H7N3 viruses collected at the beginning of the epidemic (7772/02 and 8093/02), shortly after the transmission from wild to domestic birds, when viruses still had to recover from this bottleneck event and its subsequent founder effect yet.

Bottlenecks may be speculated also during the H7N1 LPAI-I evolution, as suggested by their low population complexity, the high fixation rate and the long branches in the HA phylogeny characterizing this group. Unfortunately, the lack of H7N3 clinical samples belonging to the H7N3 LPAI-I group does not allow performing any comparison.

Differently, the high entropy values of the H7N1 HPAI samples collected at the beginning of the HPAI epidemic may suggest that this strain a) had not been selected through a narrow bottleneck, b) had undergone exceptionally high-levels of replication which could have rapidly increased its entropy, or c) had circulated for a while before its first detection.

However, based on the findings described in chapter 5, this last theory can be excluded.

The different heterogeneity among the analysed samples may also be related to tissue/organ-specific amplification. Although our data do not provide evidence of any clear association between the different origins of the analysed samples - single organs (i.e. lung, trachea, intestine) and pool of organs (table 6.1) – this possibility cannot be ruled out.

Notably, the commonly acquired mutations and the progressive fixation of several amino acid substitutions which had been observed in these two epidemics indicate that these mutations may confer a strong fitness advantage to H7 viruses in poultry. Further studies will be necessary to elucidate their precise role and to explore mutations in the remaining gene segments. Although the mechanisms involved in the acquisition of virulence determinants cannot be inferred from this study, the identification of minority variants transmitted from farm to farm provide important insights into the evolutionary dynamics of H7 viruses during an avian influenza epidemic, highlighting the importance of the application of a deep sequencing approach to predict the progress of the evolution of a viral population.

ACKNOWLEDGEMENTS

This work was financially supported by the European projects PREDEMICS (Grant Agreement No. 278433) and Epi-SEQ (research project supported under the 2nd joint call for transnational research projects by EMIDA ERA-NET [FP7 project nr 219235]). I acknowledge the receipt of a fellowship from the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems in 2013.

Chapter 7

Discussion

Emerging infectious diseases continue to represent a significant public health threat and nearly 60% of them are of zoonotic origin (Cleaveland et al., 2001). Due to their unparalleled ability to adapt to new hosts and environments, RNA viruses are considered as the most threatening emerging pathogens (Grenfell et al., 2004). The evolutionary dynamics of RNA viruses are complex and their high mutation rates, rapid replication kinetics and large population size present a challenge to traditional population genetics.

This thesis investigates the most essential aspects related to the evolution and epidemiology of RNA viruses thanks to the large amount of sequence data used to study the viral dynamics at multiple spatial-temporal scales. The two RNA viruses analysed here represent two widespread zoonotic diseases responsible for great economic losses and public health concern. Collectively, the four evolutionary studies presented in this thesis (Chapters 3 to 6), which involve the analysis of substitution rates, evolutionary history, spatial spread, selection pressure and within-host diversity, depict the evolutionary process of RNA viruses collected during four distinct epidemics: i) a fox-rabies virus epidemic occurred in north-eastern Italy between 2008 and 2011 (chapter 3), ii) highly pathogenic H5N1 avian influenza outbreaks reported in Egypt between 2006 and 2010 (chapter 4), iii) two avian influenza epidemics caused by two distinct subtypes (H7N1 and H7N3) that took place in northern Italy from 1999 to 2001 (chapter 5) and from 2002 to 2004 (chapter 6).

Through the application of bioinformatics tools, I analysed a large amount of sequences from these epidemics obtained using both first and second generation sequencing technology.

Data generated in this thesis provide important insights into a) the impact and the efficacy of surveillance strategies and control measures implemented during an outbreak, b) the differences in the evolutionary dynamics and in the pattern of viral gene flows between distinct genetic groups, c) the emergence of amino acid mutations which may provide a fitness advantage to the virus, d) the inter-host transmission of viral variants and e) the gain of virulence. All these achievements have been discussed in the following paragraphs.

7.1. IMPACT AND EFFICACY OF SURVEILLANCE CONTROL STRATEGIES

The analyses performed in this thesis can be a useful tool to better understand the impact and the efficacy of the surveillance strategies and the control measures implemented during an outbreak.

In chapter 3, the phylogenetic and phylogeographic analyses of fox rabies sequences sampled during the 2008-2011 Italian outbreaks show the insufficiency of the surveillance strategies and the inadequacy of the oral fox vaccination campaigns implemented at the beginning of the epidemic. Analyses of the spatial dispersal patterns of the two genetic groups, Italy-1 and Italy-2, indicate that after the first identification of the Italy-2 group in the Italian territory in December 2008, this was recognized only 10 months later and more than 100 km farther east, suggesting an insufficient surveillance and a limited coverage of the oral rabies vaccination of foxes at the beginning of the epidemic.

In chapter 4, a comparison of the evolutionary rates between different H5N1 avian influenza viruses collected in countries where poultry vaccination was adopted or not adopted indicates that an inappropriate application of vaccination as a control measure may contribute to the rapid evolution and antigenic changes of H5N1 viruses, creating a good opportunity for the virus to escape from vaccine protection. However, in chapter 6, I showed that when properly planned and adopted vaccination can be a powerful tool for the control and eradication of the avian influenza virus. Although the H7N3 viruses analysed here exhibited similar rates of nucleotide substitutions and a similar pattern of amino acid variability as the H7N1 viruses, which had circulated for similar periods of time in a no-vaccinated poultry population reared in the same geographic area, a possible effect of vaccination in shaping the viral evolution cannot be ruled out.

The phylogeographic analysis conducted in chapter 5 reveals that during the 1999-2001 LPAI/HPAI H7N1 Italian epidemic, the province of Verona (Veneto region) may have played

the most important role in the initial emergence of the HPAI and LPAI-I variants, while the avian population in the province of Mantova (Lombardia region) seems to have played the most important role in the spatial spread of the HPAI viruses. Although the poultry population in Mantova was markedly smaller than the one in Verona, Mantova likely acted as a bridge between the large poultry population in Verona and the other affected provinces. These findings suggests that the ban of restocking and pre-emptive culling which had started in Veneto 20 days earlier and on a greater number of farms than in Lombardia, played an important role in reducing virus transmission within and outside the Veneto area.

7.2. EVOLUTIONARY DYNAMICS OF CO-CIRCULATING GENETIC GROUPS

The studies included in this thesis have provided a better understanding of the evolutionary dynamics and viral gene flows of co-circulating genetic groups. In particular, I analysed a) two different groups of rabies viruses, which showed a different spatial diffusion (chapter 3); b) two Italian H7N1 viruses with different pathogenic properties (chapter 5), showing similar evolutionary rates, but a different behaviour in their geographic spread; and c) two HPAI H5N1 subclades circulating in Egypt between 2007 and 2010 displaying different selection profiles and rates of nucleotide substitution (chapter 4). Despite co-circulating in the same geographic area, these two groups had been subjected to different evolutionary pressures, such as different poultry sectors (backyard and commercial flocks) and different vaccination coverage (about 1-30% in the household sector and 50-60% in the commercial farms). These factors may have facilitated the action of natural selection in favouring an increased substitution rate of the main surface glycoprotein.

All these findings can be extremely useful to better characterize the viral spread and the role of environment in favouring the selection of genetic variants with fitness advantage.

7.3. EVOLUTION OF VIRAL FITNESS

Virus populations are dynamic heterogeneous ensemble of genomes. Changing environmental conditions, i.e., a new host, as in the case of H7N3 viruses (chapter 6) which were transmitted from wild to domestic birds, or immune pressures, such as those conferred by the application of vaccination in poultry in Egypt (chapter 4), can modify the fitness of a viral population. Variants with a high reproductive success in the new environments, if

present, can become dominant in the population: a situation which turns out to be advantageous for the virus, making its eradication difficult. Following, I reported three examples of selection of mutations that may increase viral fitness described in the research chapters of my thesis.

1) In chapter 5, two separate lineages with different pathogenic properties, the H7N1 HPAI and LPAI-I, which emerged from a common ancestor only a few months apart, have been described. Notably, these two viral lineages exhibited a degree of parallel evolution, including a glycosylation site in the HA protein and a truncation of the NS1 protein. This indicates that, despite their very different pathogenic effects, both viral populations may have been subjected to similar selection pressures, acquiring parallel mutations that may have conferred replication and fitness advantages upon avian influenza viruses in the poultry hosts.

2) Similarly, in chapter 6, I provided evidence of a parallel evolution between two distinct avian influenza epidemics, H7N1 and H7N3, circulating for similar periods of time in similar epidemiological conditions. In particular, I identified a total of 28 identical amino acid mutations that had been acquired independently by both subtypes in the course of the two epidemics, two of them located at the antigenic sites of the main surface glycoprotein. In addition, both strains showed evidence of a number of low frequency variants transmitted from outbreak to outbreak and occasionally fixed in the viral population. All these substitutions may have conferred a strong fitness advantage to H7 viruses in poultry.

3) In the HA protein of the Egyptian H5N1 viruses analysed in chapter 4, six of the nine individual codons under positive selection were located at antigenic sites, suggesting that these mutations may confer a fitness advantage to the viruses. In addition, comparison between viruses collected in countries where poultry vaccination was or was not adopted revealed a marked difference in the number of positively selected sites in the HA gene, with viruses collected from vaccinated poultry populations showing higher level of positive selection. This finding suggests that the circulation of H5N1 viruses in partially immunized poultry populations may contribute to increase the fixation rate of advantageous mutations, which may have a strong fitness impact and contribute to the emergence of mutants escaping vaccine pressure.

7.4. GAIN OF VIRULENCE

In chapter 5, I explored how the HPAI H7N1 strain had emerged from its progenitor virus. The HPAI lineage diverged from the LPAI viruses for 19 amino acid changes acquired in all the eight segments over a time period of only nine months. Of note, some of these substitutions have been previously demonstrated to modulate avian influenza virus pathogenicity (D139N and V136I and six amino acids truncation at the C-terminal in the NS1) or have been previously described in other H7 HPAI strains (A146T and the additional glycosylation site at position 141-143 in the HA protein and the truncation at the NS1 protein).

Interestingly, viral variants containing some of these 19 amino acid mutations typical of the HPAI viruses, including the multi-basic HA cleavage site, have been identified in the LPAI samples, suggesting that some of the amino acid changes characterizing the HP cluster were already present with low frequency within several LPAI progenitor viruses. The lack of evidence for a stepwise process of accumulation of mutations at the HA cleavage sites suggests that, in this epidemic, this insertion was most probably caused by a non-homologous recombination.

7.5. INTER-HOST TRANSMISSION OF VIRAL VARIANTS

In chapter 6, using a deep sequencing approach I recognized a number of low frequency variants transmitted from outbreak to outbreak with occasional fixation in the viral population, suggesting that the emergence of mutations can be, in some cases, predicted. This may be extremely useful to control the emergence of highly virulent strains, to detect immune escape variants, to identify subpopulations harbouring mutations associated to resistance toward antiviral drugs or responsible for host switching, and so on.

This study shows that analysis of the only consensus genome sequences provide an incomplete picture of the viral evolutionary dynamics. RNA viruses consist of huge, heterogeneous populations, containing large number of minority sequence variants which can be potentially transmitted to other susceptible hosts. Understanding the evolutionary dynamics of these viral subpopulations may help providing better control strategies of ongoing and future epidemics.

7.6. CONCLUSION

Through these studies, new light has been shed on the origin, evolutionary dynamics, acquisition of virulence determinants and spreading of several epidemic events, providing a better understanding of the complexity of RNA virus evolution. The knowledge of inter and intra-host viral evolution acquired to date will be instrumental to target future viral sampling and data collection most efficiently and effectively, providing precious samples, that, thanks to the advent of the NGS technology, may offer great opportunity for dramatic advancement in our understanding of the complicated evolutionary dynamics of these pathogens.

REFERENCES

Aamir UB, Naeem K, Ahmed Z, Obert CA, Franks J, Krauss S, Seiler P, Webster RG. 2009. Zoonotic potential of highly pathogenic avian H7N3 influenza viruses from Pakistan. Virology 390(Suppl 2):212-220.

Abbas MA, Spackman E, Swayne DE, Ahmed Z, Sarmento L, Siddique N, Naeem K, Hameed A, Rehmani S. 2010. Sequence and phylogenetic analysis of H7N3 avian influenza viruses isolated from poultry in Pakistan 1995-2004. Virol. J. 7:137.

Abdel-Moneim AS, Shany SA, Fereidouni SR, Eid BT, El-Kady MF, Starick E et al. Sequence diversity of the haemagglutinin open reading frame of recent highly pathogenic avian influenza H5N1 isolates from Egypt. Arch Virol 2009; 154:1559-62.

Abolnik C, Londt BZ, Manvell RJ, Shell W, Banks J, Gerdes GH, Akol G, Brown IH. Characterisation of a highly pathogenic influenza A virus of subtype H5N2 isolated from ostriches in South Africa in 2004. 2009. Influenza Other Respir Viruses 3:63-8. doi: 10.1111/j.1750-2659.2009.00074.x.

Alexander DJ 2000. A review of avian influenza in different bird species. Vet Microbiol 74(1-2):3-13.

Alexander DJ 2001. Ecology of avian influenza in domestic birds. Proceedings of the International Symposium on Emergence and Control of Zoonotic Ortho- and Paramyxovirus Diseases. Merieux Foundation, Veyer du Lac, France 25-34.

Alexander DJ 2008. Avian influenza manual for diagnostic tests and vaccines for terrestrial animals, 6th edition. Chapter 2.7.12. World Organisation for Animal Health, Paris, France. http://www.oie.int/eng/normes/ mmanual/a_00002.htm.

Anne-Mieke Vandamme. 2009. Basic concepts of molecular evolution. In: Lemey P, Salemi M, Vandamme AM. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, New York, pp. 1-28.

Anthony SJ, St Leger JA, Pugliares K, Ip HS, Chan JM, Carpenter ZW, et al 2012. Emergence of fatal avian influenza in New England harbor seals. MBio. Jul 31;3(4):e00166-12. doi: 10.1128/mBio.00166-12.

Arafa A, Suarez DL, Hassan MK, Aly MM. Phylogenetic analysis of hemagglutinin and neuraminidase genes of highly pathogenic avian influenza H5N1 Egyptian strains isolated

from 2006 to 2008 indicates heterogeneity with multiple distinct sublineages. Avian Dis 2010; 54(1 Suppl):345-9.

Arai YT. Epidemiology of rabies virus and other lyssaviruses]. Nippon Rinsho. 2005; 63(12): 2167 - 2172.

Arai YT. , Kuzmin IV, Kameoka Y, Botvinkin AD. New lyssavirus genotype from the Lesser Mouse-eared Bat (Myotis blythi), Kyrghyzstan. Emerg Infect Dis. 2003; 9(3): 333 – 337.

B. Dietzschold, M. Schnell, and H. Koprowski. Pathogenesis of rabies. Current Topics in Microbiology and Immunolology, 292:45–56, 2005.

Badrane, H., Tordo, N., 2001. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. J. Virol. 75, 8096-8104.

Baigent SJ, McCauley JW. 2003. Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. Bioessays. Jul;25(7):657-71.

Baillie GJ, Galiano M, Agapow PM, Myers R, Chiam R, Gall A, Palser AL, Watson SJ, Hedge J, Underwood A, Platt S, McLean E, Pebody RG, Rambaut A, Green J, Daniels R, Pybus OG, Kellam P, Zambon M. 2012. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. J. Virol. 86(Suppl 1):11-18.

Balish AL, Davis CT, Saad MD, El-Sayed N, Esmat H, Tjaden JA, Earhart KC, Ahmed LE, Abd El-Halem M, Ali AH, Nassif SA, El-Ebiary EA, Taha M, Aly MM, Arafa A, O'Neill E, Xiyan X, Cox NJ, Donis RO, Klimov AI. Antigenic and genetic diversity of highly pathogenic avian influenza A (H5N1) viruses isolated in Egypt. Avian Dis. 2010 Mar;54(1 Suppl):329-34. PubMed PMID: 20521654.

Banks J, Speidel EC, McCauley JW, Alexander DJ. 2000. Phylogenetic analysis of H7 haemagglutinin subtype influenza A viruses. Arch. Virol. 145(Suppl 5): 1047-1058.

Banks J, Speidel ES, Moore E, Plowright L, Piccirillo A, Capua I, Cordioli P, Fioretti A, Alexander DJ. 2001. Changes in the haemagglutinin and the neuraminidase genes prior to the emergence of highly pathogenic H7N1 avian influenza viruses in Italy. Arch. Virol. 146(Suppl 5): 963-973.

Bataille A, van der Meer F, Stegeman A, Koch G. 2011. Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathog. 7(Suppl 6): e1002094. doi: 10.1371/journal.ppat.1002094.

Beato MS, Mancin M, Yang J, Buratin A, Ruffa M, Maniero S, Fusaro A, Terregino C, Wan XF, Capua I. Antigenic characterization of recent H5N1 highly pathogenic avian influenza viruses

circulating in Egyptian poultry. Virology. 2013 Jan 20;435(2):350-6. doi: 10.1016/j.virol.2012.09.016. Epub 2012 Nov 2. PubMed PMID: 23123011.

Berg M, Englund L, Abusugra IA, 1990. Close relationship between mink influenza (H10N4) and concomitantly circulating avian influenza viruses. Arch Virol 113(1-2):61-71.

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

Bordería AV, Stapleford KA, Vignuzzi M. RNA virus population diversity: implications for inter-species transmission. Curr Opin Virol. 2011 Dec;1(6):643-8. doi: 10.1016/j.coviro.2011.09.012. Epub 2011 Oct 22.

Bos ME, Nielen M, Toson M, Comin A, Marangon S, Busani L. 2010. Within-flock transmission of H7N1 highly pathogenic avian influenza virus in turkeys during the Italian epidemic in 1999-2000. Prev. Vet. Med. 95(Suppl 3-4):297-300.

Bourhy H., Kissi B., Tordo N. (1993). Taxonomy and evolutionary studies on lyssaviruses with special reference to africa. Onderstepoort J Vet Res 60, 277-282.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J.F., Holmes, E.C., 1999. Ecology and evolution of rabies virus in Europe. J. Gen. Virol. 80 (Pt 10), 2545-2557.

Bourhy H, Dacheux L, Strady C, Mailles A. Rabies in Europe in 2005. Euro Surveill. 2005; 10(11): 213 - 216.

Bourhy H., Reynes J. M., Dunham E. J., Dacheux L., Larrous F., Huong V. T., Xu G., Yan J., Miranda M. E., Holmes E. C. (2008). The origin and phylogeography of dog rabies virus. J Gen Virol 89, 2673-2681.

Bullough PA, Hughson FM, Skehel JJ, Wiley DC. 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature. Sep 1;371(6492):37-43.

Busani L, Valsecchi MG, Rossi E, Toson M, Ferrè N, Pozza MD, Marangon S. 2009. Risk factors for highly pathogenic H7N1 avian influenza virus infection in poultry during the 1999-2000 epidemic in Italy. Vet. J. 181(Suppl 2):171-177.

Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. 1999. Predicting the evolution of human influenza A. Science. 286 (Suppl 5446): 1921-1925.

Butt AM, Siddique S, Idrees M, Tong Y. Avian influenza A (H9N2): computational molecular analysis and phylogenetic characterization of viral surface proteins isolated between 1997 and 2009 from the human population. Virol J. 2010 Nov 15;7:319. doi: 10.1186/1743-422X-7-

319.

Campitelli L, Mogavero E, De Marco MA, Delogu M, Puzelli S, Frezza F, Facchini M, Chiapponi C, Foni E, Cordioli P, Webby R, Barigazzi G, Webster RG, Donatelli I. Interspecies transmission of an H7N3 influenza virus from wild birds to intensively reared domestic poultry in Italy. Virology. 2004 May 20;323(1):24-36.

Capello, K., Mulatti, P., Comin, A., Gagliazzo, L., Guberti, V., Citterio, C., De Benedictis, P., Lorenzetto, M., Costanzi, C., Vio, P., Zambotto, P., Ferri, G., Mutinelli, F., Bonfanti, L., Marangon, S., 2010. Impact of emergency oral rabies vaccination of foxes in northeastern Italy, 28 December 2009-20 January 2010: preliminary evaluation. Euro Surveill. 15, 19617.

Capobianchi MR, Giombini E, Rozera G (2013) Next-generation sequencing technology in clinical virology. Clin Microbiol Infect 19 (1):15-22. doi:10.1111/1469-0691.12056.

Capua I, Mutinelli F, Marangon S, Alexander DJ. 2000. H7N1 avian influenza in Italy (1999 to 2000) in intensively reared chickens and turkeys. Avian Pathol. 29(Suppl 6): 537-543

Capua I, Marangon S, Cancellotti FM. 2003. The 1999-2000 avian influenza (H7N1) epidemic in Italy. Vet. Res. Commun. 27(Suppl 1): 123-127.

Capua I, Marangon S. Control and prevention of avian influenza in an evolving scenario. Vaccine 2007; 25:5645-52. Review.

Capua I, Alexander DJ. 2008. Avian influenza and Newcastle disease: A field and Laboratory Manual. Springer-Verlag, Italy.

Carroll SM, Higa HH, Paulson JC. 1981. Different cell-surface receptor determinants of antigenically similar influenza virus hemagglutinins. J Biol Chem. Aug 25;256(16):8357-63.

Cattoli G, Milani A, Temperton N, Zecchin B, Buratin A, Molesti E, Aly MM, Arafa A, Capua I. Antigenic drift in H5N1 avian influenza virus in poultry is driven by mutations in major antigenic sites of the hemagglutinin molecule analogous to those for human influenza virus. J Virol. 2011 Sep;85(17):8718-24.

Cattoli G, Monne I, Fusaro A, Joannis TM, Lombin LH, Aly MM, Arafa AS, Sturm-Ramirez KM, Couacy-Hymann E, Awuni JA, Batawui KB, Awoume KA, Aplogan GL, Sow A, Ngangnou AC, El Nasri Hamza IM, Gamatié D, Dauphin G, Domenech JM, Capua I. 2009. Highly pathogenic avian influenza virus subtype H5N1 in Africa: a comprehensive phylogenetic analysis and molecular characterization of isolates. PLoS One 4(Suppl 3):e4842.

Cattoli G, Fusaro A, Monne I, Coven F, Joannis T, El-Hamid HS, Hussein AA, Cornelius C, Amarin NM, Mancin M, Holmes EC, Capua I. 2011. Evidence for differing evolutionary

dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. Vaccine. 29(Suppl 50):9368-9375.

Cauthen AN, Swayne DE, Sekellick MJ, Marcus PI, Suarez DL. 2007. Amelioration of influenza virus pathogenesis in chickens attributed to the enhanced interferon-inducing capacity of a virus with a truncated NS1 gene. J Virol 81(Suppl 14): 1838-1847.

Ceballos, N.A., Morón, S.V., Berciano, J.M., Nicolás, O., López, C.A., Juste, J., Nevado, C.R., Setién, Á.A., Echevarría, J.E., 2013. Novel lyssavirus in Bat, Spain. Emerg. Infect. Dis. 19, 793–795.

CEC. 1992. Council Directive 92/40/EEC of 19 May 1992 introducing Community measures for the control of avian influenza. Off J Eur Comm L167, 1-15.

Centers for Disease Control and Prevention (CDC). (2012) Notes from the field: Highly pathogenic avian influenza A (H7N3) virus infection in two poultry workers--Jalisco, Mexico, July 2012. MMWR Morb. Mortal. Wkly Rep. 61(36): 726-727.

Chan PK 2002. Outbreak of avian influenza A (H5N1) virus infection in Hong Kong in 1997. Clin Infect Dis 34(2 Suppl):58–64.

Chen H, Smith GJD, Zhang SY, Qn K, Wang J, Li K S et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature 2005; 436:191–192.

Chen H, Smith GJ, Li KS, Wang J, Fan XH, Rayner JM, Vijaykrishna D, Zhang JX, Zhang LJ, Guo CT, Cheung CL, Xu KM, Duan L, Huang K, Qin K, Leung YH, Wu WL, Lu HR, Chen Y, Xia NS, Naipospos TS, Yuen KY, Hassan SS, Bahri S, Nguyen TD, Webster RG, Peiris JS, Guan Y. Establishment of multiple sublineages of H5N1 influenza virus in Asia: implications for pandemic control. Proc Natl Acad Sci U S A. 2006;103:2845-50.

Chen Y, Liang W, Yang S, Wu N, Gao H, Sheng J, Yao H, Wo J, Fang Q, Cui D, Li Y, Yao X, Zhang Y, Wu H, Zheng S, Diao H, Xia S, Zhang Y, Chan KH, Tsoi HW, Teng JL, Song W, Wang P, Lau SY, Zheng M, Chan JF, To KK, Chen H, Li L, Yuen KY. Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. Lancet. 2013 Jun 1;381(9881):1916-25. doi: 10.1016/S0140-6736(13)60903-4. Epub 2013 Apr 25. PubMed PMID: 23623390

Chenik M., Chebli K., Gaudin Y., Blondel D. (1994). In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): Existence of two N-binding sites on P protein. J Gen Virol 75 (Pt 11), 2889-2896.

Clancy,S. 2008. Genetics of the influenza virus. Nature Education 1(1):83.

Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Philos Trans R Soc Lond B Biol Sci 356: 991–999. doi: 10.1098/rstb.2001.0889.

Croville G, Soubies SM, Barbieri J, Klopp C, Mariette J, Bouchez O, Camus-Bouclainville C, Guérin JL. 2012. Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. J. Clin. Microbiol. 50(Suppl 9):2881-2887. doi:10.1128/JCM.01142-12.

Das K, Aramini JM, Ma LC, Krug RM, Arnold E. Structures of influenza A proteins and insights into antiviral drug targets. Nat Struct Mol Biol. 2010 May;17(5):530-8. doi: 10.1038/nsmb.1779. Epub 2010 Apr 11. Review.

De Benedictis, P., Gallo, T., Iob, A., Coassin, R., Squecco, G., Ferri, G., D'Ancona, F., Marangon, S., Capua, I., Mutinelli, F., 2008. Emergence of fox rabies in north-eastern Italy. Euro Surveill. 13, pii: 19033.

De Benedictis, P., Capua, I., Mutinelli, F., Wernig, J.M., Aric, T., Hostnik, P., 2009. Update on Fox Rabies in Italy and Slovenia. Rabies Bulletin Europe 33, 5-7.

de Mathos Carlos A, de Mattos Cecilia C, Rupprecht Charles E. Rhabdoviruses. 1245 - 1277. In: Knipe David M, Howley Peter M. Fields Virology 2001. Lippincott Williams & Wilkins, Philadelphia . Baltimore . New York . London . Buenos Aires . Hong Kong . Sydney . Tokyo.

de Wit E, Fouchier RA. Emerging influenza. J Clin Virol. 2008 Jan;41(1):1-6. Review.

de Wit E, Munster VJ, van Riel D, Beyer WE, Rimmelzwaan GF, Kuiken T, Osterhaus AD, Fouchier RA. 2010. Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. J Virol.;84(3):1597-606.

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26(Suppl 19): 2455-2457. doi:10.1093/bioinformatics/btq429.

Dietzgen RG, Calisher CH, Kurath G, Kuzmin IV, Rodriguez LL, Stone DM, et al. Family Rhabdoviridae In: King AM, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier; 2011.

Domenech J, Dauphin G, Rushton J, McGrane J, Lubroth J, Tripodi A, et al. Experiences with vaccination in countries endemically infected with highly pathogenic avian influenza: the Food and Agriculture Organization perspective. Rev Sci Tech 2009; 28:293-305. Review.

Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012 Jun;76(2):159-216. doi: 10.1128/MMBR.05023-11.

Dorigatti I, Mulatti P, Rosà R, Pugliese A, Busani L. 2010. Modelling the spatial spread of H7N1 avian influenza virus among poultry farms in Italy. Epidemics 2(Suppl 1): 29-35.

Driskell EA, Jones CA, Berghaus RD, Stallknecht DE, Howerth EW, Tompkins SM. 2012. Domestic Cats Are Susceptible to Infection With Low Pathogenic Avian Influenza Viruses From Shorebirds. Vet Pathol. Jun 25.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22(Suppl 5): 1185-1192.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7: 214.

Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet. 2008 Apr;9(4):267-76. doi:10.1038/nrg2323.

Eagles D, Siregar ES, Dung DH, Weaver J, Wong F, Daniels P. H5N1 highly pathogenic avian influenza in Southeast Asia. Rev Sci Tech 2009; 28:341-8. Review.

EFSA 2005. Epidemiology report on avian influenza in a quarantine premises in Essex http://defra.gov.uk/animalh/diseases/notifiable/disease/ai/pdf/ai-epidemrep 111105.pdf.

Ellis TM, Sims LD, Wong HK, Wong CW, Dyrting KC, Chow KW et al. Use of avian influenza vaccination in Hong Kong. Dev Biol 2006;124:133-43.

Elly SS. Vaccination Experience in AI Control in Indonesia. Seminar 5-"Vaccination against AI: Issues and Strategies Within the Context of an Overall Control Program" Organized by World Bank, FAO, OIE and Tokyo Development Learning Center March 19, 2008.

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56(Suppl 1): 152-179.

EMPRES/FAO-GLEWS. H5N1 HPAI Global overview – January 2010. 2010. Issue No 19. Available from: http://www.fao.org/docrep/012/ak725e/ak725e00.pdf.

EU European Commission, 2002. The oral vaccination of foxes against rabies, Report of the Scientific Committee on Animal Welfare Adopted on 23 October 2002.

Felsenstein J 1985. Confidence limits on phylogenies: an approach using bootstrap. Evolution 39: 783-791.

Finotello F, Lavezzo E, Fontana P, Peruzzo D, Albiero A, Barzon L, Falda M, Di Camillo B, Toppo S (2012) Comparative analysis of algorithms for whole-genome assembly of pyrosequencing data. Brief Bioinform 13 (3):269-280. doi:10.1093/bib/bbr063.

Food and Agricultural Organization of the United Nations. Animal Production and Health Division. Emergency Centre for Transboundary Animal Diseases, Socio Economics, Production and Biodiversity Unit. Poultry sector country review. 2008. Available from: ftp://ftp.fao.org/docrep/fao/011/ai355e/ai355e00.pdf.

Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, Smith D, Rimmelzwaan GF, Olsen B, Osterhaus AD. 2005. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. J. Virol. 79(Suppl 5):2814-2822.

Freuling, C.M., Beer, M., Conraths, C.D., Finke, S., Hoffmann, B., Keller, B., Kliemt, J., Mettenleiter, T.C., Mühlbach, E., Teifke, J.P., Wohlsein, P., Müller, T., 2011. Novel lyssavirus in Natterer's bat, Germany. Emerg. Infect. Dis. 17, 1519–1522.

Fusaro A, Nelson MI, Joannis T, Bertolotti L, Monne I, Salviato A, Olaleye O, Shittu I, Sulaiman L, Lombin LH, Capua I, Holmes EC, Cattoli G. 2010. Evolutionary dynamics of multiple sublineages of H5N1 influenza viruses in Nigeria from 2006 to 2008. J. Virol. 84(Suppl 7):3239-3247.

Fusaro A, Joannis T, Monne I, Salviato A, Yakubu B, Meseko C, et al. Introduction into Nigeria of a distinct genotype of avian influenza virus (H5N1). Emerg Infect Dis 2009; 15:445–47.

Gaidet N, Cattoli G, Hammoumi S, Newman SH, Hagemeijer W, Takekawa JY, Cappelle J, Dodman T, Joannis T, Gil P, Monne I, Fusaro A, Capua I, Manu S, Micheloni P, Ottosson U, Mshelbwala JH, Lubroth J, Domenech J, Monicat F. Evidence of infection by H5N2 highly pathogenic avian influenza viruses in healthy wild waterfowl. 2008. PLoS Pathog. 4:e1000127. doi: 10.1371/journal.ppat.1000127.

Garcia M, Crawford JM, Latimer JW, Rivera-Cruz E, Perdue ML. 1996. Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico. J. Gen. Virol. 77(Pt 7):1493-1504.

Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier RA, Pappas C, Alpuche-Aranda CM, López-Gatell H, Olivera H, López I, Myers CA, Faix D, Blair PJ, Yu C, Keene KM, Dotson PD, Boxrud D, Sambol AR, Abid SH, St George K, Bannerman T, Moore AL, Stringer DJ,

Blevins P, Demmler-Harrison GJ, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA, Beatrice ST, et al. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325:197–201.

Giannecchini S, Clausi V, Di Trani L, Falcone E, Terregino C, Toffan A, Cilloni F, Matrosovich M, Gambaryan AS, Bovin NV, Delogu M, Capua I, Donatelli I, Azzi A. 2010. Molecular adaptation of an H7N3 wild duck influenza virus following experimental multiple passages in quail and turkey. Virology 408(Suppl 2):167-173.

Grenfell B. T., Pybus O. G., Gog J. R., Wood J. L., Daly J. M., Mumford J. A., Holmes E. C.2004 Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332. doi:10.1126/science.1090727 (doi:10.1126/science.1090727)

Grund C, Abdelwhab ES, Arafa AS, Ziller M, Hassan MK, Aly MM et al. Highly pathogenic avian influenza virus H5N1 from Egypt escapes vaccine-induced immunity but confers clinical protection against a heterologous clade 2.2.1 Egyptian isolate. Vaccine 2011 E-pub ahead of print].

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52(Suppl 5): 696-704.

Gupta P. K., Singh R. K., Sharma R. N., Rao Y. U., Butchaiah G. (2001). Preliminary report on a single-tube, non-interrupted reverse transcription-polymerase chain reaction for the detection of rabies virus in brain tissue. Vet Res Commun 25, 239-247.

Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science. 2009 Oct 30;326(5953):734-6. doi: 10.1126/science.1178258.

Herz C, Stavnezer E, Krug R, Gurney T Jr. 1981. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell. Nov;26(3 Pt 1):391-400.

Hinshaw VS, Webster RG, Easterday BC, Bean WJ Jr 1981. Replication of avian influenza A viruses in mammals. Infect Immun 34(2):354-361.

Hoffmann B, Scheuch M, Hoper D, Jungblut R, Holsteg M, Schirrmeier H, Eschbaumer M, Goller KV, Wernike K, Fischer M, Breithaupt A, Mettenleiter TC, Beer M (2012) Novel orthobunyavirus in Cattle, Europe, 2011. Emerg Infect Dis 18 (3):469-472. doi:10.3201/eid1803.111905.

Holder MH, Lewis PA 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Gen 4: 275- 284.

Holmes EC, Lipman DJ, Zamarin D, Yewdell JW. 2006. Comment on 'Large-scale analysis of avian influenza isolates'. Science. 313: 1573b.

Horimoto T, Rivera E, Pearson J, Senne D, Krauss S, Kawaoka Y, Webster RG. 1995. Origin and molecular changes associated with emergence of a highly pathogenic H5N2 influenza virus in Mexico. Virology. 213(Suppl 1):223-230.

Horimoto T, KawaokaY. 2001. Pandemic threat posed by avian influenza A viruses. Clin Microbiol Rev 14(1):129–149.

Horimoto T, Kawaoka Y. 2005.Influenza:lessons from past pandemics, warnings from current incidents. Nat. Rev. Microbiol. 3:591– 600.

Huelsenbeck JP 1995. Performance of phylogenetic methods in simulation. Syst Biol 44: 17-48.

Hughes J, Allen RC, Baguelin M, Hampson K, Baillie GJ, Elton D, Newton JR, Kellam P, Wood JL, Holmes EC, Murcia PR. 2012. Transmission of Equine Influenza Virus during an Outbreak Is Characterized by Frequent Mixed Infections and Loose Transmission Bottlenecks. PLoS Pathog. 8(Suppl 12):e1003081. doi:10.1371/journal.ppat.1003081.

Hulse-Post DJ, Franks J, Boyd K, Salomon R, Hoffmann E, Yen HL, Webby RJ, Walker D, Nguyen TD, Webster RG. 2007. Molecular changes in the polymerase genes (PA and PB1) associated with high pathogenicity of H5N1 influenza virus in mallard ducks. J. Virol. 81(Suppl 16):8515-8524.

Hurt AC, Hansbro PM, Selleck P, Olsen B, Minton C, Hampson AW, Barr IG. Isolation of avian influenza viruses from two different transhemispheric migratory shorebird species in Australia. Arch Virol. 2006 Nov;151(11):2301-9.

Iqbal M, Essen SC, Xiao H, Brookes SM, Brown IH, McCauley JW. 2012. Selection of variant viruses during replication and transmission of H7N1 viruses in chickens and turkeys. Virology.;433(2):282-95. doi: 10.1016/j.virol.2012.08.001.

ISTAT (the Italian National Institute for Statistics). 5th Census of agriculture. Data available at: http://censagr.istat.it/.

Jackson A. C. & Wunner W. H., editors. (2002). Rabies. United States of America: Academic press An Elsevier Science Imprint.

Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, et al 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. Science.

Jul 13;337(6091):199-204.

Karasin AI, Brown IH, Carman S, Olsen CW 2000. Isolation and characterization of H4N6 avian influenza viruses from pigs with pneumonia in Canada. J Virol 74(19):9322-9327.

Karasin AI, West K, Carman S, Olsen CW 2004. Characterization of avian H3N3 and H1N1 influenza A viruses isolated from pigs in Canada. J Clin Microbiol 42(9):4349-4354.

Kaverin NV, Rudneva IA, Govorkova EA, Timofeeva TA, Shilov AA, Kochergin-Nikitsky KS, Krylov PS, Webster RG. 2007. Epitope mapping of the hemagglutinin molecule of a highly pathogenic H5N1 influenza virus by using monoclonal antibodies. J. Virol. 81(Suppl 23):12911-12917.

Keiner B, Maenz B, Wagner R, Cattoli G, Capua I, Klenk HD. 2010. Intracellular distribution of NS1 correlates with the infectivity and interferon antagonism of an avian influenza virus (H7N1). J. Virol. 84(Suppl 22):11858-11865.

Kelly, R. M. & Strick, P. L. Rabies as a transneuronal tracer of circuits in the central nervous system. J. Neurosci. Methods 103, 63–71 (2000).

Kim JK, Kayali G, Walker D, Forrest HL, Ellebedy AH, Griffin YS, et al. Puzzling inefficiency of H5N1 influenza vaccines in Egyptian poultry. Proc Natl Acad Sci U S A. 2010; 107:11044-9.

King AA. Rabies. 437 - 458. In: Palmer SR, Soulsby Lord, Simpson DIH. Zoonoses: Biology, Clinical Practice, and Public Health Control 1998. Oxford University Press, USA.

Kissi B., Tordo N., Bourhy H. (1995). Genetic polymorphism in the rabies virus nucleoprotein gene. Virology 209, 526-537.

Kochs G, García-Sastre A, Martínez-Sobrido L. 2007. Multiple anti-interferon actions of the influenza A virus NS1 protein. J Virol. Jul;81(13):7011-21.

Kosakovsky Pond SL, Frost DWS. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution. 2005; 22:1208-22.

Kuhner MK, Felsenstein J. 1995. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol May;12(3):525.

Kuzmin I. V., Hughes G. J., Botvinkin A. D., Gribencha S. G., Rupprecht C. E. (2008). Arctic and arctic-like rabies viruses: Distribution, phylogeny and evolutionary history. Epidemiol Infect 136, 509-519.

Kuzmin, I.V., Tordo, N., 2012. Genus lyssavirus. In: Dietzgen, R.D., Kuzmin, I.V. (Eds.), Rhabdoviruses: Molecular Taxonomy, Evolution, Genomics, Ecology, Host– Vector Interactions, Cytopathology and Control. Caister Academic Press, London, pp. 37–57.

Lam TT, Hon CC, Pybus OG, Kosakovsky Pond SL, Wong RT, Yip CW, Zeng F, Leung FC. 2008. Evolutionary and transmission dynamics of reassortant H5N1 influenza virus in Indonesia. PLoS Pathog. 4(Suppl 8):e1000130. doi:10.1371/journal.ppat.1000130.

Lamb RA, Lai CJ, Choppin PW. 1981. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc Natl Acad Sci U S A. Jul;78(7):4170-4.

Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog. 2010 Jul 22;6(7):e1001005. doi: 10.1371/journal.ppat.1001005.

Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol. 2013 May;11(5):327-36. doi:10.1038/nrmicro3003.

Lebarbenchon C, Stallknecht DE. 2011. Host shifts and molecular evolution of H7 avian influenza virus hemagglutinin. Virol. J. 8: 328.

Lee C. W., Senne D. A., Suarez D. L.. 2004. Effect of vaccine use in the evolution of Mexican lineage H5N2 avian influenza virus. J. Virol. 78:8372–8381.

Lee MS, Chen MC, Liao YC, Hsiung CA. 2007. Identifying potential immunodominant positions and predicting antigenic variants of influenza A/H3N2 viruses. Vaccine 25(Suppl 48):8133-8139.

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

Lin YP, Shaw M, Gregory V, Cameron K, Lim W, et al. (2000) Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci U S A 97: 9654–9658. doi: 10.1073/pnas.160270697.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Li J, Wang YY, Liang BN, Wan Y, Liao Z, Chan KH, Yuen KY, Fu X, Shang X, Wang S, Yi D, Guo B, Di B, Wang M, Che X, Wu Y. Fine antigenic variation within H5N1 influenza virus hemagglutinin's antigenic sites defined by yeast cell surface display. Eur J Immunol 2009; 39:3498-3510.

Li J, zu Dohna H, Anchell NL, Adams SC, Dao NT, Xing Z, Cardona CJ. 2010. Adaptation and transmission of a duck-origin avian influenza virus in poultry species. Virus Res. 147(Suppl 1):40-46.

Li Y, Yamakita Y, Krug RM. 1998. Regulation of a nuclear export signal by an adjacent inhibitory sequence: the effector domain of the influenza virus NS1 protein. Proc Natl Acad Sci U S A; 95(9):4864-9.

Liu, P., Yang, J., Wu, X. & Fu, Z. F. Interactions amongst rabies virus nucleoprotein, phosphoprotein and genomic RNA in virus-infected and transfected cells. J. Gen. Virol. 85, 3725–3734 (2004).

Loeffen W, De Boer-Luitze E, Koch G 2004. Transmission of a highly pathogenic avian influenza virus to swine in the Netherlands. Proceedings of the in-between congress of the International Society for Animal Hygiene, 329-330.

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30 (5):434-439. doi:10.1038/nbt.2198.

Londt BZ, Banks J, Alexander DJ. Highly pathogenic avian influenza viruses with low virulence for chickens in in vivo tests. 2007. Avian Pathol. 36:347-50.

Lojkic, I., Cac, Z., Bedekovic, T., Lemo, N., Brstilo, M., Muller, T., Freuling, C.M., 2012. Diversity of currently circulating rabies virus strains in Croatia. Berl. Munch. Tierarztl. Wochenschr. 125, 249-254.

Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21(Suppl 6): 936-939. doi:10.1101/gr.111120.110.

Lvov D 1978. Circulation of influenza viruses in natural biocoenosis. Viruses and Environment 351-380.

Manrubia SC, Lázaro E, Viral evolution, Physics of Life Reviews, Volume 3, Issue 2, June 2006, Pages 65-92, ISSN 1571-0645, doi:10.1016/j.plrev.2005.11.002.

Maddison, W.P., Maddison, D.R., 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia. Primatol. (Basel) 53, 190-202.

Maddison, W. P. and D.R. Maddison. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org

Maillard AP, Gaudin Y. Rabies virus glycoprotein can fold in two alternative, antigenically distinct conformations depending on membrane-anchor type. J Gen Virol. 2002 Jun;83(Pt 6):1465-76.

Malik Peiris JS. Avian influenza viruses in humans. Rev Sci Tech. 2009 Apr;28(1):161-73. Review. PubMed PMID: 19618624.

Mannelli A, Ferrè N, Marangon S. 2006. Analysis of the 1999-2000 highly pathogenic avian influenza (H7N1) epidemic in the main poultry-production area in northern Italy. Prev Vet. Med.;73(4):273-85.

Mansfield K. L., Racloz V., McElhinney L. M., Marston D. A., Johnson N., Ronsholt L., Christensen L. S., Neuvonen E., Botvinkin A. D., Rupprecht C. E., Fooks A. R. (2006). Molecular epidemiological study of arctic rabies virus isolates from greenland and comparison with isolates from throughout the arctic and baltic regions. Virus Res 116, 1-10.

Mänz B, Schwemmle M, Brunotte L. Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. J Virol. 2013 Jul;87(13):7200-9. doi: 10.1128/JVI.00980-13. Epub 2013 Apr 24. Review.

Marangon S, Capua I, Rossi E, Ferrè N, Dalla Pozza M, Bonfanti L, Mannelli A. 2005. The control of avian influenza in area at risk: the Italian experience 1997-2005. In Koch G, Schrijver R (Eds.), Avian Influenza, Prevention and Control. Wageningen UR Frontis Series. Kluwer Academic Publisher.

Marston, D.A., McElhinney, L.M., Johnson, N., Müller, T., Conzelmann, K.K., Tordo, N., Fooks, A.R., 2007. Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and poly adenylation motif in the G-L 30 non-translated region. J. Gen. Virol. 88, 1302–1314.

Martin K, Helenius A. 1991. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell. Oct 4;67(1):117-30.

Maurer-Stroh S, Lee RT, 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. doi: 10.1186/1743-422X-10-139.

McElhinney, L.M., Marston, D.A., Freuling, C.M., Cragg, W., Stankov, S., Lalosevic, D., Lalosevic, V., Muller, T., Fooks, A.R., 2011. Molecular diversity and evolutionary history of rabies virus strains circulating in the Balkans. J. Gen. Virol. 92, 2171-2180.

Mebatsion, T., Weiland, F. & Conzelmann, K. K. Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G. J. Virol. 73, 242–250 (1999).

Mebatsion, T., Konig, M. & Conzelmann, K. K. Budding of rabies virus particles in the absence of the spike glycoprotein. Cell 84, 941–951 (1996).

Melén K, Kinnunen L, Fagerlund R, Ikonen N, Twu KY, Krug RM, Julkunen I. 2007. Nuclear and nucleolar targeting of influenza A virus NS1 protein: striking differences between different virus subtypes. J. Virol. 81(Suppl 11):5995-6006.

Messenger SL, Smith JS, Orciari LA, Yager PA, Rupprecht CE. Emerging pattern of rabies deaths and increased viral infectivity. Emerg Infect Dis. 2003 Feb;9(2):151-4.

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. Trans. R. Soc. Lond. B. Biol. Sci. 363, 3985-3995.

Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol 2008c; 2:1459-71.

Monne, I., Fusaro, A., Valastro, V., Citterio, C., Dalla Pozza, M., Obber, F., Trevisiol, K., Cova, M., De Benedictis, P., Bregoli, M., Capua, I., Cattoli, G., 2011. A distinct CDV genotype causing a major epidemic in Alpine wildlife. Vet. Microbiol. 150, 63-69.

Morelli MJ, Wright CF, Knowles NJ, Juleff N, Paton DJ, King DP, Haydon DT (2013) Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. Vet Res 44 (1):12. doi:10.1186/1297-9716-44-12.

Morens DM, Taubenberger JK, Fauci AS. 2009. The persistent legacy of the 1918 influenza virus. N. Engl. J. Med. 361:225–229.

Moya A, Elena SF, Bracho A, Miralles R, Barrio E. The evolution of RNA viruses: A population genetics view. Proc Natl Acad Sci U S A. 2000 Jun 20;97(13):6967-73.

Mulatti P, Kitron U, Mannelli A, Ferré N, Marangon S. 2007. Spatial analysis of the 1999-2000 highly pathogenic avian influenza (H7N1) epidemic in northern Italy. Avian Dis. 51(Suppl 1): 421-424.

Mulatti P, Kitron U, Jacquez GM, Busani L, Mannelli A, Marangon S. 2010a. Evaluation of the risk of neighbourhood infection of H7N1 Highly Pathogenic Avian Influenza in Italy using Q statistic. Preventive Veterinary Medicine 95:267–274.

Mulatti P, Bos MEH, Busani L, Nielen M, Marangon S. 2010b. Evaluation of interventions and vaccination strategies for Low Pathogenicity Avian Influenza: spatial and space-time analyses and quantification of the spread of infection. Epidemiol. Infect. 138: 813–824.

Mulatti, P., Ferre, N., Patregnani, T., Bonfanti, L., Marangon, S., 2011. Geographical information systems in the management of the 2009-2010 emergency oral anti-rabies vaccination of foxes in north-eastern Italy. Geospat Health. 5, 217-226.

Mulatti, P., Muller, T., Bonfanti, L., Marangon, S., 2012. Emergency oral rabies vaccination of foxes in Italy in 2009-2010: identification of residual rabies foci at higher altitudes in the Alps. Epidemiol. Infect. 140, 591-598.

Muramoto Y, Noda T, Kawakami E, Akkina R, Kawaoka Y. Identification of novel influenza A virus proteins translated from PA mRNA. J Virol. 2013 Mar;87(5):2455-62. doi: 10.1128/JVI.02656-12.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8(Suppl 7):e1002764. doi:10.1371/journal.pgen.1002764.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013 May;30(5):1196-205. doi: 10.1093/molbev/mst030.

Mutinelli, F., Stankov, S., Hirstovski, M., Seimenis, A., Theoharakou, H., Vodopija, I., 2004. Chapter 8, Rabies in Italy, Yugoslavia, Croatia, Bosnia, Slovenia, Macedonia, Albania and Greece., in Anonymous Historical Perspectives of Rabies in Europe and the Mediterranean Basin, OIE publications ed. A.A. King, A.R. Fooks, A. Wandeler & M. Aubert, Paris, pp. 98-118.

Nadin-Davis S. A., Casey G. A., Wandeler A. (1993). Identification of regional variants of the rabies virus within the canadian province of ontario. J Gen Virol 74 (Pt 5), 829-837.

Nadin-Davis S. A., Turner G., Paul J. P., Madhusudana S. N., Wandeler A. I. (2007). Emergence of arctic-like rabies lineage in india. Emerg Infect Dis 13, 111-116

Nadin-Davis, S.A., Real, L.A., 2011. Molecular phylogenetics of the lyssaviruses--insights from a coalescent approach. Adv. Virus Res. 79, 203-238.

Nanayakkara S., Smith J. S., Rupprecht C. E. (2003). Rabies in sri lanka: Splendid isolation. Emerg Infect Dis 9, 368-371.

Nei M and Kumar S 2000. Molecular Evolution and Phylogenetics, 1st edn, Edited by Anonymous.

Niezgoda Michael, Hanlon Cathleen A, Rupprecht Charles E. Animal Rabies. 163 - 218. In: Jackson Alan C, Wunner William H. Rabies 2002. Academic Press, Amsterdam . Boston . London . New York . Oxford . Paris . San Diego . San Francisco . Singapore . Sydney .

Tokyo.

Nobusawa ET, Aoyama HK, Suzuki Y, Tateno Y, Nakajima K. Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. Virology 1991; 182: 475-85.

Nouvellet, P., Donnely, C., De Nardi, M., Rhodes, C., De Benedictis, P., Citterio, C., Obber, F., Lorenzetto, M., Dalla Pozza, M., Cauchemez, S., Cattoli, G., 2013. Rabies and canine distemper virus epidemics in the red fox population of northern Italy (2006-2010). PloS one accepted.

O'Neill RE, Talon J, Palese P. 1998. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J. Jan 2;17(1):288-96.

OIE 2011. Animal health in the World. Web portal on Avian Influenza. Available at: http://www.oie.int/animal-health-in-the-world/web-portal-on-avian-influenza/about-ai/h5n1-timeline/.

Okazaki K, Yanagawa R, Kida H 1983. Contact infection of mink with 5 subtypes of avian influenza virus. Brief report. Arch Virol 77(2-4):265-269.

Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus AD, Fouchier RA. Global patterns of influenza a virus in wild birds. Science. 2006 Apr 21;312(5772):384-8. Review.

Palese P, Tobita K, Ueda M, Compans RW. 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology. Oct;61(2):397-410.

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52 (4):413-435. doi:10.1007/s13353-011-0057-x.

Park SJ, Moon HJ, Kang BK, Hong M, Na W, Kim JK, Poo H, Park BK, Song DS. 2012. Complete genome sequence of an avian-origin H3N2 canine influenza virus isolated from dogs in South Korea. J Virol. 86(17):9548-9.

Parker J, Rambaut A, Pybus OG. 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. Infect. Genet. Evol. 8(Suppl 3): 239-246.

Pasick J, Handel K, Robinson J, Copps J, Ridd D, Hills K, Kehler H, Cottam-Birt C, Neufeld J, Berhane Y, Czub S. 2005. Intersegmental recombination between the haemagglutinin and matrix genes was responsible for the emergence of a highly pathogenic H7N3 avian influenza virus in British Columbia. J. Gen. Virol. 86(Pt 3):727-731.

Peiris JS, Yu WC, Leung CW, Cheung CY, Ng WF, Nicholls JM, Ng TK, Chan KH, Lai ST, Lim WL, Yuen KY, Guan Y. Re-emergence of fatal human influenza A subtype H5N1 disease.

Lancet. 2004 Feb 21;363(9409):617-9.

Pensaert M, Ottis K, Vandeputte J, Kaplan MM, Bachmann PA. 1981. Evidence for the natural transmission of influenza A virus from wild ducks to swine and its potential importance for man. Bull. World Health Organ. 59:75–78.

Perdue ML, Crawford JM, Garcia M 1998. Occurrence and possible mechanisms of cleavage site insertions in the avian influenza hemagglutinin gene. Proceedings of the 4th International Symposium on Avian Influenza, Athens, Georgia. US Animal Health Association, 182-193.

Peyre M, Samaha H, Makonnen YJ, Saad A, Abd-Elnabi A, Galal S, Ettel T et al. Avian influenza vaccination in Egypt: limitations of the current strategy. J Mol Genet Med 2009a; 3:198-204.

Peyre M, 2009b. Assessment of Highly Pathogenic Avian Influenza vaccination strategy in Egypt. Food and Agricultural Organization report. OSRO/EGY/701/USA.

Pinto LH, Holsinger LJ, Lamb RA. 1992. Influenza virus M2 protein has ion channel activity. Cell. May 1;69(3):517-28.

Pond SL, Frost SD. 2005. A simple hierarchical approach to modeling distributions of substitution rates. Mol. Biol. Evol.;22(2):223-34.

Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics 1998;14:817-818.

ProMED-mail. Avian influenza (54): Mexico (JA) high path H7N3, poultry. ProMED-mail 2012; 16 Sept: 20120916.1295933. Accessed on 26 October 2012. Available at: http://www.promedmail.org.

ProMED-mail. 2013-09-14. Avian influenza (90): Italy (ER) poultry, HPAI H7N7, human, public health. Archive Number: 20130914.1944512. Accessed on 22 November 2013. Available at: http://www.promedmail.org/direct.php?id=20130914.1944512.

ProMED.mail. 2014-01-22. Avian influenza, human (31): China, H7N9, FAO/CHP/WHO update. Archive number: 20140122.2223976. Accessed on 23 January 2014. Available at: http://www.promedmail.org/direct.php?id=20140122.2223976

Qian XY, Alonso-Caplen F, Krug RM. 1994. Two functional domains of the influenza virus NS1 protein are required for regulation of nuclear export of mRNA.J Virol. Apr;68(4):2433-41.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341. doi:10.1186/1471-2164-13-341.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(Suppl 6): 841-842.

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N (2012) Application of next-generation sequencing technologies in virology. J Gen Virol 93 (Pt 9):1853-1868. doi:10.1099/vir.0.043182-0.

Rambaut A. 2002. Sequence alignment editor, version 2.0. Available: http://tree.bio.ed.ac.uk/software/seal/.

Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008; 453(7195):615-9.

Rambaut A. 2008. FigTree v.1.1.2. http://tree.bio.ed.ac.uk/software/figtree/.

Rauw F, Palya V, Van Borm S, Welby S, Tatar-Kis T, Gardin Y, Dorsey KM, Aly MM, Hassan MK, Soliman MA, Lambrecht B, van den Berg T. Further evidence of antigenic drift and protective efficacy afforded by a recombinant HVT-H5 vaccine against challenge with two antigenically divergent Egyptian clade 2.2.1 HPAI H5N1 strains. Vaccine 2011; 29:2590-600.

Rihtaric, D., Hostnik, P., Grom, J., Toplak, I., 2011. Molecular epidemiology of the rabies virus in Slovenia 1994-2010. Vet. Microbiol. 152, 181-186.

Röhm C, Horimoto T, Kawaoka Y, Süss J, Webster RG. 1999. Do hemagglutinin genes of highly pathogenic avian influenza viruses constitute unique phylogenetic lineages? Virology 209(Suppl 2):664-670.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Rui-Hua Z, Hong-Yu C, Ming-Ju X, Kai L, Hua-Lan C, Cun-Lian W, Dong W, Cun-Xin L, Tong X. 2011. Molecular characterization and pathogenicity of swine influenza H9N2 subtype virus A/swine/HeBei/012/2008/(H9N2). Acta Virol.;55(3):219-26.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987 Jul;4(4):406-25. PubMed PMID: 3447015.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74: 5463-5467.

Sartore S, Bonfanti L, Lorenzetto M, Cecchinato M, Marangon S. 2010. The effects of control measures on the economic burden associated with epidemics of avian influenza in Italy. Poultry Science 89 (Suppl 6): 1115-1121. doi:10.3382/ps.2009-00556.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood

phylogenetic analysis using quartets and parallel computing. Bioinformatics. 2002 Mar;18(3):502-4.

Schnell MJ, McGettigan JP, Wirblich C, Papaneri A. The cell biology of rabies virus: using stealth to reach the brain. Nat Rev Microbiol. 2010 Jan;8(1):51-61. doi: 10.1038/nrmicro2260. Epub . Review.

Scholtissek C, Rohde W, Von Hoyningen V, Rott R. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology. 1978 Jun 1;87(1):13-20.

Selleri M, Piralla A, Rozera G, Giombini E, Bartolini B, Abbate I, Campanini G, Rovida F, Dossena L, Capobianchi MR, Baldanti F. 2012. Detection of haemagglutinin D222 polymorphisms in influenza A(H1N1) pdm09-infected patients by ultra-deep pyrosequencing. Clin. Microbiol. Infect. doi:10.1111/j.1469-0691.2012.03984.x.

Senne DA, Panigrahy B, Kawaoka Y, Pearson JE, Süss J, Lipkind M, Kida H, Webster RG. 1996. Survey of the haemagglutinin (HA) cleavage site sequence of H5 and H7 avian influenza viruses: amino acid sequence at the HA cleavage site as a marker of pathogenicity potential. Avian Dis 40(2):425-437.

Shapiro GI, Krug RM. 1988. Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer.J Virol. Jul;62(7):2285-90.

Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol 2006; 23:7-9.

Shih A. C., Hsiao T. C., Ho M. S., Li W. H.. 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc. Natl. Acad. Sci. U. S. A. 104:6283–6288.

Skehel JJ, Bayley PM, Brown EB, Martin SR, Waterfield MD, White JM, et al 1982. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc Natl Acad Sci U S A. Feb;79(4):968-72.

Skehel JJ, Wiley DC. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69: 531-69. Review.

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. Mapping the antigenic and genetic evolution of influenza virus. Science 2004; 305: 371-6.

Smith GJ, Fan XH, Wang J, Li KS, Qin K, Zhang JX, Vijaykrishna D, Cheung CL, Huang K, Rayner JM, Peiris JS, Chen H, Webster RG, Guan Y. Emergence and predominance of an H5N1 influenza variant in China. Proc Natl Acad Sci U S A. 2006; 103:16936-41.

Smith, J.S., Orciari, L.A., Yager, P.A., Seidel, H.D., Warner, C.K., 1992. Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. J. Infect. Dis. 166, 296-307.

Soares Magalhaes RJ, Pfeiffer DU, Otte J. Evaluating the control of HPAIV H5N1 in Vietnam: virus transmission within infected flocks reported before and after vaccination. BMC Vet Res 2010; 6:31.

Songserm T, Amonsin A, Jam-on R, Sae-Heng N, Meemak N, Pariyothorn N, Payungporn S, et al 2006. Avian influenza H5N1 in naturally infected domestic cat. Emerg Infect Dis 12(4):681-683

Spackman E, Senne DA, Davison S, Suarez DL. 2003. Sequence analysis of recent H7 influenza viruses associated with three different outbreaks in commercial poultry in the United States. J. Virol. 77(Suppl 24): 13399-13402.

Stack JC, Murcia PR, Grenfell BT, Wood JL, Holmes EC (2013) Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. Proc Biol Sci 280 (1750):20122173. doi:10.1098/rspb.2012.2173.

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006 Nov 1;22(21):2688-90. (a)

Stech O, Veits J, Weber S, Deckers D, Schröer D, Vahlenkamp TW, Breithaupt A, Teifke J, Mettenleiter TC, Stech J. 2009. Acquisition of a polybasic hemagglutinin cleavage site by a low-pathogenic avian influenza virus is not sufficient for immediate transformation into a highly pathogenic strain. J. Virol. 83(Suppl 11):5864-5868.

Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science 2006; 312:404-10.

Stieneke-Gröber A, Vey M, Angliker H, Shaw E, Thomas G, Roberts C, Klenk HD, Garten W.et al 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. EMBO J 11(7):2407-2414.

Suarez DL, Senne DA, Banks J, Brown IH, Essen SC, Lee CW, Manvell RJ, Mathieu-Benson C, Moreno V, Pedersen JC, Panigrahy B, Rojas H, Spackman E, Alexander DJ. 2004. Recombination resulting in virulence shift in avian influenza outbreak, Chile. Emerging Infect. Dis. 10(Suppl 4):693-699.

Suarez DL, Perdue ML. 1998. Multiple alignment comparison of the non-structural genes of influenza A viruses. Virus Res. 54(Suppl1): 59-69. doi:10.1016/S0168-1702(98)00011-2.

Suarez DL. Avian influenza: our current understanding. Anim Health Res Rev 2010; 11 :19-33. Review.

Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18(Suppl 6): 1001-1013.

Suchard, M.A., Rambaut, A., 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25, 1370-1376.

Sun Y, Sun S, Ma J, Tan Y, Du L, Shen Y, Mu Q, Pu J, Lin D, Liu J. 2012. Identification and characterization of avian-origin H3N2 canine influenza viruses in northern China during 2009-2010. Virology. Oct 11. pii: S0042-6822(12)00489-8. doi: 10.1016/j.virol.2012.09.037.

Suwannakarn K, Amonsin A, Sasipreeyajan J, Kitikoon P, Tantilertcharoen R, Parchariyanon S, et al. Molecular evolution of H5N1 in Thailand between 2004 and 2008. Infect Genet Evol 2009; 9:896-902.

Suwannakhon N, Pookorn S, Sanguansermsri D, Chamnanpood C, Chamnanpood P, Wongvilairat R, Pongcharoen S, Niumsup PR, Kunthalert D, Sanguansermsri P. 2008. Genetic characterization of nonstructural genes of H5N1 avian influenza viruses isolated in Thailand in 2004-2005. Southeast Asian J. Trop. Med. Public Health 39(Suppl 5):837-847.

Swayne DE, Alexander DJ 1994. Confirmation of nephrotropism and nephropathogenicity of three low-pathogenic chicken-origin influenza viruses for chickens. Avian Pathol 23(2):345-352.

Takano R, Nidom CA, Kiso M, Muramoto Y, Yamada S, Sakai-Tagawa Y, et al. Phylogenetic characterization of H5N1 avian influenza viruses isolated in Indonesia from 2003-2007. Virology 2009; 390:13-21.

Taubenberger JK, Kash JC. 2011. Insights on influenza pathogenesis from the grave. Virus Res. 162:2–7.

Terregino C, Toffan A, Cilloni F, Monne I, Bertoli E, Castellanos L et al. Evaluation of the protection induced by avian influenza vaccines containing a 1994 Mexican H5N2 LPAI seed strain against a 2008 Egyptian H5N1 HPAI virus belonging to clade 2.2.1 by means of serological and in vivo tests. Avian Pathol 2010; 39:215-22.

Tejero H, Marín A, Montero F. The relationship between the error catastrophe, survival of the flattest, and natural selection. BMC Evol Biol. 2011 Jan 4;11:2. doi: 10.1186/1471-2148-11-2.

Thomas E. Kienzle, I. Edward Alcamo. Rabies. Deadly diseases and epidemics. Infobase Publishing, 2007.

Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, Chen LM, Johnson A, Tao Y, Dreyfus C, Yu W, McBride R, Carney PJ, Gilbert AT, Chang J, Guo Z, Davis CT, Paulson JC, Stevens J, Rupprecht CE, Holmes EC, Wilson IA, Donis RO. 2013. New world bats harbor diverse influenza A viruses. PLoS Pathog. 9(10):e1003657. doi: 10.1371/journal.ppat.1003657.

Tordo, N. & Kouknetzoff, A. (1993) Onderstepoort J. Vet. Res. 60**,** 263-269.

Tscherne DM, García-Sastre A. Virulence determinants of pandemic influenza viruses. J Clin Invest. 2011 Jan;121(1):6-13. doi: 10.1172/JCI44947. Epub 2011 Jan 4. Review.

Tûmová B 1980. Equine influenza--a segment in influenza virus ecology. Comp Immunol Microbiol Infect Dis 3(1-2):45-59

Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solórzano A, Pappas C, et al 2007. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science. Feb 2;315(5812):655-9.

Vandegrift KJ, Sokolow SH, Daszak P, Kilpatrick AM. Ecology of avian influenza viruses in a changing world. Ann N Y Acad Sci 2010; 1195:113-28. Review.

Vey M, Orlich M, Adler S, Klenk HD, Rott R, Garten W. 1992. Hemagglutinin activation of pathogenic avian influenza viruses of serotype H7 requires the protease recognition motif R-X-K/RR. Virology. May;188(1):408-13.

Vijaykrishna D, Bahl J, Riley S, Duan L, Zhang JX, Chen H, Peiris JS, Smith GJ, Guan Y. 2008. Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. PLoS Pathog. 4(Suppl 9):e1000161.

Wagner R, Wolff T, Herwig A, Pleschka S, Klenk HD. 2000. Interdependence of hemagglutinin glycosylation and neuraminidase as regulators of influenza virus growth: a study by reverse genetics. J. Virol. 74(Suppl 14): 6316-6323.

Walker PG, Cauchemez S, Métras R, Dung Do H, Pfeiffer D, Ghani AC. A Bayesian approach to quantifying the effects of mass poultry vaccination upon the spatial and temporal dynamics of H5N1 in Northern Vietnam. PLoS Comput Biol 2010; 19;6(2): doi 10.1371.

Wasilenko JL, Lee CW, Sarmento L, Spackman E, Kapczynski DR, Suarez DL. 2008. NP, PB1, and PB2 viral genes contribute to altered replication of H5N1 avian influenza viruses in chickens. J. Virol. 82(Suppl 9):4544-4553. doi:10.1128/JVI.02642-07.

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.

Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. 1988. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature. Jun 2;333(6172):426-31.

Whitt MA, Bounocore L, Prehaud C and Rose JK. 1991. Membrane fusion activity: oligomerization, and assembly of rabies virus glycoprotein. Virology 185: 681-688.

Wiley DC, Wilson IA, Skehel JJ. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289(Suppl 5796): 373-378.

Wiley DC, Skehel JJ. 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem.;56:365-94. Review.

Wilgenbusch, J.C., Swofford, D., 2003. Inferring evolutionary trees with PAUP*. Curr. Protoc. Bioinformatics Chapter 6, Unit 6.4.

Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, O'Connor DH, Hughes AL, Neumann G, Kawaoka Y, Friedrich TC (2013) Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. Nat Commun 4:2636. doi:10.1038/ncomms3636.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 1981; 289: 366-73.

Wise HM, Foeglein A, Sun J, Dalton RM, Patel S, Howard W, Anderson EC, Barclay WS, Digard P. A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. J Virol. 2009 Aug;83(16):8021-31. doi: 10.1128/JVI.00826-09.

Wood GW, McCauley JW, Bashiruddin JB, Alexander DJ 1993. Deduced amino acid sequences at the haemagglutinin cleavage site of avian influenza A viruses of H5 and H7 subtypes. Arch Virol 130(1-2):209-217.

World Health Organization (2005) WHO Expert Consultation on rabies. World Health Organ Tech Rep Ser 931: 1–88, back cover.

World Health Organization, 2013. Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2013. Available at: http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/ (Accessed on 22nd November 2013).

World Health Organization. Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO (World Health Organization). 2011. Available from http://www.who.int/csr/disease/avian_influenza/country/cases_table_2011_08_09/en/index.html Accessed September 22, 2011].

World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization H5N1 Evolution Working Group. Toward a unified nomenclature system for highly pathogenic avian influenza virus (H5N1). Emerg Infect Dis. 2008. Available from http://www.cdc.gov/EID/content/14/7/e1.htm.

World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization H5N1 Evolution Working Group. Letter to the Editor: Continuing progress towards a unified nomenclature for the highly pathogenic H5N1 avian influenza viruses: divergence of clade 2.2 viruses. Influenza and Other Respiratory Viruses. 2009; 3:59-62. Available from http://www3.interscience.wiley.com/cgi-bin/fulltext/122211969/HTMLSTART.

Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP (2011) Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. J Virol 85 (5):2266-2275. doi:10.1128/JVI.01396-10.

Wu W. L., et al. 2008. Antigenic profile of avian H5N1 viruses in Asia from 2002 to 2007. J. Virol. 82:1798–1807.

Wu X., Franka R., Velasco-Villa A., Rupprecht C. E. (2007). Are all lyssavirus genes equal for phylogenetic analyses? Virus Res 129, 91-103.

Xu X, Subbarao K, Cox NJ and Guo Y. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. Virology 1996; 224:175–183.

Zamarin D, Ortigoza MB, Palese P 2006 Influenza A virus PB1-F2 protein contributes to viral pathogenesis in mice J Virol Aug;80(16):7976-83

Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE. 2009. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. J. Virol.;83(19):10309-13. doi: 10.1128/JVI.01109-09.

ANNEX A

Table A1. Epidemiological information of the 113 fox samples selected as representative of the Italian 2008-2011 epidemic. The virus identification reveals: the identification number, the specie of origin, the month and year of detection, the circumstances of carcass analysis (D: found death; K: killed within the framework of active surveillance) and the province of origin.

Virus identification	Collection date	Municipality
RS08-1981/fox/Oct08/K/Udine	10/10/2008	Tolmezzo
RS08-2088/fox/Oct08/D/Udine	19/10/2008	Venzone
RS08-2396/fox/Nov08/D/Udine	17/11/2008	Resia
RS08-2445/fox/Nov08/K/Udine	20/11/2008	Resia
RS08-2624/fox/Dec08/K/Udine	17/12/2008	Resia
RS08-2628/fox/Dec08/D/Udine	20/12/2008	Lusevera
RS09-25/fox/Dec08/D/Udine	30/12/2008	Resia
RS09-556/fox/Mar09/D/Udine	19/3/2009	Resia
RS09-578/fox/Mar09/K/Udine	24/3/2009	Tolmezzo
RS09-899/fox/May09/D/Udine	3/5/2009	Buia
RS09-1346/fox/Jun09/D/Udine	11/6/2009	Majano
RS09-1452/fox/Jun09/D/Udine	22/6/2009	Artegna
RS09-1964/fox/Jul09/K/Udine	25/7/2009	Majano
RS09-2051/fox/Aug09/D/Udine	7/8/2009	San Daniele del Friuli
RS09-2115/fox/Aug09/K/Udine	18/8/2009	San Daniele del Friuli
RS09-2116/fox/Aug09/D/Udine	18/8/2009	Majano
RS09-2369/fox/Sep09/D/Udine	10/9/2009	San Daniele del Friuli
RS09-2420/fox/Sep09/D/Udine	18/9/2009	Gemona del Friuli
RS09-2417/fox/Sep09/K/Udine	20/9/2009	Mereto di Tomba
RS09-2439/fox/Sep09/D/Udine	23/9/2009	San Daniele del Friuli
RS09-2172/fox/Aug09/D/Udine	25/8/2009	Majano
RS09-2472/fox/Sep09/K/Udine	25/9/2009	Majano
RS09-2474/fox/Sep09/K/Udine	26/9/2009	Mortegliano
RS09-2536/fox/Oct09/D/Udine	5/10/2009	Ragogna
RS09-2617/fox/Oct09/K/Pordenone	6/10/2009	Travesio
RS09-2616/fox/Oct09/D/Pordenone	8/10/2009	Travesio
RS09-3176/fox/Nov09/Pordenone	11/11/2009	Travesio
RS09-3234/fox/Oct09/D/Belluno	23/10/2009	Forno di Zoldo
RS09-2906/fox/Oct09/D/Pordenone	30/10/2009	Travesio

RS09-2904/fox/Oct09/K/Pordenone	31/10/2009	CastelNovo del Friuli
RS09-3042/fox/Nov09/D/Pordenone	5/11/2009	Travesio
RS09-3270/fox/Nov09/K/Belluno	20/11/2009	Pieve di Cadore
RS09-3269/fox/Nov09/D/Belluno	21/11/2009	Valle di Cadore
RS09-3310/fox/Nov09/D/Udine	22/11/2009	Treppo Grande
RS09-3342/fox/Nov09/D/Belluno	23/11/2009	San Pietro di Cadore
RS09-3425/fox/Nov09/D/Belluno	25/11/2009	Vodo di Cadore
RS09-3509/fox/Nov09/D/Belluno	30/11/2009	Sovramonte
RS09-3561/fox/Dec09/K/Belluno	2/12/2009	Forno di Zoldo
RS09-3590/fox/Dec09/D/Belluno	6/12/2009	Sovramonte
RS09-3592/fox/Dec09/D/Belluno	7/12/2009	Pieve di Cadore
RS09-3766/fox/Dec09/K/Pordenone	9/12/2009	CastelNovo del Friuli
RS09-3705/fox/Dec09/D/Belluno	9/12/2009	San Vito di Cadore
RS09-3694/fox/Dec09/D/Belluno	10/12/2009	Pieve di Cadore
RS09-3845/fox/Dec09/D/Udine	10/12/2009	Prato Carnico
RS09-3796/fox/Dec09/K/Belluno	13/12/2009	San Nicolò di Comelico
RS09-3806/fox/Dec09/D/Belluno	15/12/2009	Longarone
RS09-3811/fox/Dec09/D/Belluno	16/12/2009	Sappada
RS10-46/fox/Dec09/K/Belluno	30/12/2009	San Pietro di Cadore
RS10-25/fox/Jan10/D/Belluno	2/1/2010	Ospitale di Cadore
RS10-44/fox/Jan10/K/Belluno	3/1/2010	Longarone
RS10-52/fox/Jan10/D/Belluno	3/1/2010	Longarone
RS10-65/fox/Jan10/K/Pordenone	5/1/2010	Travesio
RS10-29/fox/Jan10/D/Belluno	5/1/2010	Lorenzago di Cadore
RS10-104/fox/Jan10/D/Belluno	6/1/2010	Vigo di Cadore
RS10-93/fox/Jan10/K/Belluno	6/1/2010	Domegge di Cadore
RS10-94/fox/Jan10/D/Belluno	8/1/2010	Lozzo di Cadore
RS10-213/fox/Jan10/D/Belluno	13/1/2010	San Pietro di Cadore
RS10-215/fox/Jan10/K/Belluno	17/1/2010	Pedavena
RS10-237/fox/Jan10/D/Belluno	17/1/2010	Lorenzago di Cadore
RS10-236/fox/Jan10/D/Belluno	18/1/2010	Valle di Cadore
RS10-394/fox/Jan10/D/Belluno	21/1/2010	Vigo di Cadore
RS10-396/fox/Jan10/D/Belluno	23/1/2010	Danta
RS10-380/fox/Jan10/D/Belluno	24/1/2010	Longarone
RS10-484/fox/Jan10/K/Belluno	25/1/2010	Santo Stefano di Cadore
RS10-476/fox/Jan10/K/Belluno	25/1/2010	Lozzo di Cadore
RS10-549/fox/Jan10/D/Belluno	28/1/2010	Seren del Grappa
RS10-843/fox/Feb10/D/Trento	5/2/2010	Mazzin

RS10-704/fox/Feb10/D/Belluno	7/2/2010	Auronzo di Cadore
RS10-1126/fox/Feb10/D/Udine	9/2/2010	Raveo
RS10-1072/fox/Feb10/D/Belluno	22/2/2010	Sovramonte
RS10-1351/fox/Feb10/D/Udine	28/2/2010	Sauris
RS10-1331/fox/Mar10/D/Trento	1/3/2010	Mezzano
RS10-1709/fox/Mar10/D/Udine	9/3/2010	Forni di Sopra
RS10-1617/fox/Mar10/D/Belluno	10/3/2010	Alleghe
RS10-1719/fox/Mar10/D/Belluno	13/3/2010	Agordo
RS10-1618/fox/Mar10/D/Belluno	14/3/2010	Rocca Pietore
RS10-1735/fox/Mar10/Belluno	15/3/2010	Selva di Cadore
RS10-1934/fox/Mar10/D/Pordenone	18/3/2010	CastelNovo del Friuli
RS10-1832/fox/Mar10/D/Belluno	19/3/2010	Rocca Pietore
RS10-1919/fox/Mar10/D/Belluno	21/3/2010	Cortina d'Ampezzo
RS10-1954/fox/Mar10/K/Udine	23/3/2010	Socchieve
RS10-2009/fox/Mar10/D/Belluno	24/3/2010	Vigo di Cadore
RS10-2008/fox/Mar10/D/Belluno	25/3/2010	Cortina d'Ampezzo
RS10-2052/fox/Mar10/D/Belluno	29/3/2010	Livinallongo del Col di Lana
RS10-2048/fox/Mar10/K/Belluno	29/3/2010	Borca di Cadore
RS10-2121/fox/Apr10/K/Pordenone	1/4/2010	Clauzetto
RS10-2151/fox/Apr10/D/Belluno	3/4/2010	Borca di Cadore
RS10-2142/fox/Apr10/D/Belluno	4/4/2010	Cortina d'Ampezzo
RS10-2137/fox/Apr10/K/Belluno	4/4/2010	Cortina d'Ampezzo
RS10-2269/fox/Apr10/D/Belluno	10/4/2010	Santo Stefano di Cadore
RS10-2469/fox/Apr10/D/Belluno	11/4/2010	Livinallongo del Col di Lana
RS10-2511/fox/Apr10/D/Belluno	16/4/2010	Livinallongo del Col di Lana
RS10-2803/fox/Apr10/D/Bolzano	19/4/2010	Dobbiaco
RS10-2520/fox/Apr10/D/Belluno	19/4/2010	Livinallongo del Col di Lana
RS10-2711/fox/Apr10/D/Udine	24/4/2010	Forni Sopra
RS10-2664/fox/Apr10/D/Belluno	26/4/2010	Rocca Pietore
RS10-2804/fox/May10/D/Bolzano	1/5/2010	Badia
RS10-2782/fox/May10/K/Belluno	3/5/2010	Livinallongo del Col di Lana
RS10-2811/fox/May10/K/Udine	4/5/2010	Sauris
RS10-2892/fox/May10/D/Belluno	10/5/2010	Rocca Pietore
RS10-2979/fox/May10/D/Belluno	14/5/2010	Livinallongo del Col di Lana
RS10-3066/fox/May10/D/Belluno	21/5/2010	Rocca Pietore
RS10-3200/fox/May10/D/Bolzano	23/5/2010	Badia
RS10-3098/fox/May10/Belluno	23/5/2010	Livinallongo del Col di Lana
RS10-3263/fox/Jun10/D/Belluno	4/6/2010	Colle Santa Lucia

RS10-3313/fox/Jun10/D/Belluno	8/6/2010	Livinallongo del Col di Lana
RS10-3417/fox/Jun10/D/Trento	8/6/2010	Campitello di Fassa
RS10-3410/fox/Jun10/K/Bolzano	9/6/2010	Badia
RS10-3636/fox/Jun10/D/Trento	12/6/2010	Transacqua
RS10-3570/fox/Jun10/D/Belluno	21/6/2010	Livinallongo del Col di Lana
RS10-3888/fox/Jul10/D/Trento	3/7/2010	Campitello di Fassa
RS10-5066/fox/Aug10/K/Belluno	28/8/2010	Zoldo Alto
RS11-819/fox/Feb11/D/Belluno	14/2/2011	Pedavena

Table A2. List of primers used for amplification and sequencing of the Italian RABVs. Nucleotide positions are numbered according to the Pasteur virus sequence (GenBank accession number M13215).

Primer name	Length (nt)	Sequence (5'-3')	Gene	Position	References
RabForPyro	17	AACACYYCTACAATGGA	N	59-75	De Benetictis et al., 2011
RabRevPyro-3	22	TCCAATTNGCACACATTTTGTG	N	662-641	De Benetictis et al., 2011
RabRevPyro-2	22	TCCARTTAGCGCACATYTTATG	N	662-641	De Benetictis et al., 2011
RabRevPyro-2	22	TCCAGTTGGCRCACATCTTRTG	N	662-641	De Benetictis et al., 2011
M 220 For	20	TGGTGTATCAACATGRAYTC	M	3000-3019	De Benetictis et al., 2011
Rab-Gly-R1	20	CGRAGCCARTGRTAGTCAGG	G	3742-3723	This study
Rab-Gly-F1	21	ATGKTTCCTCWGGYTCTYTTG	G	3318-3338	This study
G780 R	20	ACCCATGTYCCRTCCATAAG	G	4096-4077	Delmas et al., 2008
Rab-Gly-F3	21	CAATAGTAGAGGGAAGAGAGC	G	3953-3973	This study
Rab-Gly-R3	20	CCCYTTKGAGGGGATGATYT	G	4403-4384	This study
G780 F	20	CTTATGGAYGGRACATGGGT	G	4077-4096	mod. from Delmas et al., 2008
Rab-Gly-R4	20	TCAACMCCTRAGAYYTGTTT	G	4666-4647	This study
Rab-Gly-F5	20	CCTGAYGGYCAYGTTCTAAT	G	4473-4492	This study
RV-F5	21	GGGGGTGAGACTAGACTGTGA	G	4872-4892	This study
RV-G-REV	21	TGAAAGCACCGTTAGTCACTG	G-L	5119-5099	This study
L1 Rev	21	GAGTTNAGRTTGTARTCAGAG	L	5536-5516	Delmas et al., 2008

Table A3. Significantly non-zero migration rate using Bayes factor test. BF=Bayes Factor.

	Bayes factor N gene	Bayes factor N+G gene
Italy-area B to Italy-area C	BF=38.36	BF=11.99
Italy-area A to Croatia	BF=17.99	-
Croatia to Slovenia	BF=8.24	BF=9.33
Slovenia to Bosnia	BF=7.51	BF=55.37
Italy-area B to Croatia	BF=6.86	-
Slovenia to Italy-area A	BF=5.31	-

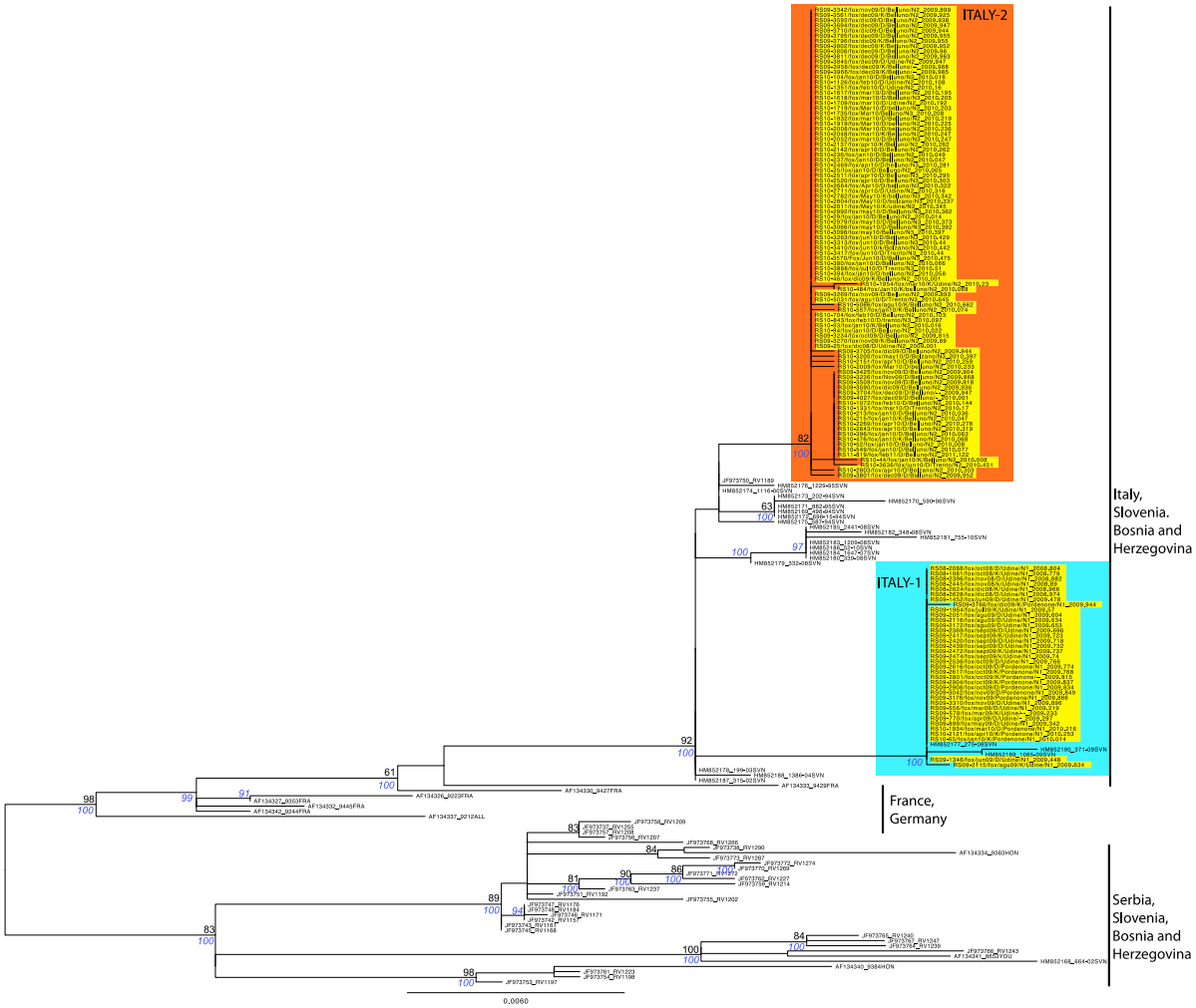


Figure A1. ML phylogenetic tree of sequences estimated from the G gene of RABVs collected in Italy 2008-2011, combined with those from neighbouring countries. The sequence identification includes: sample identification number, species of origin, month and year of collection, province where the carcass was collected. The numbers at branch points represent bootstrap values (>60, in black) or posterior probabilities (>80, in blue). The Italian samples are highlighted in yellow. Two different genetic groups are indicated in different colours: blue – Italy-1 and orange – Italy-2.

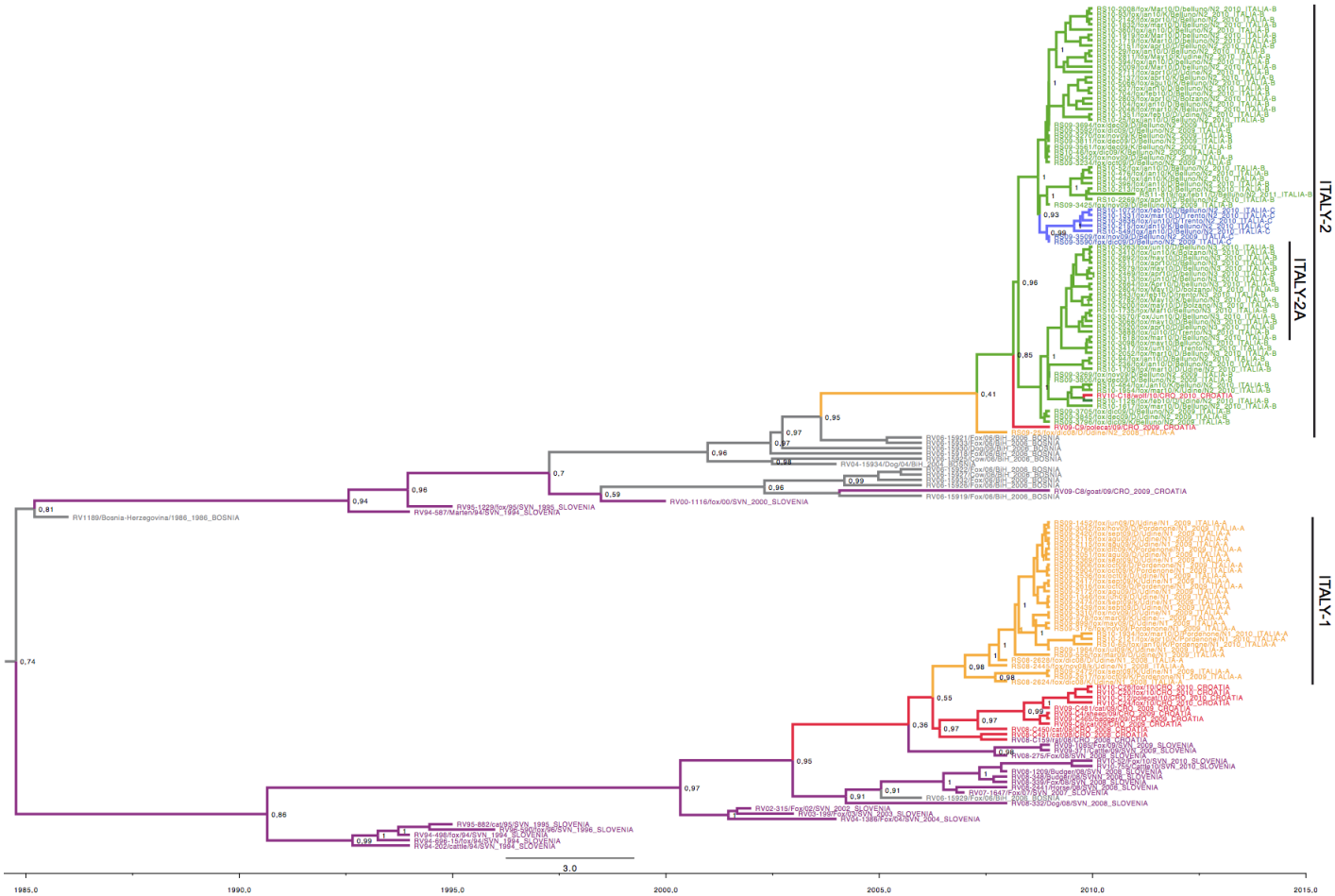


Figure A2. MCC tree inferred for the concatenated N and G sequences from Italy, BiH, Croatia and Slovenia. Sequences are coloured according to their geographic origin. Yellow: Italy A; green: Italy B; blue: Italy C; grey: BiH; red: Croatia; purple: Slovenia. The numbers at branch points represent the probability of spatial states.

ANNEX B

Table B1. Evolutionary profiles of H5N1 HPAI viruses: comparisons between strict and relaxed (uncorrelated lognormal) molecular clocks using the Bayesian skyride coalescent prior.

	Best fit clock model	**Evolutionary rates (sub/site/year) x 10^{-3}**	
		Uncorrelated log. Bayesian Skyride	**Strict clock Bayesian Skyride**
Egypt	Uncorrelated lognormal	5.36 95% HPD, 4.81 - 5.96	5.19 95% HPD, 4.66-5.71
Egypt –subclade A	Uncorrelated lognormal	4.07 95% HPD, 3.23-4.91	4.08 95% HPD, 3.31-4.95
Egypt –subclade B	Uncorrelated lognormal	8.87 95% HPD, 7.00-10.72	6.90 95% HPD, 5.52-8.32
Indonesia	Uncorrelated lognormal	6.13 95% HPD, 5.14-7.06	5.41 95% HPD, 4.62-6.21
Nigeria	Strict clock	5.20 95% HPD, 4.11-6.27	5.15 95% HPD, 4.12-6.17
Turkey	Strict clock	4.04 95% HPD, 2.16-6.04	4.67 95% HPD, 2.78-6.57
Thailand	Uncorrelated lognormal	2.52 95% HPD, 1.93-3.19	2.49 95% HPD, 1.99-3.01

Table B2. Analysis of posterior distributions of H5N1 nucleotide substitution rates among countries. The values represent the probability that the mean rate of a country is higher than the mean rate of another country.

Probability	Egypt A	Egypt B	Indonesia	Nigeria	Turkey	Thailand
Egypt	> 0.993	< 0.999	< 0.913	> 0.600	> 0.8923	>1
Egypt A		< 0.999	< 0.999	< 0.948	> 0.511	> 0.998
Egypt B			> 0.9952517	> 0.999	> 0.999	> 1
Indonesia				> 0.901	> 0.968	> 1
Nigeria					> 0.842	> 0.999
Turkey						> 0.921

ANNEX C

Table C1. Epidemiological information (collection date, city and province of collection, and pathotype) of the 144 Italian H7N1 viruses analysed in chapter 5.

Virus	Collection date (d/m/y)	City of collection	Province	Pathotype	Comments
A/turkey/Italy/12rs206-1/1999	15/3/1999	Erbè	Verona	LPAI	This study (HA)
A/turkey/Italy/12rs206-18/1999	19/3/1999	Cazzago San Martino	Brescia	LPAI	This study (HA)
A/turkey/Italy/977/1999	26/03/1999	Erbè	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/12rs206-3/1999	31/3/1999	Roverbella	Mantova	LPAI	This study (HA)
A/turkey/Italy12rs206-2/1999	1/4/1999	Villimpenta	Mantova	LPAI	This study (HA)
A/turkey/Italy/1067/1999	2/4/1999	Rocca	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/1081/1999	02/04/1999	Zimella	Verona	LPAI	Segment 8 - Dundon et al., 2006, other segments - this study
A/chicken/Italy/1082/1999	02/04/1999	Zimella	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/1083/1999	02/04/1999	Isola Rizza	Verona	LPAI	This study (complete genome)
A/turkey/Italy/1085/1999	02/04/1999	Isola Rizza	Verona	LPAI	This study (complete genome)
A/turkey/Italy/1086/1999	02/04/1999	Nogara	Verona	LPAI	Segment 8 - Dundon et al., 2006, other segments - this study
A/turkey/Italy/1265/1999	7/4/1999	Isola della Scala	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/1174/1999	7/4/1999	Porto Tolle	Rovigo	LPAI	Banks et al., 2001
A/chicken/Italy/12rs206-4/1999	8/4/1999	Poncarale	Brescia	LPAI	This study (complete genome)
A/chicken/Italy/1279/1999	9/4/1999	Nogara	Verona	LPAI	Iqbal et al., 2012
A/turkey/Italy/12rs206-12/1999	9/4/1999	Roverbella	Mantova	LPAI	This study (HA)
A/chicken/Italy/12rs206-5/1999	9/4/1999	Lonato	Brescia	LPAI	This study (HA)
A/turkey/Italy/12rs206-7/1999	9/4/1999	Poncarale	Brescia	LPAI	This study (HA)
A/turkey/Italy/12rs206-6/1999	10/4/1999	Isorella	Brescia	LPAI	This study (HA)
A/turkey/Italy/12rs206-11/1999	12/4/1999	Marcaria	Mantova	LPAI	This study (HA)
A/turkey/Italy/12rs206-13/1999	12/4/1999	Isorella	Brescia	LPAI	This study (HA)
A/turkey/Italy/12rs206-8/1999	12/4/1999	Desenzano del Garda	Brescia	LPAI	This study (HA)

A/turkey/Italy/12rs206-10/1999	12/4/1999	Marcaria	Mantova	LPAI	This study (HA)
A/turkey/Italy/4857/1999	14/4/1999	Trevenzuolo	Verona	LPAI	Banks et al., 2001
A/chicken/Italy/12rs206-16/1999	16/4/1999	Casteldario	Mantova	LPAI	This study (HA)
A/turkey/Italy/1555/1999	20/04/1999	S.Martino Buon Albergo	Verona	LPAI	This study (complete genome)
A/turkey/Italy/12rs206-14/1999	21/4/1999	Paderno Franciacorta	Brescia	LPAI	This study (HA)
A/turkey/Italy/1744/1999	26/4/1999	Isola della Scala	Verona	LPAI	This study (complete genome)
A/turkey/Italy/12rs206-17/1999	26/4/1999	Ghedi	Brescia	LPAI	This study (HA)
A/turkey/Italy/4073/1999	28/4/1999	Paderno	Brescia	LPAI	Banks et al., 2001
A/chicken/Italy/12rs206-19/1999	5/5/1999	Ospitaletto	Brescia	LPAI	This study (HA)
A/turkey/Italy/12rs206-20/1999	10/5/1999	Ospitaletto	Brescia	LPAI	This study (HA)
A/turkey/Italy/4499/1999	10/5/1999	Cazzago S. Martino	Brescia	LPAI	Banks et al., 2001
A/turkey/Italy/4603/1999	10/5/1999	Ghedi	Brescia	LPAI	Banks et al., 2001
A/turkey/Italy/2379/1999	4/6/1999	Sommacampagna	Verona	LPAI	Banks et al., 2001
A/turkey/Italy/2505/1999	6/7/1999	Palù	Verona	LPAI	Banks et al., 2001
A/turkey/Italy/2715/1999	2/8/1999	Isola della Scala	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/2716/1999	2/8/1999	Isorella	Brescia	LPAI	Segment 8 - Dundon et al., 2006, other segments - this study
A/turkey/Italy/2732/1999	3/8/1999	Isola della Scala	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/12rs206-22/1999	9/8/1999	Erbè	Verona	LPAI	This study (HA)
A/turkey/Italy/3185/1999	2/9/1999	Erbè	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/3283/1999	10/9/1999	Roverchiara	Verona	LPAI	
A/turkey/Italy/3488/1999	16/9/1999	Erbè	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/3489/1999	20/9/1999	Sorgà	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/3560/1999	27/9/1999	Roverchiara	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/3675/1999	4/10/1999	Trevenzuolo	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4042/1999	23/10/1999	Nogara	Verona	LPAI	Iqbal et al., 2012
A/turkey/Italy/4294/1999	22/11/1999	Nogarole Rocca	Verona	LPAI	The NIAID Influenza Genome Sequencing Project

A/turkey/Italy/4295/1999	22/11/1999	Verona	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4301/1999	22/11/1999	Verona	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4482/1999	2/12/1999	Villafranca	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4580/1999	10/12/1999	Salizzole	Verona	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4617/1999	12/12/1999	Salizzole	Verona	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4644/1999	13/12/1999	Villafranca	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4620/1999	13/12/1999	Sanguinetto	Verona	LPAI	Banks et al., 2001
A/turkey/Italy/4640/1999	15/12/1999	Sanguinetto	Verona	HPAI	Banks et al., 2001
A/chicken/Italy/4746/1999	17/12/1999	Guidizzolo	Mantova	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4708/1999	18/12/1999	Concamarise	Verona	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4725/1999	20/12/1999	S. Martino Buonalbergo	Verona	LPAI	Banks et al., 2001
A/chicken/Italy/4789/1999	22/12/1999	Teglio Veneto	Venezia	HPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/4845/1999	24/12/1999	Fiume Veneto	Pordenone	HPAI	This study (complete genome)
A/chicken/Italy/4911/1999	27/12/1999	Lendinara	Rovigo	HPAI	This study (complete genome)
A/turkey/Italy/4906/1999	28/12/1999	/	Brescia	HPAI	This study (complete genome)
A/quail/Italy/4992/1999	29/12/1999	Padova	Padova	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/5005/1999	30/12/1999	Lonigo	Vicenza	HPAI	This study (complete genome)
A/chicken/Italy/5079/1999	30/12/1999	Roverbella	Mantova	HPAI	This study (complete genome)
A/chicken/Italy/5093/1999	30/12/1999	Marmirolo	Mantova	HPAI	Rigoni et al., 2007
A/turkey/Italy/5033/1999	31/12/1999	Pernumia	Padova	HPAI	This study (complete genome)
A/chicken/Italy/5074/1999	31/12/1999	Mazzurega Fumane	Verona	HPAI	This study (complete genome)
A/chicken/Italy/5086/1999	31/12/1999	Castiglione delle Stiviere	Mantova	HPAI	This study (complete genome)
A/turkey/Italy/5032/1999	31/12/1999	Pernumia	Padova	HPAI	Banks et al., 2001
A/chicken/Italy/18/2000	3/1/2000	Valeggio sul Mincio	Verona	HPAI	This study (complete genome)
A/turkey/Italy/16/2000	4/1/2000	Montagnana	Padova	HPAI	This study (complete genome)
A/chicken/Italy/78/2000	4/1/2000	Valeggio sul Mincio	Verona	HPAI	This study (complete genome)
A/chicken/Italy/161/2000	5/1/2000	Gottolengo	Brescia	HPAI	This study (complete genome)

A/turkey/Italy/117/2000	7/1/2000	Pressana	Verona	LPAI	This study (complete genome)
A/turkey/Italy/140/2000	7/1/2000	Morsano al Tagliamento	Pordenone	HPAI	This study (complete genome)
A/turkey/Italy/148/2000	7/1/2000	Goito	Mantova	HPAI	This study (complete genome)
A/guineafowl/Italy/155/2000	7/1/2000	Cavriana	Mantova	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/195/2000	11/1/2000	Erbè	Verona	HPAI	This study (complete genome)
A/duck/Italy/551/2000	12/1/2000	Palosco	Bergamo	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/604/2000	13/1/2000	Borghi	Bologna	LPAI	This study (complete genome)
A/turkey/Italy/421/2000	14/1/2000	Albettone	Vicenza	HPAI	This study (HA, NS, M)
A/chicken/Italy/557/2000	14/1/2000	Roncoferraro	Mantova	HPAI	This study (complete genome)
A/guineafowl/Italy/564/2000	17/1/2000	Ceresara	Mantova	HPAI	This study (complete genome)
A/quail/Italy/396/2000	19/1/2000	Orgiano	Vicenza	HPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/501/2000	20/1/2000	San Donà di Piave	Venezia	HPAI	This study (complete genome)
A/turkey/Italy/503/2000	20/1/2000	Portugruaro	Venezia	HPAI	This study (complete genome)
A/chicken/Italy/522/2000	20/1/2000	Vighizzolo d'Este	Padova	HPAI	This study (complete genome)
A/turkey/Italy/577/2000	20/1/2000	Manerbio	Brescia	HPAI	This study (complete genome)
A/chicken/Italy/662/2000	20/1/2000	Veronella	Verona	HPAI	This study (complete genome)
A/turkey/Italy/589/2000	21/1/2000	Este	Padova	LPAI	This study (complete genome)
A/turkey/Italy/720/2000	24/1/2000	Pressana	Verona	HPAI	This study (complete genome)
A/chicken/Italy/910/2000	24/1/2000	Pavone del Mella	Brescia	HPAI	This study (complete genome)
A/chicken/Italy/914/2000	26/1/2000	Gambara	Brescia	HPAI	This study (HA)
A/guineafowl/Italy/918/2000	26/1/2000	Medole	Mantova	HPAI	This study (complete genome)
A/chicken/Italy/797/2000	31/1/2000	Noviligure	Torino	HPAI	This study (complete genome)
A/turkey/Italy/1713/2000	1/2/2000	Telgate	Bergamo	HPAI	This study (complete genome)
A/chicken/Italy/925/2000	2/2/2000	Martinengo	Bergamo	HPAI	This study (complete genome)
A/chicken/Italy/933/2000	2/2/2000	Ostiano	Cremona	HPAI	This study (complete genome)
A/chicken/Italy/1715/2000	3/2/2000	Antegnate	Bergamo	HPAI	This study (HA)
A/ostrich/Italy/984/2000	4/2/2000	Villafranca	Verona	HPAI	Rigoni et al., 2007

A/ostrich/Italy/1086/2000	4/2/2000	Mozzecane	Verona	HPAI	Segment 8 - Dundon et al., 2006, other segments - this study
A/chicken/Italy/1302/2000	4/2/2000	Castel San Giovanni	Piacenza	HPAI	This study (complete genome)
A/turkey/Italy/1009/2000	7/2/2000	Castelgomberto	Vicenza	HPAI	This study (complete genome)
A/turkey/Italy/1084/2000	7/2/2000	Cologna Veneta	Verona	HPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/1730/2000	7/2/2000	/	Como	HPAI	This study (complete genome)
A/ostrich/Italy/1038/2000	8/2/2000	Villafranca	Verona	HPAI	The NIAID Influenza Genome Sequencing Project
A/quail/Italy/1764/2000	16/2/2000	Gottolengo	Brescia	HPAI	This study (HA)
A/goose/Italy/1310/2000	18/2/2000	Padova	Padova	HPAI	This study (complete genome)
A/guineafowl/Italy/1766/2000	18/2/2000	Manerbio	Brescia	HPAI	This study (complete genome)
A/chicken/Italy/1651/2000	22/2/2000	Fusignano	Ravenna	HPAI	This study (complete genome)
A/chicken/Italy/1782/2000	24/2/2000	Palazzolo sull'oglio	Brescia	HPAI	This study (complete genome)
A/chicken/Italy/1786/2000	28/2/2000	Mairago	Lodi	HPAI	This study (complete genome)
A/chicken/Italy/1592/2000	1/3/2000	Moimacco	Udine	HPAI	This study (complete genome)
A/chicken/Italy/1870/2000	3/3/2000	San Bartolomeo in Bosco	Ferrara	HPAI	This study (complete genome)
A/chicken/Italy/1875/2000	3/3/2000	Stroncone	Terni	HPAI	This study (complete genome)
A/guineafowl/Italy/1847/2000	6/3/2000	Rivarolo Mantovano	Mantova	HPAI	This study (complete genome)
A/pekinduck/Italy/1848/2000	6/3/2000	Adro	Brescia	HPAI	The NIAID Influenza Genome Sequencing Project
A/chicken/Italy/2021/2000	6/3/2000	Almese	Torino	HPAI	This study (complete genome)
A/turkey/Italy/2023/2000	17/3/2000	Castegnero	Vicenza	HPAI	This study (complete genome)
A/turkey/italy/2330/2000	20/3/2000	Dello	Brescia	HPAI	This study (complete genome)
A/ostrich/Italy/2332/2000	22/3/2000	Inzago	Milano	HPAI	Rigoni et al., 2007
A/chicken/Italy/2335/2000	30/3/2000	Pavone del Mella	Brescia	HPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/2984/2000	10/4/2000	Prevalle	Brescia	HPAI	The NIAID Influenza Genome Sequencing Project
A/hawk/Italy/2985/2000	26/4/2000	Milano	Milano	HPAI	This study (complete genome)
A/turkey/Italy/4426/2000	14/8/2000	Roverchiara	Verona	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/4430/2000	16/8/2000	Roverchiara	Verona	LPAI	This study (complete genome)
A/turkey/Italy/4455/2000	17/8/2000	Isola Rizza	Verona	LPAI	This study (complete genome)

A/turkey/Italy/4944/2000	11/9/2000	Isola Rizza	Verona	LPAI	This study (complete genome)
A/turkey/Italy/5011/2000	13/9/2000	Roverchiara	Verona	LPAI	This study (complete genome)
A/turkey/Italy/5222/2000	27/9/2000	Nogara	Verona	LPAI	This study (complete genome)
A/turkey/Italy/5302/2000	28/9/2000	Bovolone	Verona	LPAI	This study (complete genome)
A/turkey/Italy/5662/2000	20/10/2000	Trevenzuolo	Verona	LPAI	This study (complete genome)
A/turkey/Italy/6960/2000	22/12/2000	Cologna Veneta	Verona	LPAI	This study (complete genome)
A/turkey/Italy/7001/2000	27/12/2000	Ospedaletto Euganeo	Padova	LPAI	This study (HA)
A/turkey/Italy/7002/2000	28/12/2000	Ospedaletto Euganeo	Padova	LPAI	This study (complete genome)
A/turkey/Italy/32/2001	5/1/2001	Ponso	Padova	LPAI	This study (complete genome)
A/turkey/Italy/133/2001	9/1/2001	Ponso	Padova	LPAI	This study (complete genome)
A/turkey/Italy/138/2001	9/1/2001	Ospedaletto Euganeo	Padova	LPAI	This study (complete genome)
A/turkey/Italy/223/2001	10/1/2001	Este	Padova	LPAI	This study (complete genome)
A/turkey/Italy/284/2001	13/1/2001	Cologna Veneta	Verona	LPAI	This study (complete genome)
A/chicken/Italy/322/2001	15/1/2001	Ospedaletto Euganeo	Padova	LPAI	The NIAID Influenza Genome Sequencing Project
A/turkey/Italy/1351/2001	15/2/2001	Ospedaletto Euganeo	Padova	LPAI	The NIAID Influenza Genome Sequencing Project

Table C2. Values of Association Index (AI), Parsimony Score (PS) and Monophyletic Clade (MC) obtained using BaTS program from the analysis of all the Italian H7N1 viruses.

Dataset	Statistic	Observed mean	Lower 95% CI	Upper 95% CI	Null mean	Lower 95% CI	Upper 95% CI	p- value
H7N1	AI	6,350	5,325	7,328	10,697	9,739	11,706	0
	PS	52,711	50	56	74,056	70,378	77,059	0
	MC(VR)	11,162	11	12	3,246	2,456	4,994	0,009
	MC(MN)	2,058	2	3	1,416	1,038	2,006	0,110
	MC(BS)	5,103	4	7	1,959	1,402	2,559	0,009
	MC(PN-UD)	1,006	1	1	1,020	1	1,105	1
	MC(FE-RA-RO)	1,011	1	1	1,033	1	1,094	1
	MC(BG-CO)	1,304	1	2	1,014	1	1,089	1
	MC(MI-LO)	1	1	1	1,014	1	1,041	1
	MC(VE-PD-VI)	8	8	8	1,744	1,249	2,252	0,009
LPAI H7N1	AI	1,796	1,154	2,428	4,751	4,078	5,490	0
	PS	16,331	14	18	29,817	27,875	31,352	0
	MC (VR)	11,342	11	13	3,877	2,682	5,653	0,009
	MC (MN)	2,044	2	3	1,100	1	1,586	0,029
	MC (BS)	5,268	4	7	1,689	1,118	2,361	0,009
	MC (VE-PD-VI)	8	8	8	1,309	1,000	2,047	0,009
	MC (FE-RA-BO-RO)	1	1	1	1,007	1	1,025	1
HPAI H7N1	AI	4,949	4,101	5,758	5,749	5,225	6,186	0
	PS	35,480	33	38	40,021	37,550	42,363	0
	MC (PN-UD)	1,011	1	1	1,031	1	1,084	1
	MC (BS)	1,698	1	2	1,654	1,265	2,162	0,180
	MC (FE-RA-BO-RO)	1,018	1	1	1,036	1	1,144	1
	MC (BG-CO)	1,322	1	2	1,054	1	1,260	1
	MC (MI-LO)	1	1	1	1,034	1	1,116	1
	MC (VR)	2,301	2	3	1,645	1,195	2,342	0,139
	MC (MN)	1,207	1	2	1,374	1,048	1,893	1
	MC (VE-PD-VI)	2,161	2	3	1,642	1,208	2,234	0,149

Table C3. Best-fit phylogeographic model. Log marginal likelihood estimates for four phylogeographic models. The log Bayes factor (BF) comparison between each possible model and a model with rates fixed equally is provided, with the best-fit model highlighted in bold.

	Phylogeographic model	**Log marginal likelihood**	**Log BF**
HA	Rates fixed equally	-4597.83	-
	Rate fixed to population of destination	-4593.2	**2.012**
	Rates fixed to population of origin	-4599.6	-0.769
	Rates fixed to product (destination * origin)	-4598.26	-0.188
	Rates fixed to swine flow	-4594.58	1.414
Complete genome	Rates fixed equally	-26505.95	-
	Rate fixed to population of destination	-26496.86	3.9949
	Rates fixed to population of origin	-26510.41	-1.933
	Rates fixed to product (destination * origin)	-26508-08	-0.924
	Rates fixed to swine flow	-26502.91	1.322

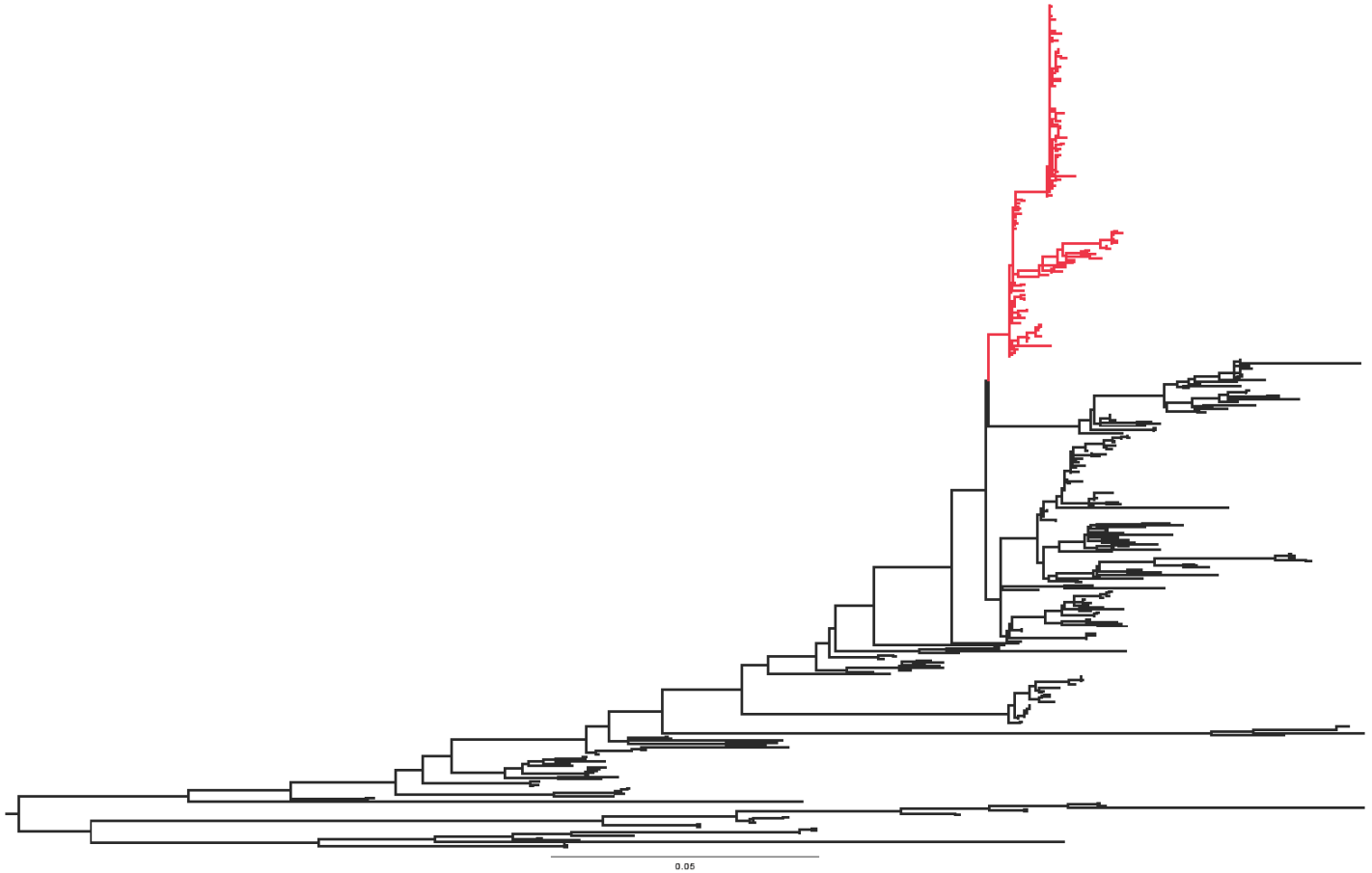


Figure C1. HA ML phylogenetic tree. This tree shows phylogenetic relationships among LPAI and HPAI H7N1 viruses (in red) and all the Eurasian H7 sequences available in GenBank (443 sequences). The tree is mid-point rooted for clarity only. Branch lengths are scaled according to the numbers of nucleotide substitutions per site. The tree is mid-point rooted for clarity only.

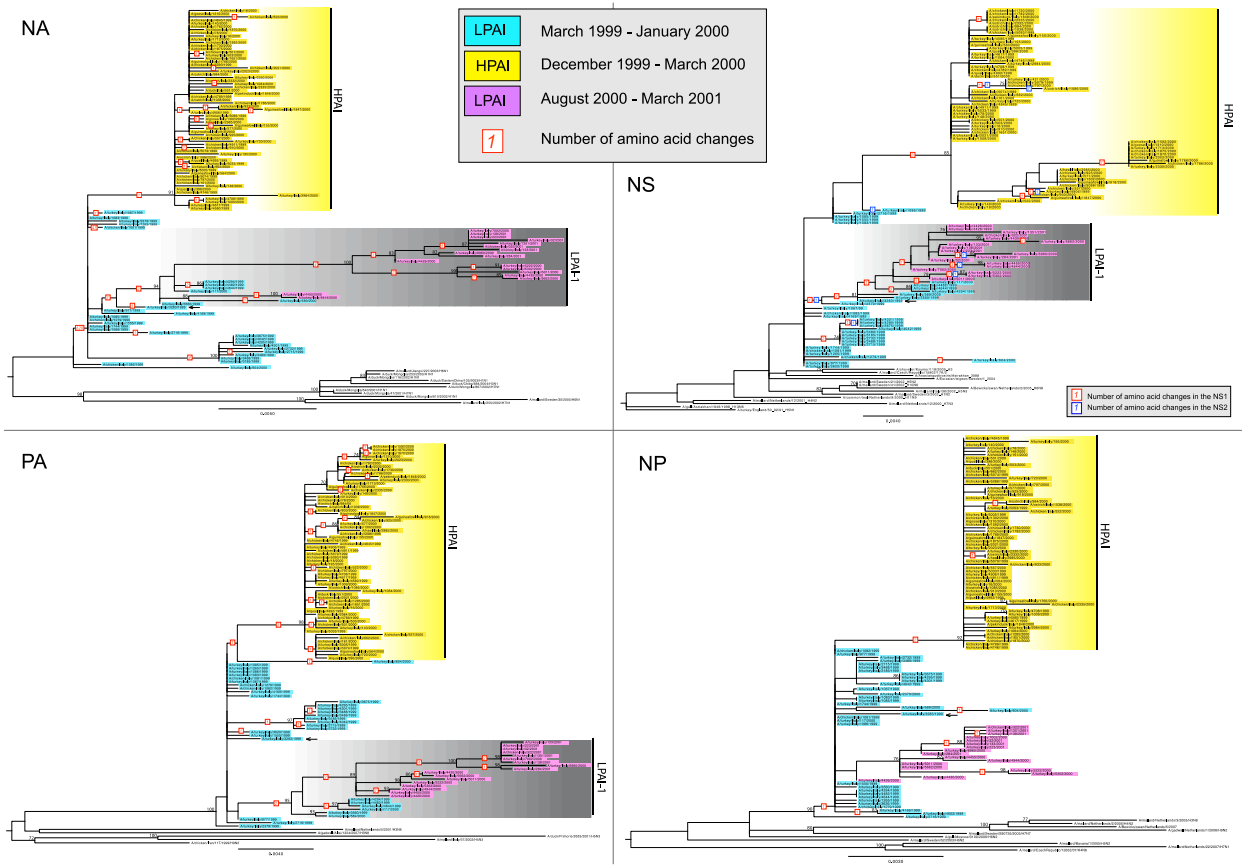


Figure C2. ML phylogenetic tree of the NA, NP, NS and PA gene segments of H7N1 avian influenza subtype. The numbers at the nodes represent bootstrap values (>60%), while branch lengths are scaled according to the numbers of nucleotide substitutions per site. The numbers highlighted in red and blue represent the number of amino acid changes identified along the main branches. The closest relative of the LPAI-I group is identified with a black arrow. The trees are mid-point rooted for clarity only.

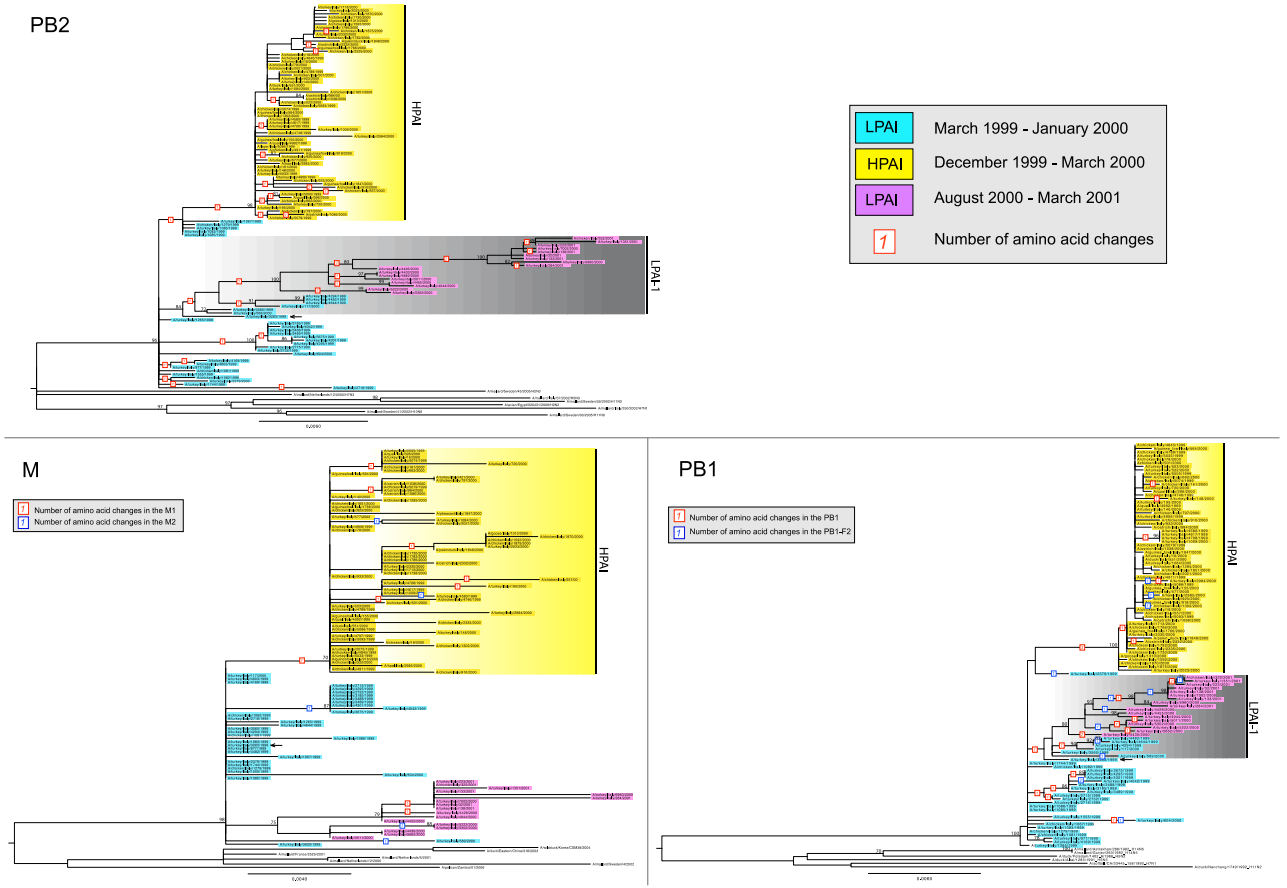


Figure C3. ML phylogenetic tree of the PB1, PB2, M gene segments of H7N1 avian influenza subtype. The numbers at the nodes represent bootstrap values (>60%), while branch lengths are scaled according to the numbers of nucleotide substitutions per site. The numbers highlighted in red and blue represent the number of amino acid changes identified along the main branches. The closest relative of the LPAI-I group is identified with a black arrow. The trees are mid-point rooted for clarity only.

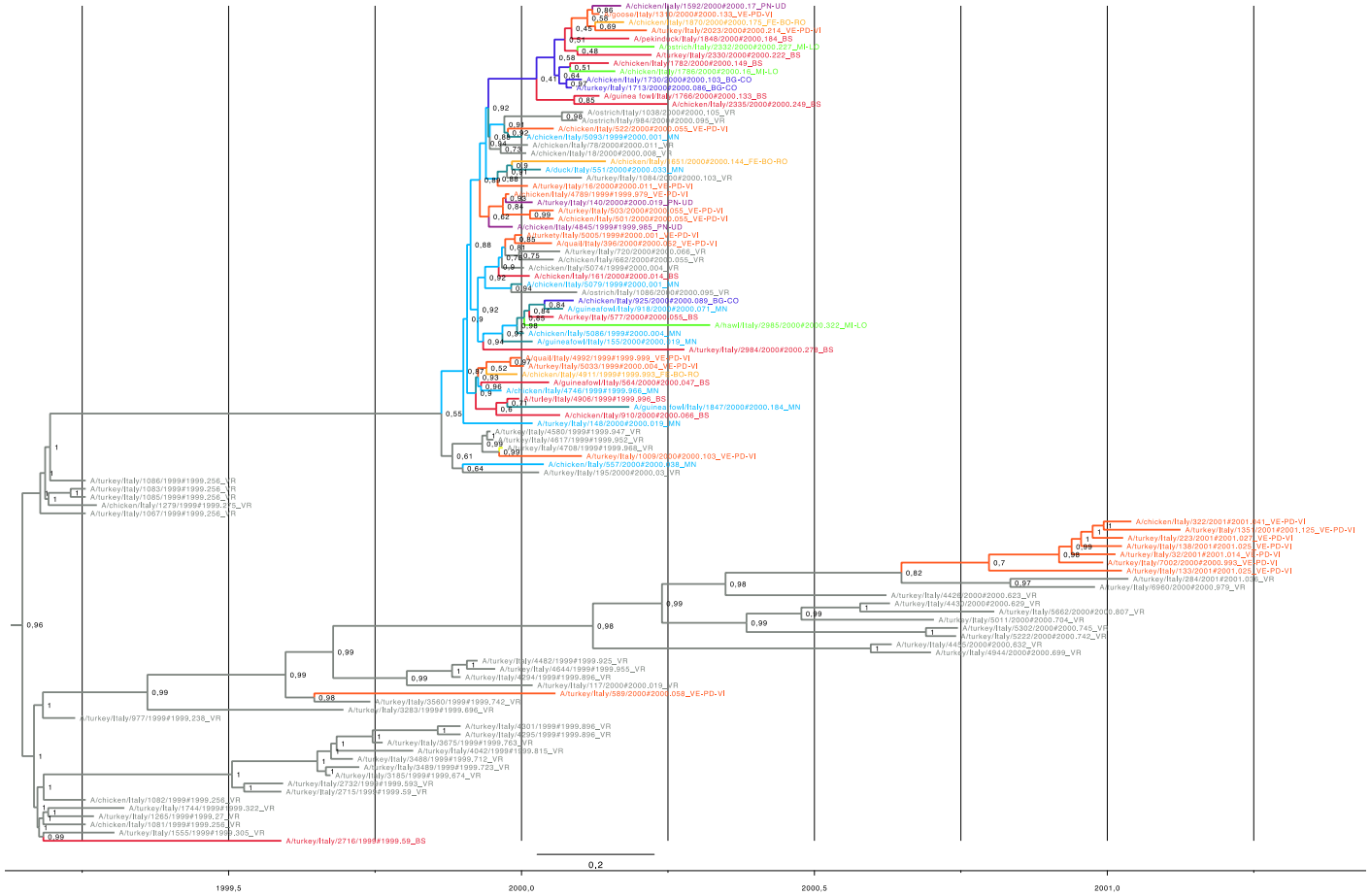


Figure C4. MCC tree inferred for the concatenated gene sequences of the Italian H7N1 viruses. Sequences are coloured according to the province of origin. Light blue: Mantova; grey: Verona; red: Brescia; yellow: Ferrara-Ravenna-Bologna-Rovigo; orange: Venezia-Padova-Vicenza; blue: Bergamo-Como; green: Milano-Lodi. The numbers at branch points represent the probability of spatial states.

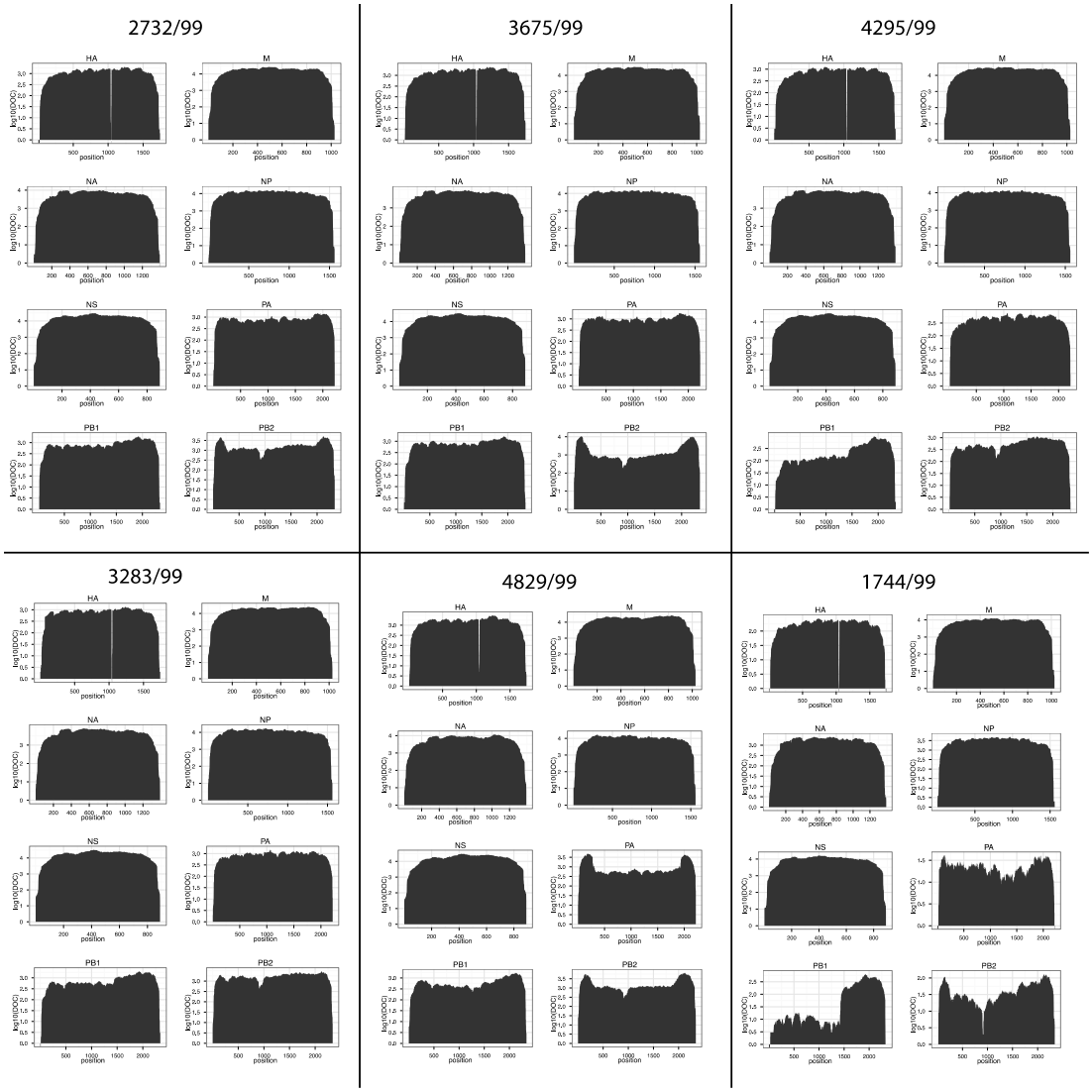


Figure C5. Depth of coverage plots. The x-axes represent nucleotide position and the y-axes represent depth for LPAI samples. Reads are aligned to the reference A/turkey/Italy/4749/99 (HPAI) virus sequences.

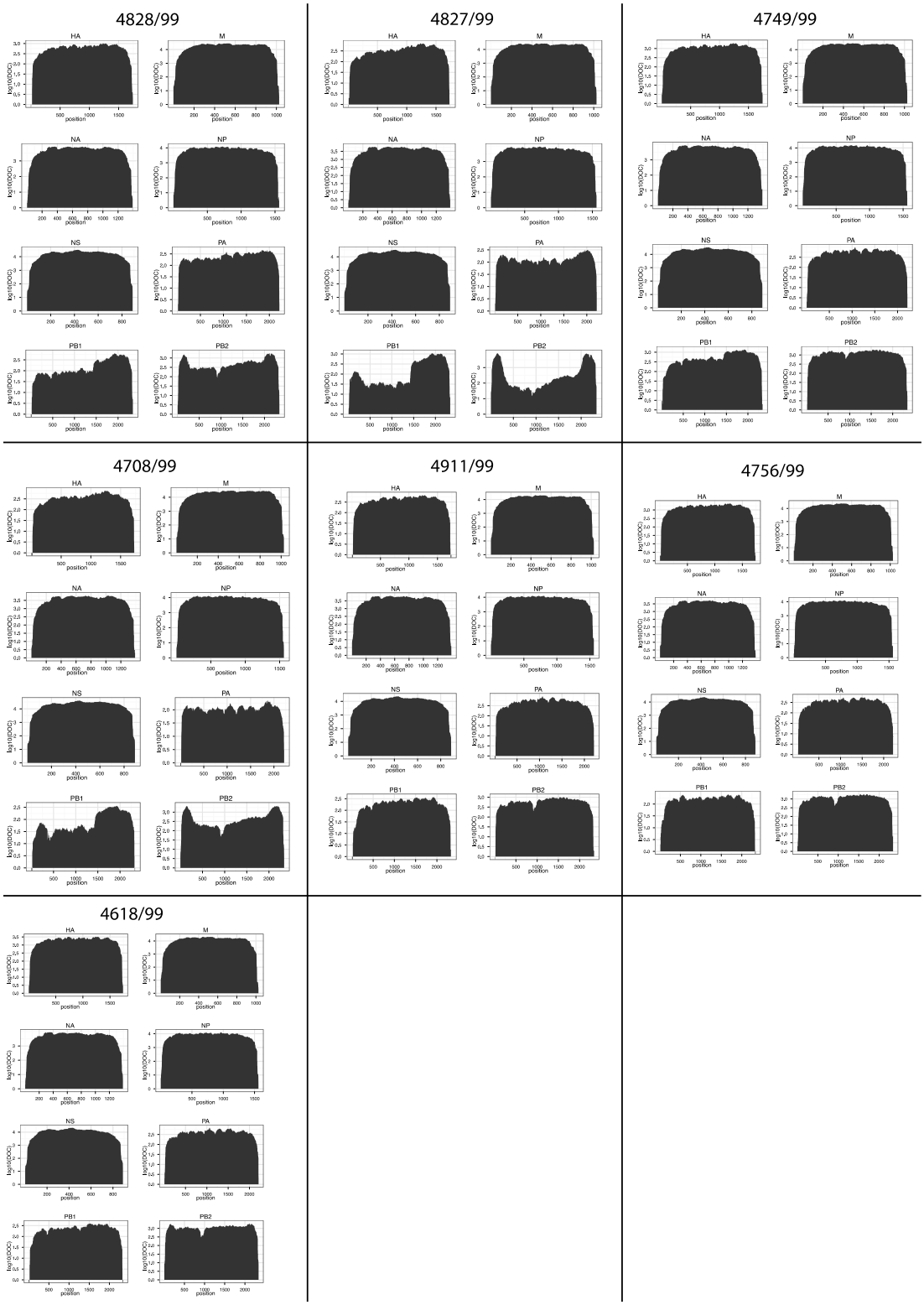


Figure C6. Depth of coverage plots. The x-axes represent nucleotide position and the y-axes represent depth for HPAI samples. Reads are aligned to the reference A/turkey/Italy/4749/99 (HPAI) virus sequences.

ANNEX D

Table D1. Epidemiological information (collection date, city and province of collection, and pathotype) of the H7N3 viruses analysed in chapter 6.

Virus	Collection date	Province	Region	Complete genome	Comments
A/turkey/Italy/214845/2002	15/10/02	na	na	Complete genome	GenBank
A/turkey/Italy/220158/2002	15/10/02	na	na	Complete genome	GenBank
A/turkey/Italy/7159/2002	16/10/02	BS	Lombardia	Complete genome	This study
A/turkey/Italy/7222/2002	18/10/02	BS	Lombardia	Complete genome	This study
A/turkey/Italy/7653/2002	29/10/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/7773/2002	31/10/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/8000/2002	06/11/02	MN	Lombardia	Complete genome	GenBank
A/turkey/Italy/8307/2002	07/11/02	VR	Veneto	Complete genome	This study
A/chicken/Italy/8093/2002	09/11/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/8303/2002	13/11/02	VI	Veneto	Complete genome	This study
A/turkey/Italy/8534/2002	13/11/02	MN	Lombardia	Complete genome	GenBank
A/turkey/Italy/8535/2002	13/11/02	BS	Lombardia	Complete genome	GenBank
A/turkey/Italy/8458/2002	14/11/02	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/8651/2002	20/11/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/8834/2002	25/11/02	VR	Veneto	Complete genome	This study
A/ostrich/Italy/8856/2002	26/11/02	BG	Lombardia	NP, HA, M, NS, NA	This study
A/turkey/Italy/8912/2002	26/11/02	VI	Veneto	Complete genome	GenBank
A/turkey/Italy/9102/2002	02/12/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/9289/2002	05/12/02	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/9314/2002	09/12/02	VR	Veneto	Complete genome	This study

A/turkey/Italy/9369/2002	10/12/02	VR	Veneto	Complete genome	This study
A/guineafowl/Italy/9360/2002	11/12/02	BS	Lombardia	Complete genome	This study
A/turkey/Italy/9374/2002	11/12/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/9441/2002	12/12/02	BO	Emilia Romagna	Complete genome	This study
A/chicken/Italy/270638/02	15/12/02	na	na	HA	GenBank
A/Guineafowl/Italy/266184/02	15/12/02	na	na	HA	GenBank
A/turkey/Italy/9504/2002	17/12/02	PD	Veneto	Complete genome	This study
A/guineafowl/Italy/9565/2002	19/12/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/9611/2002	20/12/02	PD	Veneto	Complete genome	This study
A/turkey/Italy/9739/2002	23/12/02	BS	Lombardia	Complete genome	GenBank
A/turkey/Italy/9742/2002	23/12/02	BS	Lombardia	Complete genome	GenBank
A/turkey/Italy/9692/2002	24/12/02	VR	Veneto	Complete genome	This study
A/turkey/Italy/9737/2002	30/12/02	MN	Lombardia	Complete genome	GenBank
A/turkey/Italy/4/2003	03/01/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/17/2003	07/01/03	PD	Veneto	Complete genome	This study
A/chicken/Italy/145/2003	09/01/03	PD	Veneto	Complete genome	This study
A/turkey/Italy/195/2003	09/01/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/251/2003	14/01/03	BS	Lombardia	Complete genome	GenBank
A/turkey/Italy/387/2003	17/01/03	VR	Veneto	Complete genome	This study
A/chicken/Italy/682/2003	21/01/03	PD	Veneto	Complete genome	GenBank
A/chicken/Italy/603/2003	24/01/03	BS	Lombardia	Complete genome	This study
A/turkey/Italy/1010/2003	10/02/03	MN	Lombardia	Complete genome	GenBank
A/turkey/Italy/992/2003	11/02/03	PD	Veneto	Complete genome	This study
A/guineafowl/Italy/1613/2003	12/03/03	PD	Veneto	All genes except NP	This study
A/turkey/Italy/68819/03	15/03/03	na	na	Complete genome	GenBank
A/turkey/Italy/2043/2003	24/03/03	BG	Lombardia	Complete genome	GenBank

A/turkey/Italy/97500/03	15/04/03	na	na	All genes except PB1	GenBank
A/chicken/Italy/2240/2003	16/04/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/121964/03	15/05/03	na	na	Complete genome	GenBank
A/turkey/Italy/2684/2003	15/05/03	CR	Lombardia	Complete genome	GenBank
A/turkey/Italy/2685/2003	15/05/03	BS	Lombardia	Complete genome	GenBank
A/turkey/Italy/2963/2003	23/05/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/2856/2003	26/05/03	MN	Lombardia	Complete genome	This study
A/turkey/Italy/2964/2003	27/05/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/2962/2003	04/06/03	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/2987/2003	04/06/03	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/3620/2003	01/07/03	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/4036/2003	17/07/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/4310/2003	28/07/03	VR	Veneto	Complete genome	This study
A/duck/Italy/4609/2003(H7N2)	06/08/03	MN	Lombardia	Complete genome	GenBank
A/quail/Italy/4610/2003(H7N2)	06/08/03	BG	Lombardia	Complete genome	GenBank
A/turkey/Italy/4608/2003	08/08/03	VR	Veneto	Complete genome	GenBank
A/chicken/Italy/4616/2003	11/08/03	NO	Piemonte	Complete genome	GenBank
A/chicken/Italy/4917/2003	28/08/03	VR	Veneto	All genes except PA and PB2	This study
A/turkey/Italy/5125/2003	09/09/03	VR	Veneto	Complete genome	This study
A/turkey/Italy/3337/2004	20/09/04	VR	Veneto	Complete genome	GenBank
A/quail/Italy/3347/2004	21/09/04	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/3399/2004	22/09/04	VR	Veneto	Complete genome	This study
A/turkey/Italy/3439/2004	22/09/04	VR	Veneto	All genes except M	This study
A/turkey/Italy/3477/2004	29/09/04	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/3807/2004	13/10/04	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/3829/2004	13/10/04	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/4042/2004	27/10/04	VR	Veneto	HA	This study

A/turkey/Italy/4130/2004	02/11/04	VR	Veneto	Complete genome	GenBank
A/turkey/Italy/4199/2004	05/11/04	VR	Veneto	Complete genome	This study
A/turkey/Italy/4372/2004	18/11/04	VR	Veneto	All genes except PA	GenBank
A/turkey/Italy/4479/2004	23/11/04	VR	Veneto	Complete genome	GenBank

na=not available

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisor Enrico Massimiliano Negrisolo for being an important teacher who enlightened my interests on molecular evolution and gave me invaluable direction and technical insight.

I am deeply grateful to all my colleagues from IZSVe without whom this thesis would not be possible. In particular, I want to thank my co-supervisor Giovanni Cattoli who provided me consistent guidance and support throughout my study. Special thanks to Isabella Monne for her encouragements, advice and exchange of ideas. I also appreciate the crucial work of my colleagues Alessia Schivo, Viviana Valastro, Annalisa Salviato, Adelaide Milani, Valentina Panzarin, Silvia Ormelli, Paola De Benedictis, Angela Salomoni, Angelique Angot and many others.

I am grateful for our collaborations with Martha Nelson, Eddie Holmes, Philippe Lemey, Joseph Hughes, Pablo Murcia who introduced me into the fascinated research field of viral evolution.

Francesca Ellero is gratefully acknowledged for the proof reading of many parts of this thesis.

Very special thanks go to my family and to my friends for being constantly at my side and in particular to my husband Massimo, who makes each day of my life a wonderful day. Thank you.