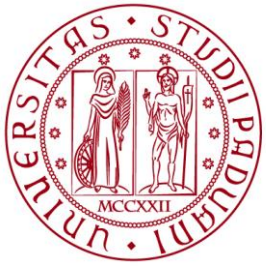


UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biomedicina Comparata e Alimentazione

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE VETERINARIE

CICLO XXVIII

Emerging H9N2 Avian Influenza virus research at the human and animal interface

Direttore della Scuola: Prof. Gianfranco GABAI

Supervisore: Dott.ssa Alessandra PICCIRIRLLO

Dottorando: Alireza HEIDARI

LUGLIO 2016

RIASSUNTO

La maggior parte delle malattie emergenti è di origine zoonotica. La globalizzazione e l'allevamento intensivo hanno contribuito ad un aumento della diffusione delle infezioni zoonotiche ed è quindi di fondamentale importanza poter conoscere ed essere preparati a gestire eventuali epidemie. L'infezione da virus influenzali di tipo A è una zoonosi che mette a rischio la sanità pubblica. Alcuni sottotipi di virus influenzali hanno superato con successo la barriera di specie e si sono adattati ai mammiferi e alla popolazione umana, causando epidemie annuali e stagionali. Il virus influenzale del sottotipo H9N2 è classificato come virus influenzale a bassa patogenicità e infetta soprattutto le specie aviarie. Dalla metà degli anni '90, questo sottotipo ha causato infezioni nei volatili selvatici e nell'industria avicola in diversi paesi dell'Asia, Europa, Nord Africa e America. L'ampia diffusione di virus H9N2 in tutta l'Eurasia, insieme con la loro capacità di causare infezione diretta nei mammiferi e nell'uomo, ha destato preoccupazione in merito a un suo potenziale impatto sulla salute pubblica come possibile candidato per la prossima pandemia influenzale.

Questa tesi di dottorato si occupa dello studio filogenetico e filogeografico, della classificazione genetica e dell'evoluzione molecolare dei virus influenzali del sottotipo H9 presenti in tutto il mondo. Particolare interesse è stato rivolto allo studio dell'emoagglutinina virale e attraverso analisi bioinformatiche è stata analizzata una grande quantità di dati e di sequenze dei virus H9 isolati nel mondo e creato un sistema di nomenclatura uniforme che fornisce al mondo scientifico una chiave di lettura e chiarisce la relazione genetica tra i diversi ceppi del sottotipo H9 dal momento della loro comparsa a oggi. L'approccio filogeografico ha inoltre contribuito a chiarire le dinamiche evolutive e la diffusione spazio-temporale tra gruppi genetici e diversi virus H9N2 dell'influenza aviaria.

Sono state elaborate in modo più dettagliato le dinamiche evolutive del virus H9N2 in un paese endemico come l'Iran. Studi filogenetici e molecolari sono stati eseguiti su tutti i virus iraniani H9N2 circolati dal 1998 al 2015, insieme alle sequenze dei virus circolanti nei paesi confinanti. I risultati hanno portato a identificare le relazioni filogenetiche delle varie introduzioni e la diffusione ad altri paesi, a evidenziare le dinamiche evolutive e la portata del suo cambiamento molecolare dal primo rilevamento in Iran e dopo l'applicazione di un programma di vaccinazione.

Inoltre, tra gli obiettivi del progetto di ricerca c'era quello di investigare la sieroprevalenza per il virus H9N2 tra i lavoratori del settore avicolo per capire se in un paese endemico come l'Iran dove, malgrado rigorosi piani di sorveglianza e vaccinazione, il sottotipo H9N2 circola da diversi anni nell'industria avicola creando molti danni economici. Questo studio è stato eseguito in collaborazione con un gruppo di ricerca iraniano per scoprire l'associazione tra l'esposizione agli avicoli e la presenza degli anticorpi verso il virus influenzale H9N2 allo scopo di rafforzare le misure di sorveglianza e controllo e monitorare il rischio biologico e quindi di preservare la salute pubblica dalla diffusione di tali virus influenzali.

SUMMARY

Most emerging diseases are of zoonotic origin. Globalization and intensive animal farming has led to an increased spread of zoonotic infections and it is therefore of crucial importance to be prepared in managing potential outbreaks. Influenza type A virus is cause of a zoonotic disease, which can affect animal and human health. A number of influenza virus subtypes have successfully crossed the species barrier and have established in mammal and human populations, causing yearly seasonal epidemics. The influenza A viruses of the H9N2 subtype are classified as low pathogenic avian influenza (AI) viruses which infect mostly avian species. The H9N2 type has caused infections both in wild birds and in the poultry population around the globe, including several countries in Asia, Europe, North Africa and Americas since the mid-1990s. The wide circulation of H9N2 viruses throughout Eurasia, along with their ability to cause direct infection in mammals and humans, raises public health concern about their potential to become candidates for the next influenza pandemic.

This Ph.D. thesis deals with phylogenetic and Bayesian phylogeographic analyses, genetic nomenclature and molecular evolutionary dynamics to elucidate the different aspects of H9 subtypes circulating worldwide. Focusing on the hemagglutinin gene through the application of bioinformatics tools a large amount of H9 subtype strains global data have been analysed and a unified nomenclature system have been designed producing a lot of data that contributed to clarify the evolution dynamics of H9 subtype viruses identified to date.

The evolutionary dynamics of H9N2 in an endemic country as Iran have been elaborated in further detail. Phylogenetic and molecular studies have been performed on all the H9N2 Iranian viruses circulating from 1998 to 2015 along with other neighboring countries to identify the phylogenetic relationships grouping the various introductions and the spread to other countries, in order to highlight the evolution dynamics and the extent of its molecular

change since the H9N2 first detection in Iran and after the application of a vaccination program.

In addition, one of the research topics includes a serological investigation of H9N2 avian influenza virus among the poultry workers in an endemic country as Iran where, despite of the control measures implemented at national level including mass vaccination of poultry, the subtype H9N2 has rapidly spread and can be considered endemic in Iranian poultry. This study has been carried in collaboration with an Iranian research group to reveal the association between professional exposure to poultry and the presence of antibodies to H9N2 viruses in order to develop surveillance and control programs to monitor the biological risk and thus to preserve public health against the spreading of the H9N2 virus.

INDEX

Chapter 1

BACKGROUND

Influenza virus

Classification

Influenza viruses belong to the *Orthomyxoviridae* family; antigenic differences in their nucleoprotein (NP) and matrix protein (M1) allow influenza viruses to be classified in five different genera: *Influenzavirus* A, *Influenzavirus* B, *Influenzavirus* C, *Isavirus*, and *Thogotovirus* (Webster *et al.*, 1992; Ducatez *et al.*, 2015). Influenza A virus is further subtyped based on the antigenic properties of the surface glycoprotein hemagglutinin (HA) and neuraminidase (NA). The HA and NA subtypes have been found in a great variety of combinations in nature. To date, eighteen HA (H1–H18) and eleven NA (N1–N11) subtypes have been identified. Aquatic birds are considered the major influenza A viruses reservoir in nature harbouring diverse combinations of the HA and NA subtypes. H17N10 and H18N11 viruses have only recently been described in bats (Tong *et al.*, 2013).

A standard naming system for the identification of these viruses is used, starting with the specific genus and followed by species of isolation, geographical location, isolate number, and year of isolation. The antigenic description of the HA and NA subtypes follows in parentheses. The naming system for most human influenza viruses follows the same format, but the species of isolation is not included, though early influenza isolates such as A/Wilson-Smith/33 (H1N1) include the names of the scientists involved rather than the place of isolation (Cox *et al.*, 2010).

Etiology

Genome and proteins

Influenza A viruses are enveloped single stranded negative sense RNA viruses. They are pleomorphic, mostly found as spherical or ovoid shaped particles, measuring 100-300 nm in diameter following *in vitro* passage, while filamentous particles up to 30 μm are typical of primary or low-passage isolates (Seladi-Schulman *et al.*, 2013).

Their genome consists of 8 segments (from 0.9 to 2.3 Kb), which encodes for at least seventeen genes (Table 1), fifteen of which are structural and non-structural proteins (Figure 1).

Gene Segment	Relative Size (nucleotides)	Gene Products	Polypeptide Length (amino acids)	Categories
1 - PB2	2341	PB2	759	Internal
2 - PB1	2341	PB1	757	Internal
		PB1-N40	718	Internal
		PB1-F2	90 (or truncated)	Internal
3 - PA	2233	PA	716	Internal
		PA-X	252	Internal
		PA-N155	568	Internal
		PA-N182	535	Internal
4 - HA	1778	HA	566	Surface
5 - NP	1565	NP	498	Internal
6 - NA	1413	NA	454	Surface
7- M	1027	M1	252	Internal
		M2	97	Surface
		M42	99	Surface
8 - NS	890	NS1	230	non-structural
		NS2 (NEP)	121	non-structural
		NS3	174	non-structural

Table 1. Influenza A virus gene segments, their relative sizes, and encoded proteins.

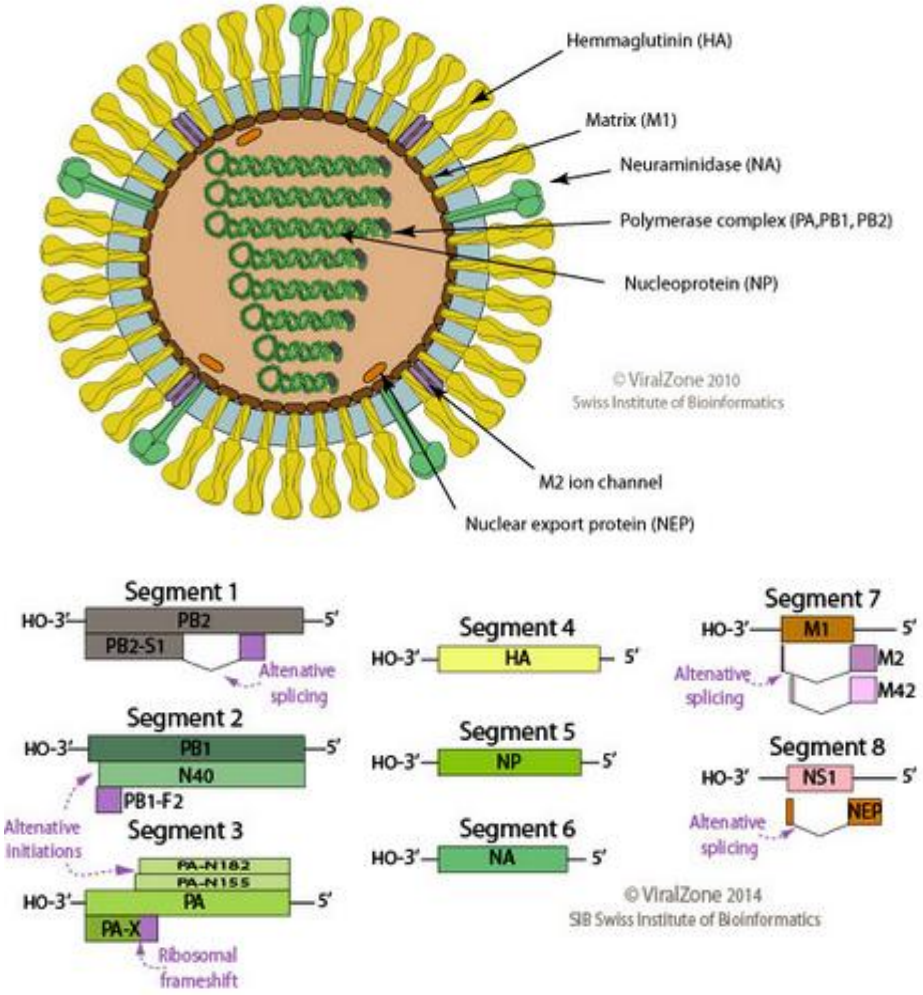


Figure. 1. Graphical representation of the spherical shape of influenza A virus and its genome. The ssRNA(-) genome is encapsidated by nucleoprotein and consists of 8 segments with a size range between 890 and 2341 nt and encodes for 12-14 proteins (depending on the strain). From ViralZone, SIB Swiss Institute of Bioinformatic (http://viralzone.expasy.org/).

Segment 1: PB2 polymerase. The basic polymerase protein 2 (PB2) is a cap-binding protein essential for the synthesis of mRNA (Plotch et al., 1979); at the beginning of the transcript it recognizes and binds the cap to the 5 'RNA end of the host, and it is partially involved in cap snatching so that it can be used as a primer for the viral RNA synthesis. (Webster et al., 1992). Associated with host range restriction.

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; 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 increases 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 new 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, 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 the most abundant antigenic glycoprotein of the viral surface encoded by the 4th viral segment of the influenza A genome. It is a type I membrane glycoprotein responsible for the binding of the virus to the receptors present in the host cell surface, for virus internalization and fusion with the endosomal host membrane. In the viral particles, each mature HA is a homotrimer which projects onto the viral envelope to form a rod-shaped structure; in infected cells, this protein is synthesized as a precursor polypeptide, called HA0, which must be cleaved into two subunits (HA1 and HA2, of 36 and 27 kDa, respectively) by host trypsin-like proteases (Copeland *et al.*, 1986). After the proteolitic cleavage of the precursor, the two subunits are covalently linked to each other by a disulfide bond, whereas each trimer is associated to the others by non covalent bonds (Klenk *et al.*, 1975). This processing is necessary for the virus infectivity because it activates the fusion of HA and it is a determinant of pathogenicity (Hamilton *et al.*, 2012). The cleavage site is a prominent surface loop near a cavity in HA0 (Chen *et al.*, 1998); the results of the cleavage process is a structural rearrangement in which the fusion peptide, formed by nonpolar N-terminus amino acids of HA2, is relocated into the interior of the trimer and buries ionizable residues involved in the acidification-induced conformational changes in the endosome. At the membrane-distal tip of each HA1 the receptor binding domain (RBD) is present, which is formed by the 130-loop, 190-helix, and 220-loop and four highly conserved sites (Weis *et al.*, 1988, Martín *et al.*, 1998); it forms the sialic acid binding pocket and contains most of the antigenic regions recognized by neutralizing antibodies (Abs). The stem-like structure HA2 has a membrane fusion activity. The 2 loops and the helix contain amino

acids that interact either with sialic acids or with internal sugars of the glycan chain associated with glycoproteins and glycolipids on the surface of epithelial cells; the base of the site contains several highly conserved residues that form an extensive hydrogen bond network. The three-dimensional structure of few HA subtypes has been resolved and characterized with respect to the localization and structure of their antigenic sites. In the H3 subtype, five antigenic sites (A, B, C, D, E) have been mapped (Wiley *et al.,* 1992), and the same structure was used to map antigenic sites on the H1 and H2. When the 3D structures of H5 and H9 HA were resolved, the H5 HA molecule was antigenically mapped. For H5 structure, the localization of two antigenic sites has been described. Site 1 includes residues 140 to 145 in HA1 (H3 numbering). Changes in the HA and NA antigenic combinations of a virus (antigenic shift) may derive from the genome segmentation. When the same cell is co-infected by distinct viral subtypes, the viral progeny may originate from a reassortment of parental genes from different viruses.

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 homotetramer of 220 kDa consisting of a head, which is an enzymatically active domain, and a tail which allows to anchor the protein to the membrane (Hausmann *et al.,* 1997). Its main functions are i) hydrolysis of sialic acid on the cellular receptor for the hemagglutinin, which allows the release of the virus from the cell surface, and ii) removal of sialic acid residues from the viral particles to prevent aggregation. The inhibitors developed for this enzymatic protein are the main antiviral drugs. In the

absence of this enzyme the virus remains attached to the cellular receptors, thus inhibiting the spread of the virus progenies in the host tissues. Therefore, antibodies directed against the neuraminidase protein do not prevent infection but reduce the spread of the virus in tissues.

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 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).

Infection and replication cycle

The influenza A virus takes advantage of the host cell machinery in almost all the phases of its replication. This process begins with receptor recognition and binding by the viral hemagglutinin (HA). Attachment is mediated by an interaction between the receptor binding site of the HA molecule and sialic acid (SA)-containing glycoconjugates on target cells. The nature of the SA linkage to its underlying galactose is considered the first major determinant for successful infection, as avian-tropic viruses preferentially recognize and bind to SAs with an alpha-2,3 linkage to galactose, while human-tropic viruses generally bind SAs with alpha-2,6 linkages. After binding to the host cell, the virus is internalized through the endosomal pathway; the acidification of the vesicle causes a conformational change of the hemagglutinin, which exposes the N-terminal hydrophobic domain of the HA2 chain (site of fusion) and causes its interaction with the endosomal membrane. The viral envelope and the endosomal membrane are then merged and the viral nucleocapsid is released into the cytoplasm of the host cell (Figure 2). This fusion occurs thanks to the cleavage of the HA0 precursor molecule into the HA1 and HA2 subunits by the 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 nucleoprotein (NP).

All the vRNA synthesis processes, including viral transcription, replication and assembly of vRNPs occur in nucleus. In the host cell nucleus, the viral RNA polymerase complex (PB1, PB2, PA) uses the negative sense vRNA as template to synthesize two different types of single-stranded positive RNA segments: cRNA (complementary RNA) used by RNA

polymerase to produce many copies of vRNA (viral RNA) and mRNA (messenger RNA) (Herz *et al.*, 1981).

In influenza virus infected cells, the nuclear export of cellular mRNAs is blocked; the cellular pre-mRNA and mRNA are degraded in the nucleus. This nuclear export block is selective: all viral mRNAs are efficiently exported; as a consequence, these newly synthesized mRNAs prevail in reaching the translation machine into the cytoplasm of infected cells, thus helping to permanently stop the gene expression of the host cell and selectively synthesize viral proteins. The nuclear export of the viral mRNA uses the host cell machinery but it is selective: it is controlled by a viral non-structural protein (NS1) that inhibits the expression of mRNA synthesized after the cell infection. Furthermore, the pre-mRNA and mRNA retained in the cell nucleus are available to the cap dependent viral endonuclease for the production of the capped RNA primers required for viral mRNA synthesis.

Viral mRNAs, after being transported into the cytoplasm, are translated to produce the corresponding proteins by the host ribosome. The membrane proteins (HA, NA and M2) are transported by the RER and the Golgi apparatus to the plasma membrane. When nucleocapsid and viral coat proteins reach the plasma membrane, they form a bud which coating contains proteins immersed in the lipids bilayer of the host cell. Subsequently the bud separates and virus particles are released outside the cell. During viral budding the host proteins present in the plasma membrane are excluded from the final viral particle. The viral proteins possessing a nuclear localization signal are transported into the nucleus (PB1, PB2, PA, NP, M1, NS1 and NS2). The progeny virions from infected cell be released from the cell upon NA cleavage of host sialic acid receptors (Baigent and McCauley, 2003; Figure 2).

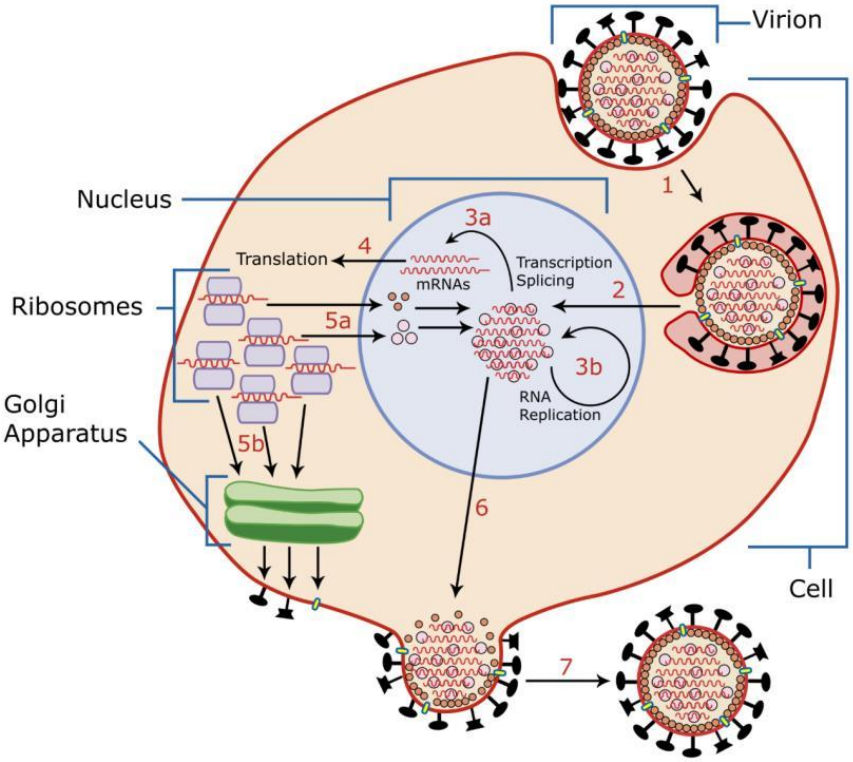


Figure 2. Influenza A virus replication. 1) A virion attaches to the host cell membrane using HA and enters the cytoplasm by receptor-mediated endocytosis, thereby forming an endosome; 2) HA2 promotes fusion of the virus envelope and the endosome membranes. M2 acts as an ion channel making the inside 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.

Evolution

Antigenic drift and shift

As for all RNA viruses, Influenza A viruses are dynamic and evolve continuously. Their evolution is mostly characterized by the combination of high mutation rate, rapid replication and infection of large populations. High mutation rate and reassortment of the segmented genome are the mechanisms that allow influenza viruses to rapidly evolve and avoid neutralization by pre-existing antibodies in the exposed population. Antigenic variation of the HA and NA surface glycoproteins occurs through two diverse mechanisms. The first is known as antigenic drift. The antigenic drift refers to the small gradual accumulation of amino acid changes that occur through point mutations; these may happen unpredictably and result in minor changes of the influenza genome. When they occur in the two genes codifying 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. Antigenic drift is caused by spontaneous mutations resulting from a lack of proof-reading by the virus polymerase, and variations are thought to occur at a rate of less than 1% each year. However, after a few years of continued accumulation a new variant strain will become predominant and cause the next rounds of seasonal epidemics (Cox *et al.*, 2010). 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). While influenza viruses change by genetic drift commonly, genetic shift happens only occasionally.

Avian influenza and pathotypes

Avian influenza viruses can be classified into two groups based on the severity of clinical disease: low pathogenic (LPAI) or highly pathogenic (HPAI) form. LPAI virus infections in poultry can be asymptomatic or cause mucosal infection, with mild to severe respiratory disease, water and feed decrease and drops in egg production; usually they do not result in high mortality of the infected hosts. To date, only H5 and H7 subtypes of influenza A viruses have shown to be able to evolve from a LPAI to a HPAI. When LPAI viruses of the H5 or H7 subtype are transmitted from wild birds to poultry, they may mutate into HPAI viruses. In poultry, HPAI viruses usually cause systemic infections and are associated with severe disease and high mortality reaching 100% within 48 hours of infection.

From a molecular point of view, this pathogenicity lies in the sequence at the HA0 cleavage site, which depends on host proteases for cleavage into its HA1 and HA2 subunits for functionality. LPAI viruses contain a single arginine at the cleavage site, limiting cleavage to a small number of extracellular host proteases and thereby restricting tissue tropism to anatomical sites where such proteases are found. Conversely, highly pathogenic viruses contain multiple basic amino acids at the HA0 cleavage site, permitting cleavage by ubiquitous intracellular proteases and ultimately allowing for systemic spread of the virus (Capua & Alexander, 2008; Taubenberger, 1998). HPAI isolates generally emerge after their low pathogenic precursors are introduced from wild birds into domestic poultry flocks. In these new host populations, viruses acquire the highly pathogenic cleavage site by either

stuttering of the polymerase during replication or rarely via non- homologous recombination with other viral gene segments (Webster, 2006).

In addition to the H5 and H7 subtypes, also some H10 viruses fall within the definition of HPAI (Alexander, 2008). The H10 subtypes have characteristics of the HPAI pathotype. However, these viruses do not possess a multi-basic cleavage site and do not cause systemic infection. When these viruses are administered intravenously into poultry, they cause the death of the host by impairing the function of the kidney (Swayne and Alexander, 1994). A recent study reported a nephritic H10N1 avian influenza virus that did not display multiple basic amino acids at the cleavage site although it showed an intravenous pathogenicity index (IVIP) of 1.9. Furthermore, when administered by a natural route (intranasal), this H10N1 virus may cause mortality (Bonfante *et al.*, 2014). To date, the World Health Organization (WHO) considers H5, H7 and H9 avian influenza subtypes as those with the greatest pandemic potential. The spread of HPAI and LPAI viruses of the H5, H7 or H9 subtypes amongst birds and sporadic infections in humans continue to pose a threat to public health (Lin *et al.*, 2000).

Host Specificity and Interspecies Transmission

To successfully cross the species barrier, the virus must overcome several constraints either at virus and host level. Binding preference of viruses to host cells is considered to be one of the most relevant determinant that prevents crossing the species barrier and influence viral tropism. Sialic acids (SAs) are present on the surface of many cell types and animal species; in general, carbon-2 of the terminal sialic acid is bound to the carbon-3 or carbon-6 of an adjacent galactose, which allows the formation of an α2,3- or α2,6-linkage configuration. The hemagglutinin of avian and equine influenza viruses usually has a binding preference for

α2,3-linked sialic acid; human isolates exhibit α2,6-linkage, whereas viruses from swine bind to both. More recently it has been demonstrated that the human epithelial cells of the lower respiratory tract harbour SAs with both α2,6- and α2,3-linkages; furthermore, the finding of α2,3-linked SAs in the human airway epithelium can explain the ability of viruses of avian origin to infect and replicate in humans, although not sufficiently enough to induce an efficient human to human virus transmission (Matrosovich et al., 1999). Neuraminidase is also involved in the interaction with host receptors; in particular, neuraminidase of avian origin prefers to hydrolyze the α2,3- rather than the α2,6-linked SA. The equilibrium between the two main surface glycoproteins HA and NA must be balanced to allow a perfect enzymatic activity and a functional viral replication.

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 (Wagner et al., 2002).

Occasionally, infection of humans and other mammals by avian influenza viruses may occur (Capua and Alexander, 2008), which proves that the species barrier is not insurmountable. Some avian or swine influenza A virus subtypes (*e.g.* swine H1 and H3, avian H5, H6, H7, H9 and H10) have occasionally infected humans, without however establishing in human population. Previous studies have shown that the H7N9 subtype is probably the result of several reassortment events involving the subtype H9N2 and the subtypes H7 and N9, all of avian origin The avian H9 influenza virus subtype can also infect humans and cause a mild influenza-like illness (Butt *et al.*, 2010).

Epidemiology of H9N2 avian influenza virus

H9N2 subtype avian influenza viruses (AIV) have widely circulated in the world since its first detection in turkeys in Wisconsin in 1966 (Homme and Easterday, 1970). In North America, H9N2 viruses were mainly found in shore birds and wild ducks (Kawaoka *et al.*, 1988; Shaw *et al.*, 2002) and no H9-associated disease was reported in chickens (Perez *et al.*, 2003). In Asia, H9N2 viruses were isolated regularly from ducks before the '90s (Shortridge, 1992). However, infections of H9 subtype AIV in chickens have been reported in many Asian countries in chickens since later '90s (Alexander, 2000).

H9N2 is enzootic in poultry across large areas of Asia, the Middle East, and North Africa where it imposes a large economic burden due to reduced growth rates in broilers and reduced fertility and egg production in breeders and layers. Although H9 viruses are classified as low pathogenicity avian influenza viruses (LPAI), mortality in the field of more than 50% has been recorded and there have been instances of field strains displaying a highly pathogenic phenotype both in the field and in the laboratory (Nili *et al.*, 2003).

The H9N2 viruses fall into a number of genetically defined lineages, with the majority of those circulating in Middle East and far East. Until now there is not a standard nomenclature system for the H9 subtype. Based on most reported genetic studies, two major lineages of H9N2 viruses have circulated in poultry and wild birds: North American and Eurasian. The Eurasian lineage is subdivided into two major sub-lineages: A/quail/Hong Kong/G1/97-like (G1-like) and A/duck/Hong Kong/Y280/97-like (Y280-like), both of which have become established in domestic poultry during the mid-1990s (Webster *et al.*, 1992).

H9N2 subtype viruses have been isolated from pigs in 1998 (Peiris *et al.*, 1999a). Subsequently, they have been isolated from humans with influenza-like illness in Hong Kong and Mainland China (Guo *et al.*, 1999; Peiris *et al.*, 1999b).

The two human H9N2 isolates from Hong Kong belong to the G1 group, while at least 1 human isolate from China and 2 isolates from pigs from Hong Kong are G9-like viruses (Guo *et al.,* 1999; Lin *et al.,* 2000).

H9N2 AIV vaccination in poultry

Vaccination is considered a promising control measure for H9N2 infection in poultry and has been adopted by many countries when the disease becomes endemic. The veterinary authorities in different countries (Iran, Korea, China) permitted the use of a single H9N2 LPAI vaccine strain to simplify the evolution of the circulating virus due to the immune pressure caused by the vaccine use (Choi *et al.,* 2010; Li *et al.,* 2005; Vasfi Marandi *et al.,* 2013).

An inactivated vaccine derived from an early isolate, A/chicken/Shandong/6/96 (CK/SD/6/96), has been used in layers and breeders in chicken farms since 1998 (Li *et al.,* 2005). However, virus infection still occurred, sometimes even in vaccinated flocks and in recent years vaccine failure in many areas has become common due to the emergence of antigenic variants which harbour mutations in the major influenza antigen, hemagglutinin (HA). Consequently, efforts have been made to identify molecular markers in the HA gene that influence the antigenic diversity of H9N2 viruses leading to vaccine failure (Park *et al.,* 2011).

Recently multiple introductions of a reassortant H5N1 avian influenza virus of clade 2.3.2.1c with PB2 gene of H9N2 subtype has been identified in Indian poultry. The virus possesses the molecular signature of high pathogenicity for chickens, which is corroborated by intravenous pathogenicity index of 2.96. The virus is a reassortant which derives its PB2 gene from H9N2 virus isolated in China during 2007-2013 (Tosh *et al.,* 2016).

H9N2 avian influenza virus a pandemic potential

In the last 20 years, novel non-seasonal influenza viruses have emerged in humans, most of which have originated from birds. In recent years, several novel reassorted influenza viruses (*e.g.* H7N9, H9N2, and H10N8) have jumped the host-species barrier and are monitored by the scientific community and public health systems. It is still unclear whether these viruses can actually cause pandemics or just isolated episodes (Trombetta *et al.*, 2015).

A significant proportion of H9N2 field isolates have acquired human virus-like receptor specificity, preferentially binding α2-6-linked sialic acid (SAα2-6) receptors, in contrast to the classic avian virus-like receptor specificity that preferentially binds α2-3-linked sialic acid (SAα2-3) receptors (Matrosovich *et al.*, 2001). Interestingly, a few of the H9N2 viruses that recognize SAα2-6 receptors have transmitted directly to humans, causing mild flu-like illness.

Domestic pigs have been confirmed as being infected with H9N2 influenza viruses, and infections have been reported also in swine along with other mammals in recent years.

The first case of human infection from H9N2 avian influenza virus was reported in Hong Kong in 1999 and further human infections were reported in 2003. To date, different human cases of H9N2 AI infection have been reported by the World Health Organization (WHO); in 2015 new cases in Egypt and Bangladesh have been described. Furthermore, H9N2 viruses can contribute with gene segments during reassortment events leading to the generation of novel avian influenza virus that can infect humans (*e.g.* recent Chinese H7N9 and H10N8 viruses). Recent transmission studies have demonstrated that some natural isolates of H9N2 viruses can acquire the ability to transmit efficiently among ferrets via respiratory droplets. In addition, it has been reported that serial passages of an H9N2 virus through guinea pigs can result in the introduction of amino acid substitutions, which increases contact transmission efficiency in this mammalian model.

The recurring presence of H9N2 infections in humans and mammals has raised concerns about the possibility of H9N2 viruses evolving into pandemic strains.

H9N2 AI in Iran

In Iran, there are 17,950 poultry farms with a yearly chicken meat production of 2 million tons (1.2 billion broilers and 14 million breeders) and a production of table eggs (40 million layers) amounting to 700,000 tons (data obtained from the Iran Veterinary Organization in 2015).

Since 1998, an epidemic of avian influenza occurred in the Iranian poultry industry. The identified agent presented low pathogenicity, and was subtyped as an H9N2 avian influenza virus. In broiler chicken farms the mortality was between 20% and 60%. Mixed infections of the influenza virus and other respiratory pathogens, particularly infectious bronchitis virus and *Mycoplasma gallisepticum*, were thought to be responsible for such high mortality, which resulted in great economic losses. Clinical signs included swelling of the periorbital tissues and sinuses, typical respiratory discharge, and severe respiratory distress (Nili *et al.*, 2003).

Unfortunately, despite the massive vaccination of breeder flocks, the H9N2 strains continued to circulate in nearly all provinces of Iran. The vaccines used for H9N2 infections in the Iranian poultry industry are produced by locally companies or imported from international pharmaceutical companies (Vasfi Marandi *et al.*, 2013).

Serological analysis of Iranian viruses revealed that the currently circulating strains are antigenically distinct from the vaccine strain. These results suggest that the commercial vaccine is not genetically and antigenically similar to the viruses currently circulating in the region. These findings also show that it is important to monitor the genetic and antigenic variations in H9N2 influenza viruses when selecting a vaccine strain (Bahari *et al.*, 2015).

Recently, the association of high mortality and case report of H5N1 and H9N2 Iranian influenza virus in wild birds raised the possibility of a new genetic modified AI virus (Majidzadeh *et al.*, 2011).

AIMS OF THE THESIS

The impact of the H9N2 AI virus with respect to public health and poultry industry has been highlighted in the introduction section; however, it can be shortly reminded here that this zoonotic agent is responsible for disease in poultry resulting in serious and huge economic losses worldwide. The extent of the genetic reassortment involving H9N2 viruses, their ability to infect humans, and their continue circulation in domestic poultry and mammals emphasizes the potential for the emergence of a virus with the ability to adapt in the human population and cause the next influenza pandemic. Therefore, continue research and surveillance programs are required to monitor the presence of the H9N2 AIVs in poultry and humans, as well as to infer their evolutionary trends. Since the second half of the ’90s, outbreaks caused by H9N2 subtype have been reported in Iran and the continued presence of this subtype may mean that it is becoming an established endemic virus in this country.

The research presented in this thesis (Chapter 2) aims to address important preliminary questions regarding the potential exposure to H9N2 AI viruses among Iranian (Fars province) poultry workers: evaluate the possibility of avian to human transmission of H9N2 AIVs; assess whether the human exposure to this influenza virus subtype occurred and whether the poultry operators in the Iranian endemic areas (Fars province, in this study) are at risk of exposure to the H9N2 infection.

Another objective (Chapter 3) of the research project focuses on phylogenetic analyses of H9N2 AI viruses to provide a better understanding of the mechanisms underlying the evolutionary dynamics along with creation of a unified nomenclature system for the identification and classification of all the H9 subtypes. 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. The adoption of an internationally

accepted nomenclature is crucial for future studies of H9N2 virus epidemiology and evolution.

Finally, the last aim (Chapter 4) is to apply the recently developed bioinformatics analyses to explore the phylogenetic evolution, virus introductions, and the presence of selection pressures of H9N2 viruses in an endemic country as Iran.

References

1. Alexander DJ. A review of avian influenza in different bird species. Vet Microbiol. 2000 May 22;74(1-2):3-13. Review.
2. 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.
3. Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature. 1994 Sep 1;371(6492):37-43. PubMed PMID: 8072525.
4. Baigent SJ, McCauley JW. Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. Bioessays. 2003 Jul;25(7):657-71. Review.
5. Bahari P, Pourbakhsh SA, Shoushtari H, Bahmaninejad MA. Molecular characterization of H9N2 avian influenza viruses isolated from vaccinated broiler chickens in northeast Iran. Trop Anim Health Prod. 2015 Aug;47(6):1195-201. doi: 10.1007/s11250-015-0848-x.
6. Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature. 1994 Sep 1;371(6492):37-43. PubMed PMID: 8072525.
7. Capua I, Alexander DJ.. Avian influenza and Newcastle disease: A field and Laboratory Manual. Springer 2008-Verlag, Italy.
8. Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ, Wiley DC. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell. 1998 Oct30;95(3):409-17.

9. Clancy,S.. Genetics of the influenza virus. Nature Education 2008, 1(1):83.
10. Cox, N. J., Neumann, G., Donis, R. O. & Kawaoka, Y. Orthomyxoviruses: Influenza. In Topley Wilson's Microbiol Microb Infect, 2008,pp. 634–698. Edited by B. W. J. Mahy, V. ter Meulen, S. P. Borriello, P. R. Murray, G. Funke, S. H. E. Kaufmann, M. W. Steward, W. G. Merz, R. J. Hay, et al. Chichester, UK: John Wiley & Sons, Ltd.
11. 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.
12. Ducatez MF, Pelletier C, Meyer G. Influenza D virus in cattle, France,2011-2014. Emerg Infect Dis. 2015 Feb;21(2):368-71.
13. Hausmann J, Kretzschmar E, Garten W, Klenk HD. Biosynthesis, intracellular transport and enzymatic activity of an avian influenza A virus neuraminidase: role of unpaired cysteines and individual oligosaccharides. J Gen Virol. 1997 Dec;78 (Pt 12):3233-45.
14. Hamilton BS, Whittaker GR, Daniel S: Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. Viruses 2012, 4:1144–1168.
15. Hausmann J, Kretzschmar E, Garten W, Klenk HD. Biosynthesis, intracellular transport and enzymatic activity of an avian influenza A virus neuraminidase: role of unpaired cysteines and individual oligosaccharides. J Gen Virol. 1997 Dec;78 (Pt 12):3233-45.
16. Homme PJ, Easterday BC. Avian influenza virus infections. I. Characteristics of influenza A-turkey-Wisconsin-1966 virus. Avian Dis. 1970 Feb;14(1):66-74.
17. Herz C, Stavnezer E, Krug R, Gurney T Jr. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell. 1981 Nov;26(3 Pt 1):391-400.I
18. Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, Dunfee RL,

Schwartzman LM, Ozinsky A, Bell GL, Dalton RM, Lo A, Efstathiou S, Atkins JF, Firth AE, Taubenberger JK, Digard P. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. Science. 2012 Jul 13;337(6091):199-204.

19. Kochs G, García-Sastre A, Martínez-Sobrido L. Multiple anti-interferon actions of the influenza A virus NS1 protein. J Virol. 2007 Jul;81(13):7011-21.
20. Kawaoka Y, Chambers TM, Sladen WL, Webster RG. Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? Virology. 1988 Mar;163(1):247-50.
21. Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, Subbarao K, Guan Y, Krauss S, Shortridge K, Webster R, Cox N, Hay A. Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci U S A. 2000 Aug 15;97(17):9654-8.
22. Majidzadeh K, Karimi V, Soleimanidor M, Estabragh AS, Barin A, Langeroudi AG. Phylogenetic study on nonstructural (NS) gene of H9N2 isolated from broilers in Iran during 1998-2007. Pak J Biol Sci. 2011 Sep 1;14(17):838-43.
23. Martin K, Helenius A. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell. 1991 Oct 4;67(1):117-30.
24. Matrosovich M, Zhou N, Kawaoka Y, Webster R. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. J Virol. 1999 Feb;73(2):1146-55.
25. Matrosovich MN, Krauss S, Webster RG. H9N2 influenza A viruses from poultry in Asia have human virus-like receptor specificity. Virology. 2001 Mar 15;281(2):156-62.
26. 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.

27. Nili H, Asasi K. Avian influenza (H9N2) outbreak in Iran. Avian Dis. 2003;47(3 Suppl):828-31.
28. O'Neill RE, Talon J, Palese P. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J. 1998 Jan 2;17(1):288-96.
29. 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. 151 Lancet. 2004 Feb 21;363(9409):617-9.
30. Perez DR, Lim W, Seiler JP, Yi G, Peiris M, Shortridge KF, Webster RG. Role of quail in the interspecies transmission of H9 influenza A viruses: molecular changes on HA that correspond to adaptation from ducks to chickens. J Virol. 2003 Mar;77(5):3148-56.
31. Pinto LH, Holsinger LJ, Lamb RA. Influenza virus M2 protein has ion channel activity. Cell. 1992 May 1;69(3):517-28.Plotch SJ, Bouloy M, Krug RM. 1979 Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. Proc Natl Acad Sci U S A. Apr;76(4):1618-22.
32. Qian XY, Alonso-Caplen F, Krug RM. Two functional domains of the influenza virus NS1 protein are required for regulation of nuclear export of mRNA. J Virol. 1994 Apr;68(4):2433-41.
33. Seladi-Schulman J, Steel J, Lowen AC. Spherical influenza viruses have a fitness advantage in embryonated eggs, while filament-producing strains are selected in vivo. J Virol. 2013 Dec;87(24):13343-53.
34. Skehel JJ, Bayley PM, Brown EB, Martin SR, Waterfield MD, White JM, Wilson IA, Wiley DC. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc Natl Acad Sci U S A. 1982 Feb;79(4):968-72.

35. Shapiro GI, Krug RM. Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. J Virol. 1988 Jul;62(7):2285-90.
36. Shortridge KF. Pandemic influenza: a zoonosis? Semin Respir Infect. 1992 Mar;7(1):11-25.
37. Swayne DE, Alexander DJ. Confirmation of nephrotropism and nephropathogenicity of three low-pathogenic chicken-origin influenza viruses for chickens. Avian Pathol. 1994 Jun;23(2):345-52.
38. 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. New world bats harbor diverse influenza A viruses. PLoS Pathog. 2013;9(10):e1003657.
39. Tosh C, Nagarajan S, Kumar M, Murugkar HV, Venkatesh G, Shukla S, Mishra A, Mishra P, Agarwal S, Singh B, Dubey P, Tripathi S, Kulkarni DD. Multiple introductions of a reassortant H5N1 avian influenza virus of clade 2.3.2.1c with PB2 gene of H9N2 subtype into Indian poultry. Infect Genet Evol. 2016 May 10;43:173-178.
40. Trombetta C, Piccirella S, Perini D, Kistner O, Montomoli E. Emerging Influenza Strains in the Last Two Decades: A Threat of a New Pandemic? Vaccines (Basel). 2015 Mar 18;3(1):172-85.
41. Vasfi Marandi M, Current Situation of Avian Influenza in Iran and around the World. Poultry Industry-2011. (http://en.engormix.com/MA-poultry-industry/health/articles/current-situation-avian-influenza-t3018/165-p0.htm)
42. Wan H, Sorrell EM, Song H, Hossain MJ, Ramirez-Nieto G, Monne I, Stevens J, Cattoli

G, Capua I, Chen LM, Donis RO, Busch J, Paulson JC, Brockwell C, Webby R, Blanco J, Al-Natour MQ, Perez DR. Replication and transmission of H9N2 influenza viruses in ferrets: evaluation of pandemic potential. PLoS One. 2008 Aug 13;3(8):e2923.

43. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992 Mar;56(1):152-79.
44. Webster, R. G. H5 Influenza Viruses. In Influ Virol Curr Top. (2006). Edited by Y. Kawaoka. Norfolk: Caister Academic Press.
45. Wiley DC: Binding of influenza virus haemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. Biochemistry 1992, 31:9609–9621.
46. 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.
47. Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature. 1988 Jun 2;333(6172):426-31.

Chapter 2

Serological evidence of H9N2 avian influenza virus exposure among poultry workers from Fars province of Iran

Heidari A.[1,2], Mancin M.[3] Nili H.[4], Pourghanbari G.H.[4,5], KB. Lankarani[6], Leardini S.[1], Cattoli G.[1], Monne I.[1], Piccirillo A.[2]

[1]Research and Innovation Department, Istituto Zooprofilattico Sperimentale delle Venezie, OIE/FAO and National Reference Laboratory for Newcastle Disease and Avian Influenza, OIE collaborating Center for Diseases at the Human-Animal Interface, Viale dell'Università 10, 35020 Legnaro (PD), Italy.

[2]Department of Comparative Biomedicine and Food Science, University of Padua, Legnaro (PD), Italy.

[3]Food Safety Department, Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe), Viale dell'Università 10, 35020 Legnaro (PD), Italy.

[4]Avian Diseases Research Center, School of Veterinary Medicine, Shiraz University Shiraz, Iran.

[5]School of Veterinary Medicine, Ardakan University Yazd, Iran.

[6]Health Policy Research Center of Shiraz University of Medical Science Shiraz, Iran.

Abstract

Background: Since the 1990s, influenza A viruses of the H9N2 subtype have been causing infections in the poultry population around the globe. This influenza subtype is widely circulating in poultry and human cases of AI H9N2 have been sporadically reported in countries where this virus is endemic in domestic birds. The wide circulation of H9N2 viruses throughout Europe and Asia along with their ability to cause direct infection in mammals and humans, raises public health concerns. H9N2 AI was reported for the first time in Iran in 1998 and at present it is endemic in poultry. This study was carried out to evaluate the exposure to H9N2 AI viruses among poultry workers from the Fars province.

Methods: 100 poultry workers and 100 healthy individuals with no professional exposure to poultry took part in this study. Serum samples were tested for antibodies against two distinct H9N2 avian influenza viruses, which showed different phylogenetic clustering and important molecular differences, such as at the amino acid (aa) position 226 (Q/L) (H3 numbering), using haemagglutination inhibition (HI) and microneutralization (MN) assays.

Results: Results showed that 17% of the poultry workers were positive for the A/chicken/Iran/10VIR/854-5/2008 virus in MN test and 12% in HI test using the titer ≥40 as positive cut-off value. Only 2% of the poultry workers were positive for the A/chicken/Iran/12VIR/9630/1998 virus. Seroprevalence of non exposed individuals for both H9N2 strains was below 3% by both tests. Statistical analyses models showed that exposure to poultry significantly increases the risk of infection with H9N2 virus.

Conclusions: The results have demonstrated that exposure to avian H9N2 viruses had occurred among poultry workers in the Fars province of Iran. Continuous surveillance programmes should be implemented to monitor the presence of avian influenza infections in humans and to evaluate their potential threat to poultry workers and public health.

Keywords: H9N2, Avian influenza, Iran, Poultry workers, Hemagglutination inhibition (HI), Microneutralization (MN).

Background

Most emerging diseases are of zoonotic origin, with wild and domestic animals acting as natural reservoirs (Kruse *et al.*, 2004). Globalization and intensive animal farming have led to an increased spread of zoonotic infections (Cutler *et al.*, 2010). Influenza type A viruses include several distinct subtypes based on the antigenic properties of the two major surface glycoproteins, the hemagglutinin (HA) and the neuraminidase (NA). To date, 18 subtypes of HA (H1-H18) and 11 subtypes of NA (N1-N11) have been described (Tong *et al.*, 2013).

A number of influenza A subtypes have successfully crossed the species barrier and have established in the mammals and human population, causing yearly seasonal epidemics or they have sporadically been directly transmitted from poultry to humans causing zoonotic infections (Riedel *et al.*, 2006, Katz *et al.*, 2009). The influenza A viruses of the H9N2 subtype are classified as low pathogenic avian influenza (LPAI) viruses. They cause infections both in wild birds and in the poultry population worldwide, including several countries in Asia, Europe, North Africa and North America (Alexander *et al.*, 2003, Brown *et al.*, 2006). A significant proportion of recent H9N2 avian influenza (AI) isolates contains the L226Q (H3 numbering) amino acid substitution in their hemagglutinins (HAs) showing preferential binding to analogs of receptors with sialic acid linked to galactose by α2,6 linkage (SAα2,6Gal), a phenotypic portrait which is characteristic of human influenza viruses. Thus, these AI viruses might possess one of the key elements for infection in humans (Lin *et al.*, 2000, Ge FF. *et al.*, 2009, Wan *et al.*, 2007). Indeed, H9N2 viruses were isolated for the first time from humans in Hong Kong in 1999 and further human infections were reported in 2003

(Saito *et al.,* 2001, Butt *et al.,* 2005). These studies have shown that avian H9N2 viruses isolated from chickens are closely related to the H9N2 viruses responsible for human infection (Cameron *et al.,* 2000). One human case of H9N2 AI was reported in Bangladesh (Shanmuganatham *et al.,* 2013) and the World Health Organization (WHO) in 2015 has reported new cases in Egypt and Bangladesh (WHO June 2015, WHO September 2015). In 1998, domestic pigs from Hong Kong were confirmed as being infected with H9N2 influenza, and infections have been reported also in recent years in swine along with other mammals (Peiris *et al.,* 2001, Peng *et al.,* 2015). Furthermore, H9N2 viruses can contribute with gene segments during reassortment events leading to the generation of novel avian influenza virus that can infect humans (e.g. recent Chinese H7N9 and H10N8 viruses) (Gao *et al.,* 2013, Zhang *et al.,* 2014). Recent transmission studies have demonstrated that some natural isolates of H9N2 viruses can acquire the ability to transmit efficiently between ferrets via respiratory droplets. In addition, it has been reported that serial passages of an H9N2 virus through guinea pigs can result in the introduction of amino acid substitutions, which increases contact transmission efficiency in this mammalian model (Li X, *et al.,* 2014, Sang *et al.,* 2015).

The wide circulation of H9N2 viruses throughout Eurasia, along with their ability to cause direct infections in mammals and humans, raises public health concerns on their potential role as candidates for the next influenza pandemic (Li KS *et al.,* 2003). H9N2 human infection is generally asymptomatic or responsible for mild clinical signs. This may explain the scarcity of evidence accounting for the circulation and transmission of this virus subtype (Liu *et al.,* 2013). Nonetheless, human sera positive for H9 subtype were identified in China, India, Iran, Thailand, Cambodia, Romania, Egypt and Pakistan (Wang *et al.*, 2015, Pawar *et al.*, 2012, Hadipour *et al.*, 2011, Alizadeh *et al.*, 2009, Anvar *et al.*, 2013, Khuntirat *et al.*, 2011, Blair *et al.*, 2013, Coman *et al.*, 2013, Gomaa, 2015, Ahad *et al.*, 2014).

In Iran, the H9N2 subtype was identified for the first time in 1998 and is still circulating in the poultry population. In the affected farms the mortality rate ranges between 20 and 60%, although this may also be attributable to co-infections with other pathogens, such as IBV or *Mycoplasma gallisepticum* (Nili *et al.,* 2003). In spite of the implemented national control measures, which include the mass vaccination of poultry, the virus has rapidly spread and can be considered endemic in the Iranian poultry (Vasfi *et al.*, 2013).

Exposure to H9N2 AI viruses in Iranian poultry workers was previously revealed by means of HI test, using serum titre ≥20 as positive cut-off (Alizadeh *et al.,* 2009, Anvar *et al.,* 2013). In previous reported studies, the H9N2 AI seroprevalence, assessed by means of HI test, in Iranian poultry workers ranged from 1.6 to 15.7% (Median 9.5). Amongst the Middle-Eastern and Southern Asia countries, the highest seroprevalence was observed in Pakistan (47.8%) by means of HI test (Ahad *et al.,* 2014) and 7.5% by means of MN test in Egypt (Gomaa, 2015). In the present study, two different serological tests were used and compared to screen 200 individuals from the Fars province, Iran to better understand the risk of infection with the H9N2 virus in the poultry sector. The two different diagnostic tests were applied to assess whether (i) the human exposure to this virus subtype can be confirmed and whether (ii) the poultry operators in the Iranian endemic areas (the Fars province, in this study) are at risk of exposure to the H9N2 infection.

Methods

Study population and sample collection

The analyses were conducted on the blood sera of 100 workers regularly exposed to poultry (exposed group) and of 100 individuals with no professional exposure to poultry (unexposed group). The individuals, belonging to the exposed group, were recruited among the operators of the poultry industry and of the University Poultry Veterinary Hospital, none of whom had been vaccinated against flu. In particular, blood samples were collected between September – December 2012 from 70 slaughterhouse workers, 30 poultry house workers and from 10 operators of the University Poultry Veterinary Hospital, all resident in the Fars province (southern Iran). In January 2013, 100 sera were also collected from subjects from the same province not professionally exposed to poultry. 98% of the individuals included in the study were males, whose mean age among the poultry workers was of 32 years (ranging between 18–62). The mean age in the professionally unexposed group was of 41 years (range between 18–67). After blood collection the serum was separated, stored at –20 °C and subsequently sent to IZSVe for the serological analyses.

Ethics statement

The use of serum samples for research purposes was approved by the Veterinary University of Shiraz in compliance with the Iranian ethical principles and institutions and according to the principles expressed in the World Medical Association Declaration of Helsinki. Informed consent was obtained from all participating individuals.

Selection of antigens and sera

To select representative H9 antigens to be used in the serological investigations, the HA gene nucleotide and amino acid sequences of five Iranian H9N2 AIV isolates available at IZSVe repository were analysed and compared with the Iranian sequences available in the

GenBank. Subsequently, a nucleotide sequence dataset of 88 Iranian strains along with representative strains of G1 lineage were aligned and phylogenetically analysed constructing a Neighbour-Joining phylogenetic tree by the MEGA 5 (http://www.megasoftware.net) program (see results). In addition, sera of all individuals were tested for the presence of antibodies to seasonal A/Minnesota/ 11/2010 H3N2 and A/California/4/2009 H1N1pdm 2009 viruses. Positive control sera containing antibodies directed against the selected H9N2 antigens were produced in SPF chickens. In addition, a panel of different H9N2, H1N1, H3N2 chicken sera and an H1N1, H3N2 positive human sera were used as controls in the serological assays.

Serological methods

Hemagglutinin inhibition (HI) test

The HI assay was applied according to the World Health Organization (WHO) and World Organisation for Animal Health (OIE) manual (WHO Manual on Animal Influenza Diagnosis and Surveillance, OIE 2015 Chapter 2.8.8). Briefly, 0.5% (vol/vol) chicken red blood cell (RBC) solution was prepared by washing with Phosphate buffered saline (PBS) containing 0.5% Bovine serum albumin (BSA) (Sigma). To remove non specific serum inhibitors, the human serum samples were treated with RDE (receptor destroying enzyme) (Sigma-Aldrich) that was reconstituted with 5 ml sterile distilled water. 50 µl of serum was added to 200 µl of RDE diluted to 100 ml with calcium saline, PH 7.2 and incubated overnight at 37 °C. Then 5 vol of 1.5% sodium citrate were added then heated at 56 °C for 30 min to inactivate remaining RDE. The treated sera were tested by haemagglutination assay to verify the presence of non-specific agglutination. After that the treated serum (50 µl of each one) was diluted in two-fold serial dilutions (1:10) with 25 µl PBS in 96-well V-bottom microtiter plates. Subsequently, 25

µl of virus antigens containing 4 HA units were added to the wells and incubated at room temperature (RT) for 30–45 min. 50 µl of 0.5% chicken RBCs were then added and the plate was incubated further for 45 min at RT before recording the agglutination titers. The HI test results were expressed as the reciprocal of the last dilution of the sample that completely inhibited haemagglutination. In all the assays control positive serum samples were included, and all assays were tested in triplicate.

Microneutralization (MN) test

The WHO MN test protocol was applied (WHO protocol of Serological diagnosis of influenza by microneutralization assay). The human serum samples were inactivated by heating at 56 °C for 30 min H9N2, H3N2 and H1N1pdm virus stocks used in our analyses were titrated in the presence of TPCK-trypsin (2 µg/ml) (Sigma-Aldrich) for determination of tissue culture infectious dose (TCID50). 10 µl of treated sera were added to 96 well cell culture plates (Costar) and 2-fold serial dilutions were performed. Inoculum was prepared in virus diluent with the addition of TPCK-trypsin (2 µg/ml) so that 50 µl contained 100 TCID50 of the respective virus. Fifty microliters of virus inoculum was added to the wells, the virus serum mixture was subsequently incubated for 1 hr at 37 °C, 5% CO2. A back-titration of the virus inoculum was performed in each assay using 2 fold serial dilutions. 100 µl MDCK cells (1.5×104 cells /well) were then added to each well and the plate was incubated overnight at 37 °C, 5% CO2 (18–20 hrs). The plate was then fixed with 100 µl/well of cold fixative (80% Acetone in PBS) for 10 min. The virus was detected with an anti-Influenza A, Nucleoprotein monoclonal antibody (Merck Millipore) and Peroxidase conjugated-goat anti-mouse IgG (γ) (KPL) as secondary antibody using ELISA. All serum samples and controls were tested in triplicate.

Criteria for seropositivity

The WHO guidelines for vaccine evaluation suggest that a neutralizing antibody titre ≥40 indicate higher than 50% protection against influenza A virus infection or disease. Based on this consideration, an individual with an antibody titer ≥40 was considered positive for different serotype in MN and HI tests (Eichelberger *et al.*, 2007).

Statistical Analyses

The Geometric Mean titers (GMTs) of the triple test per method (HI and MN) was calculated for each subject checked for the virus of interest. A binary variable was created to identify the positive/negative sample. An average titre lower than 40 identified negative samples, whereas an average titre higher than 40 identified positive samples. The Chi square test was used to verify the possible association between positivity and exposure, for each method and virus. The GLM (Generalized linear model) for binary responses was used to estimate the probability to have a positive reaction for each tested virus, considering simultaneously the methods (HI and MN), the groups (exposed and unexposed) and their interaction as variables. The Maximum Likelihood method was used to estimate the parameters of the model. The Type III F-tests were applied to evaluate the overall effect of specified variables in the model. P-values lower than 0.10 were considered as significant (Dohoo *et al.*, 2010) SAS 9.3 software was used to fit the statistical analysis.

Results

The HA gene phylogenetic analysis of 88 Iranian H9N2 strains collected between 1998-2014 shows that they belong to two different groups of G1 lineage, named sublineage A and sub-lineage B for the purpose of this study (Fig. 1). Sub-lineage A refers to the H9N2 isolates collected between 1998 and 2007, which apparently are no longer circulating among the Iranian poultry, while B includes more recent isolates collected between 2003 and 2014. The amino acid sequence analyses of the Iranian H9N2 viruses showed that sub-lineage A consists of 47% (14/30) of the strains with amino acid Q at position 226 (H3 numbering) and 53% (16/30) with L226 substitution. Interestingly, sub-lineage B encloses mostly (95%) (55/58) the strains with amino acid L226. Based on these results, two Iranian H9N2 isolates representatives of sub-lineages A (A/chicken/Iran/12VIR/9630/ 1998) and B (A/chicken/Iran/10VIR/854-5/2008) were selected as antigens for the serological study (Fig. 1). In particular, A/chicken/Iran/12VIR/9630/1998 was characterized by the amino acid Q226 while the A/chicken/ Iran/10VIR/854-5/2008 virus had amino acid L at this position. The HA amino acid sequence distance between the two selected strains was 8.1%.

The percentage of seropositive individuals against different influenza viruses obtained with the HI and MN tests along with the p-value of the chi-square test related to the association between exposure and response to each virus tested, per method are reported in Table 1. Detailed results of serological investigations obtained for the exposed and unexposed groups against different influenza viruses are described in the Table 2. Serological results showed that the prevalence of antibodies against A/ chicken/Iran/12VIR/9630/1998 H9N2 (Q 226) was of 2% in the exposed group by the HI and MN tests, while no positive results were identified in the unexposed group.

An association was observed between professional exposure to poultry species and the presence of antibodies to the A/chicken/Iran/10VIR/854-5/2008 H9N2 (L 226) virus, as revealed by the HI and MN test ($p < 0.05$). The percentage of positivity was of 12% for the exposed group and 2% for the unexposed group in the HI test, while it respectively amounted to 17% and 3% for the exposed and unexposed groups in the MN test. Antibody prevalence to H1N1pdm 2009 was of 28% in the exposed versus 53% in the unexposed groups by the HI test, and 25% versus 33%, respectively, by the MN test. For H3N2, the prevalence was 36% in the exposed group compared to 44% in the unexposed group with the HI test, while it was 28% and 24% in the exposed and unexposed groups with the MN test. The titer distribution percentage against different viruses in different tests (Fig. 2) showed that with the increase of the titre range from ≥40 to ≥80 and from ≥80 to ≥160, the percentage of antibody positivity drastically decreased. Cross MN and HI tests on homologues and different chicken H9N2 hyper immune sera used as positive controls showed the absence of cross reactivity with H1N1pdm 2009 and H3N2 seasonal viruses; the same results were observable in human immune sera for H1N1pdm 2009 and H3N2, which showed the absence of a cross reactivity with the two tested H9N2 viruses.

The probability to be positive against A/chicken/Iran/ 10VIR/854-5/2008 H9N2, H3N2 and H1N1pdm 2009 viruses in different tests was calculated. The generalized linear model for binary responses applied on the H9N2 viruses indicated that the association between professional exposure or absence of professional exposure and H9 infection was significant ($p = 0.0018$). In particular, the exposed and unexposed groups showed a significantly different positivity both in the HI test (9% vs. 1%) and in the MN test (13% vs. 2%). For the H3N2 virus, the professional exposure was not significantly associated to the presence of antibodies ($p = 0.22$). This means that in both groups the probability of testing positive is almost similar.

Differently, the unexposed group showed a significantly higher number of individuals ($p = 0.001$) seropositive to H1N1pdm 2009. If we consider that the H1N1pdm 2009 subtype does not circulate in avian species, it can be assumed that the serological data obtained for this subtype do not depend on the exposure to poultry species; other factors may have been at the origin of the distinct immunoreactivity observed against the H1N1pdm 2009 subtype in the two groups under study.

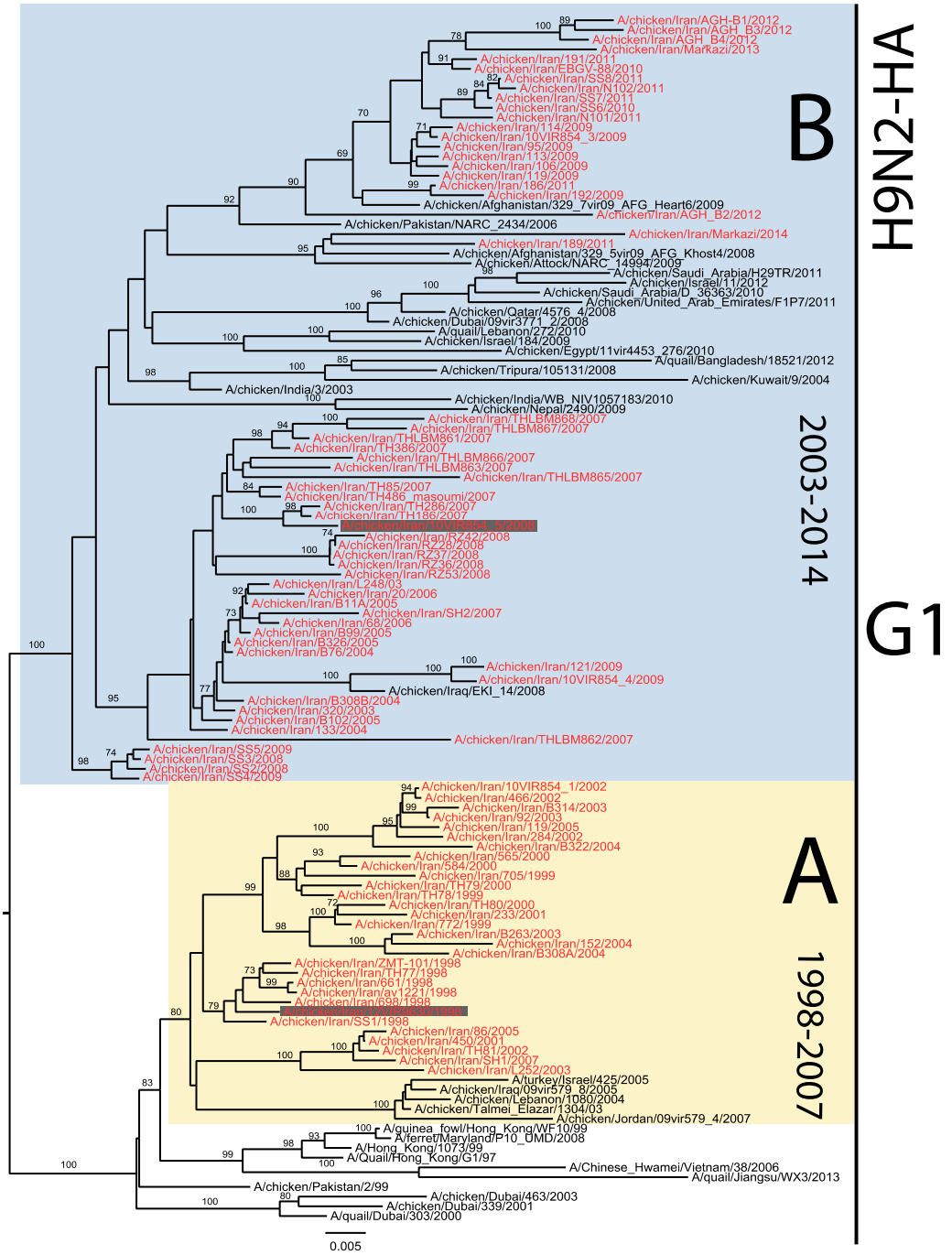


Figure 1. Neighbour-Joining Phylogenetic tree with 1000 bootstrap of the AI H9N2 HA gene of the G1 lineage. Sub-lineages A and B are highlighted with yellow and blue frames respectively, Iranian Sequences are highlighted in red and two Iranian strains used in serological analyses are highlighted with gray square. Numbers at the nodes represent bootstrap values (≥60).

Triple test average (100 sample subjects)

Virus	HI			MN		
	% Positive Exposed group	% Positive Unexposed group	p-value	% Positive Exposed group	% Positive Unexposed group	p-value
H9N2 A/chicken/Iran/12VIR/9630/1998	2	0	0.01	2	0	0.01
H9N2 A/chicken/Iran/10VIR/854-5/2008	12	2	0.005	17	3	0.001
H1N1pdm 2009 A/California/4/2009	28	53	0.0003	25	33	0.02
H3N2 A/Minnesota/11/2010	36	44	0.02	28	24	0.05

Table 1. Positive titer percentages and p-value of the chi square test related to the association between exposure and response to each virus tested, per method.

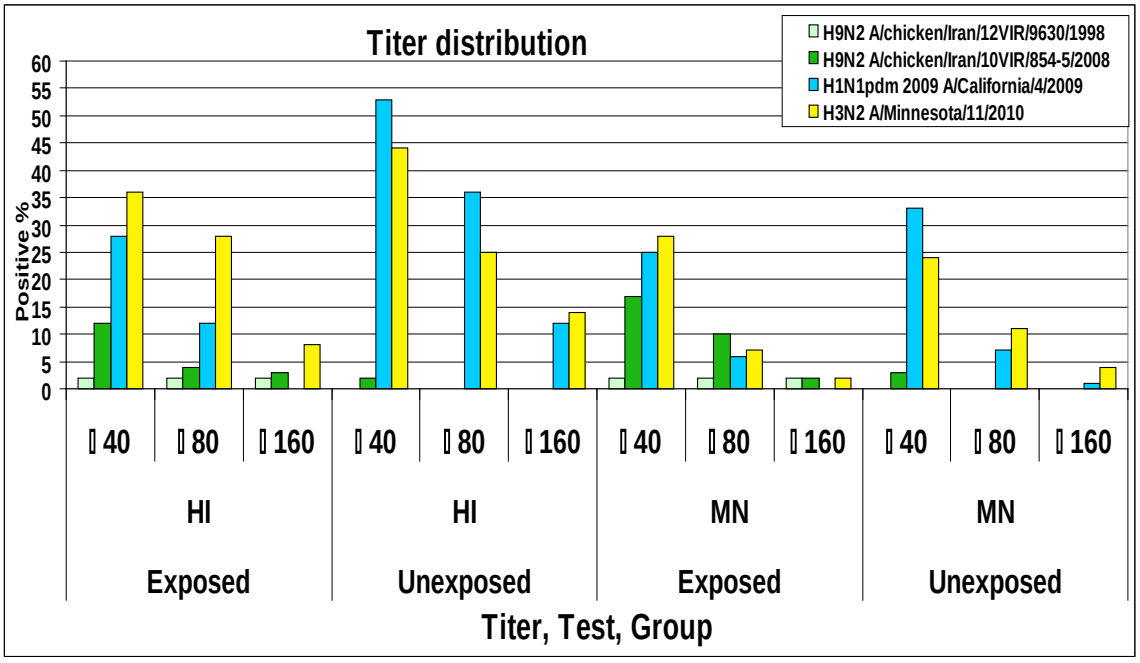


Figure 2. Graphic distribution of titer categories expressed in percentage among the subjects of the exposed and unexposed groups in the HI and MN tests.

Virus, Age group, Work group	Subjects		HI			MN		
	Exposed subjects NO.	Unexposed subjects NO.	Positivity% in Exposed (subjects NO.)	Positivity % in Unexposed (subjects NO.)	p-value	Positivity% in Exposed (subjects NO.)	Positivity% in Unexposed (subjects NO.)	p-value
H9N2 A/chicken/Iran/12VIR/9630/1998								
Positive%								
Age group (Years)								
18-39	**80**	**50**	**1 (1)**	**0**	**NS**	**1 (1)**	**0**	**NS**
40-60	**20**	**50**	**5 (1)**	**0**	**0.1**	**5 (1)**	**0**	**0.1**
Work category								
Poultry House Worker	**20**	**-**	**0**	**NO**	**NO**	**0**	**NO**	**NO**
Slaughterhouse Worker	**70**	**-**	**1.43 (1)**	**NO**	**NO**	**1.43 (1)**	**NO**	**NO**
Vet Student in poultry Hospital	**10**	**-**	**10 (1)**	**NO**	**NO**	**10 (1)**	**NO**	**NO**
Positive% ≥40	**100**	**100**	**2 (100)**	**0 (100)**		**2 (100)**	**0 (100)**	
Positive% ≥80	**100**	**100**	**2 (100)**	**0 (100)**		**2 (100)**	**0 (100)**	
Positive% ≥160	**100**	**100**	**2 (100)**	**0 (100)**		**2 (100)**	**0 (100)**	
TOT Positive%	**100**	**100**	**2 (100)**	**0 (100)**	**NS**	**2 (100)**	**0 (100)**	**NS**
H9N2 A/chicken/Iran/10VIR/854-5/2008								
Positive%								
Age (Years)								
18-39	**80**	**50**	**13.7 (11)**	**2(1)**	**0.02**	**17.5 (14)**	**2 (1)**	**0.007**
40-60	**20**	**50**	**5 (1)**	**2(1)**	**NS**	**15 (3)**	**4 (2)**	**NS**
Work category								
Poultry House Worker	**20**	**-**	**5 (1)**	**NO**	**NO**	**15 (3)**	**NO**	**NO**
Slaughterhouse Worker	**70**	**-**	**14.2 (10)**	**NO**	**NO**	**17 (12)**	**NO**	**NO**

Vet Student in poultry Hospital	**10**	**-**	**10 (1)**	**NO**	**NO**	**20 (2)**	**NO**	**NO**
Positive% ≥40	**100**	**100**	**12 (100)**	**2 (100)**		**17 (100)**	**3 (100)**	
Positive% ≥80	**100**	**100**	**4 (100)**	**0 (100)**		**10 (100)**	**0 (100)**	
Positive% ≥160	**100**	**100**	**3 (100)**	**0 (100)**		**2 (100)**	**0 (100)**	
TOT Positive%	**100**	**100**	**12 (100)**	**2 (100)**	**0.005**	**17 (100)**	**3 (100)**	**0.001**
H1N1pdm 2009								
A/California/4/2009								
Positive%								
Age (Years)								
18-39	**80**	**50**	**35 (28)**	**48(24)**	**NS**	**30 (24)**	**28(14)**	**NS**
40-60	**20**	**50**	**0 (0)**	**58(29)**	**<0.0001**	**5 (1)**	**38(19)**	**0.005**
Work category								
Poultry House Worker	**20**	**-**	**40 (8)**	**NO**	**NO**	**40 (8)**	**NO**	**NO**
Slaughterhouse Worker	**70**	**-**	**27 (19)**	**NO**	**NO**	**23 (16)**	**NO**	**NO**
Vet Student in poultry Hospital	**10**	**-**	**10 (1)**	**NO**	**NO**	**10 (1)**	**NO**	**NO**
Positive% ≥40	**100**	**100**	**28 (100)**	**53 (100)**		**25 (100)**	**33 (100)**	
Positive% ≥80	**100**	**100**	**1 (100)**	**36 (100)**		**6 (100)**	**7 (100)**	
Positive% ≥160	**100**	**100**	**0 (100)**	**12 (100)**		**0 (100)**	**1 (100)**	
TOT Positive%	**100**	**100**	**28 (100)**	**53 (100)**	**0.0003**	**25 (100)**	**33 (100)**	**NS**
H3N2								
A/Minnesota/11/2010								
Positive%								
Age (Years)								
18-39	**80**	**50**	**41 (33)**	**46 (23)**	**NS**	**32 (26)**	**24 (12)**	**NS**
40-60	**20**	**50**	**15 (3)**	**42 (21)**	**0.03**	**10 (2)**	**24 (12)**	**NS**
Work category								
Poultry House Worker	**20**	**-**	**40 (8)**	**NO**	**NO**	**20 (4)**	**NO**	**NO**
Slaughterhouse Worker	**70**	**-**	**35 (25)**	**NO**	**NO**	**32 (23)**	**NO**	**NO**
Vet Student in poultry Hospital	**10**	**-**	**30 (3)**	**NO**	**NO**	**10 (1)**	**NO**	**NO**

Positive% ≥40	**100**	**100**	**36 (100)**	**44 (100)**		**28 (100)**	**24 (100)**	
Positive% ≥80	**100**	**100**	**28 (100)**	**25 (100)**		**7 (100)**	**11 (100)**	
Positive% ≥160	**100**	**100**	**8 (100)**	**14 (100)**		**2 (100)**	**4 (100)**	
TOT Positive%	**100**	**100**	**36 (100)**	**44 (100)**	**NS**	**28 (100)**	**24 (100)**	**NS**

Table 2. Results of serological investigations obtained for the different categories of exposed and unexposed groups against different influenza viruses. NS: not significant**,** NO: No data.

Discussion and conclusion

Results demonstrated that exposure to avian H9N2 viruses had occurred in humans in Iran (Fars province). In particular, the analysis showed a significantly higher prevalence of neutralizing antibodies against the A/chicken/Iran/10VIR/854-5 H9N2 virus in poultry workers than in the professionally unexposed group. These results are in accordance with a previously reported study from Iran by Alizadeh *et al.,* 2009 confirming that poultry workers are at risk of infection from zoonotic avian influenza virus. For the first time in this study, two distinct H9N2 LPAI viruses were used as antigens for serological investigations in humans. Interestingly, neutralizing activity against just one of these two antigens was revealed in several sera, which has resulted in significant different prevalence. This is consistent with the amino acid diversity of the two H9N2 antigens used in this study. In fact, the HA amino acid sequences of A/chicken/Iran/12VIR/ 9630/1998 (Q-226) and A/chicken/Iran/10VIR/854-5/ 2008 (L-226) differed for 43 amino acids, seven of which were located in antigenic sites (S158N, S160L, S175D, E190A, S193N, N208D, Q226L (H3 numbering)) previously reported by other groups (Kaverin *et al.*, 2004, Zhu *et al.,* 2015, Okamatsu *et al.,* 2008, Yasugi *et al.,* 2013). These amino acid substitutions may account for the different serological reactivity against these antigens. However, the difference in prevalence observed in this study between two tested H9N2 viruses might have different explanations.

The receptor binding properties of A/chicken/Iran/ 12VIR/9630/1998 H9N2 (Q-226)-like viruses affect their ability to infect humans. As a consequence, the infections of this virus in the exposed population are limited. On the contrary, the A/chicken/Iran/10VIR/854-5 H9N2 (L-226) – like viruses, have increased ability to infect humans resulting in higher prevalence in the exposed population. However, it should be considered that the molecular analysis of Iranian H9N2 viruses demonstrated that those harbouring the Q-226 mutations apparently are

no longer in circulation. Furthermore, one previous study investigating the serologic response in humans exposed to avian viruses concluded that the incidence of seroconversion is low and that the antibody response after mild/asymptomatic infections is short-lived (Eichelberger *et al.,* 2007). Therefore, the reduction of the circulation of Q-226 viruses in the poultry in recent times and/or the low antibody response after infection might be responsible for the decreased prevalence of antibodies against these AI viruses.

Consistently with the results discussed so far, the unexposed group counting 100 healthy subjects showed a very low percentage of positivity for H9N2 by the HI and MN tests, while the distribution of positive titres against H1N1pdm 2009 and seasonal H3N2 influenzas were similar in the exposed and in the unexposed groups.

According to Stephenson *et al.,* 2003 the cross reacting antibodies against H2 might explain the antibody reactivity against avian H9N2, especially among people born before 1968. Although we did not test the sera for this antigen, to evaluate this hypothesis, the age of each individual who had tested H9 positive was verified. Only one sample was from a person born before 1968, while all the others were collected from younger subjects with an age between 22-46 years, who could not be positive for H2N2.

The comparison between the serological results obtained using the HI and MN tests against H9N2 showed that the MN test detected more positive samples within the same group, however, the triple test results proved that the HI titer was reasonably more consistent than the MN test. In addition, different researches have reported that intra-laboratory reproducibility by MN is lower than the one resulting from the HI test (Stephenson *et al.,* 2009), a fact that may explain the moderate discrepancy observed between the two serological assays applied for the detection of antibodies against H1N1pdm 2009 and H3N2 seasonal influenza. Both in the HI and MN tests, the percentage of positivity for H9N2 decreased with

the increase of the titre range; the obtained results were justified on the grounds that humans have a poor immune response to infections with avian influenza viruses. For instance, studies have shown that the individuals with mild or asymptotic infections had a lower antibody titer compared to individuals who had seriously developed the illness (Buchy *et al.,* 2010). H9N2 AI viruses have some of the specific molecular markers for the infection of mammals; however, they are still unable to completely adapt to mammals and to cause serious disease in humans, which may explain the low titers observed. As well, the time elapsed between the infection and sampling dates can also account for the decrease of the antibody titer.

In conclusion, this study has highlighted the potential of avian to human transmission of H9N2 AIVs, and indicated that poultry workers are at risk of infection. Avian transmission of H9N2 viruses in humans can increase the probability of human adaption, while genetic reassortment with other human seasonal viruses are means of generating an influenza virus with epidemic or pandemic potential. Hence, integrated medical-veterinary surveillance and research activities are essential in order to identify the emergence of new influenza viruses, assess the clinical significance of seropositivity and understand more on the mechanisms that favour the virus to cross the species barrier. In the poultry sector, surveillance and control programmes should be implemented to reduce the prevalence of H9N2 AIV in poultry population and minimize the risk of exposure in poultry operators.

Abbreviations

aa: amino acid;

AI: Avian Influenza;

GMTs: geometic mean titers;

HA: hemagglutinin;

HI: hemagglutinin inhibition;

MN: microneutralization assay;

MDCK: Madin-Darby canine kidney;

OIE: Organisation for Animal Health;

pdm: pandemic;

WHO: World Health Organization;

Competing interests and authors' contributions

The authors declare that they have no competing interests.

HA furnished intellectual contributions, was involved in sample testing, date analysis and drafted the manuscript. NH, LKB and PGH were involved in sample and data collection. MM provided the statistical analyses support. MI and CG supervised the laboratory analysis and the results interpretation. LS provided the technical support and PA contributed to the results interpretation. All authors read and approved the final manuscript.

Acknowledgements

Studies contained in this research project were carried out mainly at the Research & Innovation Department, Division of Biomedical Science, Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy.

This project was financially supported by the NoFlu project, Fondazione Cariplo Vaccine Program (grant number 2009-3594). The authors would like to acknowledge Dr. Angela Trocino (Department of Comparative Biomedicine and Food Science, University of Padua, Italy) for her contribution in the statistical analysis, Daniele Facco, Lorenza Boscolo and Silvia Maniero (DSBIO, IZSVe, Padova, Italy) for technical support and Francesca Ellero (DSBIO, IZSVe, Padova, Italy) for providing language help. In addition, we gratefully acknowledge the contributing authors and the originating and submitting laboratories for the sequences from the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database. This study was conducted in the framework of the Doctoral school in Veterinary Science at the University of Padua (Alireza Heidari).

References

1. Alexander DJ. Report on avian influenza in the Eastern Hemisphere during 1997-2002. Avian Dis. 2003;47(3 Suppl):792–7.
2. Ahad A, Thornton RN, Rabbani M, Yaqub T, Younus M, Muhammad K, et al. Risk factors for H7 and H9 infection in commercial poultry farm workers in provinces within Pakistan. Prev Vet Med. 2014;117(3-4):610–4.
3. Alizadeh E, Kheiri MT, Bashar R, Tabatabaeian M, Hosseini SM. Avian Influenza (H9N2) among poultry workers in Iran. Iranian Journal of Microbiology. 2009;1:3–6.
4. Anvar E, Hosseini SM, Tavasoti Kheiri M, Mazaheri V, Fazaei K, Shabani M, et al. Serological Survey of Avian Influenza (H9N2) Among Different Occupational Groups in Tehran and Qazvin Provinces in IR Iran. Jundishapur J Microbiol. 2013;6(4):e5441.
5. Blair PJ, Putnam SD, Krueger WS, Chum C, Wierzba TF, Heil GL, et al. Evidence for avian H9N2 influenza virus infections among rural villagers in Cambodia. J Infect Public Health. 2013;6(2):69–79.
6. Brown IH, Banks J, Manvell RJ, Essen SC, Shell W, Slomka M, et al. Recent epidemiology and ecology of influenza A viruses in avian species in Europe and the Middle East. Dev Biol (Basel). 2006;124:45–50.
7. Butt KM, Smith GJ, Chen H, Zhang LJ, Leung YH, Xu KM, et al. Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. J Clin Microbiol. 2005;43(11):5760–7.
8. Buchy P, Vong S, Chu S, Garcia JM, Hien TT, Hien VM, et al. Kinetics of neutralizing antibodies in patients naturally infected by H5N1 virus. PLoS One. 2010;5(5):e10864.

9. Cameron KR, Gregory V, Banks J, Brown IH, Alexander DJ, Hay AJ, et al. H9N2 subtype influenza A viruses in poultry in pakistan are closely related to the H9N2 viruses responsible for human infection in Hong Kong. Virology. 2000;278(1):36–41.
10. Coman A, Maftei DN, Krueger WS, Heil GL, Friary JA, Chereches RM, et al. Serological evidence for avian H9N2 influenza virus infections among Romanian agriculture workers. J Infect Public Health. 2013;6(6):438–47.
11. Cutler SJ, Fooks AR, van der Poel WH. Public health threat of new, reemerging, and neglected zoonoses in the industrialized world. Emerg Infect Dis. 2010;16(1):1–7.
12. Dohoo IR, Martin W, Stryhn H. Veterinary epidemiologic research. 2nd ed. University of Prince Edward Island, Prince Edward Island, Canada: Atlantic Veterinary College Inc; 2010. p. 865.
13. Eichelberger M, Golding H, Hess M, Weir J, Subbarao K, Luke CJ, et al. FDA/NIH/WHO public workshop on immune correlates of protection against influenza A viruses in support of pandemic vaccine development, Bethesda, Maryland, US, December 10-11, 2007. Vaccine. 2008;26(34):4299–303.
14. Gao R, Cao B, Hu Y, Feng Z, Wang D, Hu W, et al. Human infection with a novel avian-origin influenza A (H7N9) virus. N Engl J Med. 2013;368(20):1888–97.
15. Ge FF, Zhou JP, Liu J, Wang J, Zhang WY, Sheng LP, et al. Genetic evolution of H9 subtype influenza viruses from live poultry markets in Shanghai, China. J Clin Microbiol. 2009;47(10):3294–300.
16. Gomaa MR, Kayed AS, Elabd MA, Zeid DA, Zaki SA, El Rifay AS, et al. Avian influenza A(H5N1) and A (H9N2) seroprevalence and risk factors for infection among Egyptians: a prospective, controlled seroepidemiological study. J Infect Dis. 2015;211(9):1399–407.

17. Hadipour MM, Pazira S. Evaluation of Antibody Titers to H9N2 Influenza Virus in Hospital Staff in Shiraz, Iran. Journal of Animal and Veterinary Advance. 2011;10(7):832–4.
18. Kaverin NV, Rudneva IA, Ilyushina NA, Lipatov AS, Krauss S, Webster RG. Structural differences among hemagglutinins of influenza A virus subtypes are reflected in their antigenic architecture: analysis of H9 escape mutants. J Virol. 2004;78(1):240–9.
19. Katz JM, Veguilla V, Belser JA, Maines TR, Van Hoeven N, Pappas C, et al. The public health impact of avian influenza viruses. Poult Sci. 2009;88(4):872–9.
20. Kruse H, Kirkemo AM, Handeland K. Wildlife as source of zoonotic infections. Emerg Infect Dis. 2004;10(12):2067–72.
21. Khuntirat BP, Yoon IK, Blair PJ, Krueger WS, Chittaganpitch M, Putnam SD, et al. Evidence for subclinical avian influenza virus infections among rural Thai villagers. Clin Infect Dis. 2011;53(8):e107–16.
22. Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, et al. Avian-to human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci U S A. 2000;97(17):9654–8.
23. Li X, Shi J, Guo J, Deng G, Zhang Q, Wang J, et al. Genetics, receptor binding property, and transmissibility in mammals of naturally isolated H9N2 Avian Influenza viruses. PLoS Pathog. 2014;10(11):e1004508.
24. Li KS, Xu KM, Peiris JS, Poon LL, Yu KZ, Yuen KY, et al. Characterization of H9 subtype influenza viruses from the ducks of southern China: a candidate for the next influenza pandemic in humans? J Virol. 2003;77(12):6988–94.
25. Liu Q, Liu DY, Yang ZQ. Characteristics of human infection with avian influenza viruses and development of new antiviral agents. Acta Pharmacol Sin. 2013;34(10):1257–69.

26. Nili H, Asasi K. Avian influenza (H9N2) outbreak in Iran. Avian Dis. 2003;47(3):828–31.
27. Okamatsu M, Sakoda Y, Kishida N, Isoda N, Kida H. Antigenic structure of the hemagglutinin of H9N2 influenza viruses. Arch Virol. 2008;153(12):2189–95.
28. Organisation for Animal Health (OIE), Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2015 Chapter 2.8.8. (http://www.oie.int/ fileadmin/Home/eng/Health_standards/tahm/2.08.08_SWINE_INF.pdf)
29. Peiris JS, Guan Y, Markwell D, Ghose P, Webster RG, Shortridge KF. Cocirculation of avian H9N2 and contemporary "human" H3N2 influenza A viruses in pigs in southeastern China: potential for genetic reassortment? J Virol. 2001;75(20):9679–86.
30. Peng L, Chen C, Kai-yi H, Feng-xia Z, Yan-li Z, Zong-shuai L, et al. Molecular characterization of H9N2 influenza virus isolated from mink and its pathogenesis in mink. Vet Microbiol. 2015;176(1-2):88–96.
31. Pawar SD, Tandale BV, Raut CG, Parkhi SS, Barde TD, Gurav YK, et al. Avian influenza H9N2 seroprevalence among poultry workers in Pune, India, 2010. PLoS One. 2012;7(5):e36374.
32. Riedel S. Crossing the species barrier: the threat of an avian influenza pandemic. Proc (Bayl Univ Med Cent). 2006;19(1):16–20.
33. Saito T, Lim W, Suzuki T, Suzuki Y, Kida H, Nishimura SI, et al. Characterization of a human H9N2 influenza virus isolated in Hong Kong. Vaccine. 2001;20(1-2):125–33.
34. Sang X, Wang A, Ding J, Kong H, Gao X, Li L, et al. Adaptation of H9N2 AIV in guinea pigs enables efficient transmission by direct contact and inefficient transmission by respiratory droplets. Sci Rep. 2015;5:15928.

35. Shanmuganatham K, Feeroz MM, Jones-Engel L, Smith GJ, Fourment M, Walker D, et al. Antigenic and molecular characterization of avian influenza A(H9N2) viruses, Bangladesh. Emerg Infect Dis. 2013;19(9):1393–1402.
36. Stephenson I, Nicholson KG, Glück R, Mischler R, Newman RW, Palache AM, et al. Safety and antigenicity of whole virus and subunit influenza A/Hong Kong/1073/99 (H9N2) vaccine in healthy adults: phase I randomised trial. Lancet. 2003;362(9400):1959–66.
37. Stephenson I, Heath A, Major D, Newman RW, Hoschler K, Junzi W, et al. Reproducibility of serologic assays for influenza virus A (H5N1). Emerg Infect Dis. 2009;15(8):1252–9.
38. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, et al. New world bats harbor diverse influenza A viruses. PLoS Pathog. 2013;9(10):e1003657.
39. Vasfi Marandi M. Current Situation of Avian Influenza in Iran and around the World. Poultry Industry. 2013 http://en.engormix.com/MA-poultry-industry/health/articles/current-situation-avian-influenza-t3018/165-p0.htm
40. Wan H, Perez DR. Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells. J Virol. 2007;81(10):5181–91.
41. Wang Q, Ju L, Liu P, Zhou J, Lv X, Li L, et al. Serological and Virological Surveillance of Avian Influenza A Virus H9N2 Subtype in Humans and Poultry in Shanghai, China, Between 2008 and 2010. Zoonoses Public Health. 2015;62(2):131–40.
42. World health organization (WHO) Influenza at the human-animal interface, Summary and assessment as of 23 June2015. http://www.who.int/influenza/

human_animal_interface/Influenza_Summary_IRA_HA_interface_23_June_2015.pdf?ua=1

43. World health organization (WHO) Influenza at the human-animal interface, Summary and assessment as of 4 September 2015. http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_04_September_2015.pdf
44. World Health Organization (WHO), Manual on Animal Influenza Diagnosis and Surveillance. http://www.who.int/csr/resources/publications/influenza/en/whocdscsrncs20025rev.pdf
45. World Health Organization (WHO), protocol of Serological diagnosis of influenza by microneutralization assay. http://www.who.int/influenza/gisrs_laboratory/2010_12_06_serological_diagnosis_of_influenza_by_microneutralization_assay.pdf
46. Yasugi M, Kubota-Koketsu R, Yamashita A, Kawashita N, Du A, Misaki R, et al. Emerging antigenic variants at the antigenic site Sb in pandemic A(H1N1)2009 influenza virus in Japan detected by a human monoclonal antibody. PLoS One. 2013;8(10):e77892.
47. Zhang T, Bi Y, Tian H, Li X, Liu D, Wu Y, et al. Human infection with influenza virus A(H10N8) from live poultry markets, China, 2014. Emerg Infect Dis. 2014;20(12):2076–9.
48. Zhu Y, Yang D, Ren Q, Yang Y, Liu X, Xu X, et al. Identification and characterization of a novel antigenic epitope in the hemagglutinin of the escape mutants of H9N2 avian influenza viruses. Vet Microbiol. 2015;178(1-2):144–9.

Chapter 3

Phylogenetic, phylogeographic and structural bioinformatic approach to the evolution and spreading of H9N2 avian influenza virus.

Alireza Heidari,[1,3] Adelaide Milani,[1] Alice Fusaro,[1] Irene Righetto,[2] Isabella Monne, [1] Giovanni Cattoli,[1] Francesco Filippini[2]

[1] FAO-OIE and National Reference Laboratory for Newcastle Disease and Avian Influenza, Istituto Zooprofilattico delle Venezie (IZSVe), Legnaro, Italy

[2] Molecular Biology and Bioinformatics Unit (MOLBINFO), Department of Biology, University of Padua, Padova, Italy

[3] Department of Comparative Biomedicine and Food Science, University of Padua, Legnaro (PD), Italy

Adelaide Milani and Alireza Heidari contributed equally to this work.

Abstract

Influenza A virus is a zoonotic agent with a significant impact both on public health and poultry industry and avian influenza H9N2 virus provided the records of switch to human host. Therefore, surveillance and characterization are needed by the scientific community and public health systems and so far this was mainly based on intensive serological characterization and phylogenetic analyses aimed to infer evolutionary trends. In order to aid molecular epidemiologic assessment and support public health interventions, as well as to properly relate investigations worldwide thanks to shared nomenclature and robust guidelines, we developed a method for the clade nomenclature of all AI H9N2 hemagglutinin subtypes, based on the evolutionary dynamics of a large and non redundant viral strain dataset. This was combined to a phylogeographic analysis providing further information on the spatiotemporal evolution, correlation and spreading of H9N2 viruses from the beginning to current trends.

We found that H9N2 viruses can be clustered in five classes based on congruence of phylogenetic and phylogeographic data with structural comparison evidence. Structural analyses can properly depict structural closeness among proteins or protein domains and provide functional insights on surface regions possibly crucial to antigenicity and cell binding. Recent successful inference of surface feature fingerprints for H5N1 evolution prompted us to assess whether such fingerprints are peculiar to H5N1, or electrostatic variation could be associated to the evolution and spreading of other avian influenza viruses, e.g. the newly defined classes and clades from the H9N2 subtype.

Finally, surface feature fingerprints could be inferred that relate class and clade specific variation in electrostatic charges and isocontour to well-known hemagglutinin sites involved in modulation of immune escape and host specificity. Results from this work suggest the integration of up-to-date phylogenetic and phylogeographic analyses with sequence-based and

structural investigation of surface features as a front-end strategy for inferring trends and relevant mechanisms in influenza virus evolution.

Running Head: Phylogenetic-bioinformatic study of H9N2 evolution.

Background

Influenza A virus is a zoonotic agent with a significant impact both on public health and poultry industry. Wild waterfowl are primary reservoirs of avian influenza viruses (AIV) that in addition to birds can sporadically infect mammalian hosts including humans. This is suggestive for setting up a coordinated global surveillance network (Butler, 2012) as well as for studying viral evolution. Indeed, improving the capacity to monitor viral genetic changes to predict 'evolutionary trends' can be crucial to boost surveillance, especially when considering those viral clades for which is reported or likely to occur avian to mammals/humans host switch (Al-Tawfiq *et al.,* 2014). In addition to viral strains with well known potential to jump the host-species barrier (Nelson and Vincent, 2015), further risk for human and animal health depends on the emergence of novel reassortant viruses, especially in those regions where multiple strains and clades are known to co-circulate (Su *et al.,* 2015). Avian influenza (AI) viruses from the H5N1 subtype are unique in their ecological success, showing extremely broad host range and geographical spreading (Guan and Smith, 2013). Therefore, based on intensive phylogenetic analyses just for H5N1 clades and subclades a standard nomenclature was published (WHO/OIE/FAO H5N1 Evolution Working Group, 2008; Guan and Smith, 2013) and it is actually adopted by the scientific community. General concern for pandemic risk decreased after the peak of the H5N1 virus, but indeed novel reassorted subtypes (e.g., H7N9, H9N2, H10N8) jumped in recent years the host-species barrier and thus surveillance and characterization are needed by the scientific community and public health systems (Trombetta *et al.,* 2015). In particular, the isolation of H9N2 virus from two Hong Kong children seems to be the first record of switch to human host (Peiris *et al.,* 1999); then, these viruses could occasionally be transmitted from poultry to humans and other mammals (Lin *et al.,* 2000; Butt *et al.,* 2005; Sang *et al.,* 2015).

Even though host jump and pandemic influenza phenomena raised much attention to AI viruses because of their potential impact on human health, studying virus variation related to either low to high-pathogenicity shift or to antigenic drift is needed as well, as it is quite relevant to animal health and of special impact on poultry industry and vaccine efficacy. In fact, outbreaks of high pathogenic AI (HPAI) can result in killing hundreds millions poultry and wild birds in tens of countries (Swayne, 2012). This makes proper vaccination strategies (at least in poultry animals) crucial to prevent wide mortality and in turn viral variation potentially resulting in antigenic drift becomes a factor to be monitored in that associated to risk of impairing vaccine efficacy.

Evolution and spread of low pathogenic AI (LPAI) and HPAI viruses, belonging to H5, H7 and H9 subtypes, amongst birds and their sporadic infection in humans continues to represent a great concern for public health (Lin *et al.*, 2000). Therefore, these avian subtypes are included as top pandemic agents in the list from the World Health Organization (WHO). To date, only H5 and H7 subtypes of influenza A viruses were reported to evolve from a LPAI to HPAI form after their introduction into poultry from the wild bird reservoir (Alexander 2007). Both H5 and H7 viruses are notifiable to the Office International des Epizooties (OIE) because of the risk of LPAI becoming HPAI by mutation. H7 LPAI virus usually causes mild respiratory disease and a production decrease in infected poultry; its evolution into a HPAI form results in the generation of a virus able to cause severe disease and death in the poultry population

(http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.03.04_AI.pdf).

One example assessing the ability of LPAI to evolve into HPAI form is the H7 avian outbreaks that affected Northern Italy between 1999 and 2001. Epidemiological information sustained by phylogenetic analysis, and deep sequencing approaches helped to reveal that

HPAI strains evolved from the LPAI viruses and that both lineages shared a common ancestor (Monne *et al.*, 2014).

Current AI vaccines are based upon the elicitation of a neutralizing antibody (Ab) response against the major epitope regions of the viral surface glycoprotein, hemagglutinin (HA). However, mutations in immune-dominant regions on the HA structure may result in antigenic drift allowing the virus to escape Ab neutralization (Velkov *et al.*, 2013). Indeed, antigenic and genetic differences in HA and the other surface spike protein neuraminidase (NA) provide a rationale for classification of influenza type A virus subtypes: for instance, H9N2 viruses combine the H9 subtype HA with N2 subtype NA. Haemagglutinin plays a central role in influenza A virus evolution because it is crucial to the attachment and penetration into the host cell and as the main viral surface antigen it is also a major player in the stimulation of the neutralizing Ab response (Velkov *et al.*, 2013).

In this work, genetic diversity of H9N2 subtype was assessed through large-scale phylogenetic analysis; this resulted in a novel and updated classification scheme based on the phylogenetic topology and evolutionary distances following the same WHO standards for virus classification as for H5N1. Classic, sequence based phylogenetic analyses can be integrated by structural bioinformatic investigations that more properly depict structural closeness among proteins or protein domains. Indeed, we previously inferred molecular fingerprints for H5N1 evolution, as intriguing surface charge redistribution at the surface of the receptor binding domain (RBD) subregion of HA was found to relate to branching of still circulating clades 2 and 7 with respect to those ones that are no longer circulating (Righetto *et al.*, 2014). This prompted us to assess whether such fingerprints are peculiar to H5N1, or electrostatic variation could be associated to the evolution and spreading of other AI viruses, e.g. the newly defined clades and subclades from the H9N2 subtype.

Methods

Phylogenetic analyses

HA gene nucleotide sequences of H9N2 subtype were retrieved from the Global Initiative on Sharing Avian Influenza Data (GISAID) EpiFlu database (http://www.gisaid.org). Nucleotide sequences of at least 1500 bp length were selected.

Multiple sequence alignment of HA sequences was performed with MAFFT version 7 (http://mafft.cbrc.jp/alignment/server). Redudant isolates with 100% sequence similarity (i.e., redundant sequences) were identified and removed, giving a final HA dataset and alignment of 2842 sequences from 1960 to 2015 was subjected to phylogenetic trees reconstruction. The NJ, ML and Bayesian methods were used to construct three different phylogenetic trees for comparison. Analysis of the best-fit substitution model was performed using MEGA5 (Tamura *et al.,* 2011), and the goodness-of-fit of each model was measured by Bayesian Information Criterion and corrected Akaine Information Criterion (AICc). The General Time Reversible (GTR) model with a discrete gamma distribution (+Γ) allowing for invariant sites (+I) was selected based on AICc and used in all data analyses. MEGA5 was also used to perform phylogenetic analysis and the evolutionary history was inferred by both NJ and ML methods (Tamura and Kumar, 2002), with standard errors being calculated based on 1000 bootstrap replicates. Furthermore, PhyML (version 2.4.4) (Guindon *et al.,* 2003) was used to create ML trees. The GTR + Γ + I model of nucleotide substitution was used for the analysis, with an estimated gamma shape parameter. Robustness of the groups was assessed using the bootstrap approach with 100 replicatest. Bayesian phylogenetic tree was inferred using MrBayes software (Ronquist and Huelsenbeck, 2003) and applied to generate the dendrograms as well as to assess statistical supports for the branches from the trees generated by the original dataset. For ease of display, and also to ensure that the clade topology would be

maintained when fewer isolates are used, a small representative dataset of 360 H9N2 HA sequences was created and analyzed by the same aforementioned phylogenetic models. Phylogenetic trees were visualized using FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

The largest HA gene dataset alignment (n = 2842; length ≥ 1500 bp) used for the phylogenetic reconstruction, was also used to infer evolutionary distances (within and between groups) by pair-wise analysis. The number of base substitutions per site was calculated by two different methods. The simplest one (uncorrected pairwise distance) was performed by averaging all sequence pairs between groups, while the second method followed the Maximum Composite Likelihood model. Variation rate among sites was modelled with a Γ distribution value = 9.4 (calculated by preliminary estimation from our dataset) and the differences in the composition bias among sequences were considered in the evolutionary comparisons. The C-value ratio used in the H9N2 clades partitioning i.e. the ratio of the average pairwise distance between a particular taxon and its closest neighboring group divided by the average pairwise distance within that selected clade was used to confirm the clades partitioning.

Bayesian phylogeography reconstruction

Time-scaled phylogenies of H9N2 HA were inferred by Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.8.0 (Drummond *et al.*, 2005) using the SRD06 codon position model and the uncorrelated log-normal relaxed clock model under a Bayesian skyline coalescent tree prior to the MCMC simulations (Jin *et al.*, 2014). Spatial location reconstruction and viral migration were estimated using the discrete Bayesian phylogeographic method that utilised a continuous time Markov Chain over discrete sampling

locations, and applied a Bayesian stochastic search variable selection model (Lemey *et al.*, 2009). For our data set we performed four independent runs for 300 million generations with sampling every 30000 steps. Convergence and effective sampling size of estimates were assessed by visual inspection using Tracer v1.6 (http://beast.bio.ed.ac.uk/Tracer). Multiple chains were then combined after a 10% burn-in using LogCombiner v1.8.0 included in the BEAST package. The maximum clade credibility (MCC) tree with temporal and spatial annotation were summarized with a 10% burn-in removed using TreeAnnotator v1.8.0 in the BEAST package and presentation figures were generated using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

We also conducted Bayes factor (BF) tests to provide statistical support for transmission routes between different geographic locations using SPREAD v1.0.6 with cutoff BF = 3 (Bielejec *et al.*, 2011). BF values represent the difference between the posterior and prior probabilities that the rates between two locations are non-zero. Thus, routes with high BF have large odds that a migration exists between two locations.

To animate viral dispersal over the time, we converted annotated MCC trees into a keyhole markup language file using SPREAD v1.0.6, which can be visualized by Google Earth (http://earth.google.com) website platform.

Structural Modeling and surface analysis

Structural models for HA1 and RBD regions of target HA proteins were obtained by homology modelling as reported (Righetto *et al.*, 2014) on best available structure templates using SWISS-MODEL (Bordoli *et al.*, 2009). In particular, as a template for H9N2 HAs the currently available solved HA structure PDB 1JSH belonging to A/swine/Hong Kong/9/98 (H9N2) was used. Refinement of model structures was performed using three independent

methods as reported (Righetto *et al.,* 2014) and model quality was checked via QMEAN server (Benkert *et al.,* 2009). Protein structures were viewed using UCSF Chimera (Pettersen *et al.,* 2004) v. 1.10.2 (free download from http://www.cgl.ucsf.edu/chimera/).

Comparative analysis of electrostatic potentials was performed as reported (Righetto *et al.,* 2014), simulating physiological conditions, i.e. the spatial distribution of the electrostatic potential was calculated at ionic strength (I) = 150 mM, assuming +1/-1 charges for the counter-ions. Isopotential contours were calculated using UCSF Chimera, which allows for connecting through Opal web server to the Adaptive Poisson-Boltzmann Solver server (http://www.poissonboltzmann.org/apbs). PDB2PQR was used to assign partial charges and van der Waals radii according to the PARSE force field (Sitkoff *et al.,* 1994). Electrostatic distance was calculated using the Carbo index at the WebPIPSA server (http://pipsa.eml.org/pipsa). Rigid-body superposition was performed and electrostatic potential was computed using UCSF Chimera 1.10.2. For more details please see the methods section from Righetto *et al.,* 2014.

Results

Phylogenetic analysis and novel classification scheme for AI H9N2 HA

Since the established classification system for AI virus subtypes is based on antigenic and genetic differences in the two surface spike proteins, we assessed the genetic diversity of the HA gene to develop a unified nomenclature and classification system for AI H9 subtype genetic groups. However, given that in the different classifications there is a stronger correlation between the phylogenetic topology and the evolutionary distances within and between genetic groups, we used the genetic correlation as the base to develop objective criteria to classify strains and create a definitive unified nomenclature. The classification criteria were determined based on the phylogenetic topology and on specific evolutionary distances that reflect the diversity of the AI H9N2 subtype. In order to get robust validation for phylogenetic analyses, the evolutionary history of 2842 AI H9N2 strains was inferred for HA nucleotide sequences ≥ 1500 bp using three different algorithms: neighbor-joining (NJ), maximum likelihood (ML) and Bayesian; when the evolutionary history was inferred from a smaller dataset alignment (360 strains), this confirmed consistency of the proposed classification.

Figure 1 shows the nucleotide ML phylogenetic tree with the proposed grouping, while the full tree is depicted in Supplementary figure S1: the AI H9N2 strains clearly separate into five different monophyletic groups hereafter referred to as class A, B, C, D, and E. Within such classes, twenty-seven clades identified by numbers are separated based on inter-clade average distance ≥5% and intra-clade average distance <5% (Table 1) and separation for each identified clade is confirmed by C-value ≥ 1 (Table 2). Classes and clades were assigned when at least three isolates with different epidemiological history formed a distinct taxonomic

group with bootstrap value at the defining node ≥60%. Clades separation based on distance value cut off was confirmed using two different calculation algorithms (see methods).
Fixed guidelines for classification are resumed in Table 3; circulation data and representative H9 subtype viruses used for different analyses performed in this work are listed in Supplementary table S1 to facilitate the interpretation of relationship to the proposed numbering system.

When comparing intra-class nucleotide distance values, class A shows the highest genetic heterogeneity with a value up to 18.3%, whereas this value is around 10-11% in classes B and C. Class D just contains three strains isolated in Malaysia and class E only a few strains isolated mostly in the USA. Therefore, separate clades were not defined for the latter two classes. Few different strains that are no longer circulating and do not group within any identified class are considered as ancestral (black sequences in figure 1). Such topology distribution and grouping was fully confirmed when performing phylogenetic analysis with corresponding protein sequences (data not shown).

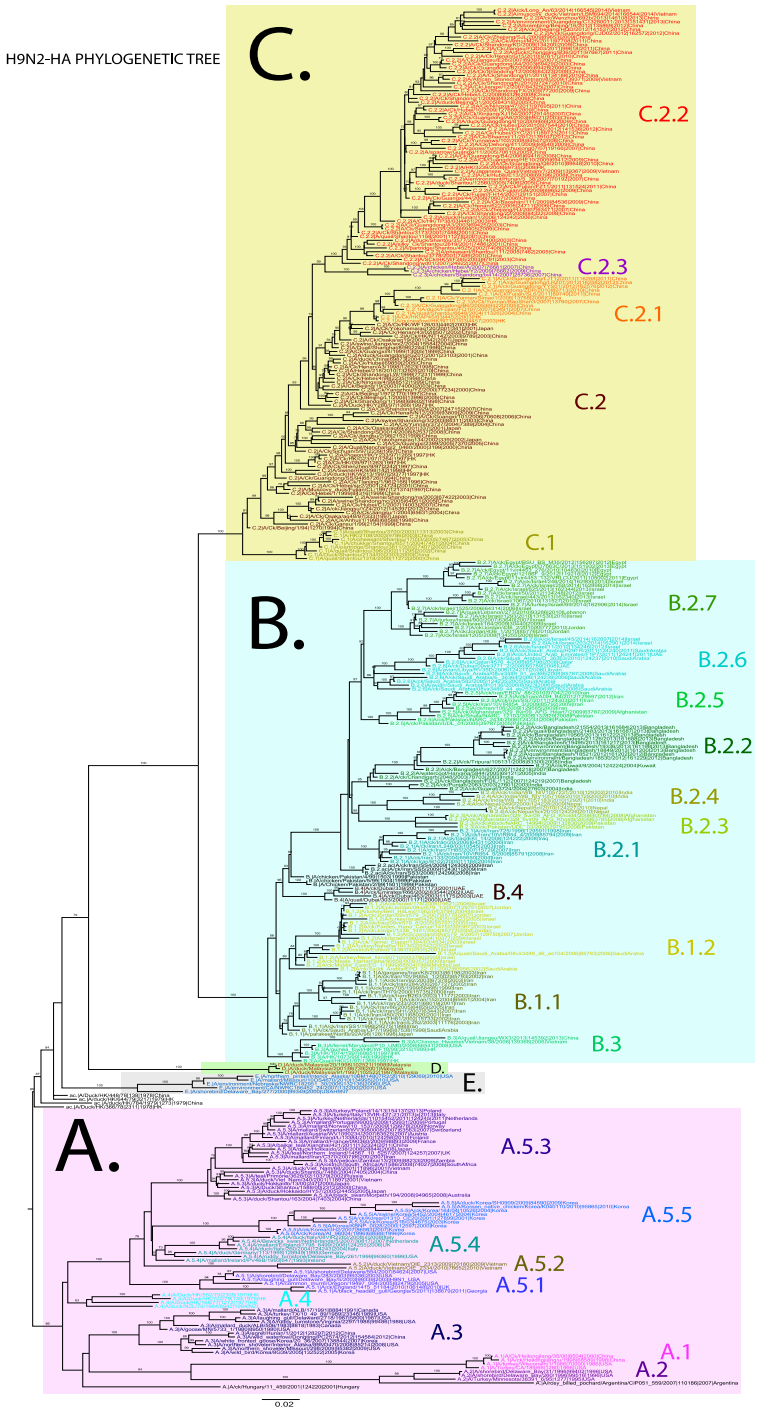


Figure 1: Maximum-likelihood trees of 360 H9N2 isolates constructed by PhyML. The different classes and clades are color coded. Estimates of the statistical significance of phylogenies were calculated by performing 100 bootstrap replicates. Numbers in the tree nodes represent the bootstrap support (≥60).

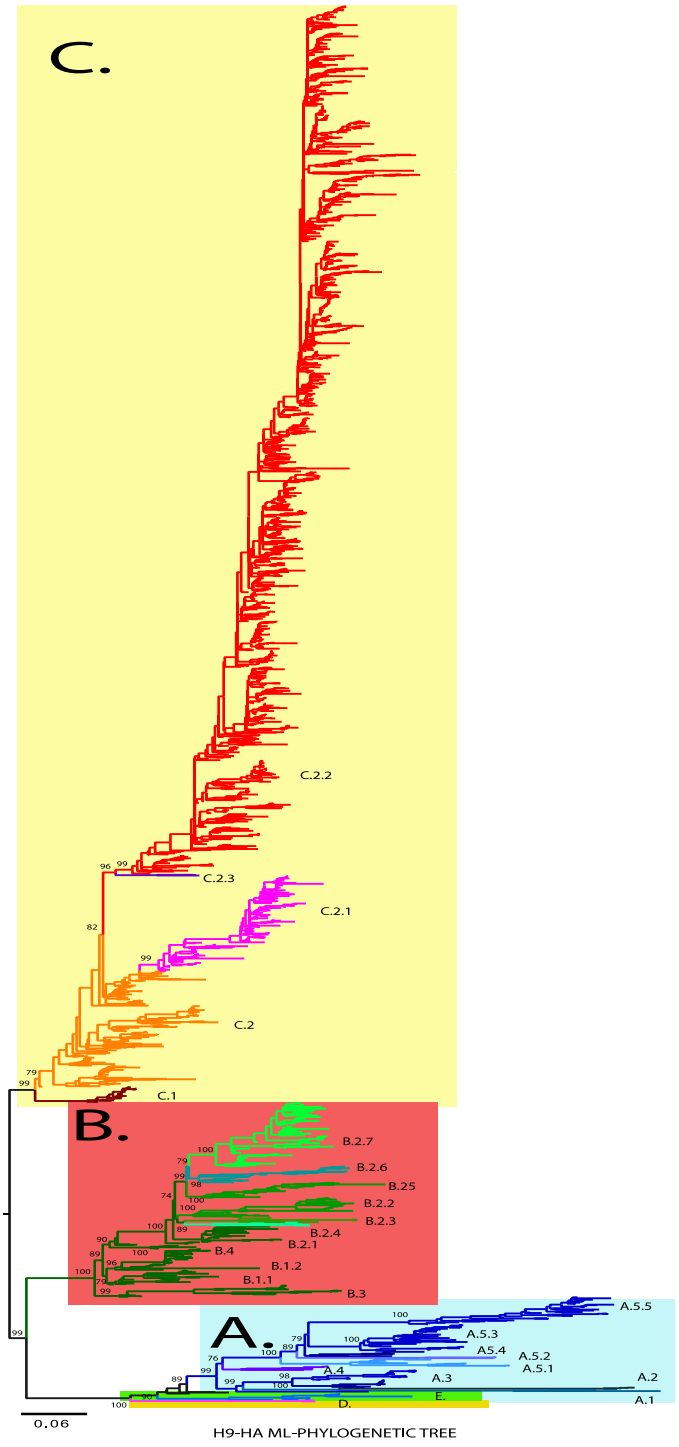


Figure S1: Maximum-likelihood trees of 2842 H9N2 isolates constructed by PhyML. The different classes and clades are color coded. Estimates of the statistical significance of phylogenies were calculated by performing 100 bootstrap replicates. Numbers in the tree nodes represent the bootstrap support (≥60).

CLADE	C.2.3	C.2.2	C.2.1	C.2	C.1	B.2.7	B.2.6	B.2.5	B.2.4	B.2.3	B.2.2	B.2.1	B.4	B.1.2	B.1.1	B.3	A.1	A.2	A.3	A.4	A.5.1	A.5.2	A..5.3	A.5.4	A.5.5	D.	E.
C.2.3	*	[0.6]	[0.6]	[0.4]	[0.7]	[0.7]	[0.8]	[0.7]	[0.8]	[0.7]	[0.8]	[0.7]	[0.7]	[0.7]	[0.7]	[0.7]	[0.8]	[0.9]	[0.7]	[0.7]	[0.8]	[0.9]	[0.8]	[0.8]	[0.8]	[0.8]	[0.7]
C.2.2	6.3	*	[0.6]	[0.5]	[0.7]	[0.7]	[0.7]	[0.7]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.7]	[0.7]	[0.8]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.8]	[0.8]	[0.8]	[0.8]	[0.9]	[0.7]
C.2.1	6.4	7.7	*	[0.5]	[0.7]	[0.7]	[0.8]	[0.7]	[0.8]	[0.8]	[0.8]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[10]	[0.9]	[0.7]	[0.7]	[0.8]	[0.9]	[0.8]	[0.7]	[0.9]	[0.8]	[0.7]
C.2	6.2	6.4	6.1	*	[0.6]	[0.6]	[0.6]	[0.6]	[0.6]	[0.6]	[0.7]	[0.6]	[0.6]	[0.6]	[0.5]	[0.6]	[0.9]	[0.9]	[0.7]	[0.6]	[0.7]	[0.7]	[0.7]	[0.7]	[0.8]	[0.8]	[0.6]
C.1	7	8.1	8.2	6.8	*	[0.6]	[0.7]	[0.7]	[0.6]	[0.6]	[0.7]	[0.6]	[0.7]	[0.6]	[0.6]	[0.6]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.8]	[0.7]	[0.8]	[0.8]	[0.8]	[0.7]
B.2.7	10.6	11.2	11.6	9.9	9.7	*	[0.4]	[0.4]	[0.6]	[0.5]	[0.5]	[0.4]	[0.6]	[0.6]	[0.6]	[0.6]	[0.9]	[0.8]	[0.7]	[0.7]	[0.9]	[0.8]	[0.8]	[0.8]	[0.9]	[0.9]	[0.8]
B.2.6	11.2	11.9	11.4	10.3	10.3	5.9	*	[0.5]	[0.6]	[0.6]	[0.5]	[0.5]	[0.6]	[0.6]	[0.6]	[0.7]	[0.8]	[0.8]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.7]	[0.8]	[0.9]	[0.7]
B.2.5	9.8	10.6	10.9	8.9	9.2	5.6	5.8	*	[0.6]	[0.5]	[0.5]	[0.5]	[0.5]	[0.6]	[0.6]	[0.6]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[0.8]	[0.8]	[0.7]
B.2.4	11.5	11.4	12.1	10.6	10.2	7.3	7.3	7	*	[0.5]	[0.5]	[0.5]	[0.6]	[0.6]	[0.5]	[0.6]	[0.9]	[0.9]	[0.8]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[0.8]	[0.8]	[0.7]
B.2.3	10.7	11.5	12	10.1	10.2	6.7	6.6	5.9	6.9	*	[0.6]	[0.5]	[0.6]	[0.6]	[0.6]	[0.6]	[0.8]	[0.8]	[0.7]	[0.6]	[0.8]	[0.7]	[0.7]	[0.7]	[0.8]	[0.8]	[0.7]
B.2.2	10.6	11.5	11.6	10.1	10.3	7.1	7.3	6.9	7.6	7.1	*	[0.5]	[0.6]	[0.6]	[0.6]	[0.6]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[0.9]	[0.7]	[0.7]
B.2.1	9.7	10.3	10.7	8.7	8.7	5.4	5.7	6	6.1	5.8	6.4	*	[0.6]	[0.5]	[0.5]	[0.6]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[0.8]	[0.8]	[0.7]
B.4	8.7	9.8	9.9	7.8	7.6	7.4	7.8	6.7	8	7.6	8	6.1	*	[0.5]	[0.4]	[0.6]	[0.9]	[10]	[0.8]	[0.8]	[0.9]	[0.8]	[0.8]	[0.8]	[0.9]	[0.9]	[0.8]
B.1.2	8.8	9.5	9.5	7.5	7.5	7.7	8.1	7.4	8.5	8.1	8.1	6.6	6	*	[0.4]	[0.5]	[0.9]	[0.9]	[0.7]	[0.7]	[0.7]	[0.7]	[0.7]	[0.7]	[0.8]	[0.8]	[0.7]
B.1.1	8.5	9.5	9.9	7.6	7.2	7.3	7.6	7	7.9	7.5	7.8	6.1	5.4	5.4	*	[0.5]	[0.9]	[0.9]	[0.7]	[0.7]	[0.8]	[0.7]	[0.7]	[0.8]	[0.9]	[0.9]	[0.7]
B.3	9.9	11.1	11.5	9.3	9.2	8.6	9.3	8.7	9.6	9.2	9.4	7.8	6.6	6.5	6.4	*	[0.9]	[10]	[0.7]	[0.7]	[0.8]	[0.8]	[0.8]	[0.8]	[0.9]	[0.8]	[0.8]
A.1	13.3	14	15.2	13.9	14.5	15.4	15.3	15	15.9	15.3	15.5	14.9	14.2	14.7	14.1	15.4	*	[0.9]	[0.8]	[0.8]	[0.9]	[0.9]	[0.9]	[0.9]	[0.9]	[10]	[0.8]
A.2	14.5	15	15.2	14.2	13.7	15.3	15.8	15.6	16.3	15.9	15.3	15.2	14.8	14.6	14.3	15.5	13.1	*	[0.8]	[0.9]	[0.9]	[0.8]	[0.9]	[0.9]	[10]	[0.9]	[0.9]
A.3	12.3	12.5	12.4	11.5	11.2	13.8	14.5	13.2	14.6	14	14.2	13.2	12.3	12.4	12.5	13.4	12.3	12.1	*	[0.5]	[0.6]	[0.6]	[0.6]	[0.6]	[0.7]	[0.8]	[0.6]
A.4	10.6	11.2	11.9	10.3	9.9	12.6	13	11.8	12.7	12.3	12.5	11.4	10.6	10.7	10.6	11.8	12.2	12.1	7.4	*	[0.6]	[0.7]	[0.6]	[0.6]	[0.8]	[0.7]	[0.7]
A.5.1	12.5	13	13.3	12.5	11.9	14.3	14.6	13.9	14.2	13.9	14.1	13	12.8	12.5	12.9	14	13.9	12.5	9.8	8.4	*	[0.7]	[0.6]	[0.5]	[0.6]	[0.8]	[0.7]
A.5.2	13.1	12.9	13.5	12.1	11.1	14.2	14.4	13.9	14.3	14.1	14.2	13.7	12.9	12.7	12.9	14.1	13.5	12.8	10.6	9.1	9.4	*	[0.6]	[0.6]	[0.7]	[0.8]	[0.7]
A..5.3	12.8	13.4	13.5	12.4	11.6	13.9	13.9	13.2	14.1	13.8	13.6	12.9	12.8	12.2	12.3	13.6	12.6	12.3	9.1	8.3	8	8.3	*	[0.5]	[0.6]	[0.8]	[0.7]
A.5.4	12.3	12.7	12.6	12	11.2	13.7	14.2	13.5	14.1	13.8	13.5	12.7	12.5	12.3	12.3	13.3	12.6	12.4	8.7	7.6	7.1	7.6	5.4	*	[0.6]	[0.8]	[0.6]
A.5.5	13.1	13.3	13.7	12.6	12.2	14.7	14.6	13.8	14.7	14.2	14.8	13.5	13.3	12.8	13.1	14	14.3	14.4	11.2	9.9	9.7	10.4	9.1	8.5	*	[0.9]	[0.7]
D.	12.3	12.3	13.2	11.6	11.4	14.1	14.2	12.9	13.6	13.5	13.5	12.8	11.8	11.8	11.4	12.8	15.2	13.4	11	9.5	12.5	12.7	12	11.6	13.7	*	[0.8]
E.	11.5	11.6	12.1	10.9	10.3	12.9	13.2	12.6	13.3	12.9	12.8	12	11.1	10.8	11	11.8	13.3	11.9	9.3	8.4	9.9	11.1	10	9.5	11.8	9.9	*

Table 1: Estimates of evolutionary distances between H9N2 identified clades. Evolutionary distances (calculated by p-distance) between identified clades were calculated using 2842 nucleotide HA sequences ≥ 1500 bp in lenght. Max/min evolutionary distance values within classes A, B and C are highlighted in respectively red and blue colors. Values between brackets are standard errors, obtained by a bootstrap procedure (500 replicates).

clades	APD Within		Closest clade	APD between clade & closest clade	C-Value
C.2.3	2.5	[0.3]	C.2	6	2.4
C.2.2	4.1	[0.2]	C.2.3	6.3	1.5
C.2.1	3.3	[0.2]	C.2	6.1	1.8
C.2	4.7	[0.2]	C.2.3	6.7	1.4
C.1	1.2	[0.1]	C.2	6.8	5.6
B.2.7	3.8	[0.2]	B.2.5	5.5	1.4
B.2.6	4.6	[0.3]	B.2.1	5.7	1.2
B.2.5	3.4	[0.2]	B.2.7	5.4	1.5
B.2.4	4.9	[0.3]	B.2.3	6.9	1.4
B.2.3	2.7	[0.2]	B.2.5	5.9	2.1
B.2.2	4.4	[0.2]	B.2.1	6.4	1.4
B.2.1	2.7	[0.2]	B.2.3	5.8	2.1
B.1.1	3.3	[0.2]	B.4	5.4	1.6
B.1.2	2.7	[0.2]	B.4	6	2.2
B.4	1.6	[0.2]	B.1.1	5.4	3.4
B.3	3.4	[0.2]	B.1.1	6.4	1.8
A.1	1.2	[0.2]	A.4	12.2	10.1
A.2	2.6	[0.3]	A.4	12.1	4.6
A.3	4.5	[0.3]	A.4	7.4	1.6
A.4	3.4	[0.3]	A.5.4	7.6	2.2
A.5.1	3.8	[0.3]	A.5.4	7.1	1.8
A.5.2	4	[0.4]	A.5.4	7.6	1.9
A.5.3	3.8	[0.2]	A.5.4	5.4	1.4
A.5.4	3.4	[0.2]	A.5.3	5.4	1.5
A.5.5	4.5	[0.3]	A.5.4	8.5	1.8
D.	0.8	[0.2]	A.4	9.5	11.8
E.	2.8	[0.2]	A.4	8.4	3

Table 2. Estimates of average pairwise distance and C-value within each identified H9N2 clade. The average pairwise distance was calculated using 2842 HA nucleotide sequences (≥ 1500 bp). Values between brackets are standard errors, obtained by a bootstrap procedure (500 replicates). The C-value is the ratio of distance between clade and its closest clade to the distance within clade.

	Criteria used for a class and clade designation
1	The nomenclature was established based on a non redudant dataset of 2842 HA sequences from H9 viruses with sequence lenght ≥ 1500 bp.
2	Classes were assigned based on phylogenetic topology distribution and confirmed by three different methods: Maximum Likelihood, NJ, Byesian.
3	Clades were assigned based on both phylogenetic topology distribution and evolutionary distances between different taxonomic branches. Clades separation was confirmed using the three aforementioned methods (see point 2).
4	New classes and clades were designated only when at least three independent isolates without a direct epidemiologic link (i.e. distinct outbreaks) were available.
5	Bootstrap values at the classes and clades defining node should be ≥60%.
6	Distinct clades should have ≥5% average distances between other clades. Distinct clades should have <5% average distances within the clade.
7	Cut-off value 5% with C value ≥ 1 was fixed to assign new clades in each class.

Table 3. H9N2 AI viruses, class and clade identification criteria. Text in the table cells is self-explaining.

Table S1. H9N2 virus strains representative for each clade of the identified classes. Class and clade descriptions with isolation period, source, and geographic location are reported. (*) B ancestral strain for B class.

Class	Clade	Circulation			Representative strain for each clade		
		Date	Countri	Host	Full	Short name	GISAID/NCBI
A.	A.1	1966-2000	USA, China	Avian	A/turkey/CA/189/66	A.1_AtkCA66	EPI_ISL_1280/AF156390/ AAD49000
	A.2	1995-1996	USA	Avian	A/turkey/Minnesota/38391-6/95	A.2_AtkMi95	EPI_ISL_1277/AF156387/ AAD48997
	A.3	1980-2012	China, Korea, USA, Canada	Avian	A/goose/MN/5733-1/1980	A.3_AgoMN80	EPI_ISL_8950/CY006042/ ABB88390
	A.4	1979-1984	Hong Kong, New Zealand	Avian	A/duck/NZL/76/1984	A.4_AdkNZL84	EPI_ISL_8942/CY005746/ ABB20444
	A.5.1	2003-2015	Singapore, USA, Georgia, UK	Avian	A/shorebird/Delaware Bay/283/2003	A.5.1_AshDB03	EPI_ISL_99336/CY102744 /AET77176
	A.5.2	2009-2010	Vietnam	Avian	A/duck/Vietnam/OIE-2334/2010	A.5.2_AdkVN10	EPI_ISL_76652/ AB569975/BAJ10561

	A.5.3	1999-2014	Japan, Russia, Australia, Chinam, Vietnam, UK, Netherlands, Italy, France, Finland, Austria, Switzerland, Norway, Portugal, South Africa, Zambia, Iran	Avian	A/duck/Hokkaido/13/00	A.5.3_AdkHo03	EPI_ISL_247/AY330340/AAQ97383
	A.5.4	1993-2008	Ireland, UK, Italy, Netherlands, Germany, USA	Avian	A/mallard/Ireland/PV46B/1993	A.5.4_AmaIRE93	EPI_ISL_647/AB303077/BAF62259
	A.5.5	1996-2009	Korea	Avian, Swine	A/chicken/Korea/AI-96004/1996	A.5.5_AckKo96	EPI_ISL_68688/GU053194/ACZ48629

D	D	1997-2001	Malaysia	Avian	A/duck/Malaysia/2001	D_AdkMa98	EPI_ISL_96739/CY073800/AEY75590

E	E	2006-2013	USA	Avian, environment	A/environment/California/NWRC186451-18/2007	E_AshDB00	EPI_ISL_132197/CY122538/AET77024

B.	B (*)	1999	Pakistan	Avian	A/chicken/Pakistan/2/1999	B_AckPa99	EPI_ISL_146703/ KF188299/AGO1796
	B.1.1	1998-2007	Saudi Arabia, Japan, Iran	Avian	A/parakeet/Narita/92A/98	B.1.1_APaNa98	EPI_ISL_128/ AB049160/
	B.1.2	1998-2007	Lebanon, Iraq, Jordan, UAE, Israel, Saudi Arabia	Avian	A/chicken/Middle East/ED-1/1999	B.1.2_AckME99	EPI_ISL_68504/GU053201 /
	B.2.1	1998-2009	Iran, Iraq	Avian	A/chicken/Iran/B102/2005	B.2.1_AckIR05	EPI_ISL_11184/ EF063733/ABO09919
	B.2.2	2003-2013	India, Bangladesh, Kuwait	Avian	A/chicken/Chandigarh/2048/2003	B.2.2_AckCh03	EPI_ISL_78703/CY068643 /ADL64047
	B.2.3	2008-2015	Afghanistan, Pakistan, Iran	Avian	A/chicken/Afghanistan/329-6vir09-AFG-Khost9/2008	B.2.3_AckAf08	EPI_ISL_63785
	B.2.4	2009-2011	Nepal, India	Avian	A/chicken/Nepal/2490/2009	B.2.4_AckNE09	EPI_ISL_124228/ JX273549/AFO83282
	B.2.5	2005-2013	Pakistan, Afghanistan, Iran	Avian	A/chicken/Pakistan/UDL-01/2005	B.2.5_AckPA05	EPI_ISL_29787/ CY038410/ACP50642
	B.2.6	2005-2014	Saudi Arabia, UAE, Qatar, Israel, Libyan	Avian	A/chicken/Saudi Arabia/582/2005	B.2.6_AckSA05	EPI_ISL_124235/JX27355 6/AFO83289
	B.2.7	2006-2015	Israel, Egypt, Lebanon, Jordan	Avian	A/chicken/Israel/1525/2006	B.2.7_AckIS06	EPI_ISL_64314/FJ464728/ ACJ68774
	B.3	1999-2015	Hong Kong, USA, Vietnam	Avian, Human	A/Quail/Hong Kong/G1/97	B.3_AquHKG19 7	EPI_ISL_1268/ AF156378/ AAF00706
	B.4	2000-2002	UAE	Avian	A/quail/Dubai/303/2000	B.4_AquDu0 0	EPI_ISL_11171/ EF063512/ABM21877
C.	C.1	2000-2005	Hong Kong, China	Avian/Human	A/quail/Shantou/1318/2000	C.1_AquSh00	EPI_ISL_11272/ EF154910/ABM46230
	C.2	1994-2014	China, Hong Kong, Japan	Avian, Swine, Human, Environment	A/chicken/Beijing/1/1994	C.2_AckBe94	EPI_ISL_146731/ KF188294/AGO17871
	C.2.1	2003-2013	China, Hong Kong, Japan	Avian	A/guinea fowl/Hong Kong/NT101/2003	C.2.1_AgfHK03	EPI_ISL_146694/ KF188382/ABB58955
	C.2.2	2001-2015	Hong Kong, Vietnam, China	Avian, Environment	A/chicken/Shandong/wd01/2007	C.2.2_AckSd07	EPI_ISL_24922/ FJ231868/ACI2260
	C.2.3	2007-2009	China	Avian	A/chicken/Hebei/A/2007	C.2.3_AckHe07	EPI_ISL_76661/ GQ202056/ACR5617

Phylogeographic analysis for AI H9N2 HA

In order to determine the worldwide dissemination of AI H9N2, the sequences of the HA gene were grouped into eight geographic areas, namely: (i) North America, (ii) Europe, (iii) Oceania, (iv) China, (v) Meddle East, (vi) South, (vii) South east and (viii) East Asia. The final dataset included 357 viruses that could be used for in-depth special analysis. Through Posterior distribution under the Bayesian framework, we reconstructed genealogical trees with time-scale and inferred ancestral locations of each branch using sequences' sampling collection dates and locations. The time-scaled phylogeographic maximum clade credibility (MCC) tree of HAs implemented in BEAST (see methods) and the root state posterior probability are illustrated in Figure 2 (phylogeography tree) and Figure 3 (phylogeography map), in which the most probable location of each branch is assigned different colours and the calibrating time-scale. Numbers at branch points in Figure 2 are reported where state probabilities with values ≥ 0.55 correspond to the most relevant events (i.e. to area-area transitions rather than to intra-areal ones). Such transition events supported by Bayes factor >3 are graphically depicted as arrows (with class-coded colors) in Figure 3 map. For graphical reasons (saving space to fit the one page format), names of the 357 individual viral strains are not reported in Figure 2; however, the same tree with all virus names is presented in Supplementary figure S2.

Our phylogeographic results suggest that the North American strains are ancestral for all H9 subtypes. Those ones introduced in China then were spreading worldwide. In particular, most probably American Class A strains reached China first and Chinese clades moved in turn to Europe, southeast and East Asia and Australia by migratory birds. Similarly, Class B (mostly present in poultry), spread from China to the Middle East, and from the Middle East to south Asia. Class C can also be referred to as China class because it evolved and expanded

mostly in China; however, different viruses of class C were introduced and circulated in east and south Asia. Class D formed a separate class in southeast Asia, while class E evolved by back migration events of Chinese viruses to North America. The overall spatiotemporal representation of phylogeographic evolution and worldwide spreading of the H9 subtypes is presented in Supplementary visual animation S1.

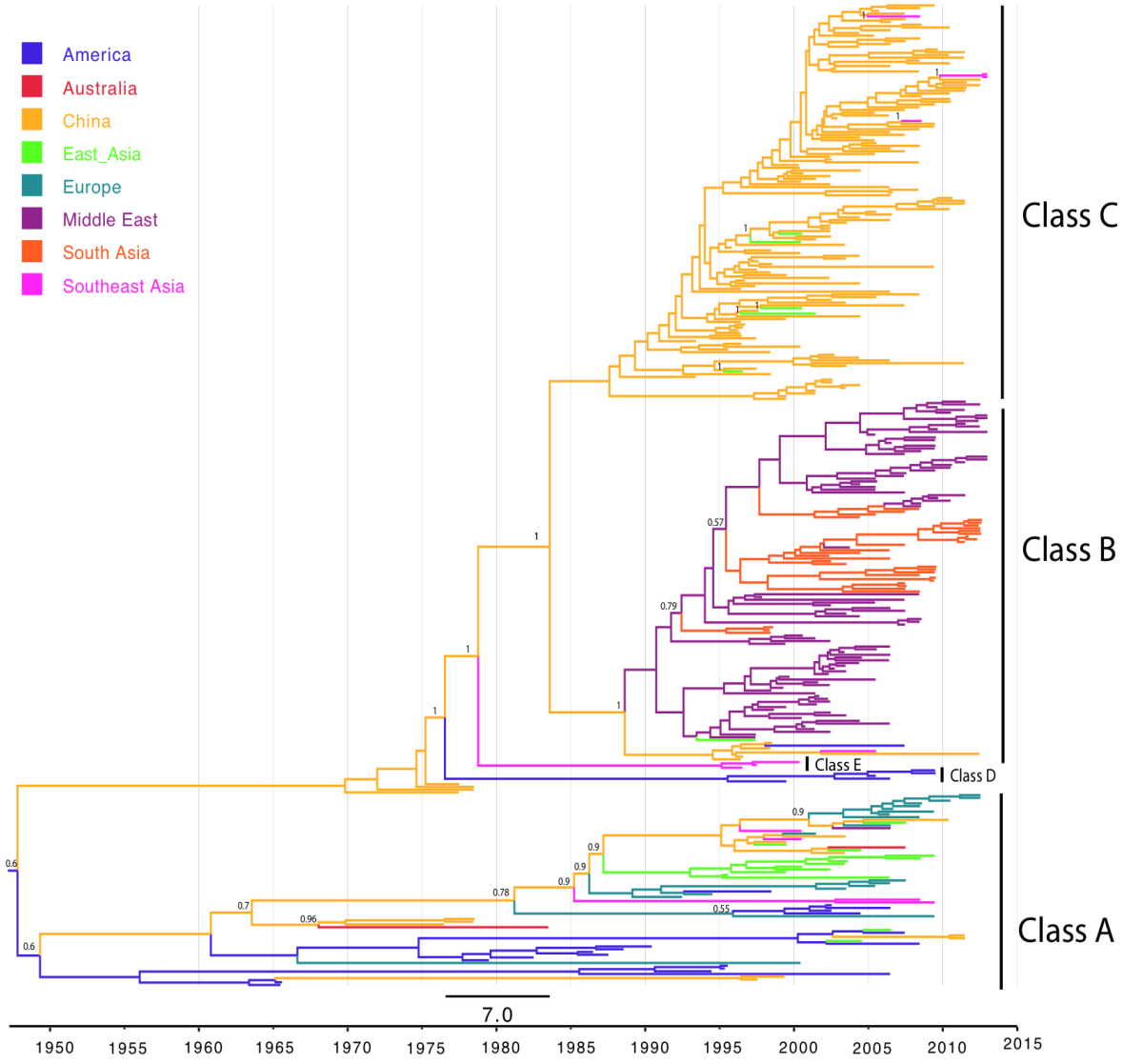


Figure 2. Maximum clade credibility (MCC) phylogenies inferred for the HA gene sequences of 357 viruses of AI H9 subtype. Branches are coloured according to the most probable ancestor location (in terms of geographic area) of their descendent nodes. Timeline at the bottom indicates the years before the most recent sampling time. Numbers are reported at branch points where state probabilities with values ≥0.55 correspond to geographic area transition events.

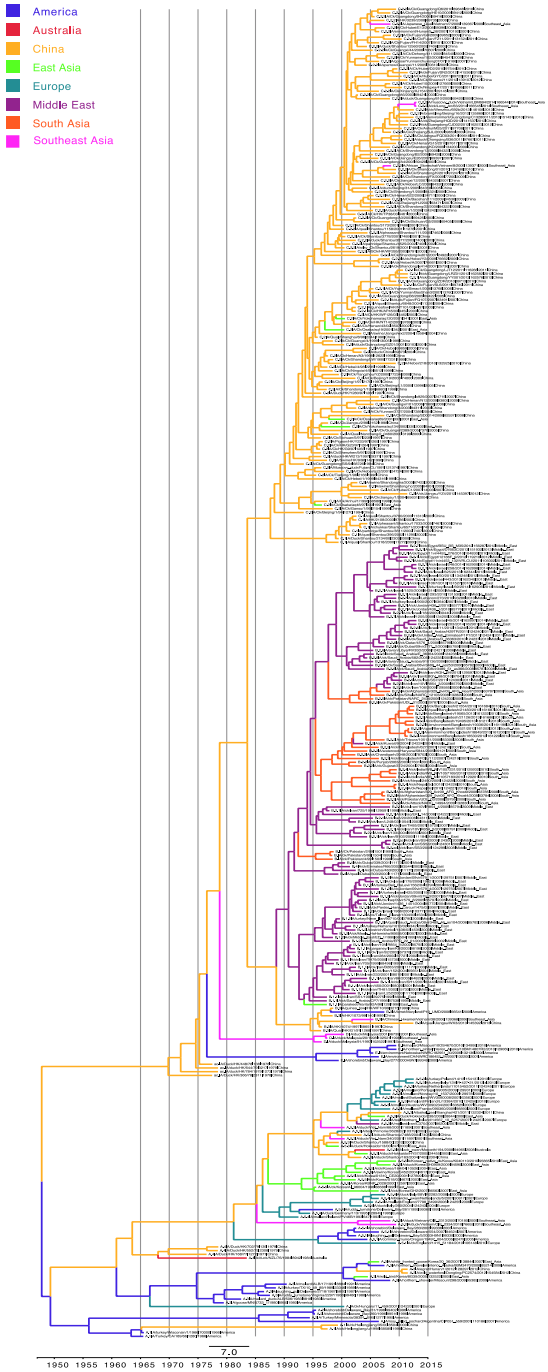


Figure S2. Maximum clade credibility (MCC) phylogenies inferred for the HA gene sequences of 357 viruses of AI H9 subtype. Branches are coloured according to the most probable ancestor location (in terms of geographic area) of their descendent nodes. Timeline at the bottom indicates the years before the most recent sampling time. Virus name information is reported in details.

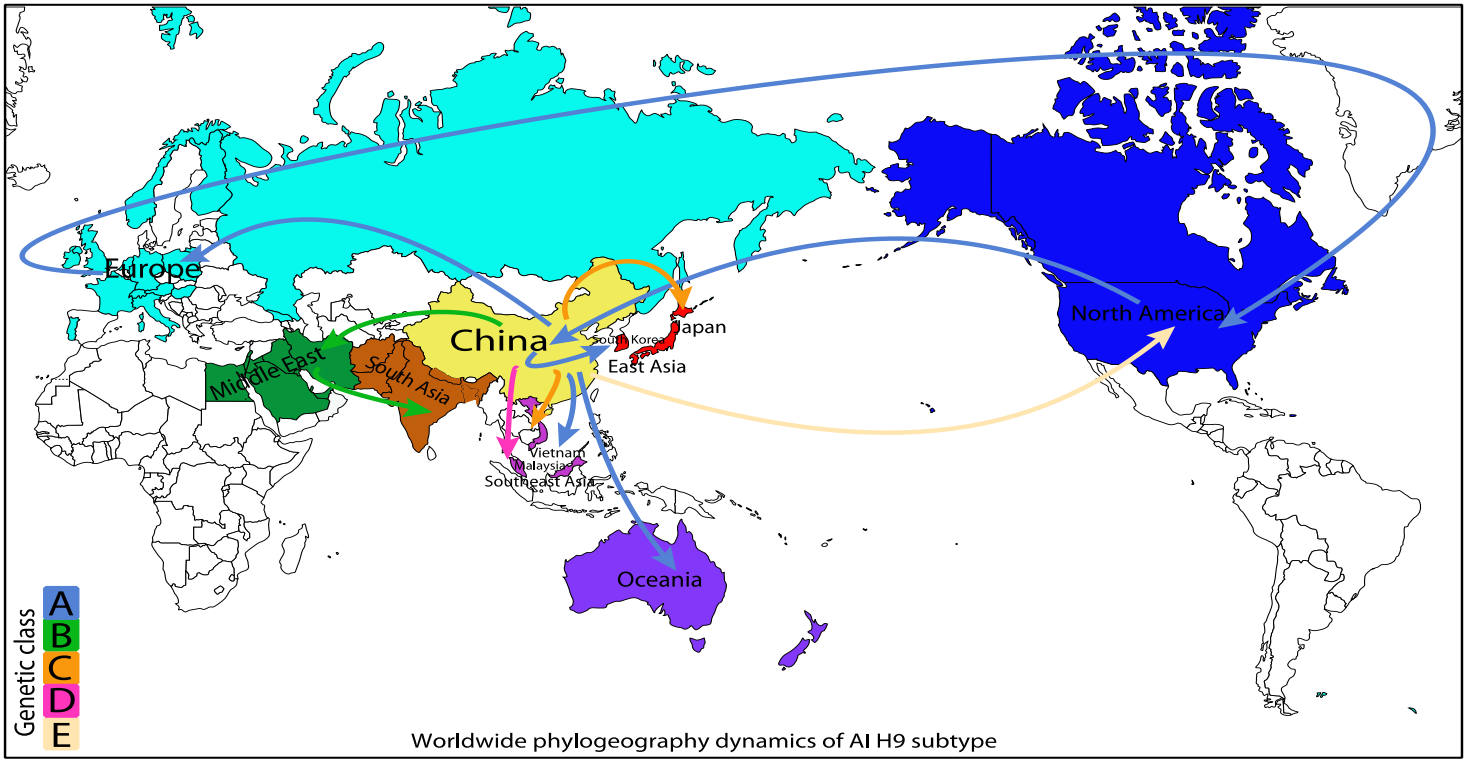


Figure 3. Phylogeography worldwide mapping and spreading of AI H9 subtype viruses. The eight geographic areas in which H9 viruses are so far known to circulate are differently coloured. Viral transition from a geographic area to another one with the well supported Bayes factor > 3 and state probabilities values ≥0.55 are represented by arrows with colour code based on corresponding H9 classes.

Clustering by electrostatic features for AI H9N2: heat maps and epograms

A recent bioinformatic work has shown that the integration of sequence and structural analyses for HA (and especially its RBD) can shed more light on the evolution and spreading of AI H5N1 viruses by unveiling surface patches as possible evolutionary fingerprints (Righetto *et al.,* 2014). Therefore, when considering findings emerged from comparative HA1 and RBD analysis, we decided to check whether variation in electrostatic features of H9N2 would relate to phylogenetic data, as observed for H5N1 (Righetto *et al.,* 2014).

Representative strains for each clade of the five H9N2 classes identified in our phylogenetic analysis are summarized in Supplementary table S1. In order to quantitatively evaluate the electrostatic distance, clustering of the spatial distributions of the electrostatic potentials was obtained by WebPIPSA (Protein Interaction Property Similarity Analysis) (Richter *et al.,* 2008). Figure 4 depicts the heat map and density plot for the RBD subregion of HAs from such representative strains. High electrostatic distance (dark blue, violet or magenta colors, see density plots) clearly separates classes A, D and E (typical of wild birds) from B and C (common to poultry birds), whereas the electrostatic distance between B and C is lower, as highlighted by prevalence of the light blue color. Therefore, clustering of H9N2 classes by electrostatic features shows substantial agreement to phylogenetic grouping, apart from a few exceptions: for instance, in terms of electrostatic distance, B3 and B4 are closer to C2 (light blue) than to B2 strains. It can be noticed that B3 and B4 clades used for this analysis were both isolated from the same host bird (quail). In addition to heat maps and corresponding density plots, the distance matrices of the electrostatic potentials were also displayed as trees referred to as ‘epograms’ (electrostatic potential diagrams). The epogram for RBD confirms grouping of the A wild bird cluster as well as homogeneity of class C and B2 clades; moreover, once again B3 sorts with C clades rather than with B ones (Figure 5).

Both heat map and epogram for HA1 subregion (not shown) confirmed clustering from the RBD analysis.

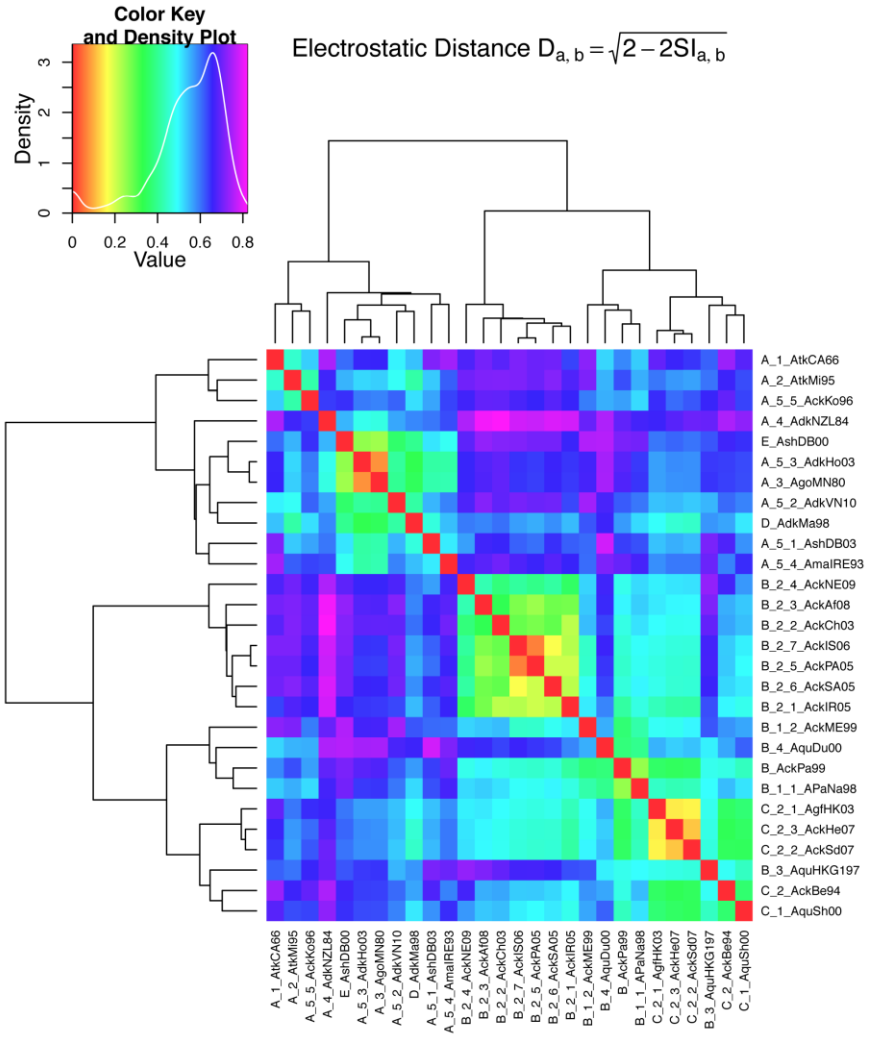


Figure 4. Heat map and density plot for the RBD subregion from representative H9N2 strains. The electrostatic distance formula is reported. In both density plot and heat map, warm to cold color shift corresponds to increasing electrostatic distance.

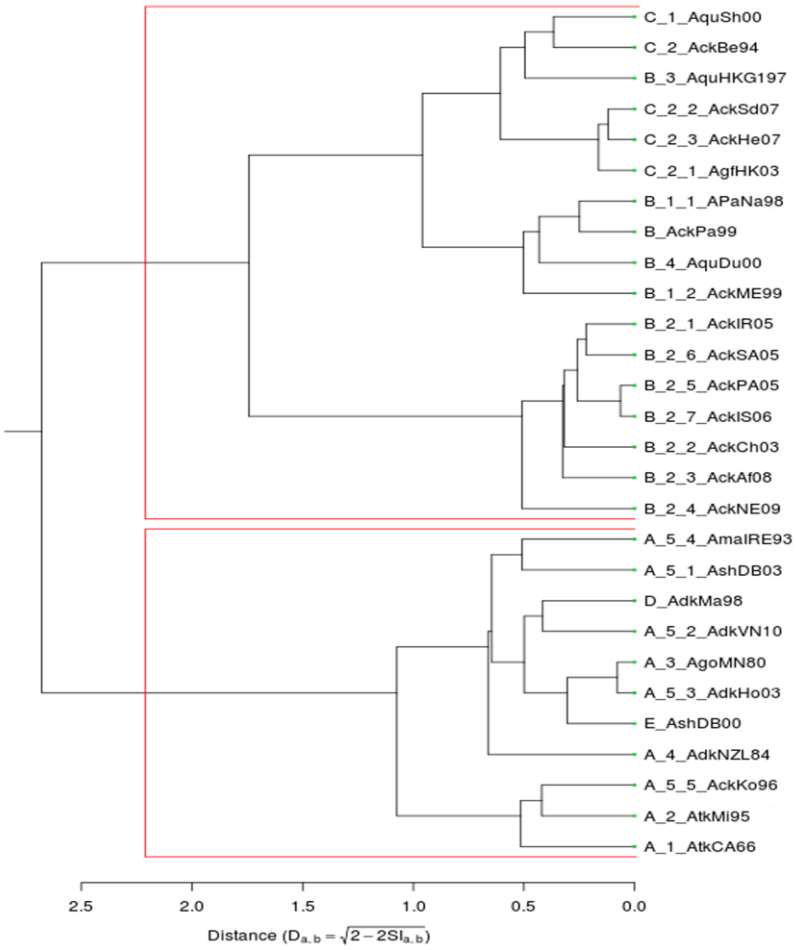


Figure 5. Electrostatic potential diagram (epogram) for the RBD subregion from representative H9N2 strains. The electrostatic distance formula is reported. The two main clusters are 3/4 bordered by red lines.

Variation in charge distribution among AI H9N2 classes and strains

In depth analysis of the distribution of charged residues, of their variation in number and position and of isocontours from the different HA subregions, further confirmed that variation especially concerns the RBD subregion and suggested possible electrostatic fingerprints are associated to different H9N2 classes.

Class-associated 'charge redistribution' is found to occur at RBD positions 135, 146 and 162: the net charge for these three positions is zero in all classes (as the sum of two opposite charges and a non charged residue), but the charge distribution pattern shared by the 'wild bird' classes A, D and E is different from the pattern conserved in viruses from the 'poultry' classes B. and C. In fact, distribution at 135-146-162 is neutral-positive-negative in A-D-E, and negative-neutral-positive in B-C (Table 4). In particular, at position 135, almost all viruses from class A share a non charged residue (167/177 strains) with prevalence (127/177) of Asn, which is 100% conserved (16/16) in classes D-E; mutation to charged residue (N135D) only concerns 10/177 viruses from clades A.5.3, A.5.4 and A.5.5. Instead, negativization at position 135 is most often observed in both classes B (311/364) and C (1011/1102), with prevalence of Asp/Glu over other amino acids in almost all B-C clades. A compensatory mechanism is observed for exceptions, i.e. for those clades that do not share a negative charge at position 135. For example, clade C.1 lacks the negative charge of classes B-C and shares instead (31/31 sampled viruses) N135 with classes A-D-E; however, this is compensated as C.1 is also the only B-C clade missing a positive charge at position 131. Similarly, B3 (showing prevalence of Gl35) is also the only B-C clade with a negative charge (Glu) instead of a non charged residue at position 180. Therefore, compensatory mutations in the RBD seem to keep class-specific fingerprints and net charge, while progressively 'sliding' positions of charged residues over RBD sites in the viral population. Residue 146 is His in

169/177 viruses from class A and in all D-E strains, while prevalence of Gln is observed in all clades from class B (363/364) and C (1075/1102). The only exception in class A is clade A.5.2, showing Q146 (like B-C) instead of H146 (common to A-D-E). However like for example above a counterbalancing unique mutation is observed: depositivization at position 146 of A.5.2 is compensated by peculiar denegativization at position 162 (E162N). In most (114/177) class A viruses and in all D-E strains, residue 162 is Glu except for clade A.5.5, showing mutation E162W in 51/51 viruses. Intriguingly, the lost negative charge is rescued at the contiguous N-terminal position 161 by the equally conserved (51/51) and peculiar mutation N161D, suggesting the negative charge at position 162 (or 161) as a landmark for A-D-E viruses. Instead, in viral strains from classes B-C the major residues are Arg and Gln, with prevalence of the former over the latter in all clades but B.2.4 and C.2.2, where reverse prevalence is observed. Therefore, ongoing positivization of position 162 seems to be a landmark as well for viruses circulating in poultry. Altogether, counterbalancing mutation pairs observed at positions 131-135, 135-180, 146-162 and 161-162 seem to support the compensatory mechanism suggested above for mantaining the overall net charge of the RBD while sliding charges over the RBD itself, i.e. for keeping class specific landmarks along with contemporary creation of novel fingerprints.

A second and different kind of variation in electrostatic features is observed at positions 180 and 186 of the RBD (Table 4). In all H9N2 classes, the net charge for this amino acid pair is zero. However, in A-D-E viruses this results from the sum of opposite charges (+1 -1 = 0), while in B-C viruses, it depends on replacement of both charged residues by neutral ones (0 + 0 = 0). Therefore in this case differently from previous examples of charge redistribution manteinance of the net charge is associated to a decrease in the percentage of charged residues in the RBD.

HA subtype		HA (mature chain) position number										
H9N		**131**	**135**	**146**	**161**	**162**	**165**	**18**	**186**	**198**	**216**	**217**
H7N		127	134	145	162	163	166	18	187	199	217	218
H5N		133	141	152	167	168	171	18	192	204	222	223
H3N		137	145	156	171	172	175	19	196	208	226	227
H1N		134	142	153	168	169	172	18	193	205	223	224
Clad	**Strain**	**Fully / most conserved amino acid for each**										
A.1	4	R	N	H	N	E	N	E	K	D	Q	Q
A.2	3	K	N	H	T	E	N	E	K	D	Q	Q
A.3	33	K	N	H	N	E	N	E	K	D	Q	Q
A.4	8	K	N	H	N	E	N	E	K	D	Q	Q
A.5.1	15	A	N	H	N	E	N	E	K	D	Q	Q
A.5.2	3	A	N	Q	N	N	S	E	K	D	Q	Q
A.5.3	48	K	N	H	N	E	N	E	K	D	Q	Q
A.5.4	12	R	N	H	N	E	N	E	E	D	Q	Q
A.5.5	51	K	G	H	D	W	N	E	K	D	Q	Q
D	3	K	N	H	N	E	S	E	K	D	Q	Q
E	13	K	N	H	N	E	S	E	K	D	Q	Q
B	3	K	D	Q	N	R	S	A	V	D	Q	Q
B.1.1	30	K	D	Q	N	R	S	A	I	D	L	Q
B.1.2	44	K	D	Q	N	R	S	A	I	D	L	Q
B.3	7	R	G	Q	N	R	S	E	I	D	L	Q
B.4	7	K	D	Q	N	R	S	A	T	D	L	Q
B.2.1	33	K	D	Q	N	R	D	A	T	N	L	I
B.2.2	36	K	D	Q	N	R	D	A	T	N	L	I
B.2.3	9	K	N	Q	N	R	D	A	T	N	L	I
B.2.4	7	K	D	Q	N	Q	D	A	T	N	L	I
B.2.5	31	K	D	Q	N	R	D	A	T	N	L	I
B.2.6	33	K	E	Q	N	R	D	T	T	N	L	I
B.2.7	124	K	D	Q	N	R	D	A	T	N	L	I
C.1	31	N	N	Q	N	R	S	V	T	D	L	T
C.2	224	K	D	Q	N	R	N	T	T	D	Q	Q
C.2.1	150	K	D	Q	N	R	N	A	T	D	L	Q
C.2.2	694	K	D	Q	N	Q	N	T	T	D	L	Q
C.2.3	3	K	D	Q	N	R	N	V	T	D	L	Q

Table 4. Class and sub-class specific variation at the RBD in H9N2 viruses. The number of strain sequences and the most represented residue for each amino acid position (columns) are reported for each clade (rows). Negatively charged, positively charged and hydrophobic residues are highlighted by red, blue or yellow background, respectively. Position numbering for four HA subtypes is reported.

Class and sub-class specific variation in electrostatic and hydrophobicity features

Table 4 also reports intriguing variation at the contiguous positions 216 and 217. Clades from classes A-D-E share at position 216 a highly conserved (176/177 strains) polar residue (Gln), while polar to hydrophobic transition is clearly apparent in classes B and C. In particular, the 'original' Gln is replaced in the most of strains (1256/1466) by the hydrophobic residue Leu, showing prevalence in all clades (from classes B-C) but 'B' (the ancestral one without numbering) and C.2. However, Gln is still present in 202/1466 B-C strains and the most represented residue in clades 'B' and C.2. In the next position (217), sub-class variation is observed: only B.2.x viruses share a hydrophobic residue (Ile), while Gln is common to all other clades in classes A-B-C-D-E. When inspecting in more detail distribution of residues among individual clades and strains, the picture is meaningfully different from position 216. In addition to classes A-D-E (176/177 strains), the 'original' Gln is highly conserved also in class C (900/1102 strains) and in C1 the major residue is anyway a polar one (Thr, in 30/31 strains). Instead, a complex picture is displayed in class B: Gln is still 100% conserved in clades B, B1.1, B1.2 and B3, whereas polar to hydrophobic transition is ongoing in clade B4 (4 hydrophobic strains out of 7) and fully fixed (100% of strains) in the whole B.2.x subgroup. Such specific variation in sub-class B.2.x is only apparently restricted to hydrophobic patches, as 'charge sliding' is observed between positions 165 and 198. In particular, all H9 clades but B2.x share a 100% negatively charged residue at position 198, which is replaced in B2.x by a polar amino acid (mostly, Asn). Such a peculiar (with respect to other H9 viruses) denegativization is however compensated in B.2.x by an equally peculiar acquisition of a negative charge at position 165, where Asp is 100% conserved.

Residues involved in changes at the H9N2 RBD are surface exposed

The RBD from the solved structure of the H9 HA was viewed to highlight the nine amino acid positions involved in class or sub-class specific variation: as shown in Figure 6, all such positions are exposed at the RBD surface. The RBD subregions (130-loop, 190-helix and 220-loop) mediating SA binding are highlighted in yellow. The three residues 135, 146 and 162 involved in class specific 'charge redistribution' are highlighted in orange; in particular, 146 is close to 190-helix, 135 is part of 130-loop and 162 is surface exposed as well. Position pair 180-186 (mediating 'charge loss' in the A-D-E to B-C transition) is highlighted in purple and is part of 190-helix. Concerning the four positions involved in class and sub-class variation (highlighted in green), positions 216 and 217 are part of 220-loop, while 165 and 198 protrude at the other 'side'. Finally, positions 131 and 161 involved in compensatory variation (see Table 4) were also confirmed to be surface exposed (not shown).

Given that surface location is confirmed for all aforementioned positions, variation in H9N2 among different classes, subclasses and clades can be viewed and analyzed in more depth by comparison of the electrostatic isocontours, which were determined as previously reported (Righetto *et al.,* 2014; see also the methods section). Prior to starting electrostatic analysis, we checked the quality of models because shaping of the isopotential contour can be influenced by the orientation of side chains. However, model refinement proved to be unnecessary because of the very high average sequence identity (around 90%) of the H9N2 target sequences to the H9 template, as previously observed with H5N1 models (Righetto *et al.,* 2014).

Figure 7 shows the isopotential contours from two viral strain that are well representative for electrostatic fingerprints from 'wild bird' viruses (classes A-D-E) and 'poultry' viruses (classes B-C). In fact, both strain A.1_AtkCA66 and C.2.3_AckHe07 match in all positions

patterns in Table 4 typical for classes A-D-E or classes B-C, respectively. At position 162, the two strains clearly show the opposite charges; concerning position 135, the expected contours are found again, as A.1_AtkCA66 shows no charge while in C.2.3_AckHe07 a negative protrusion is found in the corresponding area. However, not always expectations are respected and comparison of the 180-186 amino acid pair clearly shows that the loss of both charged residue in the A-D-E to B-C transition does not just result in 'neutralization' of the corresponding surface area. In fact, in spite of the expected red(Glu)-to-neutral(Val) shift, position 180 shows in C.2.3_AckHe07 a seeming positivization (increased blue area), possibly because of the enlargement of electrostatic clouds from neighboring occurs.

For completeness of information, the isopotential contours of the RBDs from all representative strains used for creating the heat map and epogram in figures 4 and 5 are presented in Supplementary Figure S3.

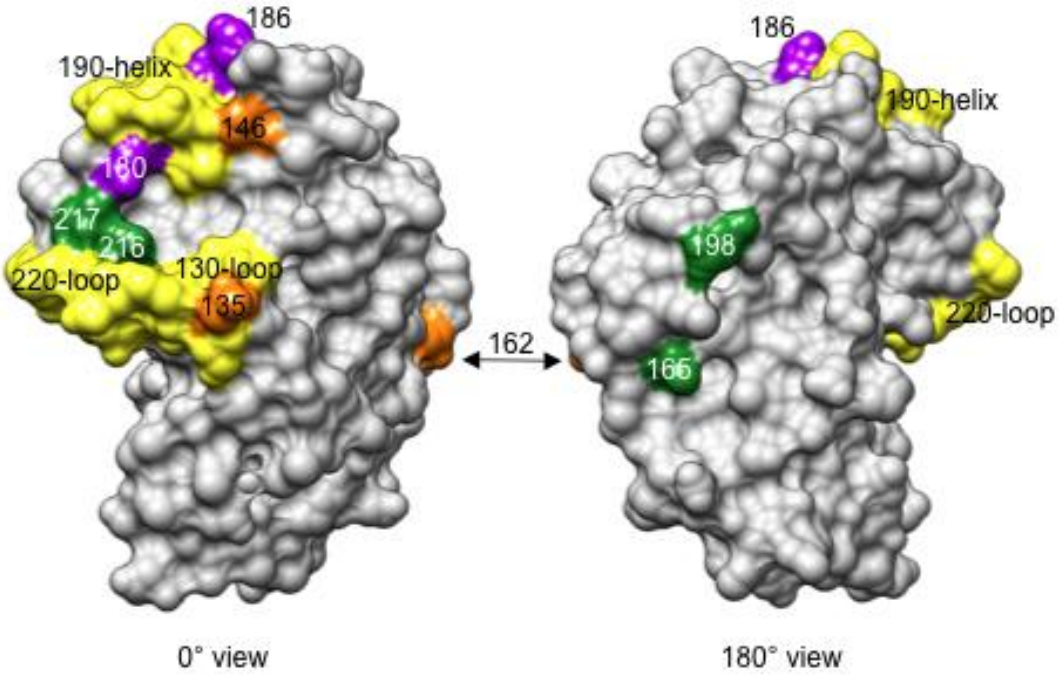


Figure 6. Location at the RBD surface of residues involved in class and sub-class specific variation in H9N2. Both 0° view (left image) and 180° view (right image) are shown. The three regions (130-loop, 190- helix and 220-loop) that mediate binding to the sialic acid (SA) moieties from the host cell (Wilson *et al.*, 1981; Kobayashi *et al.*, 2012) are highlighted in yellow. Color coding for amino acid positions is the following: 135, 146 and 162: orange; 180 and 186: purple; 165, 198, 216 and 217: green.

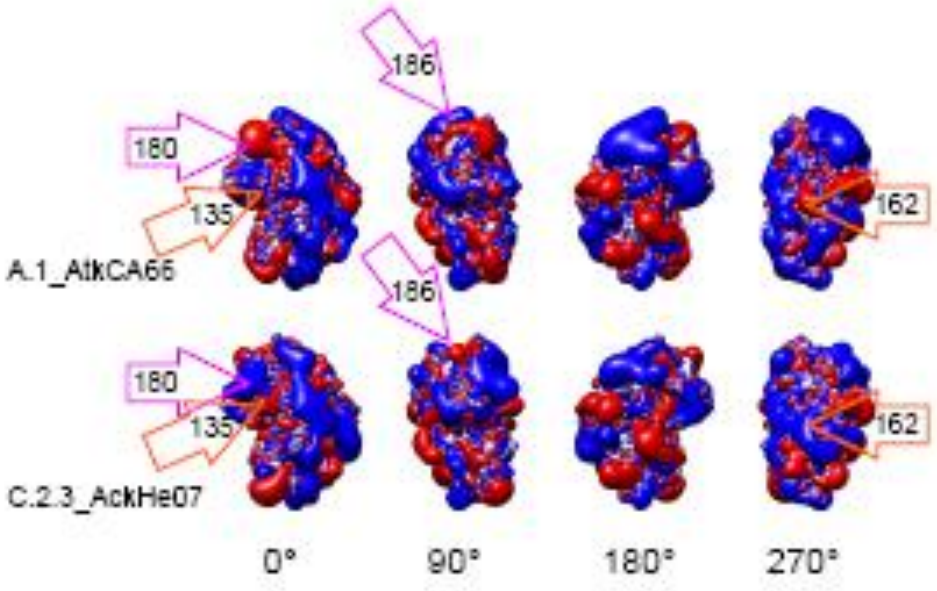
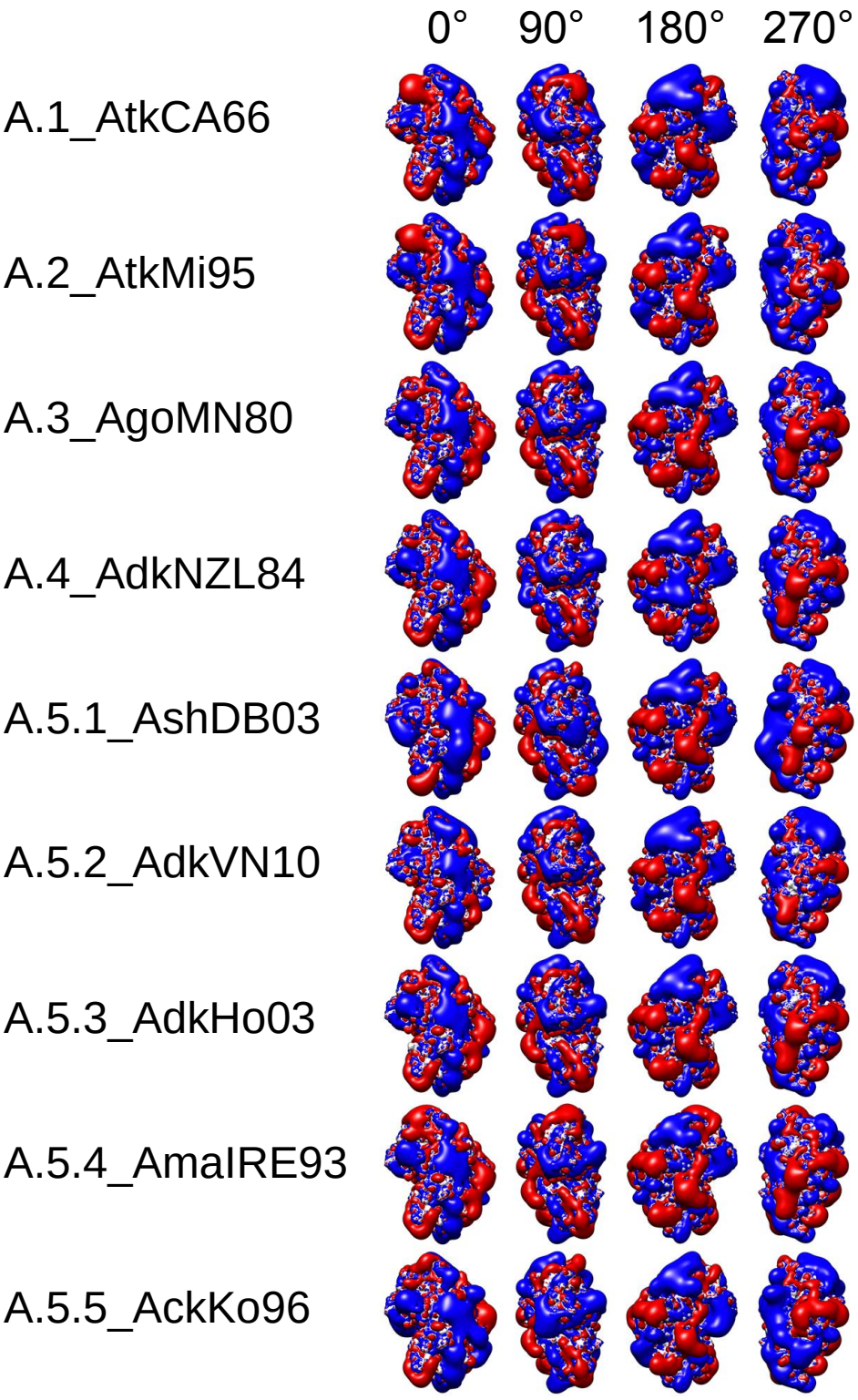
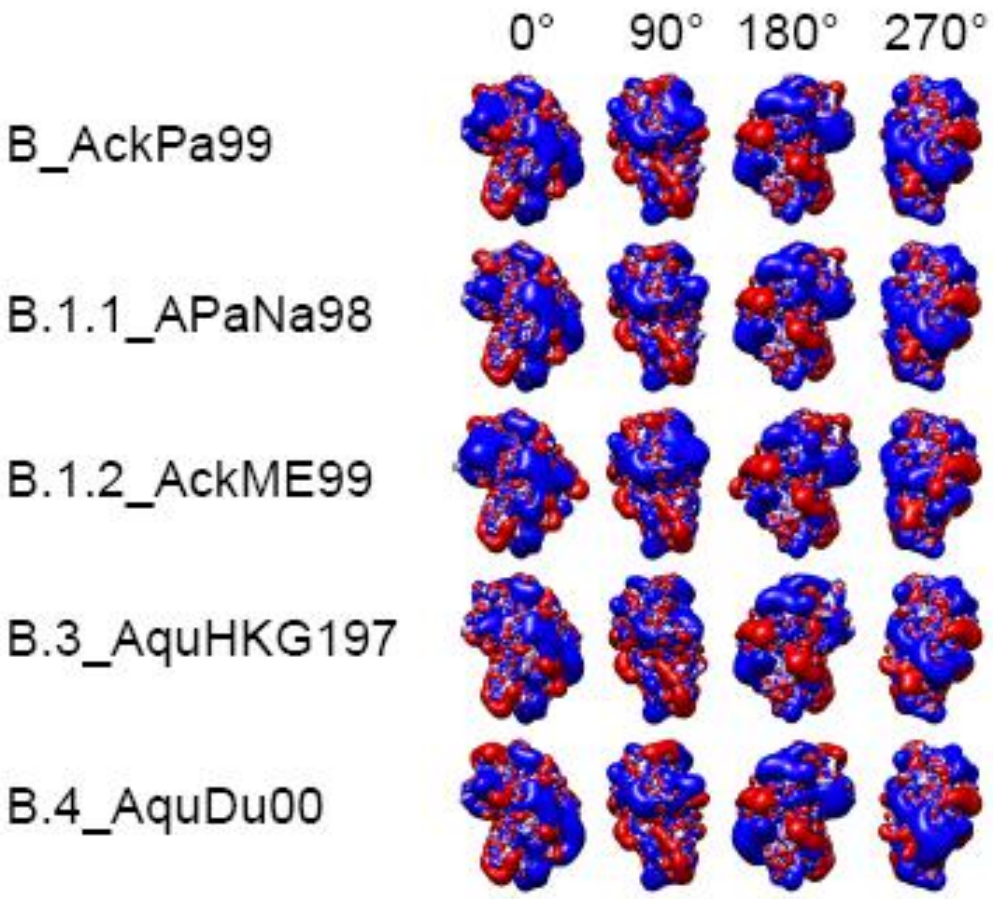
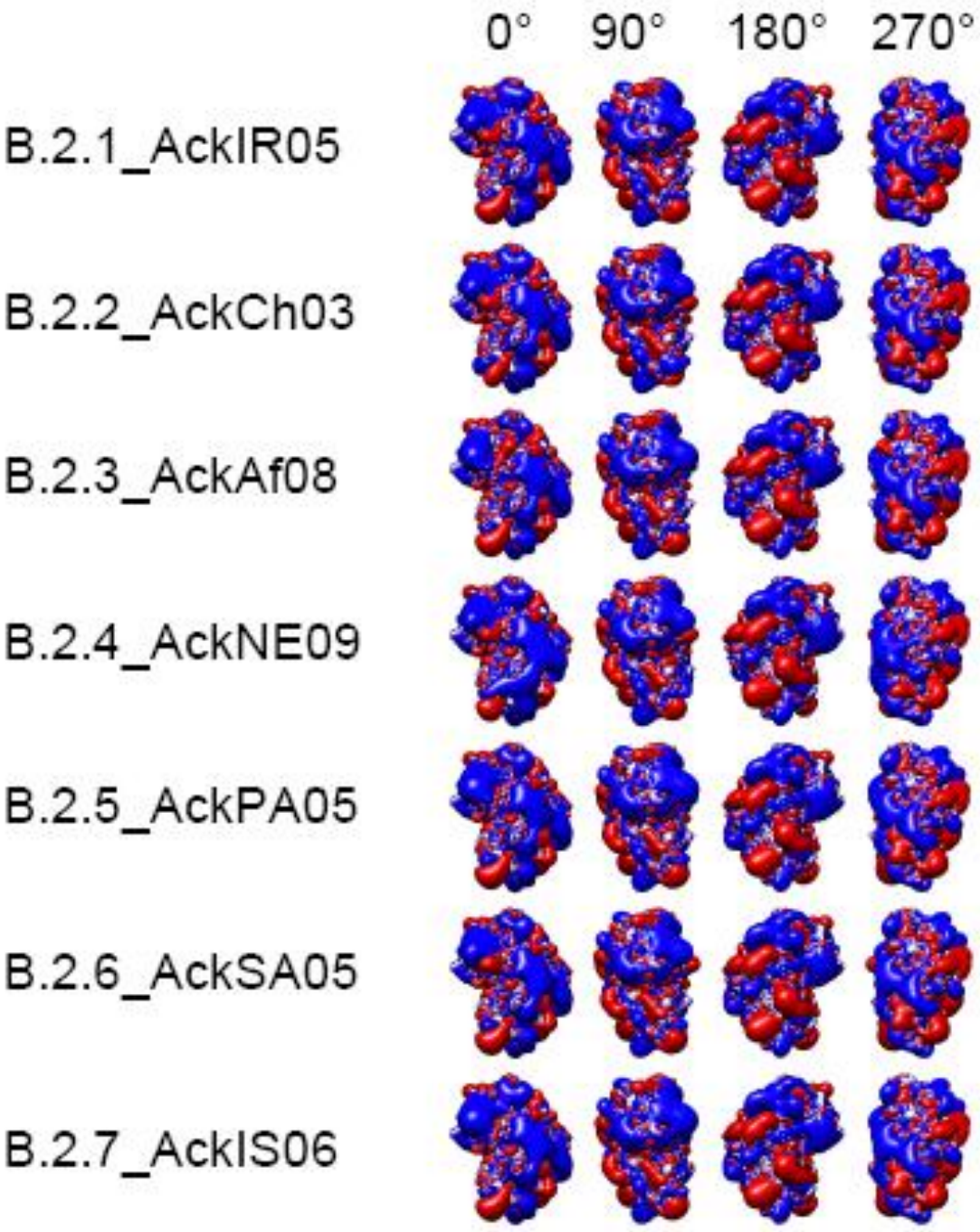


Figure 7: Comparison of representative profiles for the isopotential RBD contour in H9N2. Four 90° stepwise rotation view are presented for each representative RBD electrostatic isocontour. Names of the specific H9N2 virus strains are reported. Arrows for specific residues are color-coded according to Figure 6.


0°
90°
180°
270°
A.1_AtkCA66
A.2_AtkMi95
A.3_AgoMN80
A.4_AdkNZL84
A.5.1_AshDB03
A.5.2_AdkVN10
A.5.3_AdkHo03
A.5.4_AmaIRE93
A.5.5_AckKo96


0°
90°
180°
270°
B_AckPa99
B.1.1_APaNa98
B.1.2_AckME99
B.3_AquHKG197
B.4_AquDu00


0°
90°
180°
270°
B.2.1_AckIR05
B.2.2_AckCh03
B.2.3_AckAf08
B.2.4_AckNE09
B.2.5_AckPA05
B.2.6_AckSA05
B.2.7_AckIS06

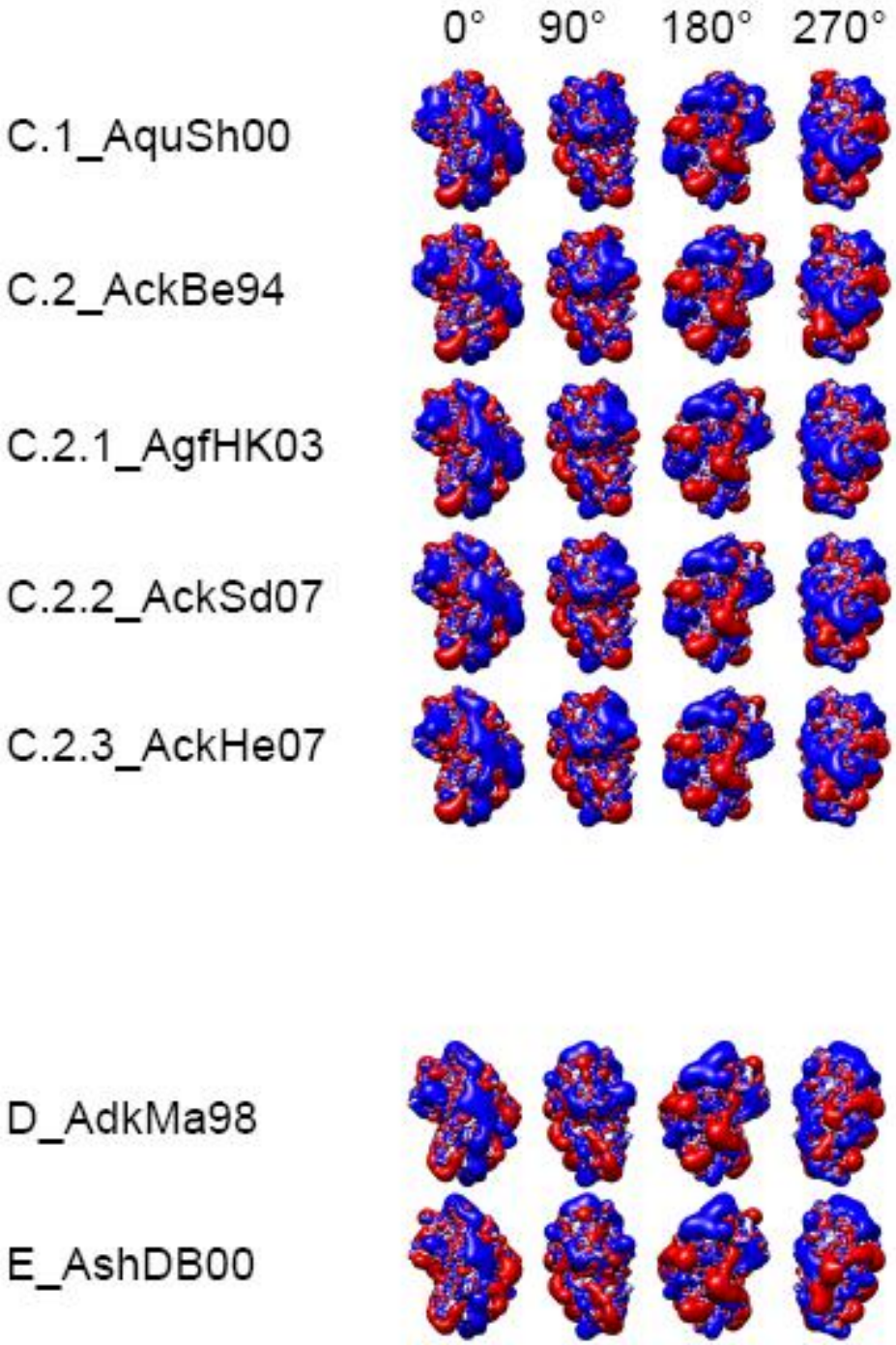


Figure S3. Isopotential contours of the RBD from representative H9N2 virus strains. Four 90° stepwise rotation views are presented for each representative RBD electrostatic isocontour. Names of the H9N2 virus strains are reported and they are the same as in figures 4 and 5.

Discussion and conclusions

Haemagglutinin has the role of main viral surface antigen in the stimulation of the neutralizing antibody response (Velkov *et al.,* 2013) and so far, classification of viral HA has been based upon antigenic and phylogenetic analyses (Stanekova and Vareckova, 2010). However, comparative structural analysis of HA can provide functional insights on surface regions possibly crucial to antigenicity and binding to the host cells. In fact, recent work on H5N1 demonstrated that electrostatic and hydrophobicity variation at the RBD surface relates to both evolution and spreading of viral clades and is able to provide fingerprints and infer trends to complement phylogeny and functional analyses (Righetto *et al.,* 2014). Intriguingly, when comparing RBD regions, H5 is structurally quite closer to H9 than H3 and H7, and when RBD electrostatic potential is considered, H5 is even closer to H9 (member of a different phylogenetic HA group) than to H2 (member with H5 of the same phylogenetic group) (Righetto *et al.,* 2014). Therefore, the possibility that similar mechanisms might underlie H5 and H9 evolution and spreading further prompted us to investigate on H9 evolution and on surface features of the H9 HA.

We developed a method for the HA clade nomenclature of all AI H9N2 subtypes, based on the evolutionary dynamics of a large and non redudant viral strain dataset. Clade assignments were made by following several criteria (Table 3), collectively used to rationally name groups by clade numbering. Based on phylogenetic topology, five different genetic classes could be distinguished, which so far consist of twenty-seven clades according to the different molecular analyses and fixed criteria. Once further clades arise along evolution of H9N2 viruses, such a nomenclature is ready to be expanded (by enlarged numbering) based on already fixed criteria. Circulation and evolution of the H9N2 HA gene show a remarkable similarity to the H5 subtype and notable difference from the typical evolution of H3. The

evolution of the human influenza viruses since 1968 is characterized by a limited diversity among circulating strains. This lack of diversity is likely the consequence of rapid extinction after the emergence of new clades and lineages. As expected, the evolutionary tree of human influenza HA genes has extended trunks and extremely short branches; conversely, AI H9N2 strains show extended branches as these viruses continue to co-circulate in different regions and host species and this allows the clades for further evolving and differentiating. Therefore, a standard nomenclature system for H9N2 classification is needed now in order (i) to provide a rationale for the AI H9N2 evolution, and (ii) to properly relate investigations worldwide thanks to robust guidelines. Moreover, we believe that the rapid and accurate annotation of clades will aid molecular epidemiologic assessment and support public health interventions. Last but not least, shared nomenclature criteria can boost correlation analyses and favour proper naming of newly identified strains along epidemiological analyses.

As a fundamental component of modern biogeography (Riddle, 2009) and an approach of great impact on the most basic of biological questions, phylogeography is actually a hot research field boosting studies aimed at clarifying evolutionary dynamics in most life sciences disciplines (Turchetto-Zolet *et al.,* 2013; Brown *et al.,* 2014; Ni *et al.,* 2014; Gräf *et al.,* 2015; Pyron, 2015; Zhang *et al.,* 2015; Maixner *et al.,* 2016; Stacy *et al.,* 2016) including analysis of influenza viruses (Lu *et al.,* 2014; Bedford *et al.,* 2015; Hill *et al.,* 2015; Pollett *et al.,* 2015; Tian *et al.,* 2015). In addition to the classification scheme and as a complement to it, phylogeographic data can provide further information on the spatiotemporal evolution, correlation and spreading of AI H9N2 viruses from the beginning to current trends.

Relationship among variation in electrostatic features, viral evolution and clades spreading observed for H9N2 in this work confirms and further strengthen previous observations on H5N1 AI viruses, in which the evolution of circulating clades is accompanied by 'charge

redistribution' at the RBD (Righetto *et al.,* 2014). Moreover, most of changes occuring at the RBD in H9N2 viruses seems to concern sites known to play a special role in RBD interactions, immune escape and host specificity. In particular, three RBD subregions mediate binding to the sialic acid (SA) moieties from the host cell and they are major antigenic determinants hence involved in immune escape/antigenic drift: 130-loop (H3 residues 135-138), 190-helix (H3 residues 190-198) and 220-loop (H3 residues 221-228) (Wilson *et al.,* 1981; Kobayashi *et al.,* 2012). For Readers' convenience, in addition to H9 and H3 numbering systems, Table 4 also shows HA mature chain numberings for subtypes H1, H5, and H7, all five numberings being based on the most recently published table of correspondence (Burke and Smith, 2014).

As described in the results section, changes at positions 162 and 217 of H9N2 HA result in either class or sub-class specific variation of charge or hydrophobicity features of the RBD. In particular, position 162 is involved together with positions 135 and 146 in class specific 'charge redistribution' and, when denegativization occurs at position 162 in two clades from class A, this is compensated by either depositivization of residue 146 (in A5.2) or negativization of residue 161 (in A5.5). Sub-class specific, polar to hydrophobic transition occurs instead at residue 217. The involvement in immune escape of mutations at both positions 162 and 217 in H9N2 has been recently reported (Peacock *et al.,* 2016). Indeed, as part of the 220-loop, position 217 is also likely involved in increased virus binding to α2-6 SA and thus in improved affinity to the human host; for instance, residue 224 in H1N1 (217 in H9) mediates hydrogen bond interactions with α2,6-SA (Chutinimitkul *et al.,* 2010). Evidence on the involvement of this RBD position in the modulation of host range is also based on previous studies on H5N1 (Gambaryan *et al.,* 2005). Position 227 (H3 numbering) is located between amino acids 226 and 228, both being part of the 220-loop and playing a key

role in receptor specificity and host range restriction of influenza A viruses (Vines *et al.*, 1998). In this work, class-specific variation in H9N2 position 216 (226 in H3) is observed (Table 4). Positions 180-186 in H9N2 (where the conserved dual opposite charges pair in classes A-D-E shifts to a non-charged pair in classes B-C, see previous sections) both belong to the 190-helix of the RBD involved in binding to SA moieties from the host cell (Wilson *et al.*, 1981; Kobayashi *et al.*, 2012). Contemporary loss of the two opposite charges is somehow 'compensatory' in that saving the original net charge of the RBD. It is noteworthy that in influenza A viruses, amino acid substitutions increasing (charge+) and decreasing (charge-) the charge of the SA binding RBD region can modulate binding avidity and affinity, and thus contemporary charge+ and charge compensatory substitutions are often observed and likely to compensate gain and loss effects and to ensure the HA-NA charge balance is kept (Kobayashi *et al.*, 2012). However, in A-D-E to B-C class transition, compensation only keeps the net charge, while the overall loss of two charged residues from the RBD in B-C class viruses occurs. This in turn is likely to favour immune escape, because of both the location of the two residues and the well-known role of charged amino acids in modulating protein antigenicity and immunogenicity. In fact, it is well known and an established evidence that both positively and negatively charged residues improve the antigenic recognition (up to several folds, depending on their number in the antigenic site) by creating further salt bridges with the recognizing antibody complementary surface (Young CR, 1984; Farber *et al.*, 2007).

As shown in Table 4, charge variation also occurs at position 146 (involved with 135 and 162 in 'charge redistribution' on the RBD, see results), which is exposed at the RBD surface close to 190-helix (Figure 6). Based on the aforementioned role of 190-helix in binding SA moieties from host cells, changes like the observed depositivization at position 146 are likely to influence binding affinity and specificity. Considering that chickens possess both α-2′3′ and

α-2′6′ SA receptors (Gambaryan *et al.,* 2002), it is tempting to speculate that such changes could be linked to host adaptation and species specificity (Perez *et al.,* 2003).

In addition to represent a valuable complement to integrate phylogenetic and antigenic studies, structural analyses are also of great help to improve sequence-based, functional comparison. Sequence comparison was able to infer class and sub-class specific fingerprints presented in Table 4 as sequence patterns; however, only once sequence analysis was complemented by the structural approach, a real-estate picture of the system emerged. Comparison of the electrostatic isocontours showed that identified mutations cannot be considered as just isolated 'point changes'. In fact, surface features that are pivotal players in regulating molecular interactions e.g. immune escape and host specificity are modulated by the direct change of any mutated residue, as well as by the effects that such a mutation may exert on the local equilibrium in the surrounding area (salt bridges or repulsions, hydrophobicity changes, decreased or increased charge density etc.), as shown by the unexpected variations observed in charge clouds.

In conclusion, although much further work is needed to clarify in details the complex network of equilibria that can be altered by specific mutations, evidence from this study supports the integration of up-to-date phylogenetic and phylogeographic analyses with sequence-based and structural investigation of surface features as a front-end strategy for inferring trends and relevant mechanisms in influenza virus evolution.

List of abbreviations

Ab: antibody;
AC: accession code;
AI: Avian influenza;
AICc: corrected Akaine Information Criterion;
BF: Bayes factor;
ED: electrostatic distance;
Epogram: electrostatic potential diagram;
GTR: General Time Reversible
GISAID: Global Initiative on Sharing Avian Influenza Data;
HA: Haemagglutinin;
HPAI: High pathogenic Avian influenza;
I: ionic strength;
LPAI: Low pathogenic Avian influenza;
MCC: maximum clade credibility;
MCMC: Markov Chain Monte Carlo;
ML: maximum-likelihood;
NA: Neuraminidase;
NJ: neighbor-joining;
OIE: Office International des Epizooties;
PDB: Protein Data Bank;
PIPSA: Protein Interaction Property Similarity Analysis;
RBD: receptor-binding domain;
SA; Sialic acid;
WHO: World Health Organization.

Competing interests and authors' contributions

The authors declare that they have no competing interests.

AM, AH, FF and GC conceived the study. FF oversaw the study. AM and AH performed most of analyses; AF and IR provided other authors with help in data interpretation. AM, AH and FF wrote the paper with inputs from AF, IM and GC. All authors read and approved the final manuscript.

Acknowledgements

The phylogenetic studies reported in this project was conducted in the framework of the Doctoral school in Veterinary Science at the University of Padua (Alireza Heidari). The studies were carried out mainly at the Research & Innovation Department, Division of Biomedical Science, Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy. In addition, this work was supported by basic funding (‘ex 60%’) from the Italian Ministry for University and Research (MIUR) to FF. We gratefully acknowledge the contributing authors and the originating and submitting laboratories for the sequences from the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database.

References

1. Alexander DJ. An overview of the epidemiology of avian influenza. Vaccine. 2007;25(30):5637-44.
2. Al-Tawfiq JA, Zumla A, Gautret P, Gray GC, Hui DS, Al-Rabeeah AA, Memish ZA. Surveillance for emerging respiratory viruses. Lancet Infect Dis. 2014;14(10):992-1000.
3. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, Daniels RS, Gunasekaran CP, Hurt AC, Kelso A, Klimov A, Lewis NS, Li X, McCauley JW, Odagiri T, Potdar V, Rambaut A, Shu Y, Skepner E, Smith DJ, Suchard MA, Tashiro M, Wang D, Xu X, Lemey P, Russell CA. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 2015; 523(7559):217-20.
4. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W510-4.
5. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011;27(20):2910-2.
6. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T: Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 2009; 4(1):1-13.
7. Brown MV, Ostrowski M, Grzymski JJ, Lauro FM. A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. Mar Genomics. 2014;15:17-28.
8. Burke DF, Smith DJ. A recommended numbering scheme for influenza A HA subtypes. PLoS One. 2014;9(11):e112302.
9. Butler D. Flu surveillance lacking. Nature. 2012 Mar 28;483(7391):520-2. doi:10.1038/483520a.

10. Butt KM, Smith GJ, Chen H, Zhang LJ, Leung YH, Xu KM, Lim W, Webster RG, Yuen KY, Peiris JS, Guan Y. Human infection with an avian H9N2 influenza A virus inHong Kong in 2003. J Clin Microbiol. 2005;43(11):5760-7.
11. Carugo O, Pongor S: A normalized root mean square distance for comparing protein three dimensional structures. Protein Sci 2001;10:1470-1473.
12. Chutinimitkul S, Herfst S, Steel J, Lowen AC, Ye J, van Riel D, Schrauwen EJ, Bestebroer TM, Koel B, Burke DF, Sutherland-Cash KH, Whittleston CS, Russell CA, Wales DJ, Smith DJ, Jonges M, Meijer A, Koopmans M, Rimmelzwaan GF, Kuiken T, Osterhaus AD, Garcìa-Sastre A, Perez DR, Fouchier RA. Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. J Virol. 2010;84(22):11802-13.
13. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22(5):1185-92.
14. Farber DL, Sleasman JW, Virella G. Immune response: Antigens, Lymphocytes and Accessory Cells. Medical Immunology, Sixth Edition, 2007; Chapter 4:35-54.
15. Gambaryan, A., R. Webster, and M. Matrosovich. Differences between influenza virus receptors on target cells of duck and chicken. Arch. Virol. 2002;147:1197-1208.
16. Gambaryan A, Tuzikov A, Pazynina G, Bovin N, Balish A, Klimov A. Evolution of the receptor binding phenotype of influenza A (H5) viruses. Virology. 2006;344(2):432-8.
17. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A: Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook Edited by Walker JM. Humana Press 2005:571-607.
18. Gräf T, Vrancken B, Maletich Junqueira D, de Medeiros RM, Suchard MA, Lemey P, Esteves de Matos Almeida S, Pinto AR. Contribution of Epidemiological Predictors in

Unraveling the Phylogeographic History of HIV-1 Subtype C in Brazil. J Virol. 2015;89(24):12341-8.

19. Guan Y, Smith GJ. The emergence and diversification of panzootic H5N1 influenza viruses. Virus Res. 2013;178(1):35-43.
20. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696-704.
21. Hill SC, Lee YJ, Song BM, Kang HM, Lee EK, Hanna A, Gilbert M, Brown IH, Pybus OG. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. Infect Genet Evol. 2015;34:267-77.
22. Hu M, Li X, Ni X, Wu J, Gao R, Xia W, Wang D, He F, Chen S, Liu Y, Guo S, Li H, Shu Y, Bethel JW, Liu M, Moore JB, Chen H. Coexistence of Avian Influenza Virus H10 and H9 Subtypes among Chickens in Live Poultry Markets during an Outbreak of Infection with a Novel H10N8 Virus in Humans in Nanchang, China. Jpn J Infect Dis. 2015;68(5):364-9.
23. Jin Y, Yu D, Ren H, Yin Z, Huang Z, Hu M, Li B, Zhou W, Yue J, Liang L. Phylogeography of Avian influenza A H9N2 in China. BMC Genomics. 2014;15:1110.
24. Kobayashi Y, Suzuki Y. Compensatory evolution of net-charge in influenza A virus hemagglutinin. PLoS One. 2012;7(7):e40422.
25. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9):e1000520.
26. Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, Subbarao K, Guan Y, Krauss S, Shortridge K, Webster R, Cox N, Hay A. Avian-to-human transmission of

H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci U S A. 2000;97(17):9654-8.

27. Lu L, Lycett SJ, Leigh Brown AJ. Determining the phylogenetic and phylogeographic origin of highly pathogenic avian influenza (H7N3) in Mexico. PLoS One. 2014;9(9):e107330.
28. Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, Kusebauch U, Vigl EE, Malfertheiner P, Megraud F, O'Sullivan N, Cipollini G, Coia V, Samadelli M, Engstrand L, Linz B, Moritz RL, Grimm R, Krause J, Nebel A, Moodley Y, Rattei T, Zink A. The 5300-year-old Helicobacter pylori genome of the Iceman. Science. 2016;351(6269):162-5.
29. Monne I, Fusaro A, Nelson MI, Bonfanti L, Mulatti P, Hughes J, Murcia PR, Schivo A, Valastro V, Moreno A, Holmes EC, Cattoli G. Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. J Virol. 2014;88(8):4375-88.
30. Nelson MI, Vincent AL. Reverse zoonosis of influenza to swine: new perspectives on the human-animal interface. Trends Microbiol. 2015;23(3):142-53.
31. Ni G, Li Q, Kong L, Yu H. Comparative phylogeography in marginal seas of the northwestern Pacific. Mol Ecol. 2014;23(3):534-48.
32. Peacock T, Reddy K, James J, Adamiak B, Barclay W, Shelton H, Iqbal M. Antigenic mapping of an H9N2 avian influenza virus reveals two discrete antigenic sites and a novel mechanism of immune escape. Sci Rep. 2016;6:18745.
33. Peiris M, Yuen KY, Leung CW, Chan KH, Ip PL, Lai RW, Orr WK, Shortridge KF. Human infection with influenza H9N2. Lancet. 1999;354(9182):916-7.
34. Perez DR, Lim W, Seiler JP, Yi G, Peiris M, Shortridge KF, Webster RG. Role of quail in the interspecies transmission of H9 influenza A viruses: molecular changes on HA that

correspond to adaptation from ducks to chickens. J Virol. 2003;77(5):3148-56.

35. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605-1612.
36. Pollett S, Nelson MI, Kasper M, Tinoco Y, Simons M, Romero C, Silva M, Lin X, Halpin RA, Fedorova N, Stockwell TB, Wentworth D, Holmes EC, Bausch DG. Phylogeography of Influenza A(H3N2) Virus in Peru, 2010-2012. Emerg Infect Dis. 2015;21(8):1330-8.
37. Pyron RA. Post-molecular systematics and the future of phylogenetics. Trends Ecol Evol. 2015;30(7):384-9.
38. Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade R. WebPIPSA: a web server for the comparison of protein interaction properties. Nucleic Acid Res 2008;36(Web Server Issue):W276-W280.
39. Riddle BR. What is modern biogeography without phylogeography? J. Biogeogr 2009;36:1-2.
40. Righetto I, Milani A, Cattoli G, Filippini F. Comparative structural analysis of haemagglutinin proteins from type A influenza viruses: conserved and variable features. BMC Bioinformatics. 2014;15:363.
41. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572-4.
42. Sang X, Wang A, Ding J, Kong H, Gao X, Li L, Chai T, Li Y, Zhang K, Wang C, Wan Z, Huang G, Wang T, Feng N, Zheng X, Wang H, Zhao Y, Yang S, Qian J, Hu G, Gao Y, Xia X. Adaptation of H9N2 AIV in guinea pigs enables efficient transmission by direct contact and inefficient transmission by respiratory droplets. Sci Rep. 2015;5:15928.

43. Sitkoff D, Sharp K, Honig B: Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 1994;98:1978-1988.
44. Stacy A, McNally L, Darch SE, Brown SP, Whiteley M. The biogeography of polymicrobial infection. Nat Rev Microbiol. 2016;14(2):93-105.
45. Stanekova Z and Vareckova E: Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virol J. 2010;7:351.
46. Su S, Bi Y, Wong G, Gray GC, Gao GF, Li S. Epidemiology, Evolution, and Recent Outbreaks of Avian Influenza Virus in China. J Virol. 2015;89(17):8671-6.
47. Swayne DE. Impact of vaccines and vaccination on global control of avian influenza. Avian Dis. 2012;56(4 Suppl):818-28.
48. Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 2002;19(10):1727-36.
49. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731-9.
50. Tian H, Zhou S, Dong L, Van Boeckel TP, Cui Y, Newman SH, Takekawa JY, Prosser DJ, Xiao X, Wu Y, Cazelles B, Huang S, Yang R, Grenfell BT, Xu B. Avian influenza H5N1 viral and bird migration networks in Asia. Proc Natl Acad Sci U S A. 2015;112(1):172-7.
51. Trombetta C, Piccirella S, Perini D, Kistner O, Montomoli E. Emerging Influenza Strains in the Last Two Decades: A Threat of a New Pandemic? Vaccines (Basel). 2015;3(1):172-85.

52. Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C. Phylogeographical patterns shed light on evolutionary process in South America. Mol Ecol. 2013;22(5):1193-213.
53. Velkov T, Ong C, Baker MA, Kim H, Li J, Nation RL, Huang JX, Cooper MA, Rockman S. The antigenic architecture of the hemagglutinin of influenza H5N1 viruses. Mol Immunol. 2013;56(4):705-19.
54. Vines A, Wells K, Matrosovich M, Castrucci MR, Ito T, Kawaoka Y. The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction. J Virol. 1998;72(9):7626-31.
55. Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature. 1981;289(5796):366-73.
56. WHO/OIE/FAO H5N1 Evolution Working Group. Toward a unified nomenclature system for highly pathogenic avian influenza virus (H5N1). Emerg Infect Dis. 2008;14(7):e1.
57. Young CR. Structural requirements for Immunogenicity and Antigenicity. in: Molecular Immunology A Textbook. CRC Press, Atassi, Van Oss, Absolom eds. 1984;1-14.
58. Zhang L, Li H, Li S, Zhang A, Kou F, Xun H, Wang P, Wang Y, Song F, Cui J, Cui J, Gouge DH, Cai W. Phylogeographic structure of cotton pest Adelphocoris suturalis (Hemiptera: Miridae): strong subdivision in China inferred from mtDNA and rDNA ITS markers. Sci Rep. 2015;5:14009.

Chapter 4

Phylogenetic and phylogeographic investigation and molecular evolution dynamics of H9N2 Avian Influenza Virus in Iran from 1998 to 2014

Alireza Heidari[1], Seyed ali Ghafouri[2], Hassan Nili[3], Gholam Hossein Pourghanbari[3,4], Alessandra Piccirillo[1]

[1]Department of Comparative Biomedicine and Food Science (BCA), University of Padua, Legnaro (PD), Italy.

[2] Iranian Veterinary Organization, Tehran Iran.

[3]Avian Diseases Research Center, School of Veterinary Medicine, Shiraz University Shiraz, Iran.

[4]School of Veterinary Medicine, Ardakan University Yazd, Iran.

Abstract

Since the ‘90s, avian influenza viruses (AIVs) of the H9N2 subtype have been causing infections in the poultry population around the globe. In Iran, H9N2 has been identified for the first time in 1998 and it is still circulating in the country, with a mortality rate ranging between 20% and 60% in the affected farms. Irrespective of the national control measures implemented, including mass vaccination of poultry, the virus has rapidly spread and currently it can be considered endemic. In this study, phylogenetic and phylogeographic analyses were carried out to obtain a comprehensive picture of the evolution dynamics of the Iranian H9N2 AIVs and to evaluate the possible effect of vaccination on the evolution and population dynamics. Analyses revealed that since 1998 four important introductions of H9N2 occurred in Iran and phylogeography indicated that the introductions had different origin. Amino acid analyses of the HA gene showed the presence of a positive selection pressure in the Iranian H9N2 viruses isolated after the implementation of the vaccination program in poultry. Additionally, analyses revealed that the first Iranian introduction was characterized by the presence of the amino acids Q and L at the position 226 with the same percentage, while the new introductions comprised mainly the amino acid L at this position. Data generated in this study provide complete and novel insights into the evolution and phylogeography of the Iranian H9N2 AIVs and highlight different important epidemiological aspects which may help the Iranian veterinary community to eradicate ongoing and future H9N2 outbreaks.

Background

Avian influenza viruses (AIVs) of the H9N2 subtype are widespread worldwide, being endemic in poultry populations in Asia and the Middle East (Capua and Alexander, 2007). A significant proportion of recent H9N2 AIV isolates contains the L226Q (H3 numbering) amino acid substitution in the hemagglutinins (HAs) which show preferential binding to analogues of receptors with sialic acid linked to galactose by α2,6 linkage (SAα2,6Gal), a phenotypic portrait characteristic of human influenza viruses. Thus, these AIVs seem to possess one of the key elements for infection in humans (Lin *et al.*, 2000). Since 1998, H9N2 AIV outbreaks have been one of the major problems of the Iranian poultry industry (Nili and Assasi, 2003). These viruses have spread in poultry flocks throughout the country; therefore, a vaccination strategy using an inactivated H9N2 vaccine was adopted. Despite the vaccination program, outbreaks have continued to occur in several broiler flocks, with great economic losses from either mortality and weight loss (Bahari *et al.*, 2015). The prolonged circulation of influenza viruses in a partially immunized population may have a direct impact on the evolution of the virus, particularly on the occurrence of antigenic drift (Shih *et al.*, 2007; Smith *et al.*, 2004). Different studies and monitoring projects have reported that many substitutions have occurred in the HA gene of the Iranian H9N2 viruses in recent years (Norouzian *et al.*, 2014), which may affect the pathogenicity and the antigenicity of the virus. Therefore, it is important to investigate and monitor whether the circulation of H9N2 AIVs have undergone further significant genetic changes, in particular under the selection pressure of vaccination. Phylogenetic and evolutionary dynamics analysis of Iranian H9N2 AIV strains can be an effective approach for vaccine development and to better understand the evolutionary trend and the introduction sources of these viruses in the Iranian geographic area, in order to prevent and eradicate the circulation of H9N2 AIVs.

In the present study, a deep investigation to obtain a comprehensive picture of the dynamics and the evolution of Iranian H9N2 viruses was performed. Furthermore, the phylogenetic relations grouping the various introductions and the spread to other countries were investigated to reveal the time and the location of the most recent ancestor, highlighting the nucleotide substitution rates and depicting the population dynamics and the extent of its molecular change since the H9N2 AIV first detection in Iran and after the application of a vaccination program.

Methods

Virus sequences

The HA gene nucleotide sequences of 88 Iranian H9N2 viruses and representative strains from G1 lineage were downloaded from publicly available databases: GenBank National Center for Biotechnology (NCBI) and Influenza Sequence Database of Global Initiative on Sharing Avian Influenza Data (GISAID). The analysis included only nearly full-length HA sequences (at least 1,500 nt in length) to ensure a robust analysis.

Phylogenetic analysis

Sequences were aligned using MAFFT version 7 (http://mafft.cbrc.jp/alignment/server/), and 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). Bootstrap analysis with 100 replicates was also employed. To confirm different genetic groups and introductions, genetic distances among Iranian H9N2 viruses by using MEGA5 software (Tamura *et al.*, 2002) were calculated. The distances were inferred by pair-wise analysis and the number of base substitutions per site was calculated by (p-distance) model. Rate of nucleotide substitutions per site, year and time to the Most Recent Common Ancestor (tMRCA) for datasets, comprising the HA gene of Iranian viruses, were estimated using the BEAST program, version 1.8 (Drummond *et al.*, 2007), which employs a Bayesian Markov chain Monte Carlo (MCMC) approach.

The dataset was analysed with the codon-based SRD06 nucleotide substitution model and the Skyline coalescent prior. Two molecular clock models, a strict (constant) and an

uncorrelated lognormal relaxed clock were compared by estimating the marginal likelihood implemented in the Tracer program (Rambaut A *et al.,* 2014) Prior to BEAST analysis, 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 and molecular-clock assumption were estimated from a maximum likelihood (ML) tree plotted against sampling date.

Phylogeography analysis

The HA gene sequence alignment including 160 H9N2 viruses from Iran and the most related clades was used for phylogeography analysis. To obtain the tree with robots, the Path-O-Gen tool was used for investigating the temporal signal and the clocklikeness of molecular phylogenies, and viruses without molecular-clock were removed from the alignment B.2 ancestor strains. Time-scaled phylogenies of the Iranian H9N2 HA gene were inferred by Bayesian Markov Chain Monte Carlo (MCMC) using BEAST v1.8.0 (Drummond *et al.,* 2007) using the SRD06 codon position model and the uncorrelated log-normal relaxed clock model under a Bayesian skyline coalescent tree prior to the MCMC simulations (Jin *et al.,* 2014). For the dataset, 500 million generations, with sampling every 50,000 steps, were performed. Convergence and effective sampling size (ESS) of estimates were assessed by visual inspection using Tracer v1.6 (Rambaut *et al.,* 2014). The Maximum Clade Credibility (MCC) trees with temporal and spatial annotation were summarised with a 10% burn-in removed, using TreeAnnotator v1.8.0 in the BEAST package and the topologies of the MCC trees were visualized using FigTree v1.3.1(http://tree.bio.ed.ac.uk/software/figtree/).

To animate viral dispersal over time, annotated MCC trees were converted into a keyhole markup language (KML) file using SPREAD v1.0.6, which can be visualised by Google Earth website platform (http://earth.google.com) (Bielejec *et al.*, 2011). A Bayes Factor (BF) comparison test (Suchard *et al.,* 2001) was used to identify significant non-zero migration rates using SPREAD v1.0.6.

Analysis of selection pressure

Gene and site-specific selection pressures for the HA protein of the Iranian H9N2 viruses were measured as the ratio of non-synonymous (dN) to synonymous (dS) nucleotide substitutions per site. In all cases, dN/dS ratios were estimated using the single-likelihood ancestor counting (SLAC) and fixed-effects likelihood (FEL). The codon-based maximum likelihood (IFEL) and FUBAR (Kosakovsky Pond *et al.,* 2005) methods were available at the DataMonkey online version of the HY-Phy package (http://www. datamonkey.org). Using phylogenetic Bayesian tree along with Mesquite software, the evolution and the character history of amino acid 226 (H3 numbering) was reconstructed. Using a Bayesian phylogenetic tree and an amino acid sequence distribution, Mesquite reconstructed the amino acid character states at ancestral nodes. Mesquite uses either parsimony, likelihood or Bayesian methods to reconstruct ancestral states and has several display methods, including "Trace Character History" which paints the branches of the tree to show the reconstruction.

Results

The maximum-likelihood phylogenetic tree (Figure 1), inferred from the H9N2 HA gene alignment of 88 Iranian AIVs and representative strains, showed that Iranian strains belong to four distinct genetic group, herein named clade B.1.1, B.2.1, B.2.3 and B.2.5, as defined by

high bootstrap values (>70%). The average distance between clades is >7% and the average distance within the clades is <5%. At population level, most of the Iranian strains isolated from 1998 to 2007 cluster into clade B.1.1, showing that the first introduction of Iranian H9N2 AIVs since 2003 was from clade B.2.1 that circulated until 2009 creating thus the second introduction. Since 2009, the third introduction from clade B.2.5 can be observed, which clusters with viruses from Pakistan and Afghanistan. In 2014, the fourth introduction from clade B.2.3 can be observed, which clusters with viruses from Pakistan and Afghanistan (Figure 1).

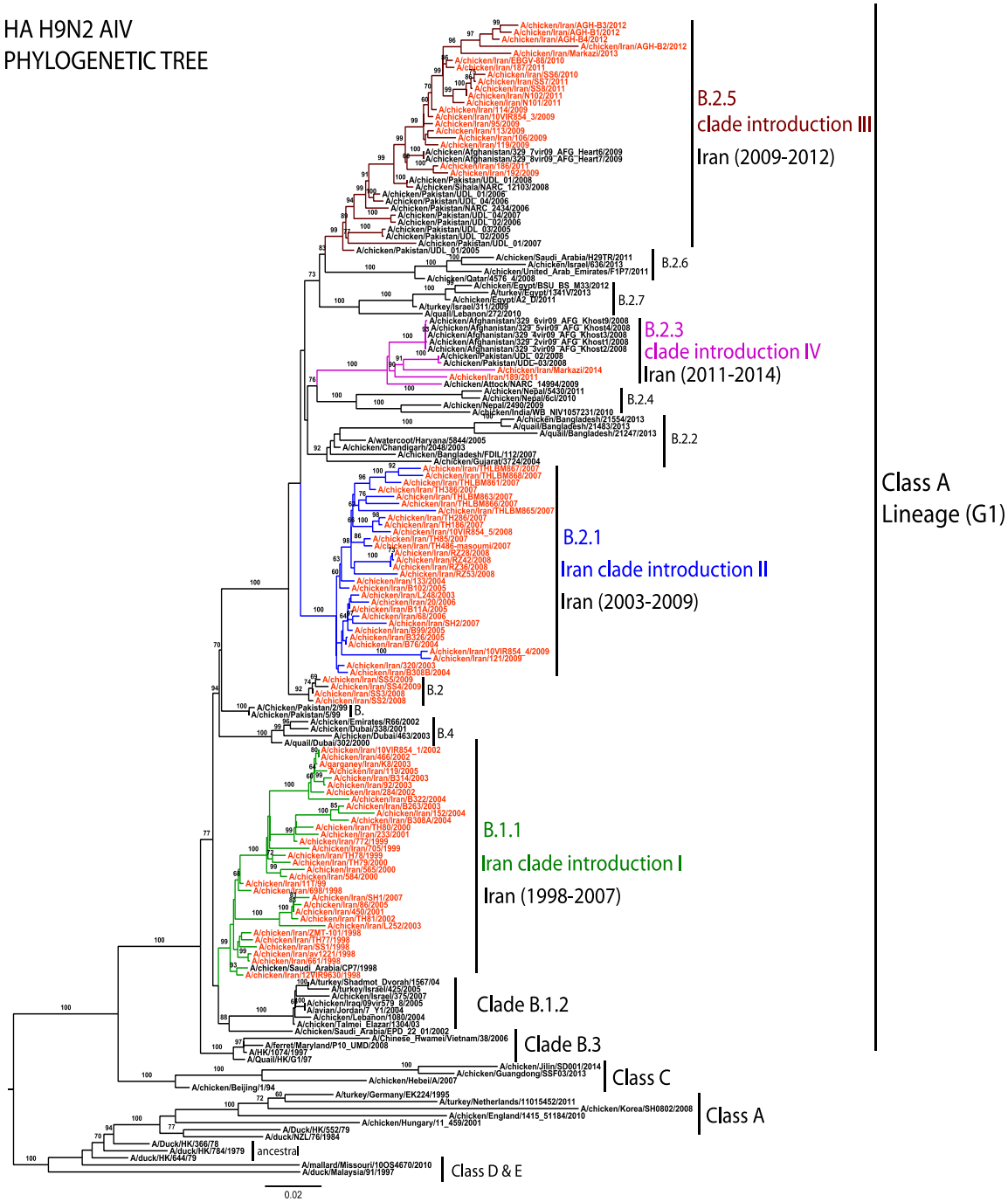


Figure 1. Maximum-likelihood tree of Iranian H9N2 AI viruses from 1998 until 2014 and Eastern countries constructed by PhyML. Colours and classification were by Heidari. Iranian isolates are showed in red and global H9N2 isolates in black. Different introductions and circulation date are reported behind of each clade.

Iranian H9N2 AIVs phylogeographic analysis

In order to determine the Iranian dissemination of H9N2 AIVs, the HA gene sequence alignment including the Iranian viruses and the most related clades was used for in-depth analysis. Through Posterior distribution under the Bayesian framework, genealogical trees with time-scale and inferred ancestral locations of each branch were reconstructed using sequences sampling collection dates and locations. The time-scaled phylogeographic maximum clade credibility (MCC) trees of HAs implemented in BEAST and the root state posterior probability are illustrated in Figure 2, in which the most probable location of each branch is assigned by different colors and by calibrating time-scale. Numbers at branch points in Figure 2 are reported where state probabilities with values ≥0.55 correspond to the most relevant events (i.e. country-country rather than intra-country transitions). Such transition events are graphically shown in Figure 3 world map distribution. Phylogeographic results suggest that the first Iranian H9N2 AIV introduction is of Chinese origin and it can be considered as ancestor of the Iranian clade B.1.1 viruses. Interpreting phylogeographic and phylogenetic trees, clades B.1.1 enclose the oldest Middle eastern H9N2 viruses, which are the oldest clades evolved in Middle East. Except for one case in Saudi Arabia, the clade B.1.1 evolved only in Iran; further, Iran may be the first country where H9N2 started to evolve. The ancestor of Iranian strains belonging to clade B.2.1 is likely to have been introduced from Pakistan. The viruses of clade B.2.1 did not circulate in any other country and, as for first Iranian introduction, evolved only in Iran. The last two Iranian introductions, belonging to clade B.2.5 and B.2.3, originated from Pakistan and, before the introduction in Iran, they had already evolved and circulated in that country. Iran contributed to the eastward spread of viruses from clade B.2.5 to the Afghanistan. The overall spatiotemporal representation of

phylogeographic evolution and spread of the H9N2 viruses is presented in Supplementary visual animation S1 (JUST ONLINE).

Figure 3. Phylogeography mapping and spreading of H9N2 AI viruses in Iranian and Middle-East countries. Viral transition from a geographic area to another one with the well supported Bayes factor > 3 and state probabilities values ≥ 0.55 are represented by lines.

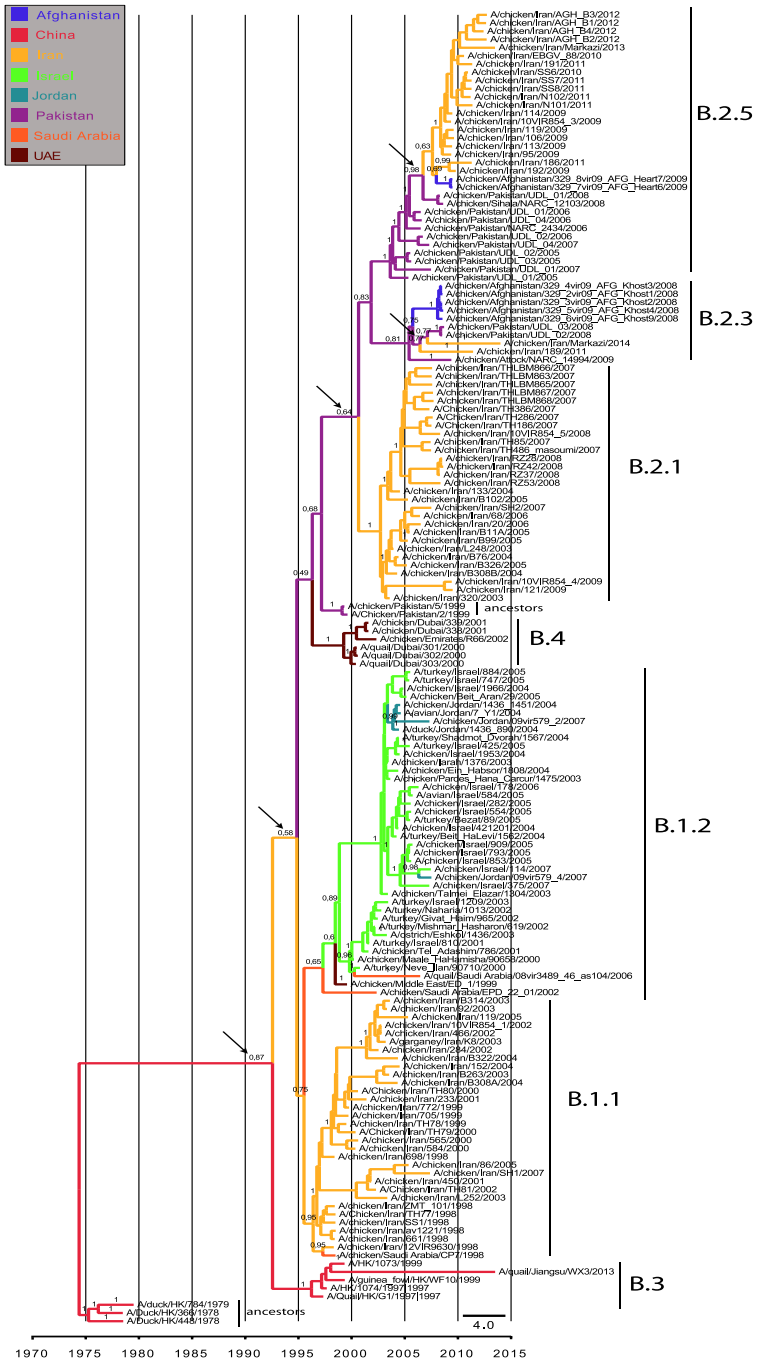


Figure 2. Maximum clade credibility (MCC) phylogenies inferred for the HA gene sequences of Iranian and Middle-East H9N2 AI viruses. Branches are coloured according to the most probable ancestor location (in terms of geographic area) of their descendent nodes. Timeline at the bottom indicates the years before the most recent sampling time. Numbers are reported at branch points where state probabilities with values ≥ 0.55 correspond to geographic area transition events.

Evolutionary rates and selection pressure

The selection pressures of the Iranian H9N2 AIVs were estimated by assessing the overall number of non-synonymous to synonymous substitutions (dN/dS) The same two sites in the HA gene were under a positive selection pressure by at least three computational models with significant statistical values (http://www.datamonkey.org), FUBAR (posterior probability 0.99), FEL, IFEL and SLAC (p-value for all 0.01-0.0005). Nevertheless, three individual codons were found under positive selection. The amino acids, which were found to be prone to changes, were 198, 234, and 267 (H9 numbering). corresponding to positions 190, 226, and 259 of the AIVs H3 numbering. The finding of the three identified amino acids, located in antigenic sites the amino acid positions 190 and 226, is extremely important since they have been previously shown to be responsible for potential antigenic and adaptive changes. substitutions E190D and Q226L at the receptor binding domain of the HA molecule (H3 numbering) are essential for the horizontal spread of H9N2 AIV in mammals (Sorrell *et al.*, 2009) Furthermore, site 234 (226 in H3 numbering) falls within epitope II of H9 and plays a role in the receptor binding specificity of HA (Kaverin *et al.*, 2009). The evolutionary rate of all Iranian viruses from 1998 to 2014 calculated by BEAST was 5.2 x 10^{-3} substitutions/site/year (95% HPD from 4.04 x 10^{-3} to 5.07 x 10^{-3} subs/site/year) and the mean time of the most recent common ancestor (TMRCA) was estimated to be around 1996 (95% HPD from 1998 to 1993). Equivalent analysis using root-to-tip regression against sampling time (program Path-O-Gen) was performed and the result was 6.58 E^{-3} substitutions/site/year (correlation coefficient: 0.98, R squared-0.96) according to BEAST results. The correlation and the possible effect of poultry vaccination program (implemented in the 2000s) on selection pressure was investigated, by analyzing the HA protein of all Iranian isolates circulating from 1998 to 2000 when vaccination had not been introduced and of the strains

from 2000 to 2014 when the vaccination was enforced. The results showed that the strains circulating from 1998 to 2000 had no any amino acids under selective pressure. Conversely, three codons were predicted to be under positive selection in the Iranian viruses circulating from 2001 to 2014. The three identified codons are the same previously described above for all Iranian viruses. Although the number of positive selected sites may be affected by sample size, the codons predicted to be under positive selection, in particular among strains circulating under vaccination programs alone, support the hypothesis that there is a positive selection due to vaccination. The comparison of the evolutionary rates of the clade B.1.1 comprising the viruses of the first introduction and the clades B.2.1 and B.2.5 reveals that it has not had a great success because the number of viruses for each clade is not enough and the Iranian clades, particularly clade B.2.5, has not an evolutionary history and consequently it is not possible to calculate the evolutionary rate separately for each introduction and compare them with each other and with different countries.

Specific amino acid evolution history

The amino acid sequence analyses of the Iranian H9N2 AIVs showed that the first introduction (clade B.1.1) consists of 50% (16/32) of strains with amino acid Q at position 226 (H3 numbering) and 50% with L226 substitution. Interestingly, other introductions comprise mainly (>97%) strains with the amino acid L226. With the phylogenetic Bayesian tree along with amino acid sequence distribution, the evolution and character history of amino acid 226 (H3 numbering) was reconstructed by Mesquite software (Figure 4). Results show that the ancestors of all Iranian H9N2 viruses had an amino acid L. The ancestors of clade B.1.1 were probably viruses with amino acid Q, whereas for the other clade there was an

amino acid L. This result suggests the presence of selection pressure from viruses with the amino acid Q to the amino acid L.

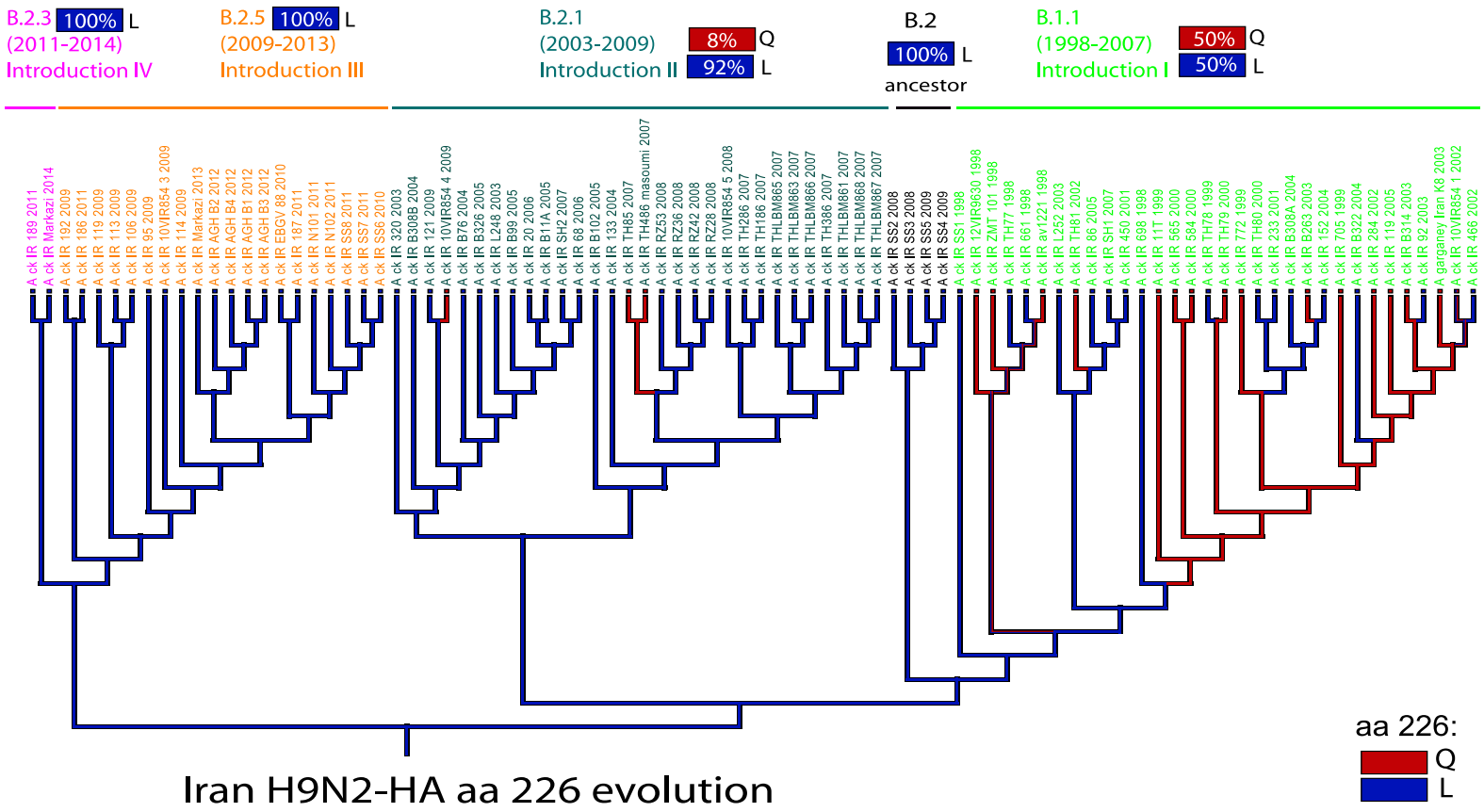


Figure 4. Amino acid 226(Q/L) evolution trend of Iranian H9N2 viruses. Mesquite tree shows the presence of AA 226 (Q/L) in each Iranian viruses and the percentage in each introduction.

Discussion and conclusion

Since 1998, avian influenza H9N2 viruses rapidly evolved and continuously circulated in the Iranian poultry population. The overall picture revealed by our analyses confirms that Iranian H9N2 AIVs evolved genetically with high evolutionary rates and in spite of the implemented national control measures, including mass vaccination of poultry, the virus rapidly spread and currently it can be considered endemic in the Iranian poultry. Phylogenetic analysis of Iranian strains, circulating from 1998 to 2014, revealed the presence of four distinct introductions, which grouped exactly within the four clades. It is worth to highlight that a new introduction seems to start when the previous one decreases. Otherwise, when the virus of a dominant clade decreases the beginning of a new introduction is observed. On the basis of the TMRCA results, the first detection of H9N2 AIVs in Iran occurred late after their common ancestor, which is estimated around 1996. However, the late detection (in 1998) probably contributed to the circulation and evolution of the virus in the country. Interestingly, the Iranian H9N2 shows evolutionary rate values similar to those of the Egyptian and the Indonesian H5N1 viruses (Cattoli *et al.*, 2010).

In this study, we also noted that the HA gene sequences of the Iranian viruses, circulating during the vaccination program, seemingly had greater numbers of positively selected sites compared to the HA sequences of viruses circulating before the implementation of vaccination programs. Unfortunately, it was impossible to evaluate the rates of nucleotide substitutions among the different introductions, before and after the vaccination program. However, the direct association between the H9N2 vaccination and evolution is difficult to establish since other factors, difficult to evaluate under field conditions, might have contributed to the apparent differences in evolutionary dynamics. For example, poultry density, various ecological factors or environmental changes, may have contributed to the

different rates of nucleotide substitution and strength of natural selection in viruses circulating in poultry (Suarez, 2010; Vandegrift *et al.*, 2010). It is reported that, when properly planned and implemented, 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 (Capua and Marangon, 2007; Ellis *et al.*, 2006; Domenech *et al.*, 2009), where it reduced the economic losses and the risk for zoonotic transmission. However, if not properly applied and not coupled with careful surveillance, robust vaccine strategy, use of recommended commercial vaccines in agreement with the dominant viruses and strict biosecurity measures, vaccination may contribute to the rapid evolution and antigenic change of AIVs by creating opportunities for the viruses to escape from vaccine protection. The molecular analysis of Iranian H9N2 viruses demonstrated that those harbouring the Q226 mutation apparently are no longer circulating, and the recent strains contain the L226 mutation. This result indicates that the Iranian H9N2 AI viruses may have specific molecular markers for the infection of mammals, and poultry vaccination may lead the viruses to human adaptation. Thus, there is a need for a continue monitoring and surveillance of H9N2 viruses to identify the evolution and the adaptation of these viruses to humans in order to prevent potential future outbreaks. In addition, the Bayesian phylogeographic analysis framework was applied to determine the source, migration patterns, and the corresponding demography history of the H9N2 AIVs circulating in Iran. The results suggest that the first introduction originated from Chinese viruses evolved in Iran by forming the clade B.1.1, and the subsequent introductions occurred in Iran had the original source from Pakistan. Contrary to that observed in the first two introductions that circulated and evolved only in Iran, the last two introductions comprising the clades B.2.3 and B.2.5 had already circulated and evolved in the neighbouring countries and subsequently introduced in Iran from Pakistan. The H9N2 virus introduction/migration

seems to have occurred basically from the eastward area, particularly between Iran and Pakistan. It is also interesting to note that the last two Iranian introductions occurred in recent years from Pakistan, the viruses were introduced in Iran and from Iran were exported to Afghanistan. In conclusion, reinforcing surveillance and control measures would be useful in order to prevent future introductions and circulation of outside H9N2 AIVs from migratory birds, poultry trade, and transportation.

Competing interests and authors' contributions

The authors declare that they have no competing interests.

HA performed most of analyses and drafted the manuscript. NH, and PGH supervised the results interpretation. GSA provided the technical support and PA contributed to the results interpretation.

Acknowledgements

We gratefully acknowledge the contributing authors and the originating and submitting laboratories for the sequences from the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database. This study was conducted in the framework of the Doctoral school in Veterinary Science at the University of Padua (Alireza Heidari).

References

1. Capua I, Alexander DJ. Avian influenza infections in birds-a moving target. Influenza Other Respir Viruses. 2007 Jan;1(1):11-8.
2. Capua I, Marangon S. Control and prevention of avian influenza in an evolving scenario. Vaccine 2007; 25:5645-52. Review.
3. Bahari P, Pourbakhsh SA, Shoushtari H, Bahmaninejad MA. Molecular characterization of H9N2 avian influenza viruses isolated from vaccinated broiler chickens in northeast Iran. Trop Anim Health Prod. 2015 Aug;47(6):1195-201.
4. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011 Oct 15;27(20):2910-2.
5. 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 Nov 21;29(50):9368-75.
6. 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.
7. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007 Nov 8;7:214.
8. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003 Oct;52(5):696-704.

9. Jin Y, Yu D, Ren H, Yin Z, Huang Z, Hu M, Li B, Zhou W, Yue J, Liang L. Phylogeography of Avian influenza A H9N2 in China. BMC Genomics. 2014 Dec 15;15:1110.
10. Kaverin NV, Rudneva IA, Ilyushina NA, Lipatov AS, Krauss S, Webster RG. Structural differences among hemagglutinins of influenza A virus subtypes are reflected in their antigenic architecture: analysis of H9 escape mutants. J Virol. 2004 Jan;78(1):240-9.
11. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005 May;22(5):1208-22.
12. Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, Subbarao K, Guan Y, Krauss S, Shortridge K, Webster R, Cox N, Hay A. Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci U S A. 2000 Aug 15;97(17):9654-8.
13. Nili H, Asasi K. Avian influenza (H9N2) outbreak in Iran. Avian Dis. 2003;47(3 Suppl):828-31.
14. Norouzian H, Bashashati M, Vasfimarandi M. Phylogenetic analysis of neuraminidase gene of H9N2 avian influenza viruses isolated from chicken in Iran during 2010-2011. Iran J Microbiol.
15. Rambaut A, Suchard MA, Xie D, Drummond AJ: Tracer v1.6. 2014. http://beast.bio.ed.ac.uk/Tracer
16. Shih AC, Hsiao TC, Ho MS, Li WH. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc Natl Acad Sci U S A. 2007 Apr 10;104(15):6283-8.

17. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. Mapping the antigenic and genetic evolution of influenza virus. Science. 2004 Jul 16;305(5682):371-6.
18. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol. 2001 Jun;18(6):1001-13.
19. Sorrell EM, Wan H, Araya Y, Song H, Perez DR. Minimal molecular constraints for respiratory droplet transmission of an avian-human H9N2 influenza A virus. Proc Natl Acad Sci U S A. 2009 May 5;106(18):7565-70.
20. Suarez DL. Avian influenza: our current understanding. Anim Health Res Rev 2010; 11 :19-33. Review.
21. Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 2002 Oct;19(10):1727-36.
22. 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.

Chapter 5

CONCLUSIONS

Emerging infectious diseases represent a significant public health threat worldwide. Influenza A virus is a zoonotic agent with a significant impact either on public health and the poultry industry. Avian influenza (AI) H9N2 viruses are currently circulating in the Middle and Far Eastern, as well as in European poultry, and the infection is endemic in several countries. H9N2 influenza viruses have human virus-like receptor specificity and potentially they could cross the species barrier and switch to human host. Therefore, surveillance and characterization of new isolates are needed by the scientific community and public health systems. One of the aims of this thesis was the serological investigation of H9N2 avian influenza virus exposure among poultry workers in one of the Middle Eastern endemic countries, as Iran. Through the application and comparison of two different serological tests human serum samples from Fars province of Iran were analysed to reveal the potential exposure of poultry workers to H9N2 AI viruses. The results of this research confirmed an association between professional exposure to poultry species and the presence of antibodies to the Iran 2008 H9N2 (human like) virus. Furthermore, these results were obtained from a collaboration among different international research groups and highlighted the potential transmission of avian H9N2 AIVs to humans and that poultry workers of Fars province (Iran) are at risk of infection. We discussed about our findings with the Iranian research group involved in this project and provided our feedback about future surveillance and research activities to minimize the risk of exposure in poultry operators and prevent the public heath from a future influenza pandemic.

Phylogenetic analyses and evolution dynamics of global H9 viruses along with the proposal of a unified nomenclature system has been another topic of this research project. The phylogenetic grouping structure, the evolutionary dynamics and the origin, herein described, provided a better understanding of the complexity of the H9 viruses evolution. In addition, phylogenetic analysis along with spatial dispersal patterns obtained by phylogeographic investigations contributed to the creation of a genetic classification of the H9 viruses. This system can be proposed to the scientific community as standard nomenclature to unify the results. In fact, recently the FAO has shown interest to the proposal of a unified standard nomenclature for the H9 subtype (as the H5 ones) and requested the collaboration of different research groups. IZSVe has collaborated to this project and provided these results that should be evaluated, approved and maybe used in the standard nomenclature of the H9 AI viruses that hopefully will be proposed.

The phylogenetic and the evolutionary history of the H9N2 viruses in Iran from 1998 to 2014 were also targeted in this research project by analyzing an endemic country as Iran where a national vaccination program is implemented. The results demonstrated the presence and the origin of different genetic group introductions. In addition, a selection pressure on circulating viruses operated by the vaccination was revealed, as well as the mutation history of a very important amino acid Q/L 226 (H3 numbering) was traced demonstrating that Iranian H9N2 viruses shifted from avian like to human like. In fact, in the first project the seropositivity of the Iranian operators was against the human like viruses which confirms the adaption of this viruses to humans.

ACKNOWLEDGEMENTS

The following work would not have been possible without the support and collaboration of numerous individuals. First and foremost, I would like to express my gratitude to my Supervisor Dr. Alessandra Piccirillo for being an important teacher who gave me invaluable direction and technical insight.

I would also like to extend a tremendous thank you to Prof. Gianfranco Gabai for his guidance and coordination during my Ph.D. program.

I am deeply grateful to all my colleagues from IZSVe. In particular, I want to thank Dr. Ilaria Capua for her valued suggestions and my co-supervisor Dr. Giovanni Cattoli who provided me consistent guidance and support throughout my study.

Special thanks to Dr. Isabella Monne for her encouragements, advice and exchange of ideas. I also appreciate the crucial work of my colleagues Samantha Kasloff, Silvia Maniero, Adelaide Milani, Viviana Valastro and many others.

I am grateful for the collaborations with Prof. Francesco Filippini, Prof. Hassan Nili, Dr. Seyed Ali Ghafouri and Prof. Gholamhosein Pourghanbari who were involved with and participated in different projects.

Very special thanks go to my family and to my friends for always being at my side and in particular to my partner, who supports me constantly.

Thank you.