ABSTRACT

Title of thesis: PHYLOGENETIC ANALYSIS OF SWINE INFLUENZA VIRUSES ISOLATED FROM HUMANS IN ALMA-ATA, KAZAKHSTAN.

Rangarajan Padmanabhan, Master of Science, 2009.

Thesis directed by: Associate Professor Dr.Daniel Perez,
Department of Veterinary Medicine,
VA-MD Regional College of Veterinary Medicine,
The University of Maryland, College Park.

Continuous surveillance of influenza becomes important considering the economic, epidemic and pandemic implications of influenza infections. This study details phylogenetic & molecular analysis of the genes of four swine influenza viruses isolated from humans in Alma-Ata, Kazakhstan. Phylogenetic analysis placed the eight segments of the four viruses in the classical H1N1 swine clade, along with the isolate A/sw/Jamesburg/1942, except for the HA of A/Alma-Ata/32/98, which was placed in the human H1N1 lineage, along with the isolate A/WS/1933. On amino acid analysis, the viruses displayed mutations on HA and ribonucleoproteins which putatively disrupt antigenic recognition of the virus by the host immune system. The presence of these

viruses relatively unchanged for 6 decades after their initial isolation could be speculated to be a combination of laboratory leaks in southern USSR in 1980s, low divergence of classical H1N1 viruses in pigs, and the low population density of Kazakhstan.

PHYLOGENETIC ANALYSIS OF SWINE INFLUENZA VIRUSES ISOLATED FROM HUMANS IN ALMA-ATA, KAZAKHSTAN

by

Rangarajan Padmanabhan

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2009

Advisory committee:

Professor Daniel Perez, Chair.

Professor Jeffrey DeStefano.

Professor Yanzin Zhang.

© Copyright by

Rangarajan Padmanabhan

2009

ACKNOWLEDGEMENTS

I first of all wish to my Amma, Appa & Thiruvengadamudayan without whom I would not be a mature independent and an amiable person that I am now. I love you Mom.

I wish to thank Dr. Daniel Perez for being a kind advisor, understanding my needs and aspirations and letting me finish my work in a short time. My experience with him has given me deep insights on independent efforts and the real world situation of working in science. I'd also like to thank all my lab members, who've been awesome colleagues giving me academic and emotional support in my stay here. I owe special thanks to Danielle & Lindomar for being my good friends.

Thanks to my committee members, Dr. DeStefano and Dr. Zhang. Dr.DeStefano has frankly been the coolest professor I've ever met. He has given me great support from the time I was inducted into the CBMG program till this moment. I also owe my gratitude to Dr. Zhang for having patiently going over my thesis and providing me recommendations and insights.

Friends are obviously an integral part of one's successes. I'd like to thank a ton of my friends, particularly Giri & Tony. They've been and are my awesome friends I'll cherish for the rest of my life. I also thank my roomies for being an amazing bunch of people. They've heard me sad and they've heard me happy. At the Vet Med, I thank Madhuri, Senthil and Flavia for being great friends who helped me throughout.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

List of tables v

List of figures vi

CHAPTER 1: INTRODUCTION 1

1.1 Overview 1

1.2 Research objectives 2

CHAPTER 2: LITERATURE REVIEW 4

2.1 Influenza A viruses- Biology and Replication 4

2.2 Transmission of Influenza 8

2.2.1 Ecology 8

2.2.2 Molecular determinants of host range 10

2.2.3 Evolution 12

2.3 Swine influenza viruses 14

2.3.1 History & Epidemiology 14

2.3.2 Outbreak, persistence and evolution 16

2.3.3 Zoonotic potential of swine influenza 17

CHAPTER 3: MATERIALS AND METHODS 20

3.1 Virus isolation and propagation 20

3.2 RNA extraction and cDNA synthesis 20

3.2 PCR amplification and gel extraction 21

3.3 DNA sequencing and phylogenetic analysis 22

CHAPTER 4: RESULTS & DISCUSSION 24

4.1PAIRWISE ALIGNMENT 24

4.2 BLAST Analysis 25

4.3 Hemagglutinin (HA) 26

4.3.1 Phylogenetic analysis 26

4.3.2 Amino acid comparison with the most identical sequence 26

4.3.3 Receptor-binding sites (RBS) 27
4.3.4 Glycosylation sites 31
4.3.5 Antigenic sites 33
4.3.6 Cleavage site 34
4.4 Neuraminidase (NA) 36
4.4.1 Phylogenetic analysis 36
4.4.2 Catalytic, Glycosylation and Antigenic sites 36
4.4.3 Virulence determinants 37
4.4.4 Phylogenetically important regions 38
4.5 Polymerases- PA, PB1 and PB2 40
4.5.1 Phylogenetic analysis 40
4.5.2 Genetic signatures associated with pathogenicity and host range 40
4.5.3 Antigenic residues 46
4.6 Nonstructural protein (NS) 47
4.6.1 Phylogenetic analysis 47
4.6.2 Functional residues 47
4.6.3 Virulence determinants 48
4.7 Nucleoprotein (NP) 51
4.7.1 Phylogenetic analysis 51
4.7.2 Functional residues and CTL epitopes 51
4.8 Matrix (M) 54
4.8.1 M1 gene 54
4.8.2 M2 gene 55
4.9 Evolutionary anachronism of the isolates 57
CHAPTER 5: SUMMARY AND CONCLUSIONS 65
REFERENCES 67

List of tables

Table 1: Influenza A viruses circulating in pigs worldwide…………………………....15

Table 2: Pairwise comparison of the genes of the sequenced viruses……………….......24

Table 3: Influenza viruses most related to the sequenced viruses………………………25

Table 4: Pairwise comparison of HA…………………………………………………....28

Table 5: Receptor binding sites of HA of the sequenced viruses………………………..30

Table 6: Glycosylation sites of HA of the sequenced viruses…………………………...32

Table 7: Phylogenetically important regions of the polymerases………………………..42

Table 8: Mutations in T-cell epitopes of polymerases…………………………………...46

Table 9: Phylogenetically important regions in NS1 and NS2…………………………..49

Table 10: Mutations in T-cell epitopes of NP……………………………………………...52

List of figures

Figure 1: Influenza A replication cycle 8

Figure 2: Ecology of influenza A viruses.. 10

Figure 3: Phylogenetic tree of HA1 gene 29

Figure 4: Surface representation of HA1 domain. 35

Figure 5: Phylogenetic tree of NA. 39

Figure 6: Phylogenetic tree of PA gene 43

Figure 7: Phylogenetic tree of PB1 44

Figure 8: Phylogenetic tree of PB2. 45

Figure 9: Phylogenetic tree of NS1 gene 50

Figure 10: Phylogenetic tree of NP gene 53

Figure 11: Phylogenetic tree of M1 gene 56

CHAPTER 1: INTRODUCTION

1.1 Overview

Of the three types of influenza viruses (influenza A, B, & C), only influenza A viruses are established in animals other than humans. Influenza A viruses infects a large variety of animal species, including swine, equine, human, avian and sea mammals (Webster et al., 1992). Influenza A viruses are classified according to the two surface molecules, hemagglutinin (HA) and neuraminidase (NA). Wild aquatic birds are the reservoir of all the subtypes of influenza A viruses (H1-H16 and N1-N9), where they are usually asymptomatic. Avian influenza A viruses from aquatic birds are frequently transmitted to domestic birds & swine, and occasionally to aquatic (seals, whales) or terrestrial mammals (horses and humans). Since pigs are susceptible to infection with both avian and human influenza A viruses, often being in close contact with humans or birds, they play an important role in the zoonosis of influenza A viruses to humans. By being susceptible to viruses infecting two different species, they create a platform for creation of new 'reassortant' viruses for which humans might have little prior immunity. For this reason, they are postulated to be the 'mixing vessels' for influenza A viruses (Webster et al., 1992). Substantiating this fact, swine viruses are classified into 'avian-like', 'human-like' and 'classical swine', indicating the different virus lineages that can infect swine. Thus it is critical for continuous surveillance of swine influenza viruses throughout the world from an economic, epidemic and pandemic perspective.

1.2 Research objectives

This study intended to genetically characterize four swine influenza A isolates from humans in Kazakhstan over a separated time period. The main research objectives are:

I. Deduce the evolutionary relationship of the four isolates to other influenza viruses through phylogenetic analysis of the nucleotide sequences. The viruses are: A/Alma-Ata/1044/83, A/Alma-Ata/1417/84, A/Alma-Ata/5/98, and A/Alma-Ata/32/98.

II. To do complete sequence analysis of the individual genes of the isolates to find out residues that might potentially play a role in pathogenicity and antigenicity.

The results of the study are:

I. The four isolates are influenza A viruses, and all the eight genes of the four isolates have 98-99% identity to the 'classical swine' isolate A/swine/Jamesburg/1942, except for the HA of A/Alma-Ata/32/98 which has 99% identity to the WS/1933 virus.

II. The virus isolates probably recognize α2,6 linked sialic acids (present on humans and swine). The isolates do not seem to carry any obvious amino acid 'signatures' observed in highly pathogenic viruses.

III. The viruses share a lot of features with the 'classical' swine viruses, but show considerable sequence variation at residues that are involved in antigenic

recognition of the host immune system. These mutations can be postulated to influence immune recognition of the viruses by the host.

IV. Considering the rate of mutation of influenza viruses, it's next to impossible that the viruses were preserved in its entirety in a separate geographic reservoir for ~50 years after their initial circulation in an another continent. The presence of these viruses in Kazakhstan could be explained by a combination of laboratory leaks of classical human/swine isolates in the southern USSR in 1980s, low divergence of classical H1N1 in pigs, and the low population density in Kazakhstan. A more solid swine and human influenza surveillance in central Asian countries is necessary to resolve the anachronism.

CHAPTER 2: LITERATURE REVIEW

2.1 Influenza A viruses- Biology and Replication

Influenza A viruses are members of the Orthomyxoviridae family (Genus Influenzavirus A). The influenza A viruses are primarily distinguished from influenza B & C viruses based on the genetic and antigenic differences in their nucleoprotein (NP) and matrix (M) proteins. Whereas influenza A infects a wide variety of animals, influenza B & C primarily infects humans (influenza C has also been isolated from pigs & dogs) (Palese et al., 2007).

Influenza A viruses are enveloped single stranded, negative sense RNA viruses with a segmented genome. There are eight RNA segments, coding for 11 proteins. The largest are the polymerase proteins, comprising the polymerase basic 2 (PB2; 2341 bp), polymerase basic 1 (PB1; 2341 bp), and polymerase acid (PA; 2233 bp). These are followed by the hemagglutinin (HA; 1778 bp), nucleoprotein (NP; 1565 bp), and neuraminidase (NA; 1413), matrix (M; 1027 bp), and non-structural protein (NS; 890 bp). PB1, M, and NS code for splice variants PB1-F2 (261 bp), M2 (366 bp), and NS2/NEP (418 bp) respectively (Palese et al., 2007).

The HA and NA spikes are embedded in the lipid envelope, along with the M2 (matrix 2) protein (Palese et al., 2007). The HA is the receptor-binding protein mediating the fusion between virus envelope with the host cell membrane. It specifically interacts with sialic acids on cell surfaces (α2,3 linked in case of avian and α2,6 in case of human

& swine), and is the major target of host humoral immune response (Wilson et al., 1990). The NA helps in the budding of virus particles after infection by removing sialic residues from HA as well as the infected cell. Like HA, NA is also under significant immunogenic pressure (Colman et al., 1983). The last component of the envelope is composed of the matrix proteins. As mentioned earlier, there are two spliced variants of the matrix gene, M1 and M2. M2 protein, serves as an ion channel, and helps in the uncoating of the virus. Additionally, it also modulates the pH of Golgi apparatus, preventing premature conformational change of the HA protein prior to virus assembly. The M1 protein lies beneath the lipid envelope, providing rigidity to the membrane (Wright et al., 2006b). This envelope encapsulates 8 RNA segments, each of which is coated by nucleoprotein (NP) and bound to the heterotrimeric polymerase complex made of PA, PB1 and PB2 at their extremities. This polymerase complex is responsible for transcription and replication of the viral RNA (Naffakh et al., 2008). Recently a splice variant of the PB1 protein was discovered (PB1-F2), and was found to have pro-apoptotic activity (Chen et al., 2001). The virus also encodes the NS gene, which also has two splice variants NS1 and NEP. The NS1 protein is multi-functional, with one of its important activity being the antagonist of IFN responses (Hale et al., 2008). The other splice variant & last influenza protein is the nuclear export protein (NEP), and is found to be required for vRNP export from nucleus into cytoplasm (Neumann et al., 2000).

The infection begins with HA binding to sialic acids on the cell surfaces and subsequent internalization by endocytosis. The acidity of the late endosome stimulates a conformation change that first involves the cleavage of the HA precursor (HA0) into

HA1 and HA2 and subsequently fusion of the virus envelope and membrane of the endosome. The host range of influenza viruses is frequently associated with these two events, with regard to the HA. Other than the sialic acid specificity of avian and human influenza viruses discussed earlier, the cleavability of HA is associated with both virulence and host range of influenza viruses. Unlike the low pathogenic viruses, the HPAI H5 and H7 subtype viruses have a stretch of basic residues at the cleavage site. So while the HA of LPAI viruses can be cleaved only by extracellular trypsin-like proteases secreted at the lung and intestines of the bird, the HA of HPAI viruses can be cleaved by furin-like proteases expressed ubiquitously in the trans-Golgi of all cells. This accounts for the systemic spread of H5 and H7 HPAI viruses (Steinhauer, 1999).

Acidification of the late endosome also activates ion channel activity of M2 protein, which results in the disassembly of M1 from the vRNPs by letting in protons to the viral interior (Colman and Lawrence, 2003). Consequently, the vRNPs are released into the cytoplasm and transported to the nucleus, where the transcription and replication of the viral RNA takes place. Transcription initiation begins with the cap-snatching mechanism, where capped cellular premessenger RNAs are bound by PB2 and cleaved by the endonuclease activity of PB1. This is followed by transcription of the vRNA by PB1, the RNA-dependent RNA polymerase. Termination of mRNA synthesis and polyadenylation occur at an oligoU sequence located near the 5' end of the vRNA. Because of the identity of viral transcripts to cellular mRNA's (polyA and 5'cap), they are translated by the cellular translation machinery. The translated polymerase proteins get back into the nucleus to prime more transcription. The polymerase complex (PA, PB1

and PB2) also executes vRNA replication. Replication initiation is primer independent and results in the synthesis of complementary copies of the viral RNA, called cRNA. The exact function of PA is not yet clear, although it's found that it is required for vRNA replication. The newly synthesized vRNAs serve as templates for more transcription and translation processes. Alternatively, they may exit the nucleus to be incorporated into progeny virions. The mechanism which defines this balance is not yet elucidated, although NP and NEP proteins have been proposed to play a role (Palese et al., 2007).

vRNPs exit from the nucleus occurs through association with M1, which in turn interacts with NEP. NEP in turn interacts with the cellular hCRM1 nuclear export machinery which facilitates the whole export process (Cros et al., 2003). The interaction of M1 with the vRNP plays a crucial role in preventing reentry of the vRNPs back to the nucleus and helps in targeting them to the assembly sites at the apical side of polarized cells, which already have the envelope proteins embedded. The exact mechanism which ensures that all 8 segments are packaged in a single virion is not completely understood, although it is known that they are incorporated according to specific mechanisms (Noda et al., 2006). Finally, as cited earlier, the NA protein ensures proper release of new viruses by cleaving sialic acid molecules present on both cellular membranes as well as the HA. The whole replication cycle is summarized in Figure 1.

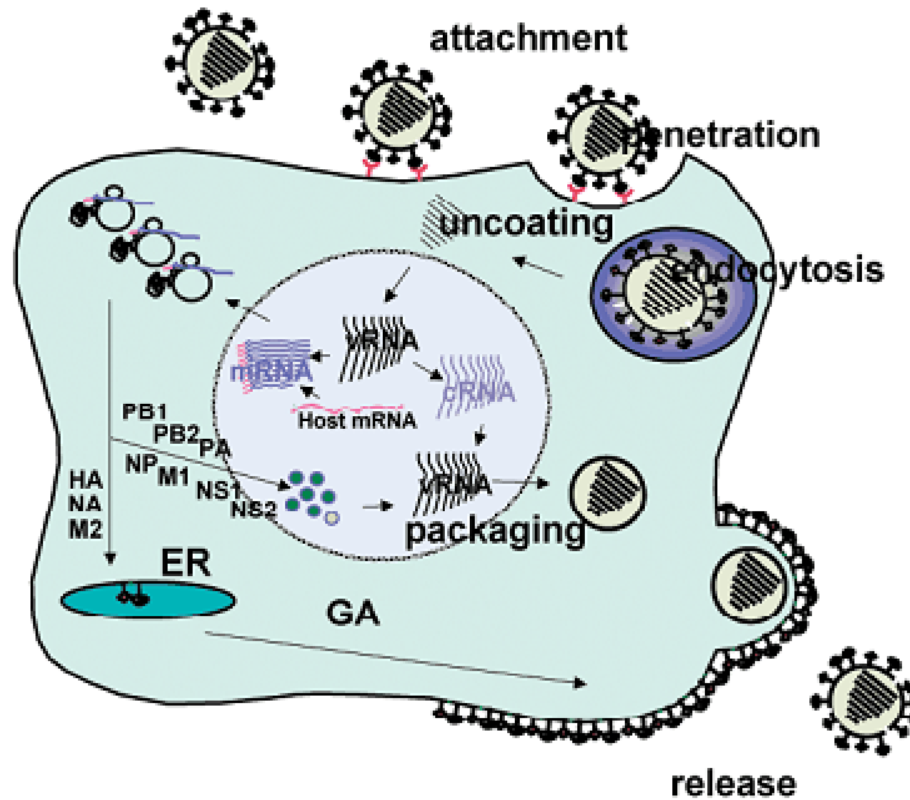

Figure 1: Influenza A replication cycle (Sorrel EM, 2003)

2.2 Transmission of Influenza

2.2.1 Ecology

Influenza A viruses have evolved to survive and propagate in a wide variety of species. In humans, only H1, H2, H3, N1 and N2 subtypes are known to cause infection (Webster et al., 1992). In horses, infections have been largely restricted to H7N7 and H3N8 subtypes (Webster et al., 1993), and pigs have been known to be infected with only H1, H3, N1 and N2 subtypes (Webster et al., 1992). But it's agreed that the avain species serve as the universal reservoir of influenza viruses, from viruses of subtypes H1 to H16 and N1 to N9. Infections in most of the avian species are asymptomatic (highly host

adapted), and particularly to wild ducks. This is associated with a evolutionary ‘stasis’ in those species, wherein the influenza viruses still undergo the same rate of mutations, but which does not lead to amino acid changes (called synonymous substitution; the converse is called non-synonymous substitution). Two notable reservoirs are wild ducks and wild waterfouls, which are susceptible to influenza viruses of most of the HA and NA subtypes. Influenza has also been sporadically isolated from shorebirds like gulls, terns, shearwaters, guillemots and sandpipers. Domestic poultry are also susceptible to a variety of subtypes, of which H5 and H7 are highly virulent, causing 100% mortality experimentally (Webster et al., 1992). Thus it's agreed that avian reservoir serves as the pervasive, vast reservoir of influenza viruses, serving to generate new variants and reassortants. This is supported by the findings that aquatic birds serve as the ancestor of all mammalian and equine lineage viruses (Subbarao et al., 2006), including the three recent human pandemics of 1918, 1957 and 1968. In addition, avian influenza viruses have also been linked to outbreaks in seals, mink and whales, and domestic species like the dog and cat (Webster et al., 1992; Beeler 2009). The complex ecological sustenance of influenza viruses is summarized in Fig 2.

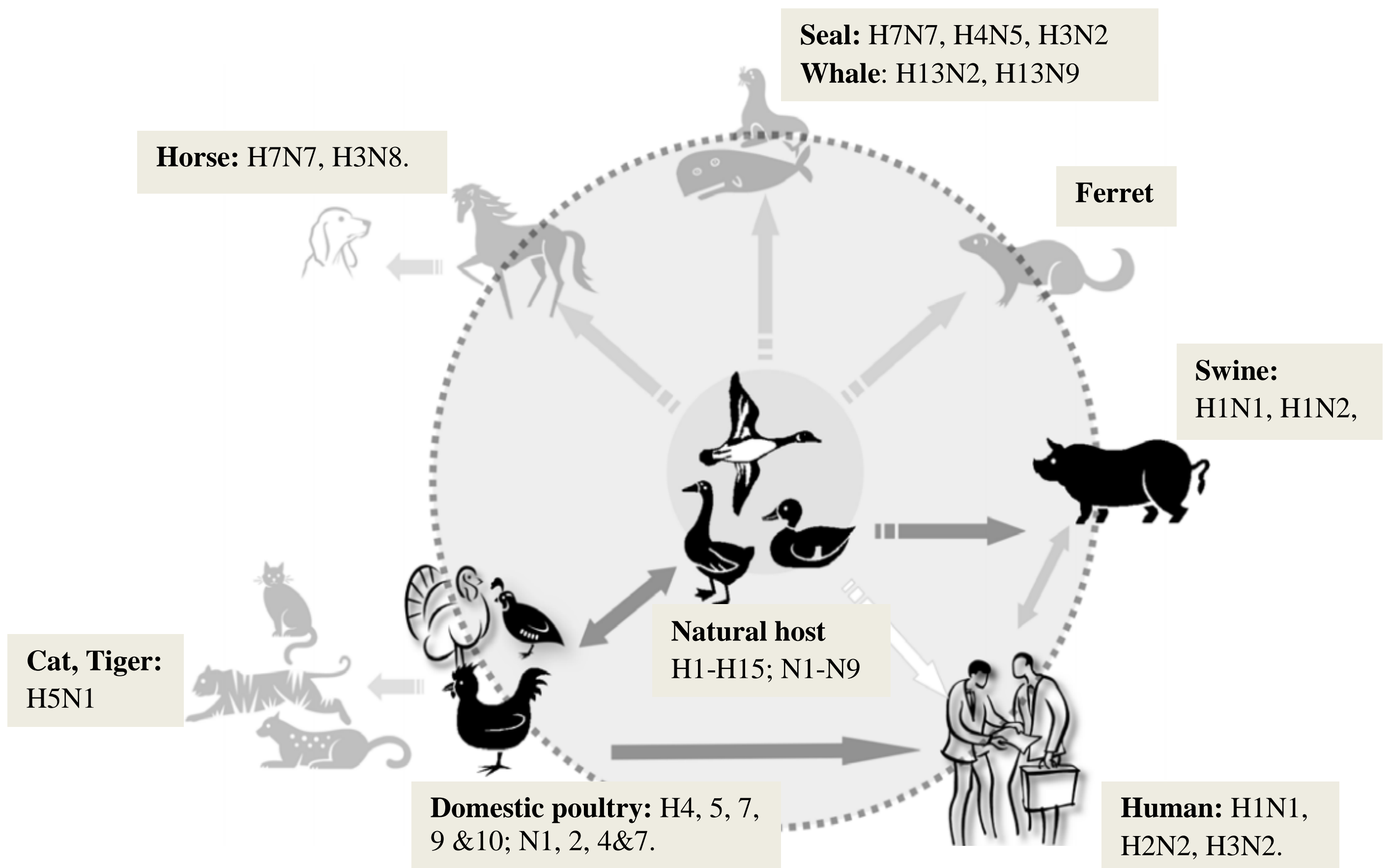

Figure 2: Ecology of influenza A viruses. Aquatic species serve as the natural reservoir which is continuously transmitted to other species through a complex network of ecological pathways. Double headed arrows indicate mutual transmission, and one headed arrow indicates a one way transfer to the pointed species (Adapted and modified from Pascua 2006).

2.2.2 Molecular determinants of host range

Since 1997, a series of harsh avian epidemics which had limited human causalities (H5N1, H9N2 and H7N7) caused widespread concern about the possibility of those avian viruses becoming a human pandemic. Despite the concerns, it was found that the human-to-human transmission of these viruses is low, a requisite for a pandemic potential (Landolt & Olsen, 2007). It highlights the presence of host range barriers on transmission of influenza viruses from one species to another. Although it's established

that species specificity is most likely a multigenic trait (Subbarao et al., 2006), the HA is considered to be the most important factor for a species 'jump' to occur. As stated earlier, mutations in HA which shift its receptor specificity from α2,3 to α2,6 linked sialic acids is usually associated with capacity to infect human and swine populations. These mutations are L226 & S228 in human H3 (Q226 & G228 in avian), and D190 in human H1 (E190 in avian) (Matrosovich et al., 2000). These mutations require significant adaptation in an intermediate host. Since pigs possess the receptors for both avian and human influenza viruses (Ito et al., 1998, Webster et al., 1992), they have been considered to be this "mixing vessel" for emergence of new reassortants. This is substantiated by the several reassortant viruses circulating in swine population worldwide (discussed below). With regard to NA, the stalk length and low pH stability is postulated to be adaptations for species specificity. Short stalk length is associated with low pathogenicity, although recent HPAI are found to possess large deletions in the stalk region. With regard to low pH stability, unlike conventional human viruses, HPAI H5N1 could replicate in human intestine, causing gastrointestinal symptoms and shed in large quantities in stool. This is attributed to the low pH stability of NA, maintaining its enzymatic activity at the acidic conditions prevailing in the stomach (de Jong et al., 2005).

With regard to internal proteins, the characteristic mutation E627K (E in human and K in avian) in PB2 is supposed to determine species specificity in mice and humans. NP is also proposed to be the determinant of host range which can attenuate or restrict virus replication (Scholtissek et al., 1985). Further, other internal proteins like PB1, NP,

and M genes have also been implicated in species specificity, although no clear picture has emerged. Additionally, as discussed earlier, the presence of multiple basic residues at the HA1-HA2 cleavage site also determine pathogenicity and probably species barrier.

2.2.3 Evolution

Evolution in influenza occurs in two major forms: antigenic drift and antigenic shift. Antigenic drift occurs when the virus undergoes mutations that results in different antigenic structure, and is thought to be driven by immunogenic pressure of the host. Antigenic shift is a major change in the viral architecture wherein new genes are acquired through reassortment. Antigenic shift occurs when the same cell of an organism is infected by two viruses which infect two different species, and wherein the genetic segments are exchanged between these co-infecting viruses. The resulting virus is totally unrelated to any virus the host has ever encountered. Notably, such reassortment events create major epidemics and pandemics, including the 1957 H2N2 and 1968 H3N2 pandemic. The third but rare form of evolution involves recombination between homologous segments within the virus; although there is ample evidence to show non-homologous recombination occurs rarely. Nevertheless, a clear picture is yet to emerge about recombination as a source of genetic variation in influenza viruses (Nelson et al., 2007).

Since the RNA polymerase of the influenza lacks proofreading activity, there is high rate of mutations during replication of the viral genome. Thus during infection of a single host cell with one or two viral strains, there theoretically exists several

quasispecies, rather than a homogenous viral population. Although not all mutants or reassortants are viable, this allows for considerable plasticity for adaptation to a new host species. Most of these changes occur in genes which are under immunogenic pressure, like the HA and the NA, rather than in the internal genes. This is termed as positive selection, theoretically selecting mutants that are most 'fit' to the infected host, and is frequently associated with changes in amino acid sequences (Webster et al., 1992). In this regard, a recent development is the concept of 'neutral' evolution, wherein it has been proposed that the mutations occurring in the surface genes need not necessarily have antigenic significance. The virus undergoes a lot of these silent 'neutral' evolutions for a considerable amount of time, exploring different options in the host immunogenic space, before 'hitting' a point wherein its optimally adapted to cause a epidemic or a possible pandemic (Koelle et al., 2006). Internal genes (like NP) on the other hand display a high degree of host-specific adaptation and a change in them might lead to attenuation of the viral infection (Webster et al., 1992). Rate of mutations of influenza A viruses is proposed to be at a range of 0.001-0.007 nucleotide substitutions per site per year. This number varies between species, and is usually indicates the length of the time the virus has been associated with a particular host. Thus, 'old' viruses evolve more slowly compared to recent entrants, a classic example of which has been 'evolutionary stases' of influenza viruses in certain avian species, as described earlier (Krasnitz et al., 2008).

2.3 Swine influenza viruses

2.3.1 History & Epidemiology

Swine influenza was first reported in 1918, as it coincided with the severe pandemic that was occurring in humans, and was first isolated in 1930 (sw/Iowa/1930) (Shope, 1931). The disease in pigs has a lot of similarities with disease in humans, and was suggested to have had common etiology. Recent sequencing of the human H1 HA revealed that the virus most probably spread from humans to pigs, and is supported by observations of veterinarians (during the 1918 pandemic) who described the appearance of the disease in pigs just after its appearance in humans (Brown, 2000).

Like humans, pigs are predominantly susceptible to three subtypes: H1N1, H3N2 and H1N2. H1N1 lineage viruses are frequently called 'classical swine' lineage, indicating their continuing presence in swine populations several decades after their initial appearance (sw/Iowa/1930). Apart from that, recent entrants including 'avian-like' H1N1, 'human-like' & 'avian-like' H3N2 viruses also continue to circulate amongst the swine herds (Wright et al., 2006a, Olsen et al., 2006). The viruses have remained endemic in pig populations worldwide, and occasionally result in local epidemics, especially when there is an antigenic drift. Table 1 summarizes the epidemiology of swine influenza viruses

Table 1: Influenza A viruses circulating in pigs worldwide. Adapted from Brown (2000).

Subtype	Location	Comments
H1N1	North America Europe Asia South America Europe Asia	Derivatives of 'Classical' virus, first isolated in 1930 in North America. 'Avian-like' virus first isolated in 1979 'Avian-like' virus first isolated in 1993
H3N2	Asia Europe North America South America Asia	'Human-like' virus first isolated in 1970 in Asia H3N2 human/ classical reassortant in 1984. 'Triple' reassortant of human, avian and classical in 1998. 'Avian-like' virus first isolated in 1978
H1N2	Asia Europe North America	Classical/'human-like' reassortant in Japan Human/'human-like' reassortant in UK in 1994 Classical/triple-reassortant reassortant (H3N2) in 1998

Classical H1N1 were the dominant cause of swine influenza in North America from the time of their first isolation in 1930 through the 1990s. Additionally, they are also isolated from South America, Europe, and Asia (Subbarao et al., 2006). The classical viruses in the US remained largely stable for several dozen of years (from 1960s) antigenically and genetically till the 1990s, after which variants and reassortants were isolated (Olsen et al., 2006).

Human-like swine viruses were first isolated after the 1968 H3N2 pandemic and have since become established in swine populations throughout the world (Brown, 2000). Human-like swine influenza viruses have been recovered from pigs in Asia, Europe, and North America. The introduction of human-like virus (H3N2) into swine in North

America in 1998 was a crucial event in emergence of reassortant viruses that presently co-dominate the swine population in the US along with the classical H1N1 swine viruses (Olsen et al., 2006).

There have been at least three wholly avian virus introductions into swine: in Europe in 1970s, in China in 1990s, and in Canada in 2004 (Pensaert et al., 1981; Guan et al., 1996; and Karasin et al., 2004). The most interesting was the introduction in Europe, which has completely replaced the other circulating lineages, leading to the proposal that 'avian-like' viruses have a selective advantage over classical swine virus (Brown, 2000). But contrasting to the European scenario, introduction of avian-like H1N1 in swine in China has not replaced the classical swine viruses, and both continue to circulate (Wright et al., 2006b). Unlike the classical swine viruses (before the appearance of H3N2), these new viruses have shown considerable genetic and antigenic diversity (Olsen et al., 2006), showing that the viruses are trying to optimally adapt to the new host.

2.3.2 Outbreak, persistence and evolution

Swine husbandry practices directly influence the evolution of influenza viruses in pigs (Brown, 2000; Easterday, 1980a). Outbreaks occur by introduction of new pigs into a herd, and the virus persists through infection of young susceptible pigs or introduction of a new stock. In large commercial farms, complete elimination thus becomes possible only by complete depopulation. The outbreak in a farm is usually associated with outbreaks in several farms within an area, due to widespread distribution of the virus

among herds in an area. Transmission is primarily through the nasopharyngeal route, and is exacerbated by commercial husbandry practices (Brown, 2000).

Though the potential of swine acting as reservoirs of interepizootic and long-term true carrier of influenza has been studied, there is no clear data to reject or accept the hypothesis. Brown (2000) considers the widespread occurrence of influenza in pigs and the husbandry practices make it likely that the virus is continually maintained by passage to young susceptible pigs.

Evolutionary rate (precisely non synonymous substitutions) of swine influenza viruses is generally lower than human influenza viruses, and is probably due to the weak immune response mounted by pigs, fuelled by their short life span and continual availability of non-immune pigs. Specifically, the evolutionary rate is lower for the classical H1N1 swine when compared with ‘avian like’ or ‘human like’ swine viruses, probably reflecting that more recent ‘avian like’ viruses have not yet completely adapted to the host. Though pigs worldwide harbor viruses (or antibodies) which are similar to ‘old’ viruses, as mentioned earlier, the possibility of pigs existing as a long term carrier of influenza is yet to be clearly established (Brown, 2000).

2.3.3 Zoonotic potential of swine influenza

Swine serve as major reservoirs of H1N1 and H3N2 influenza viruses and by being susceptible to frequent introduction of new viruses from other species, they play a major role in transmission of influenza viruses to humans. There are several cases of classical swine viruses isolated from humans in North America, with a few being fatal.

Wholly avian H1N1, reassortant H3N2 and reassortant H1N1 have also been isolated from other places (Subbarao et al., 2006; Myers et al., 2007). Most of the cases have been people who had direct contact with pigs, and there is no observable clinical distinction of swine influenza from conventional human influenza. Further evidence is the presence of viruses (or antibodies) to H1N1 and H3N2 viruses in swine similar to those circulating in the human population. The intriguing fact is that some H3N2 viruses were detected in swine populations several years after their human counterparts have disappeared, and have led to the speculation that swine may serve as the inter epidemic and long time reservoir of influenza viruses (Brown, 2000).

Similarly, there is evidence to show that there is direct transmission of influenza viruses from pigs to poultry and vice versa. As mentioned earlier, the most characteristic has been the transmission of avian viruses into pigs in 1970s, which have established a stable lineage there. It has also been found that pigs are susceptible to infection with avian viruses of all subtypes, from H1 to H13 (Kida et al., 1994). Also, considering that neither avian nor human influenza virus can directly infect each other, and that the 1957 and 1968 pandemics were both reassortants between human and avian viruses, pigs are thought to have played the intermediary role (although there is no direct evidence) (Subbarao et al, 2006). Isolation of swine H3N2 viruses from infected children in the Netherlands with genes encoding internal protein of avian origin (Class et al., 1994), isolation of 'triple reassortant' H3N2 viruses in United States (HA, NA and PB1 of human origin; NP, M and NS genes of classical swine origin; PB2 & PA of avian origin) (Zhou et al., 1999), isolation of a H1N2 virus, derived from a reassortant of classical

swine and the ‘triple reassortant’ (Karasin et al., 2000), and several other H1N2 and H1N1 reassortants (Subbarao et al., 2006) substantiate the ‘mixing vessel’ theory. It is thus important for continuous surveillance of swine influenza viruses for predicting and possibility preventing future pandemics.

CHAPTER 3: MATERIALS AND METHODS

3.1 Virus isolation and propagation

A/Alma-Ata/1044/83 was isolated from the lung of a 65 year old male patient who had succumbed to the infection, and had communal and occupational contacts with swine. A/Alma-Ata/1417/84 was isolated from the nasal wash of a 26 year old male patient, and did not have contacts with swine. A/Alma-Ata/5/98 was isolated from 27-28 week old abortive material of a 23 year old woman, and had communal contacts with swine. Finally, the A/Alma-Ata/32/98 was isolated from 11-12 week old abortive placental material of a 25 year old woman. It is interesting to note that except for the H5N1 avian influenza (Gu et al., 2007), there has not been any other case of influenza being transmitted to the fetus. The samples were inoculated into 9 to 11-day-old pathogen free chick embryos. Eggs were candled daily to monitor the embryo mortality. Allantoic fluid from dead embryos was aseptically harvested, lyophilized and propagated every five years before it was brought to the University of Maryland.

3.2 RNA extraction and cDNA synthesis

RNA was extracted from allantoic fluid of each sample using RNeasy (Qiagen, Valencia, CA) spin columns. Briefly, beta-mercaptoethanol RLT buffer solution (ratio 10:1000) was added to 200ul allantoic fluid in a 1.5ml tube. After brief mixing, 70% ethanol was added and again mixed gently. This was transferred to an RNAeasy mini column for adsorption of the RNA on the membrane, and centrifuged for 1min at

maximum speed. The contaminants were subsequently washed off two times by the RPE buffer and the RNA was eluted in 40ul of RNAse free water. RNA was stored at -70°C freezer and used for subsequent complementary DNA (cDNA) preparation.

The cDNA was prepared by first incubating 4ul RNA, 0.5ug Uni12 universal primer (AGCAAAAGCAGG), and 5.5 ul RNAse free Millipore water at 70°C for five minutes. The mixture was immediately incubated in ice. Subsequently, the mixture of 4ul 2.5mM dNTPs, 4ul of 5X Reverse transcriptase buffer (Promega, WI), 1 ul RNaseout (Invitrogen, CA) and 1 ul Reverse Transcriptase AMV (Promega, WI) was added to the denatured RNA and incubated in at 42°C for 1 hour for the cDNA synthesis to occur. Reaction was subsequently stopped by heat inactivation at 70°C for 10 minutes.

3.2 PCR amplification and gel extraction

One micro liter of cDNA was used with 0.75 ul of 100ng/ul forward and reverse primers (Invitrogen, CA) (universal and gene specific primers), 12.5ul Go-Taq green PCR master mix (Promega, WI) and 10ul nuclease free water. Polymerase chain reaction was performed by PTC-200 Peltier thermal cycler (Bio-Rad, CA) with the following amplification parameters: initial denaturation at 94°C for 4 minutes, followed by 30 cycles of amplification at 94°C for 20 seconds, 56°C for 30 seconds, and 72°C for 5 minutes. After 30 cycles, the final extension was performed at 72°C for 10 minutes before the final incubation at 4°C.

The PCR products were then run on 1.5% agarose gels stained with ethidium bromide, and subsequently the DNA gel purified using the QIAquick Gel Extraction Kit

(Qiagen, CA). Briefly, the DNA fragment was excised from the gel and three volumes of bugger QG was added to one volume of the gel. Following dissolution of the gel at 50°C, 1 gel volume isopropanol was added and transferred to the QIAquick mini spin columns, for the PCR product to adsorb on the column membrane. After two buffer washes to wash out contaminants (buffer QG and PE), the DNA is eluted in 10ul of nuclease free water. It was stored at -20°C for subsequent sequencing reactions.

3.3 DNA sequencing and phylogenetic analysis

DNA sequencing was carried out using the Big Dye Terminator V.3.0 Cycle Sequencing Ready Reaction kit (ABI, Foster City, CA), by amplification of individual strands of the DNA using 60 ng of a specific inner primer, 2ul sequence RR-100, and the sequence buffer for a total volume of 20ul. Subsequently, any unincorporated labeled ddNTPs and residual contaminants were removed by ethanol precipitation and the DNA was vacuum dried. The dry DNA pellet was resuspended in HiDi formamide (Applied Biosystems, CA). The sequences were then resolved on an Applied Biosystems 373 AI 3100 Genetic Analyzer (Perkin Elmer, Norwalk, CT), and sequences were edited using the SeqMan program (DNASTAR, Madison, WI). Genomic information was derived from overlapping sequences covered by forward and reverse primers. At least two independent RT-PCR reactions were produced for each gene and used for sequencing. Complete edited sequences were exported to the EditSeq program in the same suite.

Most identical nucleotide sequences available in the influenza sequence database were identified through the publicly available NCBI BLAST program

(http://blast.ncbi.nlm.nih.gov/Blast.cgi). A group of highly related and least related sequences were obtained, and few other historically important & geographically important sequences were also obtained. Multiple sequence alignment and phylogenetic analysis were performed using *MEGA* version 4 (Tamura et al., 2007) with the default parameters for sequence alignment. Neighbor-Joining method was used for the tree construction, and confidence values for the tree topologies were evaluated by bootstrap analysis of 1000 pseudo-replicate datasets. The trees were edited in the Tree explorer available along with the same suite. Three dimensional protein models were generated by means of Swiss-model (Arnold et al., 2006) with the identical human H1 (1918 HA pdb code: 1ruz) and swine H1 (sw/Iowa/1930 pdb code: 1ruy), and the models were visualized using UCSF Chimera (Petterson et al., 2004).

CHAPTER 4: RESULTS & DISCUSSION

4.1PAIRWISE ALIGNMENT

To ascertain the genetic relationship between the sequenced isolates, the nucleotide sequences of the eight segments of the Alma-Ata/1044/83 virus was compared to the corresponding eight segments of the other three viruses, and are shown in Table 2. It can be observed that except for the HA of Alma-Ata/32/98, which is only 87% identical to the Alma-Ata/1044/83 virus, all the other segments of all the other viruses were 99-100% identical to each other.

Table 2: Pairwise comparison of the nucleotide sequence of the gene segments Alma-Ata/1044/83 virus with the nucleotide sequences other three isolates.

	The 1983 virus (vs.)		
Gene	**Alma-Ata/1044/83 (% identity)**	**Alma-Ata/5/1998 (% identity)**	**Alma-Ata/32/1998 (% identity)**
PB2 (segment 1)	99.95	99.96	99.91
PB1 (segment 2)	100	99.97	99.73
PA (segment 3)	99.9	99.95	99.68
HA (segment 4)	99.88	99.94	87.62
NP (segment 5)	100	100	99.86
NA (segment 6)	99.79	99.86	99.79
M (segment 7)	100	100	100
NS (segment 8)	100	99.84	99.76

4.2 BLAST Analysis

Most related evolutionary ancestor to the sequenced isolates were identified through the NCBI BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and top three 'hits' for each sequences are provided in Table 3. It can be observed that except for the HA of Alma-Ata/32/98 (98% identical to the HA of WS/1933 strain), the individual genes of the other three viruses are 98-99% identical to the sw/Jamesburg/1942 virus.

Table 3: Influenza viruses most related to the sequenced viruses

Gene segment	Most homologous isolate	Percentage identity
PB2 (segment 1)	(1) sw/Jamesburg/42 (2) sw/Iowa/15/30 (3) sw/31	99% 98% 98%
PB1 (segment 2)	(1) sw/Jamesburg/42 (2) sw/31 (3) sw/OH/23/35	99% 98% 97%
PA (segment 3)	(1) sw/Jamesburg/42 (2) sw/Iowa/15/30 (3) sw/Iowa/31	99% 98% 98%
HA (segment 4) A/Alma-Ata/1044/83 A/Alma-Ata/1417/84 A/Alma-Ata/5/98	(1) sw/Almata/1417/84. (2) sw/Jamesburg/42 (3) sw/Iowa/30.	99% 99% 99%
HA (segment 4) A/Alma-Ata/32/98	(1) WS/33 (2) sw/Cambridge/39	98% 97%
NP (segment 5)	(1) sw/41/49 (2) sw/Jamesburg/42 (3) sw/Iowa/46	99% 99% 99%
NA (segment 6)	(1) sw/Jamesburg/42 (2) sw/Iowa/73 (3)sw/Iowa/31	99% 99% 99%
M (segment 7)	(1) sw/Jamesburg/42 (2) sw/Iowa/15/30 (3) sw/29/37	99% 99% 98%
NS (segment 8)	(1) sw/Jamesburg/42 (2) sw/Iowa/15/30 (3) sw/Iowa/31	98% 98% 98%

4.3 Hemagglutinin (HA)

4.3.1 Phylogenetic analysis

To deduce the evolutionary relationship of the gene segments of the isolated viruses, phylogenetic trees were generated with the nucleotide sequences of all eight segments with representative sequences in avian (several subtypes if possible), classical swine, 'avian-like' swine, 'human-like' swine, human H1N1, and the 1957 (A/Singapore/1/57) & 1968 (A/Hong Kong/1/1968 or A/Aichi/2/1968) pandemic strains. Phylogenetic analysis of the HA1 gene (Fig 3) of the four viruses clearly separates the tree into classical swine, human H1N1, and other subtypes. The tree (Fig 3) shows that the Alma-Ata/32/98 belongs to the human H1N1 clade, placed near the WS/1933 strain, while the other three viruses are related to the classical swine clade (H1N1), placed near the sw/Jamesburg/1942 strain. Thus the HA of the viruses is seen to fall under two groups, the Alma-Ata/32/98 in one and the Alma-Ata/1044/82, Alma-Ata/1417/84, and Alma-Ata/5/98 in the other (referred to as Alma-Ata/1044 cluster hereon). It is also interesting to note that the contemporary Mongolian isolates (geographically closer to Kazakhstan) are also homologous to an old human isolate, PR/1934 (arrows in Fig 3).

4.3.2 Amino acid comparison with the most identical sequence

Since phylogenetic analysis & BLAST analysis placed the viruses very near to sw/Jamesburg/1942 and WS/1933 strains, the amino acid sequence of the HA of the

isolated viruses were compared to the aforementioned strains to investigate the differences (Table 4). It was observed that the Alma-Ata/1044 cluster viruses had undergone seven mutations compared to sw/Jamesburg/1942, and Alma-Ata/32/98 had undergone six differentiating mutations compared to WS/1933. The significance of these mutations is discussed later.

4.3.3 Receptor-binding sites (RBS)

HA binds to sialic acids, and it has been observed that swine & human influenza A viruses preferentially bind to α 2,6 linked sialic acids, while avian, equine influenza A viruses preferentially recognize α 2,3 linked sialic acids. This change in receptor specificity is linked to a jump in species barrier for avian influenza viruses to infect mammalian hosts (Weis et al., 1988, Matrosovich et al., 1997, Matrosovich et al., 2000). Thus the residues important for sialic acid binding specificity of the HA were analyzed for the viruses sequenced in this study (Table 4). Alma-Ata/1044 cluster viruses possessed amino acids typically recognizing α 2,3 linked sialic acids at residues 77, 138, 155, 186, 194, 227, and 225 and amino acids typically recognizing α 2,6 linked sialic acids at residues 159 and 190. The classical swine isolate sw/Iowa/30 and sw/Jamesburg/1942 also possessed the same residues at these locations. Since it has been structurally shown (Gamblin et al., 2004) that the mutation E190D shifts from a dual specificity (both α 2,3 and α 2,6) to preferential recognition of α 2,6 linked sialic acids (in H1N1 subtype), it can be predicted that the Alma-Ata/1044 cluster viruses preferentially recognize α 2,6 linked sialic acids. On the other hand, the Alma-Ata/32/98 carried amino acids typically recognizing α 2,3 linked sialic acids at residues 77 & 194 and amino acids

typically recognizing α 2,6 linked sialic acids at 138, 155, 159, 186, and 225. The strain WS/1933 also carried the same residues at these locations. Since Alma-Ata/32/98 possesses Glu at residue 190, it can be postulated that it might have a dual specificity to both α 2,3 and α 2,6 linked sialic acids.

Table 4: Pairwise comparison of HA of the indicated viruses. Numbered according to the viruses sequenced in this study.

Amino acid	Alma-Ata/1044 cluster	sw/Jamesburg/42	Amino acid position	Alma-Ata/98	WS/1933
86	L	S	171	T	K
123	N	S	179	T	N
125	M	V	281	I	T
147	K	R	284	S	G
322	R	K	409	T	A
361	L	M	457	S	L
388	N	K			

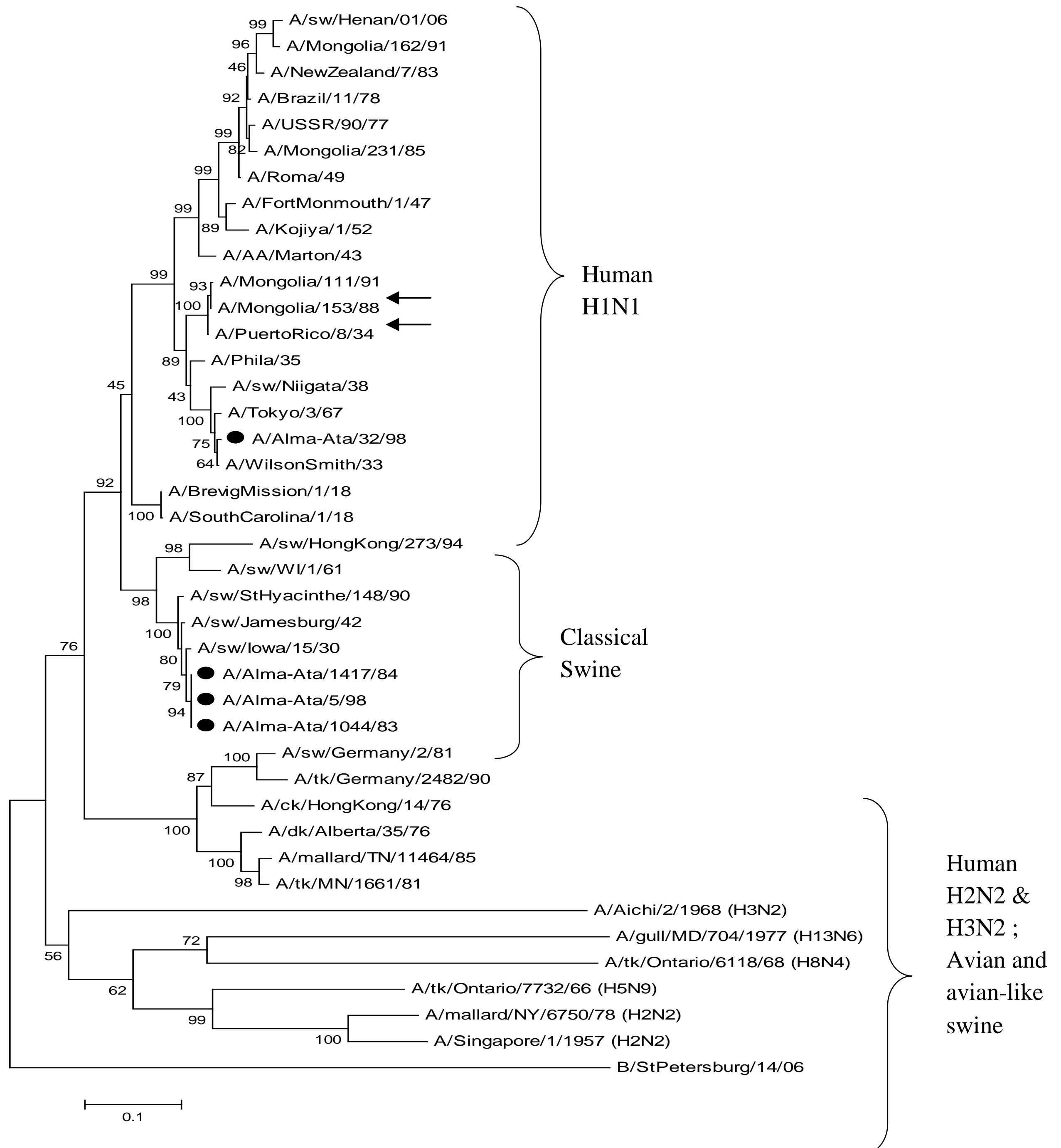

Figure 3: Phylogenetic tree of HA1 gene, aligned by Neighbor Joining method of the MEGA program, rooted to the B/StPetersburg/14/2006 HA. Bootstrap values are shown at every node (1000 replicates). Major clades are shown with large flower brackets. The Alma-Ata viruses sequenced in this study are labeled with a filled circle, and the Mongolian viruses are shown in the arrow (refer to text). A distance bar is shown below the tree. Standard influenza nomenclature is followed, with the following abbreviations: sw- Swine, ck-Chicken, dk-Duck, tk-Turkey, standard two letter abbreviations are used for US states.

Table 5: Receptor binding sites comparison between sequenced viruses and α2,3/ α2,6 binding conserved residues. Amino acid numbering is according to H3 HA.

Amino acid position	Avian consensus α2,3 binding [a, b]	α2,6 consensus for H1 [a, b]	Alma-Ata/32/98 (all RBS same as A/WS/1933)	Alma-Ata/1044/8 AlmaAta/1417/84 Alma-Ata/5/98 (all RBS same as swine/Jamesburg/ 1942)	Functional Significance of the amino acid
77	D	E	D	D	Though conserved, may not be important, as not near receptor binding site (RBS) [a]
138	A, G/T/S in very few	S	S	A	Ser allows a shallow pocket by interaction with Q226, enabling a better fit for α2,6 [c].
155	T	V/I (in classical swine & avian-like swine), T(human)	T	V	Postulated to help better binding to sialic acids in pigs compared to birds & humans [b]
159	T	N/D/S (sw & AvSw) G (human)	D	N	Minimal role in better positioning of the tip sialic acid [d]
186	P (only H1)	S (in majority)	S	P	By indirect interaction with G228, it could influence the binding to sialic acids towards α2,6 [a]
190	E	D	E	D	Crucial amino acid: E190D shifts from dual specificity to preference towards α2,6. [c,d]

194	L	I	L	L	Non polar contact to the N-acetyl methyl group. [e]
225	G	D	D	G	G- dual specificity D- preference towards α2,6. [d]. Structural reasons not clear.
226	Q	Q	Q	Q	Van der walls contact with A138 [d]
227	S	A/G	P (same in WS)	A	Near antibody binding site D (residues 225-228) [f]. Possibility of no effect on sialic acid binding [g].
228	G; R/S (in case of H3 and H13)	G	G	G	Hydrogen bond to NeuAc [e]

(a) Matrosovich et al., 1997, (b) Matrosovich et al., 2000, (c) Gamblin et al., 2004, (d) Stevens et al., 2006 (e) Martin et al., (1998) (f) Wilson et al., 1990 (g) Wu et al., 2004.

4.3.4 Glycosylation sites

HA is glycosylated on at least 4-5 sites, and it is reported that they are required for the proper function and immune evasion through antigenic variation. The acquisition of additional glycosylation sites has been associated with immune evasion mechanisms, as is seen with recent strains (Schulze 1997, Igarishi et al., 2008). The predicted glycosylation sites (Gupta et al., 2004) of Alma-Ata/1044 cluster and Alma-Ata/32/98 is given in Table 6. For a comparison, the sites of WS/1933 and sw/Jamesburg/1942 are also shown. It was observed that that Alma-Ata/32/98 has one less glycosylation site compared to its most identical sequence, WS/1933. But the glycosylation potential at this

site is not strong, as predicted by the NetNGlyc program. On the other hand, a comparison of the glycosylation sites between the Alma-Ata/1044 cluster and sw/Jamesburg/1942 reveals no difference in the number of glycosylation sites.

Table 5: Predicted glycosylation sites on HA of Alma-Ata viruses, WS/1933, and sw/Jamesburg/42 using NetNGlyc 1.0 (Gupta et al., 2004)

A/Alma-Ata/32/98		A/WS/1933	
Position & sequence	Glycosylation potential	Position & sequence	Glycosylation potential
28 NSTD	+++	28 NSTD	+++
40 NVTV	+++	40 NVTV	+++
286 NASM	+	179 NNSY	+
304 NSSL	++	286 NASM	+
498 NGTY	+	304 NSSL	++
557 NGSL	++	498 NGTY	+
		557 NGSL	++

A/Alma-Ata/1044 cluster		A/swine/Jamesburg/1942	
Position & sequence	Glycosylation potential	Position & sequence	Glycosylation potential
28 NSTD	+++	28 NSTD	+++
40 NVTV	++	40 NVTV	++
104 NGTC	+	104 NGTC	+
304 NSSL	++	304 NSSL	++
498 NGTY	+	498 NGTY	+
557 NGSL	++	557 NGSL	++

4.3.5 Antigenic sites

Four antigenic sites have been identified on HA, denoted as Ca, Cb, Sa and Sb, comprising a total of 42 amino acids (Caton et al., 1983). These are the highly variable sites identified as antibody binding sites on HA. Briefly, the sites are Ca: 155-158, 160, 184, 186-88, 222-23, 239-40, 255; Cb-86-92 & 130; Sa: 139, 142-43, 172-73, 175-82; Sb: 171, 174, 204, 207-09, 211-13 (all numbered according to A/Brevig Mission/1918 strain). A comparison with its most identical strain WS/1933 shows that Alma-Ata/32/98 has a single point mutation in Sa (amino acid 171), making it lose the glycosylation site, as well as generating a possible antigenic variant. Additional mutations are observed in the 'stem' region, which may be significant considering a recent discovery that established that universal neutralizing antibodies against HA bind to the stem region, near the fusion peptide (Sui et al., 2009) (Fig 4). On the other hand, four of the seven differentiating mutations observed between Alma-Ata/1044 cluster and sw/Jamesburg/1942 map to the surface of the molecule. Since two of these are near

antigenic sites, it could again be postulated to generate an antigenic variant. Three dimensional views of the two viruses were generated using Swiss-model (Arnold et al., 2006) using homology modeling (Fig 4). The figure shows that mutations have occurred in and contiguous to antigenic residues.

4.3.6 Cleavage site

HA is synthesized as an inactive precursor HA0 which has to be cleaved to HA1 and HA2 before they become active. While this cleave site is an Arg in several influenza viruses, it has been observed that highly pathogenic H5 and H7 avian influenza viruses isolates possess a stretch of basic amino acids at those locations, that has been shown to increase the virulence of the virus by increasing the possibility of a systemic infection of the host (Horimoto et al., 1994). The Alma-Ata viruses don't show any multiple basic amino acids at the cleavage site.

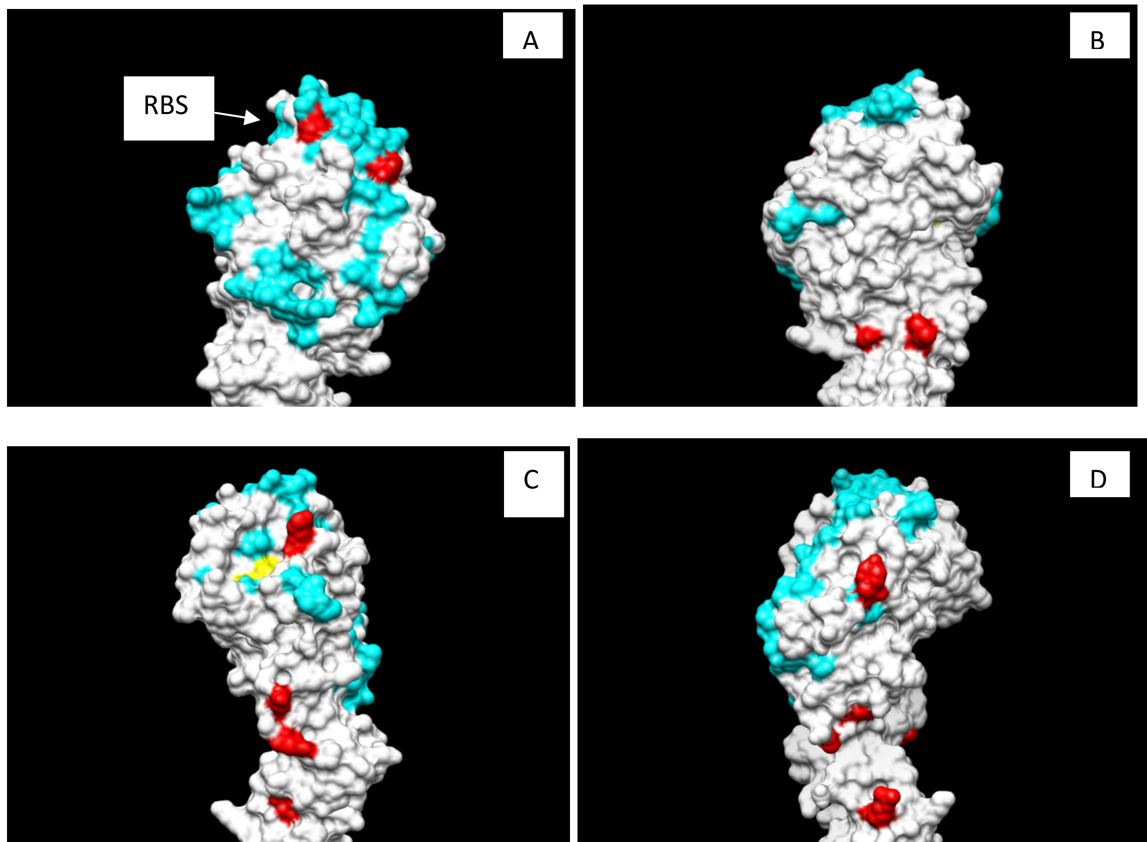

Figure 4: Surface representation of HA1 domain generated by Swiss-model (Arnold et al., 2006) and visualized by UCSF Chimera (Petterson et al., 2004), with the sialic acid binding site colored yellow (A) Representation of Alma-Ata/32/98 and (B) it's 360° lateral anticlockwise view, with red representing the amino acids differing from WS/1933, and cyan representing the antigenic sites Ca, Cb, Sa, and Sb. (C) Representation of some of the surface mutations in Alma-Ata/1044/84 cluster, and (D) it's 90° lateral anticlockwise view, showing amino acids different from sw/Jamesburg/42 in red (Table 2) and antigenic sites Ca, Cb, Sa and Sb in cyan.

4.4 Neuraminidase (NA)

4.4.1 Phylogenetic analysis

The phylogenetic tree places the neuraminidase of all the four isolated viruses near the root of the swine lineage, similar to the phylogenetic root of HA. As seen in the tree (Fig 2), there is high sequence identity between the NA of the Alma-Ata viruses and sw/Jamesburg/42, placing them near each other.

4.4.2 Catalytic, Glycosylation and Antigenic sites

Influenza neuraminidase is a tetramer, with the active site present at the exposed tip of the molecule. All the sequenced viruses have the active site residues conserved as present across different subtypes (Colman et al., 1983).

Just like HA, neuraminidase is also glycosylated, and it has been reported that the number of glycosylation sites in N1 avian (domesticated and wild birds) species is 7, and additional sites serving as antigenic variants are observed in human strains from 1947-1986 (from 365-367) and also in strains after 1954 (amino acids 454-456) (Reid et al., 2000). Like the avian strains, the NA of the Ama-Ata viruses has 7 predicted glycosylation sites. Also, there was no loss of glycosylation site at amino acid 146, which has been postulated to extend the tropism of neurotropic influenza strains (Li et al., 1993 and Reid et al., 2000).

The complete antigenic structure of N2 & N9 subtype neuraminidase has been determined (Colman et al., 1983 and Colman et al., 1987), and is described to be 22 amino acids spread over seven epitopes on the surface exposed part of the macromolecule. As antigenic structures of N1 subtype viruses are not mapped, the Alma-Ata N1 subtype NA proteins were aligned with N2, and the homologous 22 amino acids were compared. It was observed that NA of the Alma-Ata viruses has their antigenic sites closely resembling the classical virus (including sw/Jamesburg/1942), except for a mutation S369N, which interestingly is a critical residue in the HB site 364-375 (Webster et al., 1987). Although the HB site is prevalent in avian viruses, it's usually lost when it adapts to a human or swine host (Kobasa et al., 1997). This specific differentiating mutation between A/swine/Jamesburg/42 and the human Alma-Ata viruses can thus be extrapolated as adaptation of the swine virus to humans. Since this mutation also lies near the antigenic residue 367, it could also influence antibody recognition of the virus.

Also, on comparison of the cytotoxic T-lymphocyte recognizing epitopes on neuraminidase of sw/Iowa/30 and Alma-Ata viruses, the epitope SLCPIS**R**WAI (residues75-84) in Iowa/30 was mutated to SLCPIS**G**WAI. Such mutations are interesting as they are supposed to play a role in immune evasion mechanisms (Eisenlohr et al., 1992).

4.4.3 Virulence determinants

The length and sequence of the NA stalk is seen to vary between different influenza viruses (Blok et al., 1982). Shorter stalks are seen to result in less efficient viral

release (active site in the head region cannot access the sialic acid substrate) (Luo et al., 1993). However, it was observed that several HPAI viruses, including the 1995 HPAI viruses isolated in Hong Kong and early human strains like WS/1933 (but not the 1918 virus) had deletion in the stalk region (Li et al., 2003) and still exhibited virulence. The Alma-Ata viruses do not show any large stalk deletions, as observed in such highly pathogenic viruses. They carry the typical 4 amino acid deletion observed in a large number of classical swine isolates.

Avian influenza viruses have been shown to replicate in the intestinal tract in avian hosts, whereas in humans, replication is largely restricted to the respiratory tract. This low pH environment stability was experimentally shown to be primarily by two amino acids in NA: R344 and F466, both of which are located near known calcium binding sites (in N2 subtype NA) (Takahashi et al., 2003). Alma-Ata and several human viruses possessed N and F at those positions.

4.4.4 Phylogenetically important regions

Fanning et al., (2003) identify phylogenetically informative amino acid positions in the influenza neuraminidase, which they propose might be involved in important virus-host interactions. They mention about three noteworthy regions which might be involved in host adaptation of influenza viruses of the N1 subtype viruses: amino acids 67-86, 285-289, and 339-344. A comparison of the regions in Alma-Ata viruses shows that amino acids in this region are identical to conserved residues in human and swine viruses.

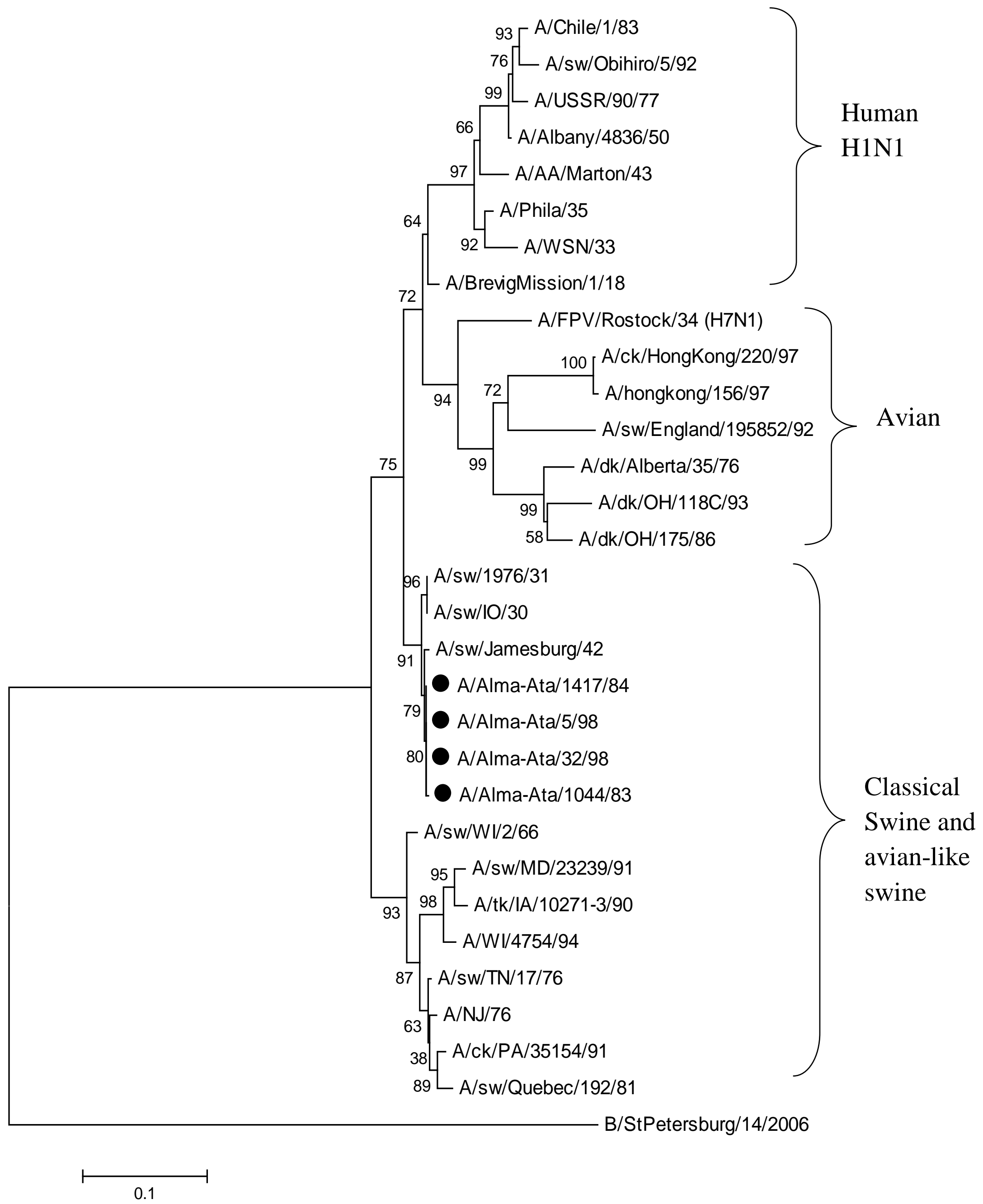

Figure 5: Phylogenetic tree of NA; analyzed in the same way as described for Fig 1.

4.5 Polymerases- PA, PB1 and PB2

4.5.1 Phylogenetic analysis

Phylogenetic trees of the polymerase genes (Fig 3, 4, 5) consistently place the polymerase genes of the isolated viruses at the root of the swine lineage, along with the classical swine virus sw/Jamesburg/42.

4.5.2 Genetic signatures associated with pathogenicity and host range

Chen et al., (2006) have identified genetic signatures of influenza viruses that might be associated with an avian virus becoming a human influenza virus. They identify 52 'species-associated' genetic signatures, of which 35 were located on the ribonucleoproteins (NP- 15, PA-10, PB1-2, and PB2-8). Interestingly, the genes of the ribonucleoproteins of the 1918 pandemic influenza share 12 of these 35 genetic signatures. Taubenberger et al., (2005) also identify some amino acid differences observed between the human, swine, avain and equine influenza viruses based on the sequence of the 1918 pandemic influenza strain (Brevig Mission/18). A summary of the important amino acids is provided in Table 7. It is seen that the polymerase is largely relates to the classical swine viruses.

Of these amino acids, the only experimentally proved and probably the most interesting residue is amino acid 627 in PB2. Several studies have documented its roles in replicative ability in mice and probably humans (Subbarao et al., 1993). The presence of K627 in Alma-Ata can thus be assumed to be involved in replication in human

respiratory tract, consistent with the human origin of these viruses. Nevertheless, it has been reported that the ability of K627 to modulate virulence is dependent on additional features of the viral genome (reviewed in Nadia et al., 2008). Another species specific residue is amino acid 375 in PB1: Asn predominates in avian isolates, while Ser in most human isolates (Taubenberger et al., 2005). Alma-Ata viruses, consistent with their human isolation, have mammalian 'signature' at those residues. In addition, there are several mutations in the ribonucleoproteins implicated with virulence in mice- PB1: L13P, D538G, K578Q, R614G, S678N; PB2: T333I, K482A, E627K, D701N, S714R; PA: S65Y, E133G, Q556R, K615N; NP: D34N, N319K, D480N; PB1-F2: N66S (Gabriel et al., 2005; Shinya et al., 2007; Brown et al., 1999; Brown et al., 2001; Conenello et al., 2007). The Alma-Ata viruses do not possess any of those mutations.

Table 7: Amino acid comparison of polymerases (numbered according to H1N1) between Alma-Ata viruses and viruses of other subtypes.

Protein	Residue no	Avian	Human H1N1	Human H2N2	Human H3N3	Classical swine	Alma-Ata viruses	Functional domains associated with the amino acid (a)
PB2	199	A	S	S	S	S	S	NP binding
	271	T	A/T	A	A	T	T	Cap binding
	475	L	M	M	M	M	M	Nuclea localization signal
	567	D	N	N	N	D	D	Cap binding
	627	E	K	K	K	K	K	NP binding
	702	K	R	R	R	R	R	Importin-α
PB1	375	N/S/T	S	S	S	S	S	cRNA interaction
PA	55	D	N	N	N	N	N	Induction of proteolysis
	100	V	A	A	A	V	V	
	382	E	D	D	D	D	D	
	409	S	N	N	N	S	S	
	552	T	S	S	S	S	T	

(a) Boulo S et al., 2007

Figure 6: Phylogenetic tree of PA gene; analyzed in the same way as described for Fig 1

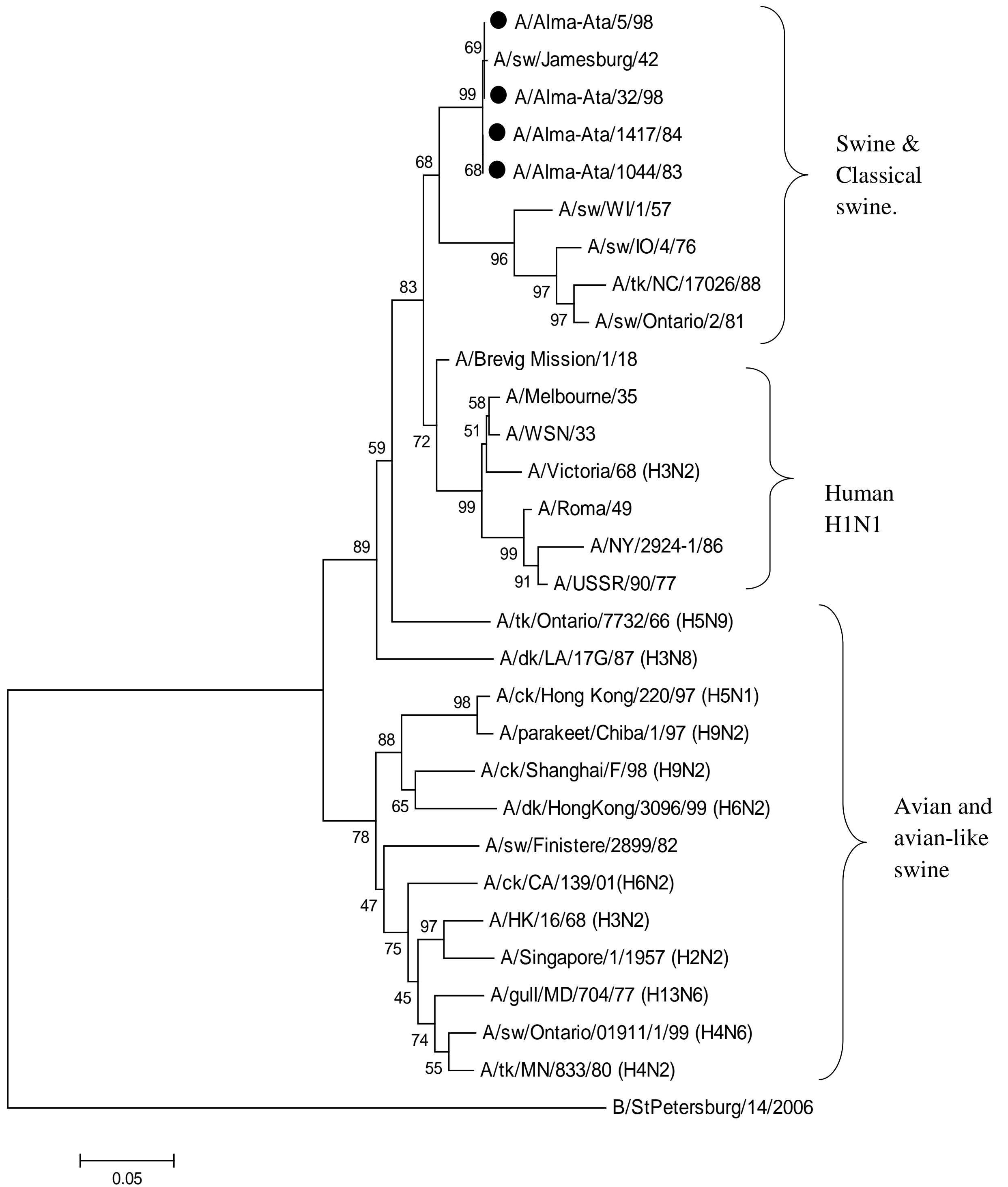

Figure 7: Phylogenetic tree for PB1; analyzed in the same way as described for Fig 1

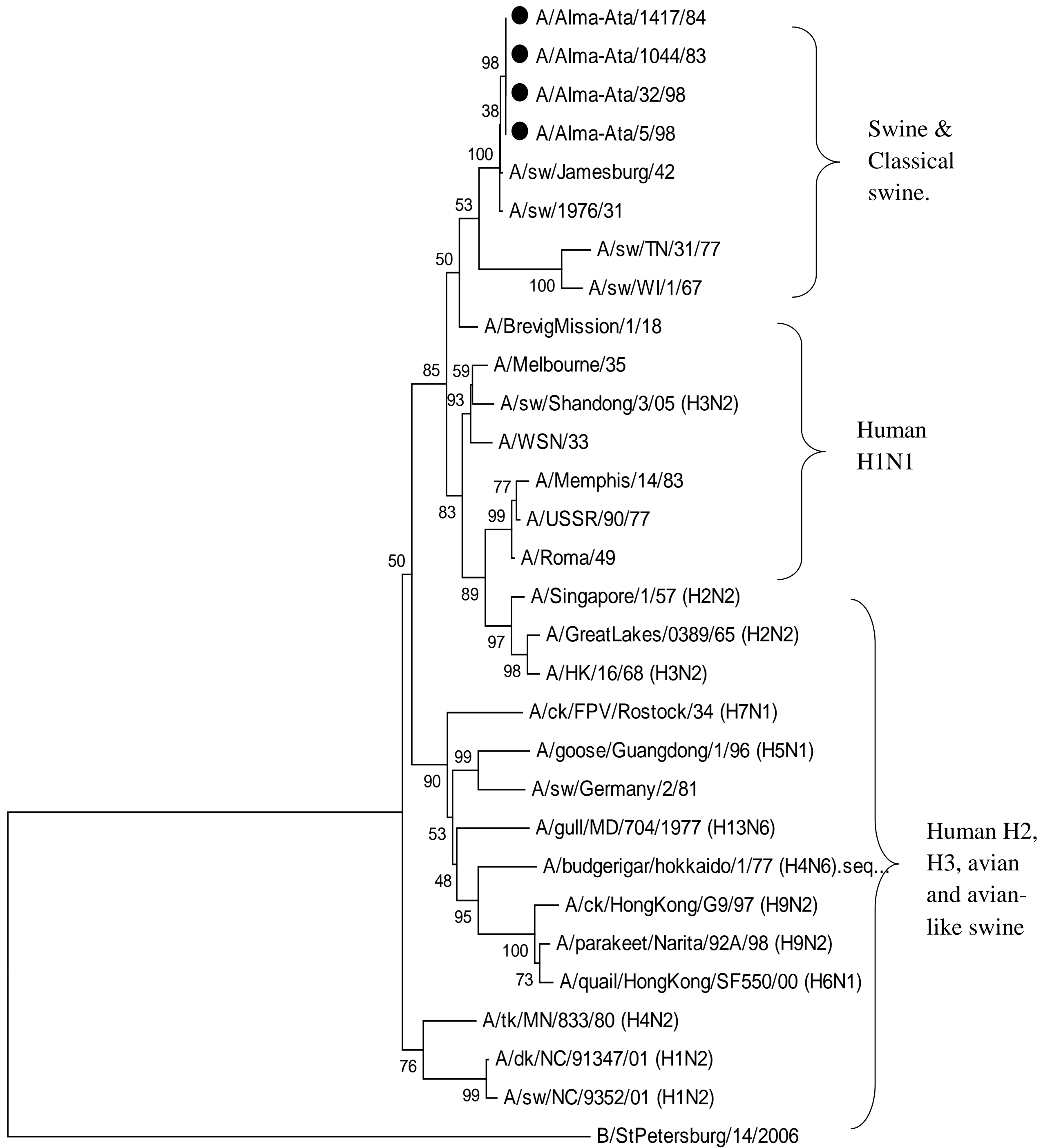

Figure 8: Phylogenetic tree for PB2; analyzed in the same way as described for Fig1.

4.5.3 Antigenic residues

While the major targets of the humoral immune response have been the surface proteins (HA and NA), T-cell responses have often been directed towards internal proteins like NP, Matrix, and polymerase proteins (Gotch et al., 1987, Assarsson et al., 2008), and several epitopes have been identified. In addition to change inside the epitope sequences, it has been reported that mutations flanking the epitopes might also affect antigen processing and presentation of these antigens (Eisenlohr et al., 1992). Interestingly, all the positions at which the Alma-Ata viruses varied from their closest relative sw/Jamesburg/42 is at epitopes or the residues flanking certain epitopes, as summarized in the Table 8.

Table 8: Mutations differences in and around T-cell epitopes in the Alma-Ata viruses and sw/Jamesburg/42 isolate (numbered according to BrevigMission/1918 virus).

Gene (amino acid number) sequence in some cases	Sequence of Sw/Jamesburg/1942	Sequence in Alma-Ata viruses	Reference for epitope
PB1 591-599	VSDGG**P**NLY	VSDGG**S**NLY	DiBrino et al., 1993
PB1 571-579	RRSFEI**T**KL	RRSFEI**K**KL	Cheuk et al., 2005
PB1 741-749 AEIMKICST	Flanking sequence: IELD**R**	Flanking sequence: IELD**K**	Assarsson et al., 2008
PB2 322—331 SFSFGGFTFK	Flanking sequence: RTSGSVK**K**	Flanking sequence: RTSGSVK**R**	Assarsson et al., 2008
PB2 607–621 LGTFDTVQIIKLLPF	Flanking sequence: AAAPPKQ**N**	Flanking sequence: AAAPPKQ**S**	Assarsson et al., 2008
PB2 645-659 MRILVRGNSPAFNYN	Flanking sequence (before & after):	Flanking sequence (before & after):	Assarsson et al., 2008

	AVNVRGSG…**R**AT	**T**VNVRGSG…**K**AT	
PA 46–54 FMYSDFHFI	Flanking sequence (before & after): **M**EVC….NER**S**	Flanking sequence (before & after): **L**EVC…NER**G**	Gianfrani et al. 2000

4.6 Nonstructural protein (NS)

4.6.1 Phylogenetic analysis

Just as observed in the other gene segments, the phylogenetic analysis of the nonstructural protein places it at the root of the classical swine lineage (Fig 6), contiguous to the sw/Jamesburg/42 virus.

4.6.2 Functional residues

The reported RNA-binding domain (residues 19-38), effector domain (134-161), nuclear localization signals (R35, R38 and K41) are all conserved in the Alma-Ata viruses (reviewed in Hale et al., 2008). Though not experimentally proved, NS has also been reported to be post-translationally modified by phosphorylation at two residues, S195 and T197 (Bornholdt & Prasad 2006) (the physiological role of which is not yet clear). Alma-Ata viruses have the conserved Ser at 195 but have Asn at 197, instead of threonine. This is also observed among some of the classical swine viruses. The significance of this substitution is not clear.

4.6.3 Virulence determinants

It was shown that a natural deletion (amino acids 191-195) variant of the NS protein was less pathogenic in chickens compared to a virus with a normal NS protein (Zhu et al., 2008). Li et al., (2006) have shown that the substitution A149V attenuated a goose influenza virus in chickens. Amino acid 42 is also implicated with virulence modulation. The presence of serine of glycine has been associated with high pathogenicity (reviewed in Hale et al., 2008). None of these characteristics were observed in the Alma-Ata nonstructural proteins. Like the aforementioned study by Chen et al., 2006, and a similar study by Finkelstein et al., (2007), Table 9 shows the 'species associated' amino acids in NS1 and NS2 protein of the Alma-Ata viruses. It shows that the Alma-Ata viruses are identical to the swine viruses in the 'species associated' signatures.

Interestingly, the NS1 T-cell recognizing epitope GEISPLPSL (residues 158-166) (DiBrino et al., 1993) has the flanking residues PG**Y**T in the Alma-Ata viruses instead of the consensus PG**H**T. As mentioned earlier, this may have a role in immune evasion.

It is important to note that specific amino acid changes in NS required for host adaptation have not yet been identified, although several studies have identified certain residues capable of modulating virulence. But it has also been underscored that the observed effects might also be highly dependent upon the host under study (reviewed in Hale et al., 2008).

Table 9: Comparison of the phylogenetically important regions of NS1 and NS2 proteins between the Alma-Ata viruses and viruses of different subtypes.

	Amino acid	Avian	Swine	Human H1N1	Human H2N2	Human H3N2	Alma-Ata viruses	Functional significance
NS1	81	I	I	I	M	M	I	eIF4γ1 binding (a)
	215	P	P	P	T	T	P	CrK interaction through SH3 domain (b). P215 is reported to hyper phosphorylate PI3K.
	227	E	R	K	R	R	R	PDZ domain (c)
NS2	70	S	G				G	M1, NEP domerization domain (d)
	107	L	F				L	M1, NEP dimerization domain (d)

(a) Aragon T et al., 2000 (b) Heikkinen LS et al., 2008 (c) Obenauer JC et al., 2006 (d) Akarsu et al., 2003.

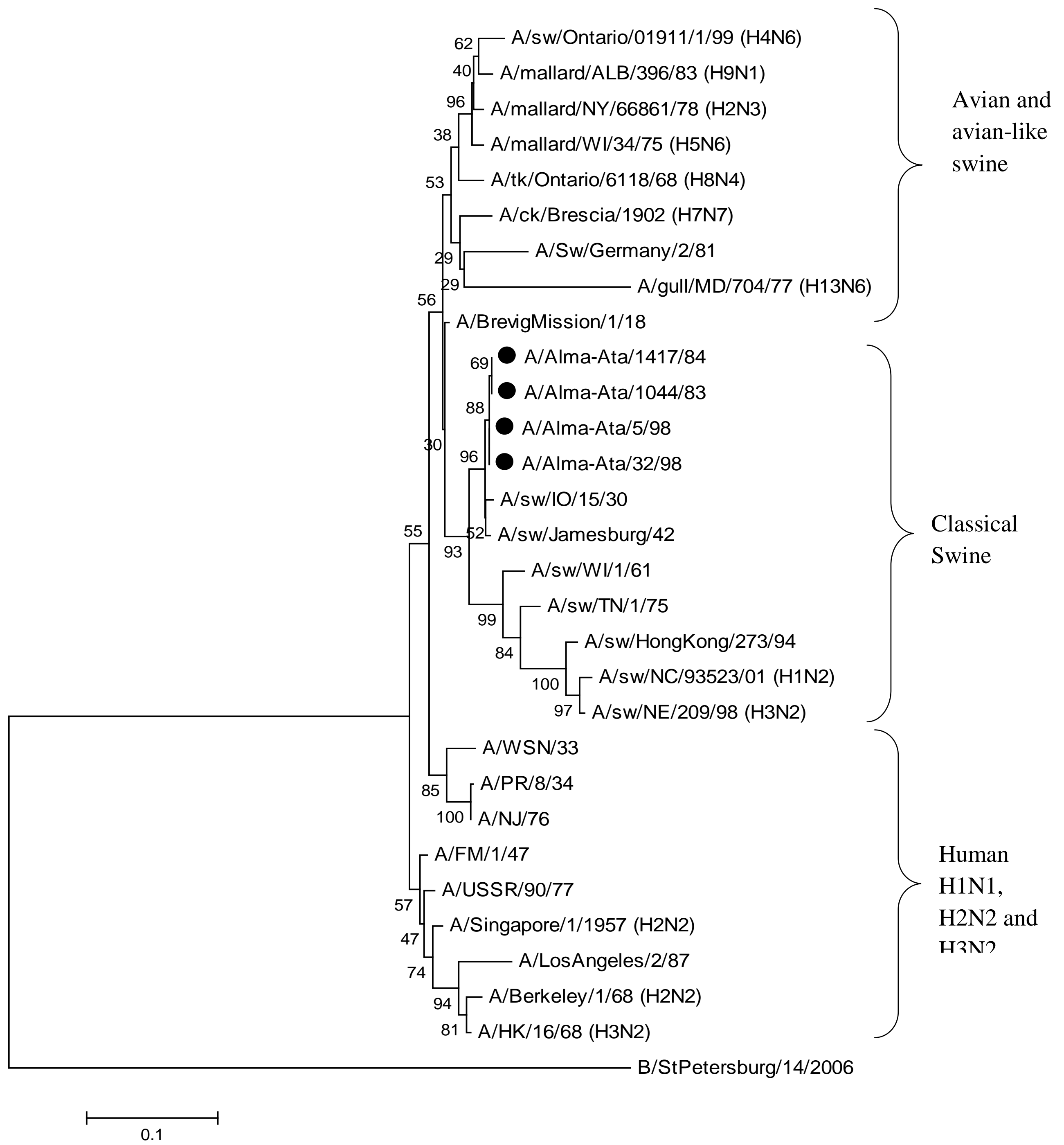

Figure 9: Phylogenetic tree for NS1 gene; analyzed in the same way as described for Fig1

4.7 Nucleoprotein (NP)

4.7.1 Phylogenetic analysis

The phylogenetic tree (Fig 7) shows that the NP genes of the Alma-Ata viruses once again fall at the root of the classical swine lineage, along with the sw/Jamesburg/42 isolate. Interestingly, the NP genes of the contemporary Mongolian viruses were also identical to the PR/8/34 strain, as seen for the HA (Fig 1).

4.7.2 Functional residues and CTL epitopes

The reported RNA binding residues (Elton et al., 1999), nuclear localization signals (residues 327-345, 179-193 and 3-13 (Bui et al., 2002), and cytoplasmic accumulation signals (F338, E339 and R342; Digard et al., 1999) are all conserved in the Alma-Ata viruses.

As mentioned earlier, the majority of the T-cell responses on influenza are upon the nucleoprotein, as it's one of the least diverging proteins of the influenza genome. Table 10 shows the epitopes that were different in Alma-Ata viruses from the sw/Jamesburg/42 as well as the classical swine viruses.

Table 10: Mutational differences in and around T-cell epitopes between Alma-Ata viruses and sw/Jamesburg/42 isolate (numbered according to BrevigMission/1918 virus).

Epitopes (amino acid number)	Sequence in sw/Jamesburg/42	Alma-Ata viruses	Reference
NP 44-52	CTELKL**S**DY	CTELKL**D**DY	DiBrino et al., 1993
NP 91-98	KTGGP**I**YRR	KTGGP**L**YRR	Cerundolo V et al., 1991
NP 103-111	KW**I**RELILY	KW**K**RELILY	Berkhoff EG et al., 2007
NP 338-347 FEDLRVSSF	Flanking region: K**K**VV	Flanking region: K**R**VV	Assarsson E et al., 2008

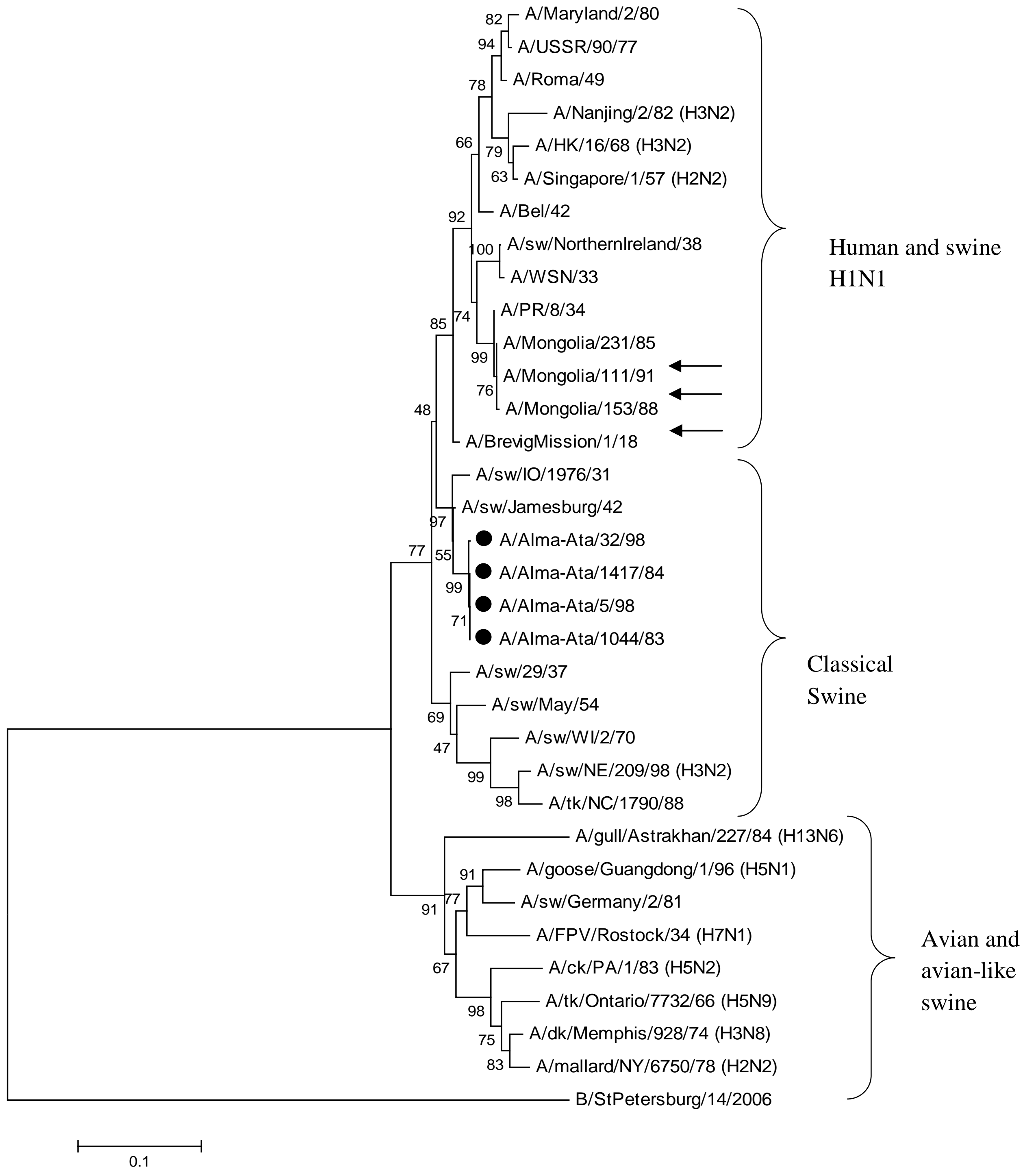

Figure 10: Phylogenetic tree for NP gene; analyzed in the same way as described for Fig1. Arrows show the contemporary Mongolian isolates identical to another lab isolate, PR/8/34.

4.8 Matrix (M)

4.8.1 M1 gene

4.8.1.1 Phylogenetic analysis

Phylogenetic analysis places the M1 gene of the sequenced viruses at the root of the classical swine lineage, near the isolate sw/Jamesburg/1942 (Fig 8).

4.8.1.2 Functional residues and virulence determinants

The RNA-binding and nuclear localization signal (residues 101-105) differs from the consensus RKLKR to GKLKR in Alma-Ata viruses. It has been reported that mutating residues 101 and 105 reduced RNA binding (Ye et al., 1999). Interestingly, one study observed that a double mutant of R101S and E105S resulted in attenuated replication of WS/33 in mice (Liu & Ye 2005). Thus the significance of this mutation in the Alma-Ata viruses remains unknown. The putative zinc binding motif (residues 148-162) is also conserved in the Alma-Ata viruses (Elster et al., 1994).

It has been reported that the mutation T139A is responsible for mouse adaptation of FM47 virus (Smeenk et al., 1994). In a similar study, it was observed that adaptation of Hong kong/1968 virus was facilitated by one change in M1 (167A) and one in M2 (D44N) (Brown et al., 2001). Adaptation of H2 viruses to mice was enabled by three changes in the matrix proteins: N30D, Q214H, and M179K (Govorkova et al., 2000). Another study comparing sequence differences between high pathogenicity viruses with

low pathogenicity viruses observed that one of the changes responsible was V15I in the M1 protein (Val for low path and Isoleu for high path) (Katz et al., 2000). None of these changes were found in Alma-Ata viruses. Additionally, it was also found that the M1 protein was not antigenically different from its closest relative sw/Jamesburg/42.

4.8.2 M2 gene

4.8.2.1 Functional residues and drug resistance

M2, the ion channel of the influenza virus is potentially palmitoylated at C50 (Sugrue et al., 1990) and phosphorylated at S63 (Holsinger et al., 1995), and the M2 protein of Alma-Ata viruses have the two post-translational modifications. Drug resistance of M2 proteins occur with point mutations in any one of the 6 amino acids: 26, 27, 30, 31, 34, or 38 (Belshe et al., 1988, Hay et al., 1986). Barring the mutation A28 instead of V28 observed in all human viruses, none of the mutations involved in drug resistance are observed in the M2 protein of the Alma-Ata viruses. Thus the Alma-Ata viruses are drug sensitive strains.

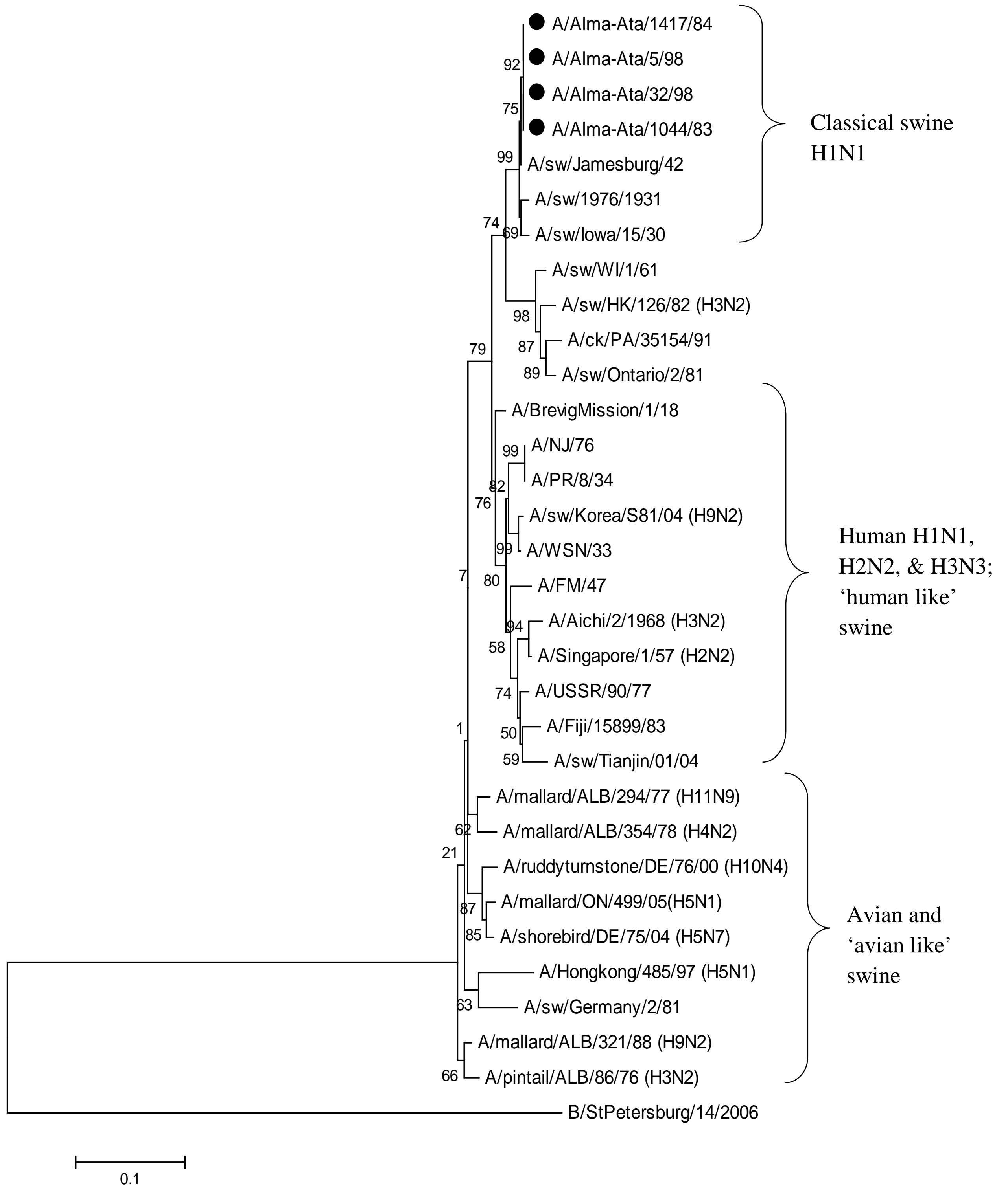

Figure 11: Phylogenetic tree for M1 gene; analyzed in the same way as described for Fig1.

4.9 Evolutionary anachronism of the isolates

In light of antigenic shift of swine viruses, it has been estimated that swine H1 HA accumulates non-synonymous mutations at a rate three times lesser than H1 HA of human influenza viruses (Sugita et al., 1991), and others have reported the numbers to be between 0.001 to 0.007 changes per nucleotide per year (the number varies between subtypes and genes) (Krasnitz et al., 2008). This is attributed to the weak immune system present in swine, compared to humans. Additionally, pigs have a short life span, which doesn't cover more than one influenza epidemic, and vaccines are only infrequently applied. Importantly, husbandry practices play a major role in the evolution of swine influenza viruses (Brown 2000). Paradoxical to this established dogma, these human viruses from Kazakhstan are seen to be 98-99% identical to the classical swine virus sw/Jamesburg/1942 (except for the HA of Alma-Ata/32/98 which is 98% identical to WilsonSmith/1933 isolate). Surprisingly, even considering a modest mutation rate of 0.001 per nucleotide per year, the number of mutations that should have occurred in a span of 50 years (first isolation in 1983), and a subsequent gap of 15 years (second isolation in 1998) has not occurred in these viruses. Though skeptical to begin with, on close analysis, several facts probably add credentials about the legitimacy of these viruses.

To begin with, this is not an isolated occurrence of anachronistic virus being discovered (in terms of genetic change), and there have been a handful of cases wherein 'old' viruses were either discovered from swine populations or resuscitated to become epidemic viruses. The best example was the latter case, wherein it was discovered that the

1977 ‘Russian’ influenza was closely related to viruses circulating in 1950, and ‘resuscitated’ after 27 years of ‘dormancy’ (Nakajima et al., 1978). Second, viruses to the 1968 H3 human influenza virus were isolated from pigs for up to at least 10 years without much antigenic variation (Shortridge et al., 1977). The authors proposed that pigs could serve as a reservoir of long term conservation of viruses. The third case is isolation of H3 from swine in 1995 which were highly identical to human strains circulating in 1975 (Bikour et al., 1995b). The fourth and the most striking case is a similar isolation of swine virus highly identical to the sw/Iowa/1930 virus (Bikour et al., 1995a). In addition to these biotic reservoirs of influenza viruses, evidence is accumulating on abiotic reserves of influenza virus (Zhang et al., 2006, Lang et al., 2008).

Although influenza viruses mutate rapidly, there are some points that can be put forth to explain long time persistence of H1 influenza, particularly with reference to swine. First, it has been reported by several authors that human H1 viruses have a lower fixation rate compared to human H3 viruses, resulting in lesser antigenic change. This is reflected by the occasional need to change the H1 component in the universal influenza vaccine recommended by WHO (reviewed in Hay et al., 2001). Wolf et al., (2006) did large scale phylogenetic analysis of HA of recent human H1 and H3 viruses, and highlight that compared to H3, H1 viruses are under less positive pressure, such that genetically diverse lineages are able to coexist. Thus if the rate of fixation of human H1 is slow, then it can be extrapolated that swine H1 viruses face a much lesser antigenic pressure to evolve, because of the aforementioned reasons. Indeed, until 1998, all swine isolates in the US were derivatives of the ‘classical’ swine isolates, with a low divergence

rates (Webster et al., 1992 & Sugita et al., 1991). Even if divergence occurs, this is seen in regions unrelated to antigenic sites (Sheerar et al., 1989; Brown, 2000). But contemporary swine viruses (H3 and H1) in other countries might be evolving at differing rates, higher or lower, due to different husbandry practices (de Jong et al., 1999). Corroborating it, Easterday (1980) associated the slow rate of mutation of swine influenza in the US (till 1998) to the identical husbandry practices followed in the country over the decades. On similar lines, it might be assumed that the husbandry practices followed in Kazakhstan might have led to the persistence of the virus in swine herds over the decades, before transmitting it to humans. The observation that the sequences are identical to swine influenza virus substantiates this possibility. It could be that the virus was maintained in small geographical 'pockets', which may be restricted to a small farm or a part of a continent. Minimization of contact between farms, coupled with the short life span and low immunity of pigs could have led to the long-term conservation of these viruses (Bikour et al., 1995). In that case, such old isolates should still be circulating in Kazakhstan, and should have been occasionally transmitted to humans through the years. Emphasizing that the influenza surveillance in general is very weak in the central Asian countries, a search in the influenza gene database [http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html] does not yield any anachronistic isolates. So there is no definite conclusion as to whether or not Kazakh swine harbor such classical viruses. Even in the assumption that they do harbor the isolates, this possibility becomes too implausible as no animal can harbor influenza viruses largely unchanged at the nucleotide level for over 50 years. So there should be an alternate explanation. Nevertheless, it is also important to mention that while there has been lot of studies on

evolution of human influenza viruses, there is dearth of complete understanding in the evolution of swine influenza viruses, and there are evidences both for & against pigs being long-term true carrier state of influenza viruses (Brown 2000).

The alternate possibility is the Mongolian influenza epidemics in the 1980s and how they relate to the influenza outbreaks in Kazakhstan. Anchlan et al., (1996) describe a set of H1N1 viruses causing severe outbreaks in the Mongolian human population from 1980-1993. It was later found that the outbreaks were largely caused by an incompletely inactivated reassortant vaccine between the PR/1934 and USSR/1977 strains. But, one of the isolated viruses, A/Mongolia/111/91 was closely related to the PR/8/34 strain in its entirety (arrows in Fig 1), a commonly used lab isolate. The authors mention that the PR/8/1934 strain might have been accidentally used for vaccination in southern USSR in 1980s (through personal communication to the author), a time period which *exactly* coincides with several outbreaks of influenza in Kazakhstan (Chuvakova et al., 1984 and 1985; Dem'ianenko et al., 1987), and other places in the former USSR (Degitarev et al., 1987, Ivanova et al., 1984, and Petrov et al., 1990). In Alma-Ata, several isolates were obtained during the outbreaks of 1980-85, including the two viruses sequenced in this study (Alma-Ata/1044/83 and Alma-Ata/1417/84). So the possibility of an accidental laboratory leak of these viruses becomes an attractive explanation. If one observes the characteristic of the Mongolian and Kazakhstan outbreaks, both occurred at more or less the same time, infecting significant number of people. So if a virus suddenly infects a significant number of people, it can be extrapolated that its antigenic make up cannot be old. This is because of the general observation that a pathogen would be severe (infecting

several hosts) only when it encounters a population which has not previously seen the pathogen. So the possibility of an accidental laboratory leak becomes highly probable. But if they were laboratory leaks, the next important question is how did the virus sw/Jamesburg/42, which is not a commonly used laboratory strain (unlike PR/1934 or WS/1933), got into the human population? Another question is why did the viruses die out? The reason might be due to herd immunity, as the population probably had antibodies to an antigenically related H1 (USSR/77) virus. So with the available data, there is no definite conclusion that can be reached about the origin of the virus. But the accidental leak theory can explain the origin of WS/1933 HA in the reassortant Alma-Ata/32/98. Chuvakova et al., (1984) report the isolation of A/Alma-Ata/13/81 from a 3 year old infant (lethal case), and describe the HA to be identical to the WS/1933 strain and NA to be identical to the PR/8/34 strain, through serological tests. Thus it can be hypothesized that the WS/1933 strain was also accidentally released along with PR/1934, enabling it to cause infections. Such speculations will remain open, and can only be clarified with more frequent surveillance among the swine and human population in the country.

The next mystery is regarding the 1998 isolates: after their initial "introduction" in 1980s, how is that the 1998 isolates, after a period of 15 years, show low rates of divergence from the 1983 isolates? There are two explanations which can be provided to explain this anomaly. First, it is important to note that the isolated viruses are not absolutely conserved compared to the 'classical' isolates. On detailed sequence analysis, it was observed that both the surface proteins and internal genes have undergone several

mutations, most of which possibly play a role in evading immune recognition. For instance, on comparison of HA of Alma-Ata/1044 cluster viruses with their most related evolutionary ancestor sw/Jamesburg/1942, the differentiating mutations are seen to be on the surface of the molecule, in or contiguous to antigenic sites, and at the 'stem' region of HA1 (Fig 2). On comparison of Alma-Ata/32/98 with its most related evolutionary ancestor WS/1933, it was observed that Alma-Ata/32/98 has lost a glycosylation site, which also lies contiguous to an antigenic site. It has been shown that variation in glycosylation patterns can lead to immune evasion (Schulze 1997, Igarishi et al., 2008). All of the differentiating mutations (of the sequenced isolated from sw/Jamesburg/42) in the internal genes, notably the ribonucleoproteins (NP, PA, PB1 and PB2) are in or on the residues flanking T-cell recognition epitopes (Table 6 &7). As mentioned earlier, these possibly help in immune evasion mechanisms. This is also seen in the differentiating mutations in NA and NS, also falling into T-cell recognition epitopes. Thus it can be speculated that the viruses have undergone just enough mutations to escape immune detection in the host. Another factor that has to be considered while giving an explanation to the persistence of the viruses was that these strains were isolated from a city with a population density of just 4.15 persons per km^2 [from a country of population density of only 5.5 people per km^2 living in a land of the size of Western Europe]. So it becomes obvious that the human-to-human and swine-to-human spread is going to be limited. Thus in order for the virus to survive in this extremes of ecological constraints, it can be postulated that virus might undergo minimal replication cycles by needing to keep the host healthy enough and long enough, to spread to another host, consequently resulting in reduced rate of mutations. This is reinforced by the minimal mutations the

viruses have undergone, probably just enough to escape immune detection. But swine can fit in both these constraints to explain this mystery. A virus that has undergone mutations just enough to escape immune detection can persist genetically unchanged for a long time in an immunologically weak host like swine. Thus it could be theorized that after the initial introduction of the viruses into Kazakhstan in the 1980s (by whatever means), the viruses spread to swine, persisted in them due to a multitude of factors (mentioned earlier), and infected humans in 1998. Since the virus did not have any immune pressure to evolve, it did not undergo lots of mutations and so this could be the reason behind low of the 1998 viruses. Once again, to resolve such questions, better surveillance efforts are needed in the aforementioned countries, not only in humans but also in other animals.

Finally, there are reasons to suggest that this might not be laboratory contamination. To begin with, none of the labs in Kazakhstan nor did our lab at the University of Maryland possessed the sw/Jamesburg/42 virus. Next is the serological evidence from Russian papers published during the time Kazakhstan and Mongolia were experiencing outbreaks. Through hemagglutinin inhibition assays, it was reported that the strains causing the outbreaks were antigenically related to classical swine viruses (Chuvakova et al., 1984 and 1985; Dem'ianenko et al., 1987). Additionally, after their initial isolation from clinical samples, the viruses were propagated only once in five years. These exclude the possibility that the viruses got contaminated in routine laboratory passages through the years before they were brought to the University of Maryland. Next, the viruses Alma-Ata/5/98 and Alma-Ata/32/98 were re-isolated from the original placental samples at the Ivanovsky institute of virology at Moscow,

confirming the nature of the virus. Since our lab at the University of Maryland possesses the WS/1933 strain, it could be possible that the isolate Alma-Ata/32/98 was an accidental laboratory reassortant. This cannot be because the personnel who grew and handled the Kazakhstan viruses at UMD did not possess the WS/1933 strain. Next, it could be possible that these were sequencing errors. As mentioned in the Materials & Methods section, every segment of every virus was sequenced with at least two different RNAs and with several different primers. Finally, the pre and post-PCR activities were physically separated. These collectively stand a proof that the viruses might not have been laboratory contaminants.

CHAPTER 5: SUMMARY AND CONCLUSIONS

This study details about the sequencing and phylogenetic analysis of all the eight segments of four swine influenza viruses isolated from humans in Alma-Ata, Kazakhstan in early 1980s and 1998. Except for the HA of Alma-Ata/32/98, all the other genes of all the other viruses were 99-100% identical to each other at the nucleotide level. Corroborating this, the phylogenetic analysis placed all the genes of all the other viruses in the classical H1N1 swine clade, near the isolate sw/Jamesburg/1942 (except for the HA of Alma-Ata/32/98 which was placed in the human H1N1 lineage near WS/1933). On detailed molecular analysis of the genes, several conclusions were reached:

1. It could be speculated that while the HA of Alma-Ata/32/98 had dual specificity in recognizing α 2,3 linked and α 2,6 linked sialic acids, the HA of the other three isolates probably recognize α 2,6 linked sialic acids preferentially. On comparison of the glycosylation sites, it was found that Alma-Ata/32/98 had lost putative glycosylation site residue 171, which also lies near an antigenic site. On comparison of the NA antigenic structure with sw/Jamesburg/1942, it was found that the Alma-Ata viruses had a single amino acid change at residue 371, which also lies near an antigenic site.
2. The Alma-Ata viruses do not seem to possess any obvious pathogenicity "signatures" observed in other pathogenic influenza viruses.
3. Going by the amino acids present in the NA active site and M2 membrane channel, it could be postulated that the Alma-Ata viruses would be a drug sensitive strain.

4. On nucleotide comparison of the viruses isolated in 1998 versus the viruses isolated in 1983 & 84, it was found that while there are significant number of mutations in the internal genes (particularly the polymerase proteins), the surface antigens HA & NA had surprisingly no/little mutations.
5. In addition to the mutations in HA and NA present near antibody binding sites, the internal genes had undergone several mutations in and contiguous to T-cell recognizing epitopes. Collectively, these mutations can be speculated to play a role in immune evasion in the host.
6. Excluding these mutations, the nucleotide sequences of the four isolates were seen to share a lot of features with classical swine viruses.
7. The presence of an antigenically unchanged classical swine virus in Kazakhstan 50 years after its initial circulation in New Jersey, USA (sw/Jamesburg/1942) could be possibly explained by a combination of laboratory leaks, persistence of H1N1 viruses in pigs, and low population density in Kazakhstan.
8. More solid influenza surveillance in central Asian countries is necessary to resolve the mystery of these anachronistic isolates.

REFERENCES

Akarsu H, Burmeister WP, Petosa C, Petit I, Muller CW, Ruigrok RW, et al. (2003) Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). *EMBO J.* **22**:4646–55.

Anchlan D, Ludwig S, Nymadawa P, Mendsaikhan J, and Scholtissek C. (1996). Previous H1N1 influenza A viruses circulating in the Mongolian population. *Arch. Virol.* **141:**1553-1569.

Aragon T, de la Luna S, Novoa I, Carrasco L, Ortin J, Nieto A. (2000). Eukaryotic translation initiation factor 4GI is a cellular target for NS1 protein, a translational activator of influenza virus. *Mol. Cell. Biol.* **20**:6259–6268.

Arnold K, Bordoli L, Kopp J, Schwede T (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**,195-201.

Assarsson E, Bui HH, Sidney J, Zhang Q, Glenn J, Oseroff C, Mbawuike IN, Alexander J, et al., (2008). Immunomic Analysis of the Repertoire of T-Cell Specificities for Influenza A Virus in Humans *J. Virol.* **82:** 12241-12251

Baum LG, Paulson JC. (1990). Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity. *Acta Histochem. Suppl.* **40**, 35-38.

Beeler E. (2009). Influenza in dogs and cats. *Vet Clin North Am Small Anim Pract.* **39(2):**251-264.

Belshe RB, Hall Smith M, Hall CB, Betts R, Hay AJ. (1988). Genetic basis of resistance to rimantadine emerging during treatment of influenza virus infection. *J.Virol* **62**: 1508-1512.

Berkhoff EG, Geelhoed-Mieras MM, Fouchier RA, Osterhaus AD, Rimmelzwaan GF. (2006). Assessment of the extent of variation in influenza A virus cytotoxic T-lymphocyte epitopes by using virus-specific CD8+ T-cell clones. *J Gen Virol.* **88:**530-535.

Bikour MH, EH Frost, S Deslandes, B Talbot, Y Elazhary. (1995a). Persistence of a 1930 swine influenza A (H1N1) virus in Quebec. *J. Gen. Virol.* **76:**2539-2547.

Bikour MH, Frost EH, Deslandes S, Talbot B, Weber JM, Elazhary Y. (1995b). Recent H3N2 swine influenza virus with haemagglutinin and nucleoprotein genes similar to 1975 human strains. *J. Gen. Virol.* **76:** 697-703

Blok J, Air GM. (1982). Variation in the membrane-insertion and “stalk” sequences in eight subtypes of influenza type A virus neuraminidase. *Biochemistry* **21:**4001–7.

Bornholdt ZA, Prasad BV (2006). X-ray structure of influenza virus NS1 effector domain. *Nat Struct Mol Biol* **13**, 559–560.

Boulo S, Akarsu H, Ruigrok RW, Baudin F. (2007). Nuclear traffic of influenza virus proteins and ribonu-cleoprotein complexes. *Virus Res.* **124**:12–21

Bui M, Myers JE, and Whittaker GR. (2002). Nucleo-cytoplasmic localization of influenza virus nucleoprotein depends on cell density and phosphorylation. *Virus Res.* **84**:37–44.

Brown IH, Ludwig S, Olsen CW, Hannoun C, Scholtissek C, Hinshaw VS, Harris PA, McCauley JW, Strong I, Alexander DJ. (1997). Antigenic and genetic analyses of H1N1 influenza A viruses from European pigs. *J. Gen. Virol.* **78:** 553-562.

Brown EG, Bailly JE. (1999). Genetic analysis of mouse-adapted influenza A virus identifies roles for the NA, PB1, and PB2 genes in virulence. *Virus Res*. **61:**63–76

Brown IH. (2000). The epidemiology and evolution of influenza viruses in pigs. Vet. Microbiol. **74:**29-46

Brown E, Liu H, Kit L, Baird S, Nesrallah M. (2001). Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. *Proc. Natl. Acad. Sci. USA* **98:**6883–6888.

Brown EG, Liu H, Kit LC, Baird S, Nesrallah M. (2001). Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. *Proc. Natl. Acad. Sci. USA* **98:**6883–88

Caton AJ, Brownlee GG, Yewdell JW & Gerhard W. (1983). The Antigenic structre of the Influenza virus A/PR/8/34 Hemagglutinin (H1 subtype). *Cell* **31**: 417–427.

Cerundolo V, Tse AG, Salter RD, Parham P, Townsend A. (1991). CD8 independence and specificity of cytotoxic T lymphocytes restricted by HLA-Aw68.1. *Proc Biol Sci.* **244(1310):**169-77.

Chen GW, Chang SC, Mok CK, Lo YL, et al., (2006). Genomic signatures of human versus avian influenza A viruses. *Emerg. Infect. Dis.* **12:**1353-1360

Chen W, Calvo PA, Malide D, et al. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med.* **7(12)**:1306-1312

Cheuk E, Chamberlain JW. (2005). Strong memory CD8(+) T cell responses against immunodominant and three new subdominant HLA-B27-restricted influenza A CTL epitopes following secondary infection of HLA-B27 transgenic mice. *Cell Immunol***.** **234**:110–23.

Chuvakova ZK, Fursova LM, Kim EV, Dotsenko GN, Isaeva EI (1984). Antigenic and biochemical analysis of the proteins of influenza virus A/USSR/Alma-Ata/13/81 with hemagglutinin H1. *Vopr Virusol.* **29(6):**667-72. (Article in Russian)

Chuvakova ZK, Rovnova ZI, Isaeva EI, Dem'ianenko IV, Isaeva ES. (1987). Virologic and seroepidemiologic analysis of the circulation of influenza virus A (H1N1), similar to serovariant A (Hsw1N1), in 1984-1985 in Alma-Ata. *Zh Mikrobiol Epidemiol Immunobiol* **10:** 30-36 (Article in Russian)

Claas ECJ, Kawaoka Y, De Jong JC, Masurel N, Webster RG. (1994). Infection of children with avian human reassortant influenza virus from pigs in Europe. *Virology* **204:** 453-457.

Compans RW, Content J, Duesberg PH. (1972). Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10:**795–800

Conenello GM, Zamarin D, Perrone LA, Tumpey T, Palese P. (2007). A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. *PLoS Pathog.* **3:**1414–21.

Connor RJ, Kawaoka Y, Webster RG, Paulson JC. (1994). Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205:**17-23.

Colman PM, Lawrence MC. (2003). The structural biology of type I viral membrane fusion. *Nat Rev Mol Cell Biol* **4**(4): 309-19.

Colman PM, Varghese JN, Laver WG (1983) Structure of the catalytic and antigenic sites in Influenza virus neuraminidase. *Nature (London)* **303**, 41–44.

Colman PM, Laver WG, Varghese JN, Baker AT, Tulloch PA, Air GM, Webster RG. (1987). The three-dimensional structure of a complex of influenza virus neuraminidase and an antibody. *Nature (London)* **326**:358-363.

Cros JF, Palese P. (2003). Trafficking of viral genomic RNA into and out of the nucleus: influenza, Thogoto and Borna disease viruses. *Virus Res*. **95:**3–12

Degtiarev AA, Furgal SM, Zakharov VE. (1987). Signs of the circulation of influenza A(HON1) virus among the general population and in military collectives. *Voen Med Zh*. **(3):**38-41 (article in Russian).

de Jong JC, van Nieuwstadt AP, Kimman TG, Loeffen WL, Bestebroer TM, et al., (1999). Antigenic drift in swine influenza H3 haemagglutinins with implications for vaccination policy. *Vaccine* **17:**1321-1328

de Jong, Bach VC, Phan TQ, Vo MH, Tran TT, Nguyen BH, et al. (2005). Fatal avian influenza A (H5N1) in a child presenting with diarrhea followed by coma. *N Engl J Med.* **352:**686–6891.

Dem'ianenko IV, Chuvakova ZK, Isaeva EI, Sklianskaia EI, Komarov IuS. (1987). Immunological analysis of the surface components of the influenza virus similar to the serovariant A (Hsw1N1) isolated in Alma-Ata 1984-1985. *Vopr Virusol.* **32(5):**533-8 (Article in Russian)

DiBrino M, Tsuchida T, Turner RV, Parker KC, Coligan JE, Biddison WE. (1993). HLA-A1 and HLA-A3 T cell epitopes derived from influenza virus proteins predicted from peptide binding motifs. *J Immunol* **151**: 5930-5935

Digard P, Elton D, Bishop K, Medcalf E, Weeds A, Pope B. (1999). Modulation of nuclear localization of the influenza virus nucleoprotein through interaction with actin filaments. *J. Virol*. **73**:2222–2231.

Easterday BC (1980). Animals in the influenza world. *Philos. Trans. R. Soc., London* **B288:** 433-437

Eisenlohr LC, Yewdell JW, Bennink JR. (1992). Flanking sequences influence the presentation of an endogenously synthesized peptide to cytotoxic T lymphocytes. *J. Exp. Med.* **175:**481-487

Elster C, Fourest E, Baudin F, Larsen K, Cusack S, Ruigrok RW. (1994). A small percentage of influenza virus M1 protein contains zinc but zinc does not influence in vitro M1-RNA interaction. *J. Gen. Virol*. **75:**37–42

Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, Stech J. (2005). The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. USA* **102:**18590–95

Gamblin SJ, Haire LF, Russell RJ et.al. (2004). The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**:1838–1842.

Gianfrani C, Oseroff C, Sidney J, Chesnut RW, Sette A. (2000). Human memory CTL response specific for influenza A virus is broad and multispecific. *Hum Immunol* **61(5):** 438–452.

Gibbs MJ, Armstrong JS, Gibbs AJ. (2001). Recombination in the hemagglutinin gene of the 1918 "Spanish Flu". *Science* **293:**1842

Govorkova EA, Gambaryan AS, Claas EC, Smirnov YA. (2000). Amino acid changes in the hemagglutinin and matrix proteins of influenza A (H2) viruses adapted to mice. *Acta Virol.* **44:**241–248

Gotch F, McMichael A, Smith G, Moss B. (1987). Identification of viral molecules recognized by influenza-specific human cytotoxic T lymphocytes. *J. Exp. Med*. **165:**408–416.

Gu J, Xie Z, Gao Z, et al. (2007). H5N1 infection of the respiratory tract and beyond: a molecular pathology study. *Lancet*. **370**(9593):1137-1145.

Guan Y, Shortridge KF, Krauss S, Li PH, Kawaoka Y, Webster RG. (1996). Emergence of avian H1N1 influenza viruses in pigs in China. *J. Virol.* **70:** 8041-8046.

Gupta R, Jung E, Brunak S. (2004) Prediction of N-glycosylation sites in human proteins. In preparation [http://www.cbs.dtu.dk/services/NetNGlyc/]

Hay AJ, Wolstenholm AJ, Skehel JJ, Smith MH. (1985). The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.* **4:** 3021–3024.

Hay AJ, V Gregory, AR Douglas, YP Lin. (2001). The evolution of human influenza viruses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356:**1861-1870

Hayden FG, Belshe RB, Clover RD, Hay AJ, Oakes MG, Soo W. (1989). Emergence and apparent transmission of rimantadine-resistant influenza A virus in families. *N. Engl. J. Med.* **321:**1696-1702

Hale BG, Randall RE, Ortin J, Jackson D. (2008). The multifunctional NS1 protein of influenza A viruses. *J. Gen. Virol.* **89:** 2359-2376

Heikkinen LS, Kazlauskas A, Melen K, Wagner R, Ziegler T, Julkunen I, Saksela K. (2008). Avian and 1918 Spanish influenza a virus NS1 proteins bind to Crk/CrkL Src homology 3 domains to activate host cell signaling. *J Biol Chem* **283**, 5719–5727.

Holsinger LJ, MA Shaughnessy, A Micko, LH Pinto, RA Lamb. (1995). Analysis of the posttranslational modifications of the influenza virus M2 protein. *J. Virol*. **69:**1219–1225.

Horimoto T, Kawaoka Y. (1994) Reverse genetics provides direct evidence for a correlation of hemagglutinin cleavability and virulence of an avian influenza A virus. *J Virol*. **68:**3120–8

Igarashi M, Ito K, Kida H, Takada A. (2008) Genetically destined potentials for N-linked glycosylation of influenza virus hemagglutinin.*Virology* **376(2)**:323-9

Ito T, Couceiro JS, Kelm S, Baum LG, Krauss S, Castrucci MR, Donatelli I, Kida H, Paulson JC, Webster RG, Kawaoka Y. (1998). Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J. Virol*. **72:**7367-7373

Ivanova NA, Smorodintsev AA, Grinbaum EB, Kramskaia TA, Murina EA. Isolation of influenza A viruses in 1981 in Leningrad with and antigenic formula (H0N1) related to strain A/PR8/34. *Vopr Virusol*. **29(1):**35-8. (Article in Russian)

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

Karasin AI, Olsen CW, and Anderson GA. (2000). Genetic characterization of an H1N2 influenza virus isolated from a pig in Indiana. *J. Clin. Microbiol.* **38:**2453-2456.

Katz J, Lu X, Tumpey T, Smith C, Shaw M, Subbarao K. (2000). Molecular correlates of influenza A H5N1 virus pathogenesis in mice. *J. Virol.* **74:**10807–10810.

Kida H, Ito T, Yasuda J, Shimizu Y, Itakura C, Shortridge KF, Kawaoka Y, Webster RG. (1994). Potential for transmission of avian influenza viruses to pigs. *J. Gen. Virol.* **75:** 2183- 2188.

Klimov AI, Rocha E, Hayden FG, Shult PA, Roumillat LF, and Cox NJ. (1995). Prolonged shedding of amantadine-resistant influenzae A viruses by immunodeficient patients: detection by polymerase chain reaction-restriction analysis. *J. Infect. Dis.* **172:**1352-1355

Koelle K, Cobey S, Grenfell B & Pascual M. (2006). Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314:** 1898–1903

Krasnitz M, Levine AJ, Rabadan R. (2008). Anomalies in the Influenza Virus Genome Database: New Biology or Laboratory Errors? *J. Virol.* **82:** 8947-8950

Landolt GA and Olsen CW. (2007). Up to new tricks – A review of cross-species transmission of influenza A viruses. *Animal Health Research Review.* **8**: 1-21

Lang AS, Kelly A, Runstadler JA. (2008). Prevalence and diversity of avian influenza viruses in environmental reservoirs. *J. Gen. Virol.* **89:** 509-519

Li KS, Guan Y, Wang J, Smith GJ, Xu KM, Duan L, et al (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*.**430**:209–13.

Li S, Schulman J, Itamura S, Palese P. (1993). Glycosylation of neuraminidase determines the neurovirulence of influenza A/WS/1933 virus. *J Virol.* **67**(11):6667–6673

Li Z, Jiang Y, Jiao P, Wang A, Zhao F, Tian G, Wang X, Yu K, Bu Z & Chen H. (2006). The NS1 gene contributes to the virulence of H5N1 avian influenza viruses. *J Virol* **80:** 11115–11123.

Liu T, and Ye Z. (2005). Attenuating mutations of the matrix gene of influenza A/WS/33 virus. *J. Virol*. **79:**1918-1923.

Luo G, Chung J, Palese P. (1993) Alterations of the stalk of the influenza virus neuraminidase: deletions and insertions. *Virus Res*. **29**:141–53.

Martin J, Wharton SA, Lin YP, Takemoto DK, Skehel JJ, Wiley DC, Steinhauer DA. (1998). Studies of the binding properties of influenza hemagglutinin receptor-site mutants. *Virology* **241:**101-111.

Matrosovich, MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS, Karlsson KA. (1997) Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* **233:**224-234

Matrosovich MN, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y. (2000). Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol*. **74:**8502-8512

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

Myers KP, Olsen CW, Gray GC. (2007). Cases of swine influenza in humans: a review of the literature. *Clin Infect Dis*. **44:**1084–1088.

Naffakh N, Tomoiu A, Rameix-Welti MA, van der Werf S. (2008) Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu Rev Microbiol.* **62:**403-24.

Nakajima K, Desselberger U, Palese P. (1978). Recent human influenza A (HIN1) viruses are closely related genetically to strains isolated in 1950. *Nature* 274: 334-339

Nelson MI, Holmes EC. (2007) The evolution of epidemic influenza. *Nat Rev Genet* **8:**196–205.

Neumann G, Hughes MT & Kawaoka Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J.* **19:** 6751–6758.

Noda T, Sagara H, Yen A, Takada A, Kida H et al., (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439:**490–92

Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, Finkelstein DB., Xu X, Wang J, Ma J & other authors. (2006). Large-scale sequence analysis of avian influenza isolates. *Science* **311:** 1576–1580

Olsen CW, Brown IH, Easterday BC, Van Reeth K. (2006). Swine influenza. In: Straw BE, Zimmerman JJ, D'Allaire S, Taylor DJ, editors. Diseases of swine. 9th ed. Ames (IA): Blackwell Publishing; 469–82.

Palese P, Shaw M. (2007). Orthomyxoviridae: the viruses and their replication. In Fields Virology, ed. D Knipe, P Howley, D Griffin, R Lamb, M Martin, et al., **2:**1647–89. Philadelphia/New York: Lippincott Williams &Wilkins. 5th ed

Pascua AM. (2006) Molecular epidemiology and surveillance of avian influenza in wild and domestic birds (MS thesis, University of Maryland, College park, USA).

Pensaert M, Ottis K, Vandeputte J, Kaplan MM, Bachmann PA. (1981). Evidence for the natural transmission of influenza Avirus from wild ducks to swine and its potential importance for man. *Bull. World Health Organisation* **59:** 75-78.

Petrov NA, Kiselev OI, Grinbaum EB, et al. (1990). The possibility of the circulation of vaccinal influenza A viral strains in the biosphere. *Dokl Akad Nauk SSSR*. **315(3)**:725-8. (article in Russian)

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE. (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25**(13):1605-1612. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

Reid AH, Fanning TG, Hultin JV, and Taubenberger JK. (1999). Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. USA* **96:**1651-1656.

Reid AH, Fanning TG, Janczewski TA, Taubenberger JK. (2000). Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. USA* **97:**6785-6790.

Reid AH, TG Fanning, TA Janczewski, RM Lourens, and JK Taubenberger. (2004). Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. *J. Virol*. **78:**12462-12470.

Rogers GN, Paulson JC. (1983). Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology*. **127**, 361-373

Rohrlich PS, Cardinaud S, Firat H, Lamari M, Briand P, Escriou N, and Lemonnier FA. (2003). HLA-B*0702 transgenic, H-2KbDb double-knockout mice: phenotypical and functional characterization in response to influenza virus. *Int. Immunol.* **15:**765-772

Scholtissek C, Burger H, Kistner O, Shortridge KF. (1985). The nucleoprotein as a possible major factor in determining host specfcity of influenza H3N2 viruses. *Virology* **147:** 287-294.

Schulze IT (1997) Effects of glycosylation on the properties and functions of the Influenza virus Hemagglutinin. *J. Infect. Dis.* **176**, Suppl. 1, S24–S28.

Sheerar MG, Easterday BC, Hinshaw VS. (1989). Antigenic conservation of H1N1 swine influenza viruses. *J. Gen. Virol.* **70** (Pt 12): 3297–3303.

Shinya K, Watanabe S, Ito T, Kasai N, Kawaoka Y. (2007). Adaptation of an H7N7 equine influenza A virus in mice. *J. Gen. Virol.* **88:**547–53

Shope RE. (1931). Swine influenza III. Filtration experiments and etiology. *J. Exp Med*. **54:** 373-385.

Shortridge KF, Webster RG, Buttereield WK & Campbell CH (1977). Persistence of Hong Kong influenza virus variants in pigs. *Science* **196:** 1454-1455.

Smeenk CA, and EG Brown. (1994). The influenza virus variant A/FM/1/47 MA possesses single amino acid replacements in the hemagglutinin, controlling virulence, and in the matrix protein, controlling virulence as well as growth. *J. Virol*. **68:**530–534

Sorrell EM (2003). Adaptation of A/mallard/Potsdam/178-4/83 (H2N2) in Japanese quail leads to replication and transmission in chickens (MS thesis, University of Maryland, College park, USA).

Steinhauer, DA. (1999). Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* **258:**1-20

Stevens J, Blixt O, Glaser L, Taubenberger JK, Palese P, et al. (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol* **355:**1143–1155.

Subbarao EK, London W, Murphy BR. (1993). A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**:1761–64.

Subbarao K, Swayne DE, Olsen CW. (2006). Epidemiology and control of human and animal influenza. In: Kawaoka Y, editor. Influenza Virology: Current Topics. Norfolk: Horizon Scientific Press; 229-280.

Sugita S, Yoshioka Y, Itamura S, Kanegae Y, Oguchi K, Gojobori T, Nerome K & Oya A. (1991). Molecular evolution of hemagglutinin genes of H1N1 swine and human influenza A viruses. *J. Mol. Evol*. **32:**16–23.

Sugrue RJ, Belshe RB, Hay AJ. (1990). Palmitoylation of the influenza A virus M2 protein. *Virology* **179:**51–56.

Sui J, William CH, Sandra P, Ge W, Daniel A, Li-mei C, Eugenio S, Boguslaw S, Greg C, Maryam A, Hongquan W, Akikazu M, Anuradha Y, Thomas H, Nancy JC, Laurie AB, Ruben OD, Robert CL & Wayne AM. (2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Str. Mol. Bio*. **16:** 265-273

Stevens J, Corper AL, Basler CF, Taubenberger JK, Palese P, Wilson IA**.** (2004). Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303:**1866-1870

Takahashi T, Suzuki T, Hidari KI-PJ, Miyamoto D, Suzuki Y. (2003). A molecular mechanism for the low-pH stability of sialidase activity of influenza A virus N2 neuraminidases *FEBS Lett.* **543:** 71—75

Tamura K, Dudley J, Nei M, Kumar S (*2007*) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24:**1596-1599.

Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. (2005). Characterization of the 1918 influenza virus polymerase genes. *Nature* **437:**889–93

Voeten JTM, Bestebroer TM, Nieuwkoop NJ, Fouchier RAM, Osterhaus ADME, and Rimmelzwaan GF. (2000). Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J. Virol.* **74:**6800–6807.

Webster RG (1993). Are equine 1 influenza viruses still present in horses? *Equine Veterinary Journal* **25:** 537–538.

Webster RG, Bean WJ, Gorman OT, Chambers TM & Kawaoka Y (1992). Evolution and ecology of Influenza A viruses. *Microbiological Reviews* **56:** 152-179.

Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* **333***:*426–431

Wilson AI, Cox NJ (1990). Structural basis of immune recognition of influenza virus hemagglutinin. *Annu Rev Immunol.* **8**:737–71.

Wolf YI, Viboud C, Holmes EC, Koonin EV & Lipman DJ. (2006). Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct.* **1**: 34.

Wright PF, Neumann G, Kawaoka Y. (2006a). Orthomyxoviruses. In: Fields BN, Knipe DN, Howley PH, editors. Fields virology, 6th ed. Philadelphia: Lippincott–Raven.

Wright PF and Webster RG (2006b). Orthomyxoviruses. Philadelphia, PA: Lippincott Williams and Wilkins

Wu W, Air RG. (2004). Binding of influenza viruses to sialic acids: reassortant viruses with A/NWS/33 hemagglutinin bind to a2,8-linked sialic acid. *Virology* **325;**2:340-50

Ye Z, Liu T, Offringa DP, McInnis J, and Levandowski RA. (1999). Association of influenza virus matrix protein with ribonucleoproteins. *J. Virol.* **73:**7467-7473

Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L , Yoon KJ, Krauss S, and Webster RG. (1999). Genetic reassortment of avian, swine, and human influenza A viruses in American pigs. *J. Virol.* **73:**8851-8856

Zhu Q, Yang H, Chen W, Cao W, Zhong G, Jiao P, Deng G,Yu K. et al., (2008). A naturally occurring deletion in its NS gene contributes to the attenuation of an H5N1 swine influenza virus in chickens. *J Virol* **82,** 220–228.

Zhang G, Shoham D, Gilichinsky D, Davydov S, Castello JD & Rogers SO (2006). Evidence of influenza A virus RNA in Siberian lake ice. *J Virol* **80**: 12229–12235