Role of Positive Selection Pressure on the Evolution of H5N1 Hemagglutinin

Venkata R.S.K. Duvvuri[1]*, Bhargavi Duvvuri[2], Wilfred R. Cuff[3], Gillian E. Wu[2], and Jianhong Wu[1,4]

[1] *Center for Disease Modeling, York Institute of Health Research, York University, Toronto, Ontario M3J 1P3, Canada;* [2] *School of Kinesiology and Health Sciences, York University, Toronto, Ontario M3J 1P3, Canada;* [3] *National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba R3E 3R2, Canada;* [4] *Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada.*

*Corresponding author. E-mail: dvenkata@yorku.ca
DOI: 10.1016/S1672-0229(08)60032-7

The surface glycoprotein hemagglutinin (HA) helps the influenza A virus to evade the host immune system by antigenic variation and is a major driving force for viral evolution. In this study, the selection pressure on HA of H5N1 influenza A virus was analyzed using bioinformatics algorithms. Most of the identified positive selection (PS) sites were found to be within or adjacent to epitope sites. Some of the identified PS sites are consistent with previous experimental studies, providing further support to the biological significance of our findings. The highest frequency of PS sites was observed in recent strains isolated during 2005–2007. Phylogenetic analysis was also conducted on HA sequences from various hosts. Viral drift is almost similar in both avian and human species with a progressive trend over the years. Our study reports new mutations in functional regions of HA that might provide markers for vaccine design or can be used to predict isolates of pandemic potential.

Key words: positive selection, H5N1, hemagglutinin, epitope, bioinformatics

Introduction

Influenza A viruses are immuno-epidemiologically significant from an evolutionary standpoint (*1*). They belong to the Orthomyxoviridae family and have negative-sense single-stranded RNA with a genome composed of eight different gene segments (*2*). Antigenic diversity of influenza A viruses occurs primarily at two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). HA is a homotrimer; each monomer is synthesized as a single polypeptide (HA0) that is cleaved by host proteases into two subunits (HA1 and HA2). HA1 mediates the initial contact with the cell membrane and HA2 is responsible for membrane fusion. The HA gene, as a target of neutralizing antibodies, is a classic example of an antigenically drifting protein (*3*). This gene accumulates an exceptional number of point mutations in the epitope regions or antibody combining regions (*4*). Antigenic drift of these recognized epitope regions presents a major obstacle to vaccine development since an effective vaccination is possible only when the epidemic strain matches with the vaccine strain (*5*). HA recognizes the host cell receptors containing glycans with terminal sialic acids, and the precise linkage between them determines the species specificity. Nonsynonymous mutations in HA can alter the recognition from α-2,3-linked sialic acid to α-2,6-linked sialic acid, which can permit the binding of avian viruses to human receptors (*6–8*). This conversion in receptor-binding from avian to human could be a major cause of pandemics (*8*). Hence understanding the evolution of HA is of great relevance for public health.

In the present study, we employed QUASI (*9*), a selection mapping algorithm, for the identification of positive selection (PS) sites on HA sequences of influenza A virus subtype H5N1. The cornerstone of selection mapping is the testing of each observed replacement mutation at each codon to identify those particular replacement mutations that are overabundant relative to silent mutations at that codon (*9*). We mainly focused on the mutational events of HA and addressed the following questions: How do mutations change over time in HA under the influence

of PS pressure and what is their functional impact? What are the probable B-cell and T-cell epitopes? How do known epitopes correspond with positively selected substitutions? What are the phylogenetic relationships within the H5N1 (and to a broader extent H5N2, H5N3, H5N8 and H5N9) isolates of avian, human and other animals?

Results and Discussion

Identification and analysis of PS sites

Distribution of PS sites

The selection mapping algorithm QUASI (*9*) was used to analyze the avian and human HA sequences isolated during 1996–2007 to determine positively selected sites. As a result, 33 and 34 PS sites were found throughout avian and human HA sequences, respectively (Figure 1; Tables 1 and 2). Some of the identified PS sites were reported in earlier studies (*10–12*).

The HA gene has five types of epitope sites labeled A to E (Figure 1). The epitope sites seem to be under high PS pressure, with more replacement mutations than silent mutations compared with the rest of the HA1 and the HA2 region. Changes in 41 codons within or near the epitope sites A and B of H3 HA have been shown to be associated with antigenic drift (*13*). In the current study, several changes were also observed in the epitope sites C, D and E besides A and B of H5N1 HA. The predicted PS sites were found to be mostly within or adjacent to epitope sites while some in receptor-binding and glycosylation sites. This noteworthy overlap of newly predicted PS sites within the functional regions of HA is a causal explanation of our findings and further emphasizes the antigenic plasticity of H5N1.

Frequency of PS sites

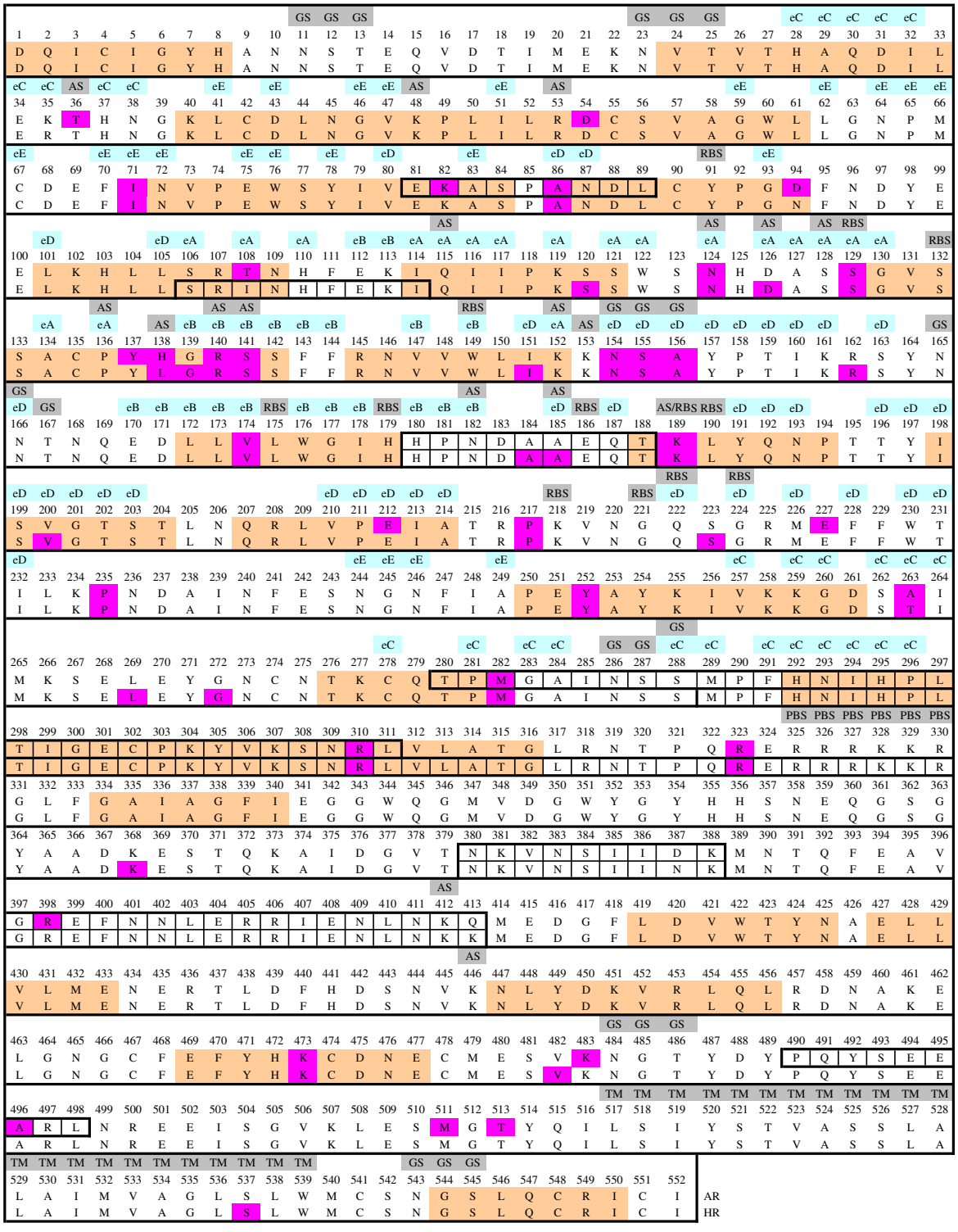
The occuring frequency of PS sites in HA increased from 1996 to 2007, which is possibly due to accumulation of replacement mutations over time (Figure 2). The increase of PS pressure in recent strains corroborates the epidemiological observation of the spread of the virus from birds to humans and other species, signifying the effect of PS towards the diversity, adaptability and lethality of the virus. Specifically, an increasing number of amino acids under PS was observed in 2001, and some of the mutations persisted for several years (Tables 1 and 2). The biological relevance of these respective mutations has to be further studied. The H5N1 isolates during 2004–2006 have several amino acid changes that alter their antigenicity, compared with those in previous years (Tables 1 and 2). This observation is in accordance with prior genetic and antigenic studies on H5N1 isolates (*14*). Some of the identified host-specific PS sites might be associated with viral adaptability to the selected hosts.

Mutational changes of PS sites

We also analyzed the mutational changes of these identified PS sites and had some interesting observations: (1) Amino acid substitutions 138L→Q, 140R→K and 156A→T continued in these years. The amino acids at sites 138 and 140 are involved in the carbohydrate specificity of HA molecule towards host receptor (*15*). Substitution 156A→T is of particular interest, since the mutation introduces a new glycosylation site at N154, consistent with a strategy commonly used by influenza virus to mask or unmask antigenic sites from the immune system. This change was found to be associated with viral adaptation to terrestrial poultry that increased its virulence for these birds (*16–18*). (2) Substitution 129S→L occurred in 2003 and continued thereafter. Serine at site 129 is a receptor-binding site for α-2,3-linked sialic acids. Structural study on the H5N1 HA has reported that this change from serine to leucine facilitates access to α-2,6-linked sialic acids (*19*). (3) An amino acid change at receptor-binding site 189K→R observed in 2003 was reported as an escape mutant (*20*) and was further clearly shown to be significantly antigenical (*21*). (4) Substitution 223S→N appeared in 2003 from human host was shown through genetic and structural studies to increase the receptor-binding capacity of HA to human cell receptor (*22*). (5) Substitution 94D→N observed in 2004–2005 from avian host was found to alter the interaction of Y91, H179 and L190 with sialosides, resulting in increased affinity of HA for human receptors (*14*). In addition, the significance of non-epitope PS sites with respect to antigenicity has to be investigated structurally, since some of the replacement mutations might influence the key regions of HA, although to a lesser extent.

Amino acid property changes of PS sites

Some of the PS sites were observed in the solvent-exposed region of HA. This is expected because the



Explanation of codes used in the Figure 1

GS	Glycosylation Site
AS	Antigenic Site
RBS	Receptor-Binding Site
PBS	Polybasic Cleavage Site
TM	Transmembrane Site
AR	Avian Root Sequence A/Goose/Guangdong/1/96
HR	Human Root Sequence A/Hong Kong/156/97

	Epitopic regions identified by RANKPEP
	Antigenic peptides predicted by Kolaskar and Tongaonkar method
	Positive selection sites predicted by QUASI algorithm
eA	Epitope A
eB	Epitope B
eC	Epitope C
eD	Epitope D
eE	Epitope E

Figure 1 Comparative representation of the distribution of PS sites in avian and human H5N1 HA sequences.

Table 1 QUASI identified positive selection sites in H5N1 HA sequences isolated from avian hosts

Year	Positive selection sites	
	HA1	HA2
1997	108 (I, T); 138 (L, H); 263 (T, A)	–
2000	140 (R, K); 217 (S, P)	–
2001	108 (I, T); 138 (L, H, Q); 156 (A, T); 174 (V, I); 212 (K, E); 217 (S, P)	398 (R, K); 513 (T, I)
2002	108 (I, T); 156 (A, T)	483 (R, K); 511 (I, M)
2003	129 (L, S); 138 (L, Q); 140 (K, R); 155 (S, N); 156 (A, T); 189 (R, K)	–
2004	86 (V, A); 94 (D, N); 108 (I, T); 124 (D, N); 129 (L, S); 138 (L, Q); 140 (K, R); 156 (A, T); 189 (R, K); 212 (R, K); 227 (E, D); 263 (T, A)	513 (T, I)
2005	71 (I, L); 82 (A, D); 86 (V, A); 94 (D, N); 108 (I, T); 124 (D, N); 129 (L, S); 140 (K, R); 141 (S, P); 154 (N, D); 155 (S, N, D); 156 (A, T); 189 (R, K); 217 (S, P); 227 (E, D); 252 (Y, N); 263 (T, A); 323 (R, G)	473 (K, R); 511 (I, M); 513 (T, I)
2006	36 (T, A); 71 (I, L); 129 (L, S); 141 (S, P); 154 (N, D); 155 (S, N, D); 156 (A, T); 252 (Y, N); 263 (T, A); 282 (I, M); 310 (R, K); 323 (G, R)	473 (K, R); 496 (A, S)
2007	36 (T, A); 54 (D, N); 71 (I, L); 129 (L, S); 137 (Y, H); 155 (S, N, D); 235 (P, S); 323 (G, R)	473 (K, R); 511 (I, M)

Table 2 QUASI identified positive selection sites in H5N1 HA sequences isolated from human hosts

Year	Positive selection sites	
	HA1	HA2
1997	140 (R, K); 156 (A, S); 156 (A, T)	368 (Q, K)
2003	120 (S, N); 124 (N, S); 126 (D, E); 129 (L, S); 138 (L, Q); 140 (R, K); 155 (S, N); 156 (A, T); 189 (K, R); 217 (P, S); 223 (S, N)	–
2004	124 (N, S); 126 (D, E); 129 (L, S); 138 (L, Q); 139 (G, E); 140 (R, K); 156 (A, T); 217 (P, S)	–
2005	124 (N, D); 126 (D, E); 129 (L, S); 140 (R, S, K); 141 (S, P); 156 (A, T); 162 (R, K); 217 (P, S)	–
2006	71 (L, I); 86 (A, T); 124 (N, S); 126 (D, E); 129 (L, S); 140 (R, S, T); 141 (S, P); 154 (N, D); 156 (A, T); 162 (R, K); 174 (V, I); 189 (R, K); 200 (V, I); 217 (P, S); 235 (P, S); 252 (Y, N); 263 (T, A); 269 (L, V); 282 (M, I); 310 (R, K); 323 (R, G)	473 (R, K); 482 (V, I); 537 (S, T)
2007	86 (A, T); 124 (N, S); 126 (D, E); 129 (L, S); 141 (S, P); 151 (I, T); 154 (N, D); 156 (A, T); 162 (R, K); 184 (A, E); 185 (A, E); 200 (V, I); 263 (T, A); 272 (G, S); 282 (I, M)	473 (R, K); 482 (V, I); 537 (S, F)

solvent accessible areas of HA provide essential adaptations required for the virus. Some of the epidemic strains with the largest changes in solvent-exposed regions have been reported to be lethal (*4*). Changes of amino acid property at PS sites in avian HA occurred in the following decreasing order: non-polar to uncharged polar (NP→UP); UP→NP; acidic to basic (A→B); NP→NP; UP→A. Whereas in humans the majority of the alterations at PS sites resulted in NP→NP, followed by UP→UP and finally NP→UP. The distinct changes in amino acid properties of avian H5N1 HA might aid in its adaptability to diverse species. To verify this observation, phylogenetic analysis was conducted.

Phylogenetic analysis

A phylogenetic tree was constructed based on HA amino acid sequences from various hosts and clustered into two specific groups (Figure 3A). Cluster A has H5N1 strains from avian along with human, swine and other animals. Cluster B contains H5N1 avian strains

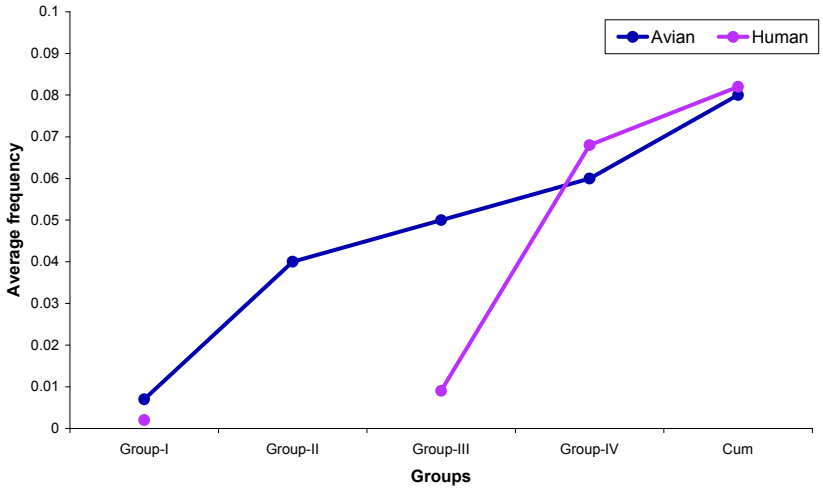


Figure 2 Average frequencies of positively selected amino acids of avian and human isolates of H5N1 virus from 1996 to 2007. Group I: Avian (1996, 1997) and Human (1997, 1998); Group II: Avian (1999, 2000, 2001); Group III: Avian (2002, 2003, 2004) and Human (2003, 2004); Group IV: Avian (2005, 2006, 2007) and Human (2005, 2006, 2007); Cum: Cumulative avian and human groups.

of 1991, 2001 and 2005 as well as H5N3 and H5N9 avian strains. Interestingly, the H5N1 avian strains of Cluster B have lost the consensus polybasic cleavage site from ERRRKKR to E– – ––TR, which is responsible for the high pathogenicity of H5N1. The phylogenetic tree constructed using HA nucleotide sequences also clustered them into two specific groups (Figure 3B). Cluster A has H5N1 strains from swine and other animals while Cluster B contains human and avian strains. The clustering of the nucleotide sequences is interesting because we observed that the nucleotide sequences of Cluster A have more silent mutations than those of Cluster B. Maximum wobble-based replacements were observed at the segment packing region of HA (positions 540–565) (*23*) from swine and other animals when compared with the avian and human strains and that might influence the viral adaptability at host level. It would be worthwhile to study the biological role of observed silent mutations, since recent evidence suggests that synonymous mutations may not be selectively neutral (*24*).

The differences in clustering of sequences with respect to nucleotide and amino acid level were expected since given a sufficient evolutionary time, conservation of a protein sequence will not necessarily be matched at the nucleotide level for those amino acids with multiple codons (*23*). The mean genetic distances between avian and human H5N1 HA sequences from 1996 to 2007 were determined and presented in Figure 4. This figure clearly elucidates that the viral drift is almost similar in both avian and human species with a progressive trend over the years.

Epitope analysis

A recent report found significant knowledge gaps in the studies concerned with the identification of influenza A epitopes and suggested the need for a more systematic and comprehensive collection of influenza immune epitopes (*25*). With this in mind, we conducted an epitope analysis on both avian and human H5N1 HA sequences to identify probable major histocompatibility complex (MHC) class I restricted T-cell epitopes (CTL) and antibody (Ab) epitopes. The peptide binders for HLA-B*2705 and HLA-B*3501 class I MHC alleles of those H5N1 HA sequences during 1996–2007 were predicted (Table S1). Epitope analysis revealed conserved stretches in the HA region, comprising important sites like receptor-binding, glycosylation, polybasic cleavage and other antigenic sites that might involve in the determination of host specificity of H5N1. In addition to CTL epitopes, antigenic peptide regions of HA that might elicit antibody response (Ig) were predicted using the Kolaskar and Tongaonkar method (*26*) (Figure 1; Table S2).

To assess selection intensity, d_N/d_S ratio was determined for all the predicted epitope regions, where d_N and d_S stand for non-synonymous and synonymous substitution rate, respectively. A d_N/d_S ratio >1 would be interpreted as evidence of positive selection (*13*). Results showed that the d_N/d_S ratios for these

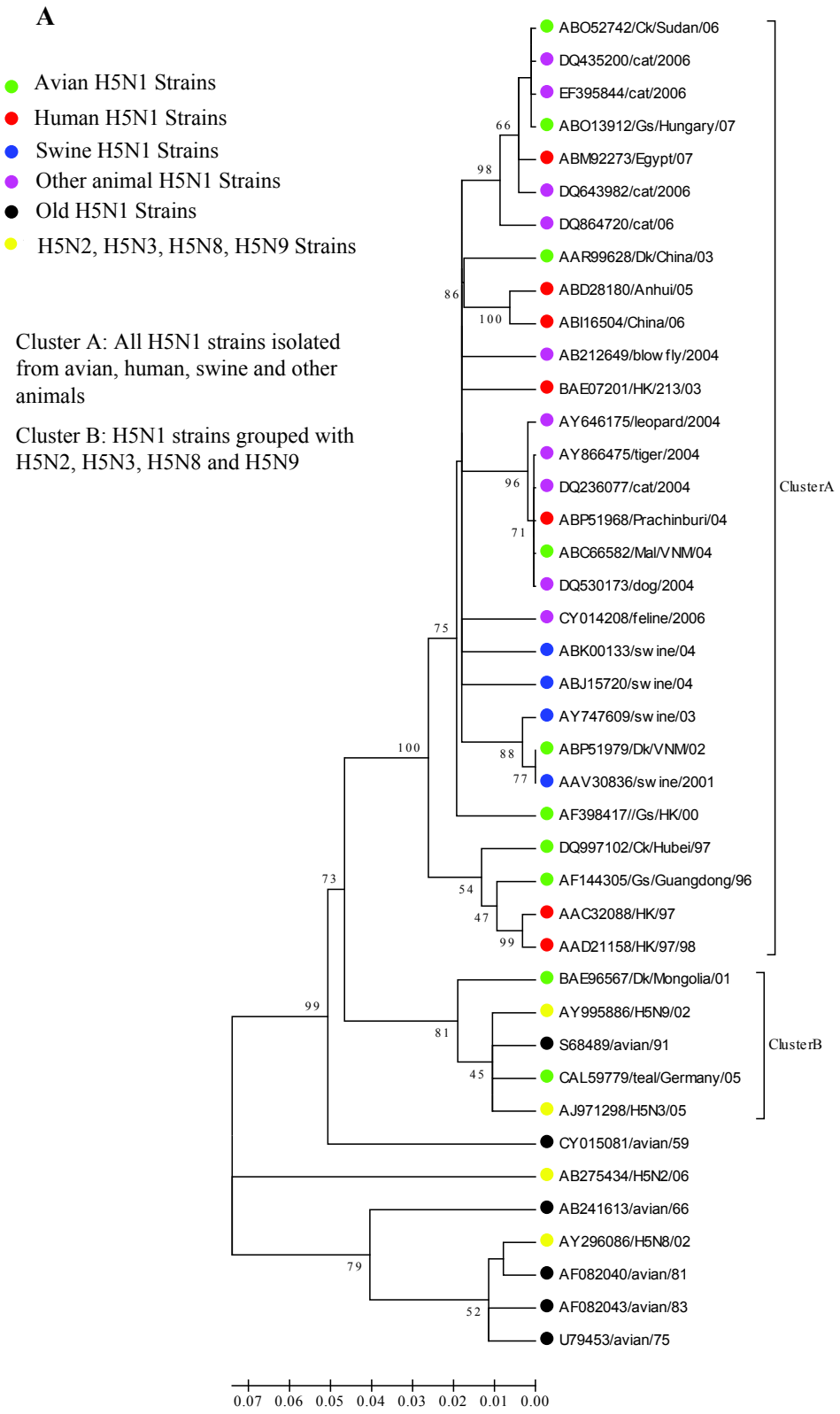


Figure 3 **A**. Protein-based evolutionary relationship of H5N1 (avian, human, swine and other animals), H5N2, H5N3, H5N8 and H5N9 (avian) HA genes using neighbor-joining tree construction. Bootstrap values of 2,500 replications are shown on major branches.

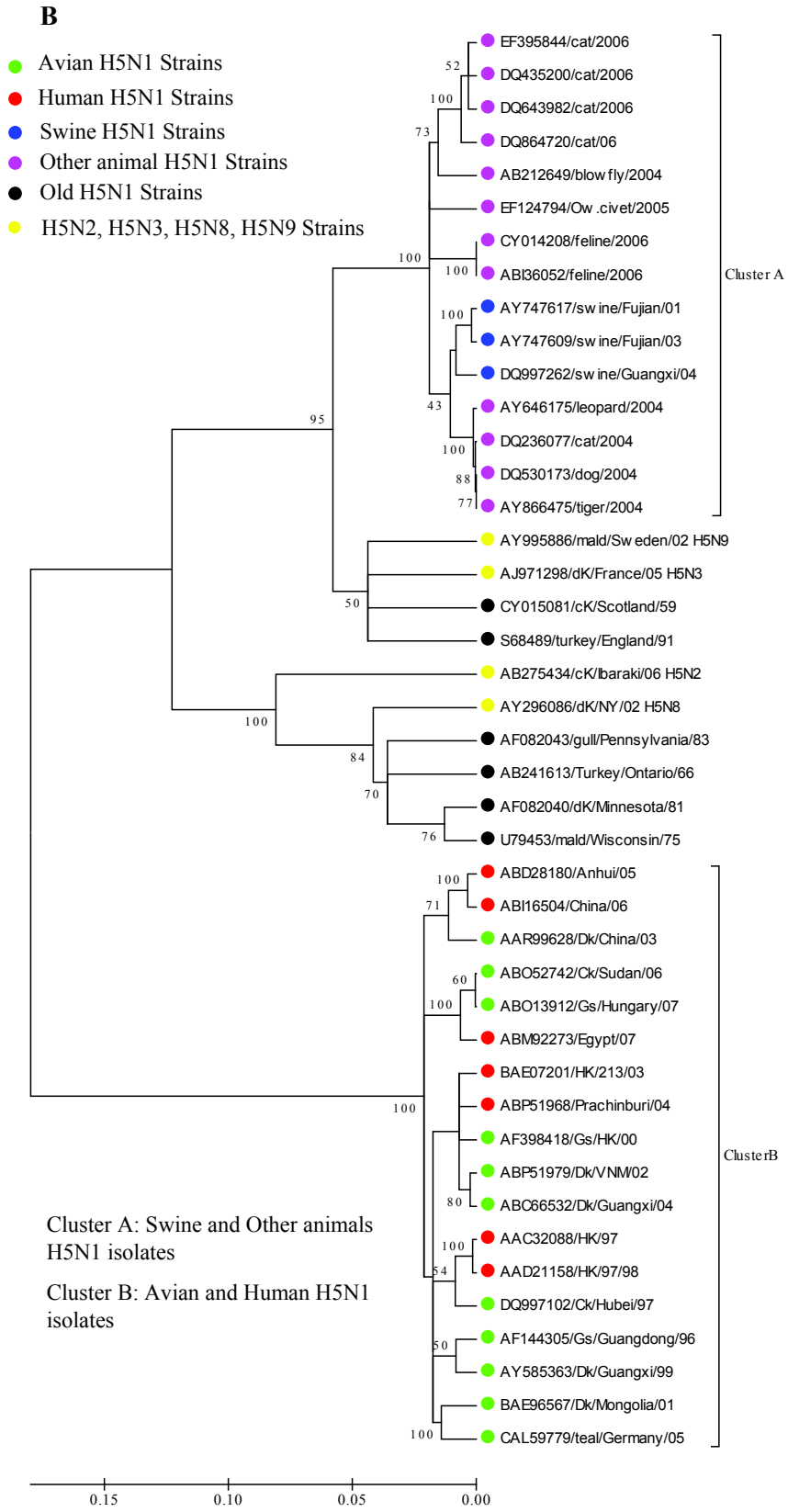


Figure 3 **B**. Nucleotide-based evolutionary relationship of H5N1 (avian, human, swine and other animals), H5N2, H5N3, H5N8 and H5N9 (avian) HA genes using neighbor-joining tree construction. Bootstrap values of 2,500 replications are shown on major branches.

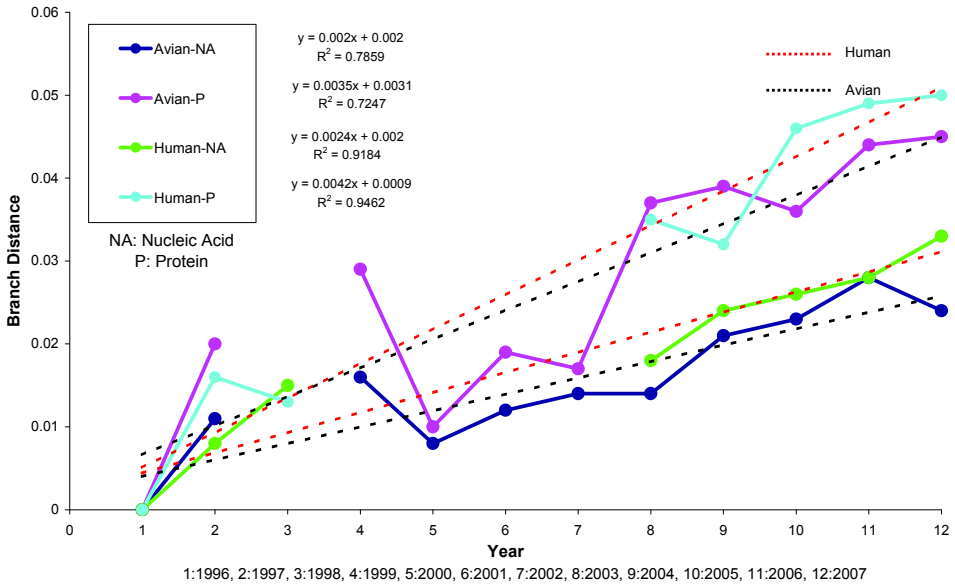


Figure 4 Evolutionary rates for nucleic acids and proteins of H5N1 HA isolated from avian and human hosts. Evolutionary rate is estimated by regression of year of isolation against branch distance from the common origin node of the nucleotide and amino acid.

epitope regions were mostly <1, except for the CTL epitope SRLVLATGL (positions 309–317) from human and the CTL epitope GREFNNLER (positions 397–405) conserved in both avian and human (Table S1). The low evolution rate of these epitope regions might be due to functional constraints. Most of the epitope regions seem to be conserved in both avian and human species, suggesting that there could be possibility of inter-strain cross-reactivity for these epitope regions, although species-specific epitopes exist. Only a few PS sites were identified in these epitope regions (Figure 1), signifying locally focused selection. Therefore, responses to conserved epitopes might be most useful with respect to broad spectrum vaccine development, whereas species-specific epitopes might be most useful for monitoring purposes.

Conclusion

To summarize, we predicted and analyzed PS sites and epitope sites of the HA sequences of H5N1 influenza A virus, with the object of clarifying the role of PS pressure on the evolution of H5N1 hemagglutinin. Analysis of the distribution, frequency and mutation of the PS sites led to the observation that known and newly identified PS sites are mostly distributed within or adjacent to epitope sites, thus we concluded that the epitope sites are under high PS pressure. To verify this observation, phylogenetic analysis was conducted from HA sequences of various hosts and a number of suggestive relationships were observed. We also noted that the frequency of PS sites increased from 1996 to 2007. Recent (2005–2007) circulating strains had greater PS frequency compared with the preceding strains and this might implicate the role of these PS sites in increased diversity. These findings are relevant in vaccine development since effective vaccination is possible only if the vaccine strain is antigenically similar to the current circulating strains. In addition, some epitope regions were found to have low rates of evolution and are conserved in both avian and human species, suggesting that there could be possibility of inter-strain cross-reactivity for these epitope regions. Epitope analysis showed that the CTL and B-cell epitopes in conserved stretches of HA region are rarely targeted by PS sites and hence may be good targets for effective vaccine development for influenza A virus. Nevertheless, this observation requires experimental verification. The epitopes with PS sites also draw attention as possible emerging escape variants.

Materials and Methods

HA sequences

The HA nucleotide sequences and corresponding protein sequences of H5N1 influenza A virus from avian, human, swine and other animals (feline, dog, civet and blowfly) isolated during 1996–2007

were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/) and the Influenza Sequence Database (http://www.flu.lanl.gov). An avian isolate A/Goose/Guangdong/1/96 (Accession No. AF144305) and a human isolate A/HongKong/156/97 (Accession No. AF036356) were used as root sequences for all the analyses. The human HA sequences used in the analysis include those isolated from Asian patients (*19*). After removing redundant sequences, 85 unique HA sequences from avian and human were used in the analysis (Table S3).

Identification of PS sites using QUASI algorithm

Sequence data were organized into four avian groups [A-I (1996, 1997), A-II (1999, 2000, 2001), A-III (2002, 2003, 2004) and A-IV (2005, 2006, 2007)] and three human groups [H-I (1997, 1998), H-II (2003, 2004) and H-III (2005, 2006, 2007)] to investigate the PS pressure on HA region over time. The QUASI algorithm (*9*) was used to identify the positively selected variants of each data group.

Sequence alignment and analysis

BioEdit, a biological sequence alignment editor (*27*), was used for sequence analysis (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Multiple sequence alignments were performed using ClustalW (*28*) with default parameters offered by BioEdit as an external program.

Phylogenetic analysis and tree building

Phylogenetic analysis was performed by MEGA software version 4.0 (http://www.megasoftware.net/) (*29*) and phylogenetic trees were generated employing the neighbor-joining method. Tree topology was evaluated by bootstrap analysis with 2,500 replicates for estimation of branch length errors and confidence intervals (*30*).

Estimating synonymous and non-synonymous nucleotide substitutions

SNAP program (http://www.hiv.lanl.gov) was used to calculate variance and covariance values of the number of synonymous (d_S) and non-synonymous (d_N) substitutions per site based on a set of codon-aligned nucleotide sequences (*13*).

Epitope analysis

The Kolaskar and Tongaonkar method (http://bio.dfci.harvard.edu/tools/antigenic.html) (*26*) was applied to predict B-cell antigenic peptides based on the occurrence of amino acid residues in experimentally determined epitopes. The reported accuracy of this method is about 75% (*26*). RANKPEP program (http://bio.dfci.harvard.edu/tools/rankpep.html), which uses position-specific scoring matrices or profiles from sets of aligned known peptides to bind to a given MHC molecule (*31*), was employed as the predictor of MHC peptide binding. Epitope analysis was conducted to identify CTL epitopes specific for HLA-B*2705 and HLA-B*3501, which were identified as preferred alleles in the influenza A virus specific CTL response.

Acknowledgements

We thank Mr. Jeffrey J. Stewart, BioGenetic Ventures, Seattle, USA for the QUASI analysis. This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Research Chairs Program (CRC) and the Mathematics for Information Technology and Complex Systems (MITACS). VD and JW also would like to acknowledge the support from the Canadian Network of Centers of Excellence, MITACS.

Authors' contributions

VD collected the datasets. VD and BD conceived the idea, conducted data analyses, and prepared the manuscript. WC suggested suitable methodology, cowrote and revised the manuscript. GW and JW supervised the project and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Earn, D.J.D., *et al.* 2002. Ecology and evolution of the flu. *Trends Ecol. Evol.* 17: 334-340.

2. Voyles, B.A., *et al.* 2002. *The Biology of Viruses* (second editon), pp.147-149, 338-341. McGraw-Hill, New York, USA.
3. Webster, R.G., *et al.* 1982. Molecular mechanisms of variation in influenza viruses. *Nature* 296: 115-121.
4. Wilson, I.A. and Cox, N.J. 1990. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* 8: 737-771.
5. Stöhr, K. 2002. Influenza—WHO cares. *Lancet Infect. Dis.* 2: 517.
6. Rogers, G.N. and Paulson, J.C. 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.
7. Auewarakul, P., *et al.* 2007. An avian influenza H5N1 virus that binds to a human-type receptor. *J. Virol.* 81: 9950-9955.
8. Matrosovich, M., *et al.* 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.
9. Stewart, J.J., *et al.* 2001. An algorithm for mapping positively selected members of quasispecies-type viruses. *BMC Bioinformatics* 2: 1.
10. Campitelli, L., *et al.* 2006. H5N1 influenza virus evolution: a comparison of different epidemics in birds and humans (1997–2004). *J. Gen. Virol.* 87: 955-960.
11. Smith, G.J., *et al.* 2006. Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology* 350: 258-268.
12. Kongchanagul, A., *et al.* 2008. Positive selection at the receptor-binding site of haemagglutinin H5 in viral sequences derived from human tissues. *J. Gen. Virol.* 89: 1805-1810.
13. Korber, B. 2000. HIV sequence signatures and similarities. In *Computational and Evolutionary Analysis of HIV Molecular Sequences* (eds. Rodrigo, A.G. and Learn, G.H.), pp.55-72. Kluwer Academic Publishers, Dordrecht, the Netherlands.
14. World Health Organization. 2005. Evolution of H5N1 avian influenza viruses in Asia. *Emerg. Infect. Dis.* 11: 1515-1521.
15. Ilyushina, N.A., *et al.* 2004. Receptor specificity of H5 influenza virus escape mutants. *Virus Res.* 100: 237-241.
16. Baigent, S.J. and McCauley, J.W. 2001. Glycosylation of haemagglutinin and stalk-length of neuraminidase combine to regulate the growth of avian influenza viruses in tissue culture. *Virus Res.* 79: 177-185.
17. Mitnaul, L.J., *et al.* 2000. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* 74: 6015-6020.
18. Perdue, M.L. and Suarez, D.L. 2000. Structural features of the avian influenza virus hemagglutinin that influence virulence. *Vet. Microbiol.* 74: 77-86.
19. Li, G., *et al.* 2006. Sequence and epitope analysis of surface proteins of avian influenza H5N1 viruses from Asian patients. *Chin. Sci. Bull.* 51: 2472-2481.
20. Philpott, M., *et al.* 1990. Hemagglutinin mutations related to attenuation and altered cell tropism of a virulent avian influenza A virus. *J. Virol.* 64: 2941-2947.
21. Kaverin, N.V., *et al.* 2007. Epitope mapping of the hemagglutinin molecule of a highly pathogenic H5N1 influenza virus by using monoclonal antibodies. *J. Virol.* 81: 12911-12917.
22. Yamada, S., *et al.* 2006. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* 444: 378-382.
23. Gog, J.R., *et al.* 2007. Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Res.* 35: 1897-1907.
24. Sauna, Z.E., *et al.* 2007. The sounds of silence: synonymous mutations affect function. *Pharmacogenomics* 8: 527-532.
25. Bui, H.H., *et al.* 2007. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc. Natl. Acad. Sci. USA* 104: 246-251.
26. Kolaskar, A.S. and Tongaonkar, P.C. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276: 172-174.
27. Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.
28. Thompson, J.D., *et al.* 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
29. Tamura, K., *et al.* 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
30. Dopazo, J. 1994. Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *J. Mol. Evol.* 38: 300-304.
31. Reche, P.A., *et al.* 2004. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 56: 405-419.

Supporting Online Material

Tables S1–S3

DOI: 10.1016/S1672-0229(08)60032-7