Natural Selection on the Influenza Virus Genome

Yoshiyuki Suzuki

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima-shi, Shizuoka-ken, Japan

Influenza viruses are the etiological agents of influenza. Although vaccines and drugs are available for the prophylaxis and treatment of influenza virus infections, the generation of escape mutants has been reported. To develop vaccines and drugs that are less susceptible to the generation of escape mutants, it is important to understand the evolutionary mechanisms of the viruses. Here natural selection operating on all the proteins encoded by the H3N2 human influenza A virus genome was inferred by comparing the numbers of synonymous (d_S [D_S]) and nonsynonymous (d_N [D_N]) substitutions per site. Natural selection was also inferred for the groups of functional amino acid sites involved in B-cell epitopes (BCEs), T-cell epitopes (TCEs), drug resistance, and growth in eggs. The entire region of PB1-F2 was positively selected, and positive selection also appeared to operate on BCEs, TCEs, and growth in eggs. The frequency of escape mutant generation appeared to be positively correlated with the d_N/d_S (D_N/D_S) values for the targets of vaccines and drugs, suggesting that the amino acid sites under strong functional constraint are suitable targets. In particular, TCEs may represent candidate targets because the d_N/d_S (D_N/D_S) values were small and negative selection was inferred for many of them.

Introduction

Influenza viruses are the etiological agents of influenza (Smith et al. 1933) and are classified into types A–C, among which type A viruses are the most pathogenic to humans (Suzuki and Nei 2002). Influenza A viruses possess a single-stranded (negative sense) and 8-segmented (segments 1–8) RNA genome in an enveloped virion (Noda et al. 2006). Segments 1, 3, 4, 5, and 6 each encode a single protein, namely, polymerase basic 2 (PB2; a total of 759 amino acid sites), polymerase acidic (PA; 716 sites), hemagglutinin (HA; 566 sites), nucleoprotein (NP; 498 sites), and neuraminidase (NA; 469 sites), respectively. Segments 2, 7, and 8 each encode 2 proteins, namely, PB1 (757 sites) and PB1-F2 (90 sites), matrix 1 (M1; 252 sites) and M2 (97 sites), and nonstructural 1 (NS1; 230 sites) and NS2 (121 sites), respectively. Codon positions 32–123 of PB1 and the entire region (positions 1–90) of PB1-F2, positions 239–252 of M1 and positions 9–24 of M2, and positions 168–230 of NS1 and positions 11–74 of NS2 overlap in different reading frames (fig. 1). Influenza A viruses are classified into subtypes H1–H16 and N1–N9 according to the antigenic properties of HA and NA, respectively (World Health Organization 1980).

Influenza A viruses have caused pandemics in humans during the 20th century, such as "Spanish Flu" caused by the H1N1 virus that killed 25–50 million people worldwide in 1918, "Asian Flu" by the H2N2 virus that killed 1–4 million people in 1957, and "Hong Kong Flu" by the H3N2 virus that killed 0.75–2 million people in 1968. The H1N1 and H3N2 viruses still continue to circulate and cause annual epidemics that kill 0.25–0.5 million people worldwide. Vaccines and drugs are, however, available for the prophylaxis and treatment of influenza virus infections. Vaccines are composed of either inactivated or live attenuated virions of the H1N1 and H3N2 human influenza A viruses as well as those of influenza B viruses. Because the antigenicities of the wild viruses evolve, vaccines are reformulated annually by updating the seed strains. However, when the antigenicities of the seed strains and wild viruses do not match, vaccines fail to protect the vaccinees (Mostow et al. 1970). In addition, even when they do match, escape mutants are often generated (Kilbourne et al. 2002; Jin et al. 2005; Zharikova et al. 2005; Venkatramani et al. 2006). The drugs include amantadine, which inhibits the uncoating of virions by interfering with M2, and oseltamivir, which inhibits the release of virions from infected cells by interfering with NA. However, escape mutants are often generated for the former drug (Webster et al. 1986) and less frequently for the latter drug (Kiso et al. 2004).

To develop vaccines and drugs that are less susceptible to the generation of escape mutants, it is important to understand the evolutionary mechanisms of the viruses. Of particular interest is natural selection operating on viral proteins. Negatively selected amino acid sites may be suitable targets for vaccines and drugs because many of the substitutions at these sites are likely to be intolerable (Suzuki 2004a). Positively selected sites may be useful for identifying the epitopes involved in the elimination of viruses from infected patients (Suzuki and Gojobori 2001), although they may not be suitable as targets because many of the substitutions at these sites are likely to be tolerable. The purpose of the present study was to infer the natural selection operating on all the proteins encoded by the H3N2 human influenza A virus genome.

Materials and Methods

Sequence Data

For each protein of the H3N2 human influenza A virus, all the nucleotide sequences encoding the entire region were extracted from the International Nucleotide Sequence Database (INSD: DNA Data Bank of Japan [DDBJ]/European Molecular Biology Laboratory/GenBank) (DDBJ release 63). After elimination of sequences derived from laboratory and vaccine strains, sequences derived from the same strains as others, and sequences containing ambiguous nucleotides, minor insertions, minor deletions, or premature termination codons, 259, 256, 76, 268, 284, 246, 345, 173, 113, 164, and 100 sequences were used for natural selection analysis for PB2, PB1, PB1-F2, PA, HA, NP, NA, M1, M2, NS1, and NS2, respectively. The INSD accession numbers and strain names for these

Key words: influenza virus, genome, positive selection, negative selection, vaccine, drug.

E-mail: yossuzuk@lab.nig.ac.jp.

Mol. Biol. Evol. 23(10):1902–1911. 2006
doi:10.1093/molbev/msl050
Advance Access publication July 3, 2006

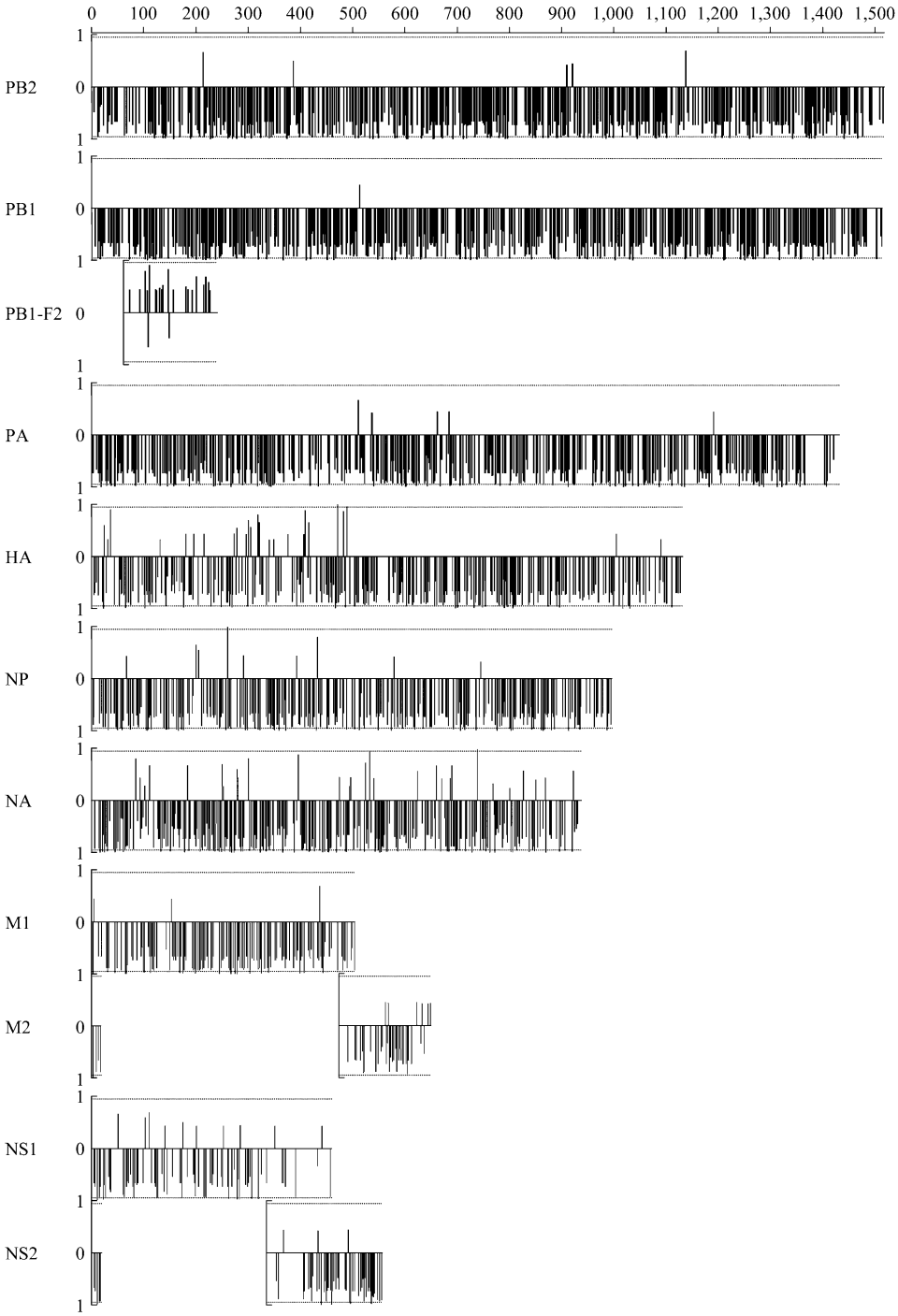

FIG. 1.—Selection profiles of all the proteins encoded by the H3N2 human influenza A virus genome. The abscissa indicates the codon positions, with the scale bar at the top. The ordinate indicates the $(1 - P)$ value for each position, which is indicated above and below the horizontal line when $d_N/d_S > 1$ and $d_N/d_S < 1$, respectively. The dotted lines represent 0.95, so that the positions where the bars cross the dotted lines above and below the horizontal line indicate the positively and negatively selected sites, respectively. M2 and NS2 are divided into 2 parts, and the positions of PB1-F2, M2, and NS2 relative to PB1, M1, and NS1 are arranged to indicate the overlapping regions.

sequences are listed in the supplementary table S1, Supplementary Material online.

Data Analysis

For each protein, a multiple alignment of the nucleotide sequences, which did not contain any gaps, was constructed using the computer program ClustalW (version 1.81) (Thompson et al. 1994). Natural selection operating on each amino acid site was inferred using ADAPTSITE (version 1.3) (Suzuki et al. 2001). In this method (parsimony method) (Suzuki and Gojobori 1999), a phylogenetic tree of the nucleotide sequences was constructed by the Neighbor-Joining method (Saitou and Nei 1987) using the p distance.

Table 1
Statistics for the Data Analyzed in the Present Study

Protein	4-Fold Degenerate Site[a]	k[a]	Total Branch Length[b]	Average Branch Length[b]	D_N/D_S[b]
PB2	267	8.8	0.5	0.0009	0.09
PB1	216	9.5	0.5	0.0009	0.09
PB1-F2	0	NA[c]	0.4	0.003	13.5
PA	209	5.9	0.5	0.0009	0.1
HA	158	6.8	0.6	0.001	0.3
NP	176	7.4	0.5	0.001	0.1
NA	141	4.1	0.8	0.001	0.3
M1	109	6.2	0.4	0.001	0.1
M2	15	2.9	0.5	0.002	0.5
NS1	59	16.9	0.4	0.001	0.5
NS2	8	7.1	0.4	0.002	0.2

[a] Only the nonoverlapping region was taken into account.
[b] Both the overlapping and nonoverlapping regions were taken into account.
[c] Not applicable.

It should be noted that the p distance is known to produce reliable phylogenetic trees, when a large number of closely related sequences is analyzed (Nei and Kumar 2000). Indeed, reliable results were obtained in the computer simulations and real data analyses using the parsimony method with the p distance (Suzuki 2004b). For each codon site, the total numbers of synonymous (c_S) and nonsynonymous (c_N) substitutions as well as the average numbers of synonymous (s_S) and nonsynonymous (s_N) sites per codon over the phylogenetic tree were computed using the maximum parsimony (MP) method (Fitch 1971). Here the transition/transversion rate ratio (κ) of nucleotide mutations was required for computing s_S and s_N. To estimate κ, the ratio (k) of the transitional/transversional nucleotide diversities at the 4-fold degenerate site was computed for each of the entire regions of PB2, PA, HA, NP, and NA as well as the nonoverlapping regions of PB1, M1, M2, NS1, and NS2, using the 2-parameter method (table 1) (Kimura 1980). κ was estimated as the average of k, weighted by the number of 4-fold degenerate sites, and found to be 7.5. The total numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site over the phylogenetic tree were computed as c_S/s_S and c_N/s_N, respectively. The null hypothesis of selective neutrality ($d_S = d_N$) was tested for each codon site by computing the probability (P) of obtaining the observed or more biased values for c_S and c_N, which were assumed to follow a binomial distribution with the probabilities of occurrence of synonymous and nonsynonymous substitutions given by $s_S/(s_S + s_N)$ and $s_N/(s_S + s_N)$, respectively. Sites where $d_N/d_S > 1$ and $d_N/d_S < 1$ with $P < 0.05$ were inferred as positively and negatively selected, respectively (2-tailed test) (Hughes and Nei 1988). It should be noted that because multiple substitutions are not corrected in this method, c_S and c_N may be underestimated, especially if the branch lengths of the phylogenetic trees constructed are large. However, the degree of underestimation appeared to be negligible for all data sets examined in the present study because the branch lengths were generally very small (table 1) (Saitou 1989).

When the biological functions of amino acid sites are known, it may be useful to group sites with similar functions for inferring natural selection (Hughes and Nei 1988). The B-cell epitope (BCE) is a group of typically 15–22 continuous or discontinuous amino acid sites that are recognized by B-cells, which generate antibodies against BCEs and neutralize the infectivity of viruses (Klein and Horejsi 1997). From the analysis of 3-dimensional structures of antigen–antibody complexes and escape mutants from monoclonal antibodies, 5 (epitopes A, B, C, D, and E) and 3 (epitopes A, B, and C) BCEs have been identified in HA (Wiley et al. 1981) and NA (Air et al. 1985), respectively. One BCE was also identified in M2 from the analysis using monoclonal antibodies (Zebedee and Lamb 1988). The $CD8^+$ and $CD4^+$ T-cell epitopes (TCEs) are groups of typically 9 and 13–18 continuous amino acid sites, respectively (Klein and Horejsi 1997). $CD8^+$ TCEs are presented on infected cells together with the human leukocyte antigen (HLA) class I (including HLA-A, -B, and -C) and are recognized by cytotoxic T-lymphocytes (CTLs), which exert cytotoxicity to infected cells. $CD4^+$ TCEs are presented on infected cells together with the HLA class II (including HLA-DQ and -DR) and are recognized by helper T-cells (Th cells). Th cells are divided into Th1 and Th2 cells, which help CTLs and B-cells activate, respectively. Th cells may also exert cytotoxicity on infected cells. $CD8^+$ and $CD4^+$ TCEs in the proteins of influenza A viruses have been identified by experiments, such as the 3-dimensional structure analysis of HLA–epitope complexes and peptide-binding and lytic assays (Macken et al. 2001). In addition, amino acid sites involved in the resistance to amantadine (Hay et al. 1985) and oseltamivir (Gubareva et al. 2000) have been determined by inhibition assays. Furthermore, influenza A viruses are usually isolated in the allantoic cavity of embryonated chicken eggs, and amino acid sites involved in the adaptation to growth in eggs have been identified by comparing the sequences derived from the isolates passaged in eggs and control cells (Hardy et al. 1995). Each of these groups of functional amino acid sites was examined for natural selection (table 4). In practice, s_S, s_N, c_S, and c_N were summed over the grouped sites to obtain S_S, S_N, C_S, and C_N, respectively (Suzuki 2004c). Here $D_S = C_S/S_S$ and $D_N = C_N/S_N$. The test of selective neutrality was conducted in a similar manner to the analysis of individual sites, where s_S, s_N, c_S, and c_N were replaced with S_S, S_N, C_S, and C_N, respectively. Positive and negative selection were inferred when $D_N/D_S > 1$ and $D_N/D_S < 1$ with $P < 0.05$, respectively (2-tailed test).

Results

Selection Profiles

The selection profiles for all the proteins of the H3N2 human influenza A virus are shown in figure 1. For PB2, 501 (66.0%) of all (759) codon sites were variable, among which $d_N/d_S < 1$ for the majority of sites (452 sites; 90.2%) and $d_N/d_S > 1$ for the minority of sites (49 sites; 9.8%) (table 2; supplementary fig. S1, Supplementary Material online). No sites were identified as positively selected, whereas 89 sites were identified as negatively selected.

The selection profiles for all the other proteins were similar to that for PB2, with the exception of PB1-F2 (fig. 1; table 2). For PB1-F2, d_N/d_S was >1 for the majority of codon sites (56 sites; 94.9%) and d_N/d_S was <1 for the

Table 2
Summary of the Selection Profiles for All the Proteins of the H3N2 Human Influenza A Virus

Protein	Total	Variable[a]	$d_N/d_S > 1$[b]	Positively Selected	$d_N/d_S < 1$[c]	Negatively Selected
PB2	759	501 (66.0%)	49 (9.8%)	0	452 (90.2%)	89
PB1	757	512 (67.6%)	55 (10.7%)	0	455 (88.9%)	77
PB1-F2	90	59 (65.6%)	56 (94.9%)	0	3 (5.1%)	0
PA	716	482 (67.3%)	63 (13.1%)	0	418 (86.7%)	68
HA	566	405 (71.6%)	95 (23.5%)	2	306 (75.6%)	34
NP	498	332 (66.7%)	44 (13.3%)	1	288 (86.7%)	62
NA	469	377 (80.4%)	84 (22.3%)	1	292 (77.5%)	86
M1	252	154 (61.1%)	20 (13.0%)	0	133 (86.4%)	21
M2	97	69 (71.1%)	26 (37.7%)	0	42 (60.9%)	1
NS1	230	155 (67.4%)	67 (43.2%)	0	87 (56.1%)	9
NS2	121	78 (64.5%)	21 (26.9%)	0	57 (73.1%)	6

[a] The percentage of variable codon sites among all sites is indicated in parentheses.
[b] The percentage of codon sites with $d_N/d_S > 1$ among variable sites is indicated in parentheses.
[c] The percentage of codon sites with $d_N/d_S < 1$ among variable sites is indicated in parentheses.

minority of sites (3 sites; 5.1%) among 59 variable sites. It should be noted that the entire region of PB1-F2 overlaps with PB1 in different reading frames, as described above. In general, when multiple genes overlap in different reading frames, nucleotide substitution is expected to be suppressed for each gene due to the superimposition of functional constraints operating on the overlapping genes. Suppression of nucleotide substitution may be detected by comparing the D_S values between the overlapping and nonoverlapping regions of a gene, under the assumptions that the mutation rates for these regions are similar and the synonymous mutations are close to selectively neutral. Indeed, in the case of M1, which overlaps with M2, D_S (= C_S/S_S = 12/12) for the overlapping region was smaller than that (C_S/S_S = 261/213) for the nonoverlapping region, although the difference was not statistically significant (χ^2 = 0.6 with 1 degree of freedom [df]; P = 0.5) (table 3). Similar results were obtained for M2, NS1, and NS2, which overlap with M1, NS2, and NS1, respectively, and these differences were statistically significant. Because PB1-F2 did not contain a nonoverlapping region, D_S for the nonoverlapping region of PB1 was used as a surrogate and compared with that for the entire (overlapping) region of PB1-F2, using the same set of 258 sequences used for the natural selection analysis for PB1. D_S (= C_S/S_S = 6/76) for PB1-F2 was significantly smaller than that (C_S/S_S = 690/578) for the nonoverlapping region of PB1 (χ^2 = 78.6 with df = 1; $P = 8.2 \times 10^{-19}$), suggesting that functional constraints operated on the overlapping region of PB1. Surprisingly, however, when the D_S values for the overlapping and nonoverlapping regions of PB1 were compared, the former (C_S/S_S = 125/82) was significantly greater than the latter (C_S/S_S = 690/578) (χ^2 = 6.5 with df = 1; P = 0.01). This observation implied that positive selection operated on PB1-F2, rather than functional constraints. To further characterize the natural selection operating on PB1-F2, the number (S_{SS}) of nucleotide sites where the mutations were expected to be synonymous in both PB1 and PB1-F2 (synonymous–synonymous sites) and that (S_{SN}) of synonymous–nonsynonymous sites as well as the numbers of corresponding substitutions (C_{SS} and C_{SN}) were computed for the overlapping region of PB1 and PB1-F2. Because D_{SN} (= C_{SN}/S_{SN} = 116/78) was significantly greater than D_{SS} (= C_{SS}/S_{SS} = 0/3) (P = 0.02), positive selection was inferred for the entire region of PB1-F2.

Among all the amino acid sites of all the proteins of the H3N2 human influenza A virus, positions 220 and 229 of HA, position 131 of NP, and position 370 of NA were inferred as positively selected (fig. 1; table 2).

Natural Selection on Groups of Functional Amino Acid Sites

When the amino acid sites involved in the BCEs were grouped, their D_N/D_S values were generally large (table 4) compared with the averages for the entire regions of the proteins (table 1). In particular, D_N/D_S was >1 for epitopes A and B of HA, although the differences were not statistically

Table 3
Comparison of C_S, S_S, and D_S between the Overlapping and Nonoverlapping Regions

			Overlapping Region			Nonoverlapping Region				
Segment	Protein	Position	C_S	S_S	D_S	C_S	S_S	D_S	χ^{2a}	P^b
2	PB1	32–123	125	82	1.5	690	578	1.2	6.5	0.01
	PB1-F2	1–90	6	76	0.08	690[c]	578[c]	1.2	78.6	8.2×10^{-19}
7	M1	239–252	12	12	1.0	261	213	1.2	0.6	0.5
	M2	9–24	3	13	0.2	70	74	0.9	6.9	0.009
8	NS1	168–230	15	56	0.3	130	150	0.9	20.2	0.000005
	NS2	11–74	38	56	0.7	68	49	1.4	13.7	0.0003

[a] χ^2 value in the comparison of D_S values between the overlapping and nonoverlapping regions.
[b] P value of obtaining the observed or greater χ^2 values in the χ^2 distribution with df = 1.
[c] C_S and S_S for the nonoverlapping region of PB1 were used as surrogates for those of PB1-F2.

Table 4
Natural Selection Inferred for the Groups of Functional Amino Acid Sites

Function	Protein	Position	Reference	D_N/D_S	P
BCE (epitope A)	HA	122, 124, 126, 130–133, 135, 137, 138, 140, 142–146, 150, 152, 168	Wiley et al. (1981)	1.2 (2.3 and 0.7)[a]	0.4 (0.06 and 0.4)[a]
BCE (epitope B)	HA	128, 129, 155–160, 163–165, 186–190, 192–194, 196–198	Wiley et al. (1981)	1.1 (1.1 and 1.0)	0.8 (0.8 and 1.0)
BCE (epitope C)	HA	44–48, 50, 51, 53, 54, 273, 275, 276, 278–280, 294, 297, 299, 300, 304, 305, 307–312	Wiley et al. (1981)	0.6 (1.0 and 0.4)	0.03 (0.8 and 0.003)
BCE (epitope D)	HA	96, 102, 103, 117, 121, 167, 170–177, 179, 182, 201, 203, 207–209, 212–219, 226–230, 238, 240, 242, 244, 246–248	Wiley et al. (1981)	0.5 (0.4 and 0.6)	0.00004 (0.0002 and 0.02)
BCE (epitope E)	HA	57, 59, 62, 63, 67, 75, 78, 80–83, 86–88, 91, 92, 94, 109, 260–262, 265	Wiley et al. (1981)	0.7 (0.9 and 0.5)	0.2 (0.8 and 0.1)
BCE (epitope A)	NA	383–387, 389–394, 396, 399–401, 403	Air et al. (1985)	0.5	0.02
BCE (epitope B)	NA	197–200, 221, 222	Air et al. (1985)	0.8	0.7
BCE (epitope C)	NA	328–332, 334, 336, 338, 339, 341–344, 346, 347, 357–359, 366–370	Air et al. (1985)	0.7	0.1
BCE	M2	2–24	Zebedee and Lamb (1988)	0.3[b]	0.003[b]
TCE (HLA-A*0201)[c]	PB1	413–421	Gianfrani et al. (2000)	0.0	0.000005
TCE (HLA-A*01)	PB1	591–599	DiBrino et al. (1993)	0.1	0.0003
TCE (HLA-A*0201)	PA	46–54	Gianfrani et al. (2000)	0.0	0.02
TCE (HLA-A*0201)	PA	225–233	Gianfrani et al. (2000)	0.3	0.06
TCE (HLA-A*1101)	HA	55–62	Gianfrani et al. (2000)	0.4	0.06
TCE (HLA-A*1101)	HA	135–144	Gianfrani et al. (2000)	1.1	0.9
TCE (HLA-DQA1*0102; HLA-DQB1*0602; HLA-DRB1*1001)	HA	255–270	Jones et al. (1994)	0.4	0.03
TCE (HLA-DRB1*0101; HLA-DRB1*0401)	HA	306–318	Stern et al. (1994); Carmichael et al. (1997)	0.2	0.0002
TCE (HLA-DRB5*0101)	HA	306–319	O'Sullivan et al. (1991)	0.2	0.00005
TCE (HLA-DRB1*1101)	HA	306–324	Carmichael et al. (1997)	0.1	< 0.0000005
TCE (HLA-A*1101)	HA	435–445	Gianfrani et al. (2000)	0.05	0.0003
TCE (HLA-A*01)	NP	44–52	DiBrino et al. (1993)	0.2	0.02
TCE (HLA-A*6801)	NP	91–99	Guo et al. (1992)	0.3	0.09
TCE (HLA-B*1402)	NP	146–154	DiBrino et al. (1994)	0.1	0.000026
TCE (HLA-B*2705)	NP	174–184	Jameson et al. (1999)	0.02	< 0.0000005
TCE (HLA-A*1101)	NP	188–198	Gianfrani et al. (2000)	0.1	0.00009
TCE (HLA-A*03)	NP	265–274	DiBrino et al. (1993)	0.07	0.005
TCE (HLA-B*44)	NP	338–346	DiBrino et al. (1995)	0.1	0.002
TCE (HLA-B*3701)	NP	339–347	Townsend et al. (1986)	0.09	0.0002
TCE (HLA-DQA1*0501; HLA-DQB1*0201)	NP	365–379	Vartdal et al. (1996)	0.3	0.008
TCE (HLA-B*44)	NP	379–387	DiBrino et al. (1995)	0.1	0.005
TCE (HLA-B*08)	NP	380–388	Sutton et al. (1993)	0.1	0.002
TCE (HLA-B*2702)	NP	381–388	Tussey et al. (1995)	0.1	0.003
TCE (HLA-B*2705)	NP	383–391	Huet et al. (1990)	0.07	0.000001
TCE (HLA-B*0702; HLA-B*3501)	NP	418–426	Boon et al. (2002); Rohrlich et al. (2002)	0.1	0.000001
TCE (HLA-CW*0102)	NP	470–479	Andersen et al. (1999)	0.7	0.6
TCE (HLA-A*0201)	NA	90–99	Gianfrani et al. (2000)	0.2	0.002
TCE (HLA-A*0201)	NA	232–240	Wedermeyer et al. (2001)	0.2	0.002
TCE (HLA-A*1101)	M1	13–21	Gianfrani et al. (2000)	0.0	0.0004
TCE (HLA-DRB1*0101)	M1	18–29	Rothbard et al. (1988)	0.09	0.002
TCE (HLA-CW*0102)	M1	51–59	Andersen et al. (1999)	0.0	0.000008
TCE (HLA-CW*0102)	M1	51–60	Andersen et al. (1999)	0.0	< 0.0000005
TCE (HLA-A*0201)	M1	58–66	Gotch et al. (1988)	0.0	0.000006
TCE (HLA-A*0201)	M1	59–68	Gotch et al. (1988)	0.1	0.0007
TCE (HLA-B*3501)	M1	125–132	Dong et al. (1996)	0.0	0.00005
TCE (HLA-DQw3)	M1	234–243	Adler et al. (1994)	0.1	0.003
TCE (HLA-DQw1)	M1	239–250	Adler et al. (1994)	0.1	0.0004
TCE (HLA-B*44)	M2	7–15	Jameson et al. (1999)	0.7	0.7
TCE (HLA-CW*0102)	M2	83–92	Andersen et al. (1999)	1.8	0.6
TCE (HLA-CW*0102)	M2	85–94	Andersen et al. (1999)	1.3	0.8

Table 4
Continued

Function	Protein	Position	Reference	D_N/D_S	P
TCE (HLA-DR*03)	NS1	34–42	Jameson et al. (1999)	0.1	0.0005
TCE (HLA-A*0201)	NS1	122–130	Man et al. (1995)	0.8	0.8
TCE (HLA-A*0201)	NS1	123–132	Man et al. (1995)	0.4	0.1
TCE (HLA-B*44)	NS1	158–166	DiBrino et al. (1995)	0.4	0.2
Resistance to amantadine	M2	26, 27, 30, 31, 34	Hay et al. (1985)	0.5	0.4
Resistance to oseltamivir	NA	119, 152, 274, 292	Gubareva et al. (2000)	0.0	0.00001
Growth in eggs	HA	111, 126, 137, 138, 144, 145, 155, 156, 158, 159, 185, 186, 193, 194, 199, 219, 226, 229, 246, 248, 276, 290	Hardy et al. (1995)	1.4 (1.3 and 1.4)	0.2 (0.4, 0.2)

[a] D_N/D_S and P values for (interior and exterior) branches are indicated in parentheses.

[b] D_N/D_S and P values were computed taking into account the fact that M2 overlaps with M1, that is, S_S, S_N, C_S, and C_N for the nonoverlapping region were summed with S_{SS}, S_{SN}, C_{SS}, and C_{SN} for the overlapping region to obtain S_S', S_N', C_S', and C_N', respectively, where S_{SS} and S_{SN} denote the numbers of nucleotide sites where the mutations are expected to be synonymous in both M1 and M2 (synonymous–synonymous sites) and that of synonymous–nonsynonymous sites, respectively, and C_{SS} and C_{SN} the numbers of corresponding substitutions. D_N/D_S and P values were computed by replacing S_S, S_N, C_S, and C_N with S_S', S_N', C_S', and C_N', respectively.

[c] For TCEs, the restricting HLA alleles are indicated in parentheses.

significant. On the other hand, the D_N/D_S values for both the $CD8^+$ and $CD4^+$ TCEs tended to be small. D_N/D_S was >1 only for one $CD8^+$ TCE (positions 135–144) of HA and 2 overlapping $CD8^+$ TCEs (positions 83–92 and 85–94) of M2, where the differences were not statistically significant. However, most (7 out of 10) of the amino acid sites involved in the former TCE were also involved in a BCE (epitope A of HA) and growth in eggs (table 4). Because the D_N/D_S values for TCEs were generally small whereas those for BCEs and growth in eggs were generally large, $D_N/D_S > 1$ observed for this TCE was not likely to be due to positive selection operating on it but due to that operating on the BCE and growth in eggs. However, the latter TCEs did not overlap with other groups of functional amino acid sites, suggesting that positive selection operated on them.

Natural selection was also examined for the groups of amino acid sites involved in resistance to amantadine and oseltamivir (table 4). The D_N/D_S value for the former group was large (0.5) and roughly comparable to those for BCEs. In contrast, negative selection was inferred for the latter group. In particular, all 9 nucleotide substitutions observed were synonymous, such that $D_N/D_S = 0$. When natural selection was examined for the group of amino acid sites involved in growth in eggs, D_N/D_S was >1, although the difference was not statistically significant (table 4). It should be noted that most (18 out of 22) of the amino acid sites involved in growth in eggs were also involved in BCEs, where D_N/D_S values were generally large (table 4). To examine whether positive selection operated on BCEs or growth in eggs, or on both, it may be useful to compute the D_N/D_S values for interior and exterior branches of the phylogenetic tree separately. Here s_S, s_N, c_S, and c_N were computed for interior and exterior branches separately and were summed over the grouped sites to obtain S_S, S_N, C_S, and C_N, respectively. The amino acid substitutions involved in growth in eggs should be observed only on exterior branches because they occur during isolation, whereas those involved in BCEs should be observed on both the interior and exterior branches. It should be noted that amino acid substitutions associated with the sampling bias for antigenic variation are also expected to be observed on exterior branches for BCEs (Bush et al. 2000). For the group of amino acid sites involved in BCEs, the D_N/D_S values (0.4–2.3) for interior branches were generally large, suggesting that positive selection operated on BCEs. However, the D_N/D_S values (0.4–1.0) for exterior branches were mostly smaller than those for interior branches, suggesting that the sampling bias did not affect the D_N/D_S values to a large extent in the present study. For the group of amino acid sites involved in growth in eggs, D_N/D_S was >1 for interior branches probably because most of them were also involved in BCEs. However, the D_N/D_S value for exterior branches was greater than that for interior branches, which was contrary to the relationship observed for most BCEs, suggesting that positive selection also operated on growth in eggs.

Discussion

Prevalence of Negatively Selected Sites

In the present study, natural selection was inferred for amino acid sites by comparing d_S (D_S) and d_N (D_N) substitutions per site that had accumulated over the phylogenetic tree. Because the direction and magnitude of natural selection may vary during evolution, the d_N/d_S (D_N/D_S) values obtained should be regarded as average values of heterogeneous selection and also as underestimates for the magnitudes of the positive and negative selection that operated over the phylogenetic tree on the sites where d_N/d_S (D_N/D_S) > 1 and d_N/d_S (D_N/D_S) < 1, respectively.

$d_N/d_S < 1$ for the majority of codon sites in all the proteins, with the exception of PB1-F2. However, relatively small numbers of sites were inferred as negatively selected. This was probably because the numbers of nucleotide substitutions ($c_S + c_N$) observed at many codon sites were insufficient for detecting statistically significant differences between c_S and c_N, although most of the sites were negatively selected in reality (Nei 1983). Therefore, more negatively selected sites should be detected as more sequence data are collected.

Positive Selection on PB1-F2

Positive selection was inferred for the entire region of PB1-F2. In fact, positive selection on PB1-F2 was previously indicated by Obenauer et al. (2006), based on the observation that $D_N/D_S > 1$ for the entire region of PB1-F2 in an analysis of avian influenza A viruses. However, PB1-F2 overlaps with PB1 in different reading frames, and D_S for PB1-F2 is suppressed by the superimposition of functional constraints operating on PB1, as clarified in the present study. Therefore, the fact that $D_N/D_S > 1$ for PB1-F2 does not necessarily support positive selection.

PB1-F2 is located in the inner mitochondrial membrane of infected cells and induces apoptosis (Chen et al. 2001). Amino acid positions 65–87 of PB1-F2 are known to exert these functions. When the nucleotide sequence of PB1-F2 was determined for many strains of influenza A viruses, premature termination codons were frequently observed (Ghedin et al. 2005; Obenauer et al. 2006). Indeed, when the reading frame of PB1-F2 was examined in the overlapping region of the 256 PB1 sequences used for the natural selection analysis in the present study, 2, 3, and 4 nonsense substitutions were inferred at codon positions 12, 80, and 88, respectively. For the first 2 cases, the function of PB1-F2 appeared to be reduced because positions 65–87 were partially or entirely eliminated. It should be noted here that when the reading frames of PB1, M1, M2, NS1, and NS2 were examined in the overlapping regions of 76 PB1-F2, 113 M2, 173 M1, 100 NS2, and 164 NS1 sequences, only 1 nonsense substitution was inferred in the reading frame of NS1. Furthermore, the function of PB1-F2 only appears to be required in vivo, where immune responses may be downregulated by inducing apoptosis of monocytes. However, it may be deleterious in cell lines and eggs, where the immune responses are absent and apoptosis is only induced in the cells containing proliferating viruses. Therefore, it is speculated that positive selection operated on the deterioration of the function of PB1-F2 during isolation. It should be noted that positive selection operating during isolation is detectable, as demonstrated for positive selection operating on the growth in eggs in the present study (Bush, Fitch, et al. 1999). Because many of PB1-F2 sequences analyzed in the present study were derived from viruses isolated in the allantoic cavity of embryonated chicken eggs and primary rhesus monkey kidney cells (Holmes et al. 2005), it will be interesting to examine whether PB1-F2 induces apoptosis of these cells.

Characteristics of Positively Selected Sites

In a previous study, positively selected amino acid sites were identified in all the proteins of the H5N1 avian and human influenza A viruses, by using the Bayesian method for detecting positively selected sites (Campitelli et al. 2006). It was found that $d_N/d_S > 1$ for positions 17, 82, 199, 334, 336, 355, and 727 of PB2, positions 138, 140, 155, 156, 218, and 227 of HA, and positions 171, 205, and 209 of NS1, by analyzing 16, 192, and 31 sequences, respectively. However, although the amino acid sites identified in HA are all involved in BCEs, TCEs, or growth in eggs, the functions for almost all the sites identified in PB2 and NS1 are unknown. It should be noted that the Bayesian method is known to generate many false positives, especially when the number of sequences analyzed is small (Suzuki and Nei 2004; Berlin and Smith 2005), implying that the positively selected sites identified in PB2 are false positives. In addition, NS1 overlaps with NS2 in different reading frames and D_S is suppressed for the overlapping region, as indicated above, suggesting that $d_N/d_S > 1$ does not necessarily support positive selection. Because all the positively selected sites identified in NS1 are located in the overlapping region, they are also likely to be artifacts.

In the present study, positions 220 and 229 of HA, position 131 of NP, and position 370 of NA were identified as positively selected. The biological function of position 220 of HA is unknown. However, its adjacent site (position 219) is involved in a BCE (epitope D) and growth in eggs, suggesting that position 220 is also involved in these functions. Position 229 of HA and position 370 of NA are also involved in BCEs (epitope D of HA and epitope A of NA, respectively). The biological function for position 131 of NP is unknown. However, this site has also been reported as positively selected in an analysis of 84 NP sequences using the Bayesian method (Forsberg and Christiansen 2003). Because the number of sequences analyzed was large, this result may be reliable to some extent. It will, therefore, be interesting to examine the function of this site experimentally using site-directed mutagenesis (Hoffmann et al. 2000). It should be noted that the positively and negatively selected sites identified in the present study may remain tentative because the multiple tests were not corrected. However, the number of effective tests that should be corrected appeared to be relatively small because the numbers of nucleotide substitutions ($c_S + c_N$) observed at many codon sites were insufficient for detecting statistically significant differences between c_S and c_N, as mentioned above.

d_N/d_S (D_N/D_S) as an Indicator for the Generation of Escape Mutants against Vaccines and Drugs

The existence of positive selection operating on BCEs, TCEs, and growth in eggs was supported by the observations that $D_N/D_S > 1$ for some of the groups of amino acid sites involved in these functions. These observations suggest that both the humoral and cellular immune systems contribute to the elimination of influenza A viruses from infected patients (Thomas et al. 2006), similar to the case for hepatitis C viruses (Suzuki and Gojobori 2001). In particular, the D_N/D_S values were generally large (0.3–1.2) for all the BCEs. Vaccines currently available for influenza viruses mainly induce immune responses against BCEs (Cox et al. 2004), although responses are also induced against TCEs (Powers and Belshe 1993). However, the generation of escape mutants has often been reported (Kilbourne et al. 2002; Jin et al. 2005; Zharikova et al. 2005; Venkatramani et al. 2006). In contrast, in the case of polioviruses, negative selection was inferred for almost all the amino acid sites involved in the neutralization epitopes (BCEs), namely, N-AgI (consisting of positions 91–101 and 144 of viral protein 1 [VP1]), N-AgIB (positions 96–104 and 141–152 of VP1), N-AgII (positions 221–226 of VP1 and positions 138,

142, 164–170, and 270 of VP2), N-AgIII (positions 286–290 of VP1 and positions 58–60, 70–73, 76, 77, and 79 of VP3), and N-AgIV (position 72 of VP2 and position 76 of VP3) (Suzuki 2004a). The D_N/D_S values were small for all the BCEs, with the exception of N-AgIV—0.1 ($P < 0.0000005$) for N-AgI, 0.1 ($P < 0.0000005$) for N-AgIB, 0.2 ($P < 0.0000005$) for N-AgII, 0.1 ($P < 0.0000005$) for N-AgIII, and 0.4 ($P = 0.1$) for N-AgIV. However, the D_N/D_S value for N-AgIV may be unreliable because only 2 sites were included and the number of nucleotide substitutions observed was small. Interestingly, vaccines against polioviruses are known to be highly effective.

The D_N/D_S value for the group of amino acid sites involved in resistance to amantadine was large (0.5) and roughly comparable to those for BCEs. It has been reported that escape mutants, which were not attenuated, were often generated against this drug (Webster et al. 1986). In contrast, $D_N/D_S = 0$ and negative selection were inferred for the group of amino acid sites involved in resistance to oseltamivir. Escape mutants were also reported to be generated against this drug, although less frequently than those against amantadine (Kiso et al. 2004). In addition, these escape mutants were often attenuated (Yen et al. 2005).

These observations indicate that the frequency of escape mutant generation is positively correlated with the d_N/d_S (D_N/D_S) values of the targets of vaccines and drugs. Specifically, frequent generation of escape mutants is associated with large d_N/d_S (D_N/D_S) values (>0.3) for the targets, whereas infrequent generation is associated with small d_N/d_S (D_N/D_S) values (<0.2). It should be noted that amino acid mutations are expected to be advantageous even for the latter case because they would also generate escape mutants. In reality, however, negative selection was inferred, suggesting that these targets were under strong functional constraints that may outcompete the advantageous effect of generating escape mutants. These results suggest that, in order to develop vaccines and drugs that are less susceptible to the generation of escape mutants, the amino acid sites under strong functional constraints, as judged by small d_N/d_S (D_N/D_S) values (<0.2) and detection of negative selection, are suitable targets, although it might also be possible to use positively selected sites as targets if the future sequences could be predicted accurately (Bush, Bender, et al. 1999; Plotkin et al. 2002; Ferguson et al. 2003; Smith et al. 2004). It should be noted that by using the former sites as targets, strong immune responses may be directed against them and, as a consequence, the advantageous effect of generating escape mutants may become larger and outcompete the functional constraints. However, even if escape mutants are generated, their absolute fitness should be small. Therefore, the proliferation of escape mutants may be prevented by using multiple vaccines and drugs in combination.

It should be noted that among the BCEs of H3N2 human influenza A viruses, positions 2–24 of M2 was associated with the smallest D_N/D_S value. In fact, this region has been proposed as a candidate target for universal vaccines against influenza virus infections (Neirynck et al. 1999). However, such vaccines may not be as effective as those against polioviruses because the D_N/D_S value (0.3) for this region was still greater than those (0.1–0.2) for the BCEs of polioviruses (except for N-AgIV).

Although positive selection may operate on some of the TCEs, the D_N/D_S values were small (<0.2) and negative selection was inferred for many of them. It is known that the cellular immune system does not affect the susceptibility to influenza virus infections and may even exert detrimental effects occasionally (Crowe et al. 2006). However, amino acid mutations in many of TCEs are expected to be advantageous, and generation of escape mutants has actually been reported (Moskophidis and Kioussis 1998; Boon et al. 2002; Gog et al. 2003). Because the recognition of TCEs by T-cells is restricted by the haplotype of HLA whereas that of BCEs by B-cells is not, the D_N/D_S values for TCEs may tend to be smaller than those for BCEs in analyses of viruses isolated from unrelated patients (Berkhoff et al. 2005). However, strong functional constraints actually operate on many TCEs, as clarified in the present study, suggesting that they represent candidate targets for vaccines and drugs for prophylaxis and treatment of influenza virus infections.

Supplementary Material

Supplementary table S1 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The author thanks Masatoshi Nei for valuable suggestions and 3 anonymous reviewers for valuable comments on an earlier version of the manuscript. This work was supported by KAKENHI 17770007.

Literature Cited

Adler S, Frank R, Lanzavecchia A, Weiss S. 1994. T cell epitope analysis with peptides simultaneously synthesized on cellulose membranes: fine mapping of two DQ dependent epitopes. FEBS Lett 352:167–70.

Air GM, Els MC, Brown LE, Laver WG, Webster RG. 1985. Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. Virology 145:237–48.

Andersen MH, Sondergaard I, Zeuthen J, Elliott T, Haurum JS. 1999. An assay for peptide binding to HLA-Cw*0102. Tissue Antigens 54:185–90.

Berkhoff EGM, de Wit E, Geelhoed-Mieras MM, Boon ACM, Symons J, Fouchier RAM, Osterhaus ADME, Rimmelzwaan GF. 2005. Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes. J Virol 79: 11239–46.

Berlin S, Smith NGC. 2005. Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. BMC Evol Biol 5:65.

Boon ACM, de Mutsert G, Graus YMF, Fouchier RAM, Sintnicolaas K, Osterhaus ADME, Rimmelzwaan GF. 2002. Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. J Virol 76:2567–72.

Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. 1999. Predicting the evolution of human influenza A. Science 286:1921–5.

Bush RM, Fitch WM, Bender CA, Cox NJ. 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol Biol Evol 16:1457–65.

Bush RM, Smith CB, Cox NJ, Fitch WM. 2000. Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution. Proc Natl Acad Sci USA 97:6974–80.

Campitelli L, Ciccozzi M, Salemi M, Taglia F, Boros S, Donatelli I, Rezza G. 2006. H5N1 influenza virus evolution: a comparison of different epidemics in birds and humans (1997-2004). J Gen Virol 87:955–60.

Carmichael P, Copier J, So A, Lechler R. 1997. Allele-specific variation in the degeneracy of major histocompatibility complex (MHC) restriction. Hum Immunol 54:21–9.

Chen W, Calvo PA, Malide D, et al. (13 co-authors). 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nat Med 7:1306–12.

Cox RJ, Brokstad KA, Ogra P. 2004. Influenza virus: immunity and vaccination strategies. Comparison of the immune response to inactivated and live, attenuated influenza vaccines. Scand J Immunol 59:1–15.

Crowe SR, Miller SC, Woodland DL. 2006. Identification of protective and non-protective T cell epitopes in influenza. Vaccine 24:452–6.

DiBrino M, Parker KC, Margulies DH, Shiloach J, Turner RV, Biddison WE, Coligan JE. 1994. The HLA-B14 peptide binding site can accommodate peptides with different combinations of anchor residues. J Biol Chem 269:32426–34.

DiBrino M, Parker KC, Margulies DH, Shiloach J, Turner RV, Biddison WE, Coligan JE. 1995. Identification of the peptide binding motif for HLA-B44, one of the most common HLA-B alleles in the Caucasian population. Biochemistry 34:10130–8.

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 151:5930–5.

Dong T, Boyd D, Rosenberg W, Alp N, Takiguchi M, McMichael A, Rowland-Jones S. 1996. An HLA-B35-restricted epitope modified at an anchor residue results in an antagonist peptide. Eur J Immunol 26:335–9.

Ferguson NM, Galvani AP, Bush RM. 2003. Ecological and immunological determinants of influenza evolution. Nature 422:428–33.

Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20:406–16.

Forsberg R, Christiansen FB. 2003. A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. Mol Biol Evol 20:1252–9.

Ghedin E, Sengamalay NA, Shumway M, et al. (19 co-authors). 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 437:1162–6.

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 61:438–52.

Gog JR, Rimmelzwaan GF, Osterhaus AD, Grenfell BT. 2003. Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A. Proc Natl Acad Sci USA 100:11143–7.

Gotch F, McMichael A, Rothbard J. 1988. Recognition of influenza A matrix protein by HLA-A2-restricted cytotoxic T lymphocytes. J Exp Med 168:2045–57.

Gubareva LV, Kaiser L, Hayden FG. 2000. Influenza virus neuraminidase inhibitors. Lancet 355:827–35.

Guo HC, Jardetzky TS, Garrett TP, Lane WS, Strominger JL, Wiley DC. 1992. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. Nature 360:364–6.

Hardy CT, Young SA, Webster RG, Naeve CW, Owens RJ. 1995. Egg fluids and cells of the chorioallantoic membrane of embryonated chicken eggs can select different variants of influenza A (H3N2) viruses. Virology 211:302–6.

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

Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proc Natl Acad Sci USA 97:6108–13.

Holmes EC, Ghedin E, Miller N, et al. (11 co-authors). 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol 3:e300.

Huet S, Nixon DF, Rothbard JB, Townsend A, Ellis SA, McMichael AJ. 1990. Structural homologies between two HLA B27-restricted peptides suggest residues important for interaction with HLA B27. Int Immunol 2:311–6.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–70.

Jameson J, Cruz J, Terajima M, Ennis FA. 1999. Human CD8+ and CD4+ T lymphocyte memory to influenza A viruses of swine and avian species. J Immunol 162:7578–83.

Jin H, Zhou H, Liu H, Chan W, Adhikary L, Mahmood K, Lee M-S, Kemble G. 2005. Two residues in the hemagglutinin of A/Fujian/411/02-like influenza viruses are responsible for antigenic drift from A/Panama/2007/99. Virology 336:113–9.

Jones CM, Lake RA, Lamb JR, Faith A. 1994. Degeneracy of T cell receptor recognition of an influenza virus hemagglutinin epitope restricted by HLA-DQ and –DR class II molecules. Eur J Immunol 24:1137–42.

Kilbourne ED, Smith C, Brett I, Pokorny BA, Johansson B, Cox N. 2002. The total influenza vaccine failure of 1947 revisited: major intrasubtypic antigenic change can explain failure of vaccine in a post-World War II epidemic. Proc Natl Acad Sci USA 99:10748–52.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–20.

Kiso M, Mitamura K, Sakai-Tagawa Y, Shiraishi K, Kawakami C, Kimura K, Hayden FG, Sugaya N, Kawaoka Y. 2004. Resistant influenza A viruses in children treated with oseltamivir: descriptive study. Lancet 364:759–65.

Klein J, Horejsi V. 1997. Immunology. 2nd ed. Oxford: Blackwell Science Ltd.

Macken C, Lu H, Goodman J, Boykin L. 2001. The value of a database in surveillance and vaccine selection. In: Osterhaus ADME, Cox N, Hampson AW, editors. Options for the control of influenza IV. Amsterdam: Elsevier Science. p 103–6.

Man S, Newberg MH, Crotzer VL, Luckey CJ, Williams NS, Chen Y, Huczko EL, Ridge JP, Engelhard VH. 1995. Definition of a human T cell epitope from influenza A non-structural protein 1 using HLA-A2.1 transgenic mice. Int Immunol 7:597–605.

Moskophidis D, Kioussis D. 1998. Contribution of virus-specific $CD8^+$ cytotoxic T cells to virus clearance or pathologic manifestations of influenza virus infection in a T cell receptor transgenic mouse model. J Exp Med 188:223–32.

Mostow SR, Schoenbaum SC, Dowdle WR, Coleman MT, Kaye HS, Hierholzer JC. 1970. Studies on inactivated influenza vaccines. II. Effect of increasing dosage on antibody response and adverse reactions in man. Am J Epidemiol 92:248–56.

Nei M. 1983. Genetic polymorphism and the role of mutation in evolution. In: Nei M, Koehn RK, editors. Evolution of genes and proteins. Sunderland, MA: Sinauer. p 165–90.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford: Oxford University Press.
Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W. 1999. A universal influenza A vaccine based on the extracellular domain of the M2 protein. Nat Med 5:1157–63.
Noda T, Sagara H, Yen A, Takada A, Kida H, Cheng RH, Kawaoka Y. 2006. Architecture of ribonucleoprotein complexes in influenza A virus particles. Nature 439:490–2.
Obenauer JC, Denson J, Mehta PK, et al. (17 co-authors). 2006. Large-scale sequence analysis of avian influenza isolates. Science 311:1576–80.
O'Sullivan D, Sidney J, del Guercio MF, Colon SM, Sette A. 1991. Truncation analysis of several DR binding epitopes. J Immunol 146:1240–6.
Plotkin JB, Dushoff J, Levin SA. 2002. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. Proc Natl Acad Sci USA 99:6263–8.
Powers DC, Belshe RB. 1993. Effect of age on cytotoxic T lymphocyte memory as well as serum and local antibody responses elicited by inactivated influenza virus vaccine. J Infect Dis 167:584–92.
Rohrlich PS, Cardinaud S, Firat H, Lamari M, Briand P, Escrious N, Lemonier FA. 2002. HLA-B*0702 transgenic, H-2KbDb double-knockout mice: phenotypical and functional characterization in the response to influenza. Int Immunol 15: 765–72.
Rothbard JB, Lechler RI, Howland K, Bal V, Eckels DD, Sekaly R, Long EO, Taylow WR, Lamb JR. 1988. Structural model of HLA-DR1 restricted T cell antigen recognition. Cell 52:515–23.
Saitou N. 1989. A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. Syst Zool 38:1–6.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–25.
Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305:371–6.
Smith W, Andrewes CH, Laidlaw PP. 1933. A virus obtained from influenza patients. Lancet 225:66–8.
Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368:215–21.
Sutton J, Rowland-Jones S, Rosenberg W, et al. (12 co-authors). 1993. A sequence pattern for peptides presented to cytotoxic T lymphocytes by HLA B8 revealed by analysis of epitopes and eluted peptides. Eur J Immunol 23:447–53.
Suzuki Y. 2004a. Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene 328:127–33.
Suzuki Y. 2004b. New methods for detecting positive selection at single amino acid sites. J Mol Evol 59:11–9.
Suzuki Y. 2004c. Three-dimensional window analysis for detecting positive selection at structural regions of proteins. Mol Biol Evol 21:2352–9.
Suzuki Y, Gojobori T. 1999. A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315–28.
Suzuki Y, Gojobori T. 2001. Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. Gene 276:83–7.
Suzuki Y, Gojobori T, Nei M. 2001. ADAPTSITE: detecting natural selection at single amino acid sites. Bioinformatics 17:660–1.
Suzuki Y, Nei M. 2002. Origin and evolution of influenza virus hemagglutinin genes. Mol Biol Evol 19:501–9.
Suzuki Y, Nei M. 2004. False-positive selection identified by ML-based methods: examples from the *Sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. Mol Biol Evol 21:914–21.
Thomas PG, Keating R, Hulse-Post DJ, Doherty PC. 2006. Cell-mediated protection in influenza infection. Emerg Infect Dis 12:48–54.
Thompson JD, Higgins DG, Gibson TJ. 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 22:4673–80.
Townsend AR, Rothbard J, Gotch FM, Bahadur G, Wraith D, McMichael AJ. 1986. The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. Cell 44:959–68.
Tussey LG, Rowland-Jones S, Zheng TS, Androlewicz MJ, Cresswell P, Frelinger JA, McMichael AJ. 1995. Different MHC class I alleles compete for presentation of overlapping viral epitopes. Immunity 3:65–77.
Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanovic S, Eriksen JE, Sletten K, Thorsby E, Rammensee HG, Sollid LM. 1996. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. Eur J Immunol 26:2764–72.
Venkatramani L, Bochkareva E, Lee JT, Gulati U, Laver WG, Bochkarev A, Air GM. 2006. An epidemiologically significant epitope of a 1998 human influenza virus neuraminidase forms a highly hydrated interface in the NA-antibody complex. J Mol Biol 356:651–63.
Webster RG, Kawaoka Y, Bean WJ. 1986. Vaccination as a strategy to reduce the emergence of amantadine- and rimantadine-resistant strains of A/Chick/Pennsylvania/83 (H5N2) influenza virus. J Antimicrob Chemother 18:157–64.
Wedermeyer H, Mizukoshi E, Davis AR, Bennink JR, Rehermann B. 2001. Cross-reactivity between hepatitis C virus and influenza A virus determinant specific cytotoxic T cells. J Virol 75:11392–400.
Wiley DC, Wilson IA, Skehel JJ. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289:373–8.
World Health Organization. 1980. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull WHO 58:585–91.
Yen HL, Herlocher LM, Hoffmann E, Matrosovich MN, Monto AS, Webster RG, Govorkova EA. 2005. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. Antimicrob Agents Chemother 49:4075–84.
Zebedee SL, Lamb RA. 1988. Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. J Virol 62:2762–72.
Zharikova D, Mozdzanowska K, Feng J, Zhang M, Gerhard W. 2005. Influenza type A virus escape mutants emerge in vivo in the presence of antibodies to the ectodomain of matrix protein 2. J Virol 79:6644–54.

Jianzhi Zhang, Associate Editor

Accepted June 28, 2006