Genomics 104 (2014) 545–553

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Adaptive evolution of a novel avian-origin influenza A/H7N9 virus

Ramaiah Arunachalam [a,b]

[a] Sri Paramakalyani Centre for Environmental Sciences, Manonmaniam Sundaranar University, Alwarkurichi 627 412, Tamil Nadu, India
[b] Centre for Infectious Disease Research, Indian Institute of Science, Bangalore 560 012, India [1]

ARTICLE INFO

Article history:
Received 23 January 2014
Accepted 29 October 2014
Available online 1 November 2014

Keywords:
Influenza A/H7N9 virus
Diversifying selection
Purifying selection
Functional annotation
Vaccine design

ABSTRACT

In China, the recent outbreak of novel influenza A/H7N9 virus has been assumed to be severe, and it may possibly turn brutal in the near future. In order to develop highly protective vaccines and drugs for the A/H7N9 virus, it is critical to find out the selection pressure of each amino acid site. In the present study, six different statistical methods consisting of four independent codon-based maximum likelihood (CML) methods, one hierarchical Bayesian (HB) method and one branch-site (BS) method, were employed to determine if each amino acid site of A/H7N9 virus is under natural selection pressure. Functions for both positively and negatively selected sites were inferred by annotating these sites with experimentally verified amino acid sites. Comprehensively, the single amino acid site 627 of PB2 protein was inferred as positively selected and it function was identified as a T-cell epitope (TCE). Among the 26 negatively selected amino acid sites of PB2, PB1, PA, HA, NP, NA, M1 and NS2 proteins, only 16 amino acid sites were identified to be involved in TCEs. In addition, 7 amino acid sites including, 608 and 609 of PA, 480 of NP, and 24, 25, 109 and 205 of M1, were identified to be involved in both B-cell epitopes (BCEs) and TCEs. Conversely, the function of positions 62 of PA, and, 43 and 113 of HA was unknown. In conclusion, the seven amino acid sites engaged in both BCEs and TCEs were identified as highly suitable targets, as these sites will be predicted to play a principal role in inducing strong humoral and cellular immune responses against A/H7N9 virus.

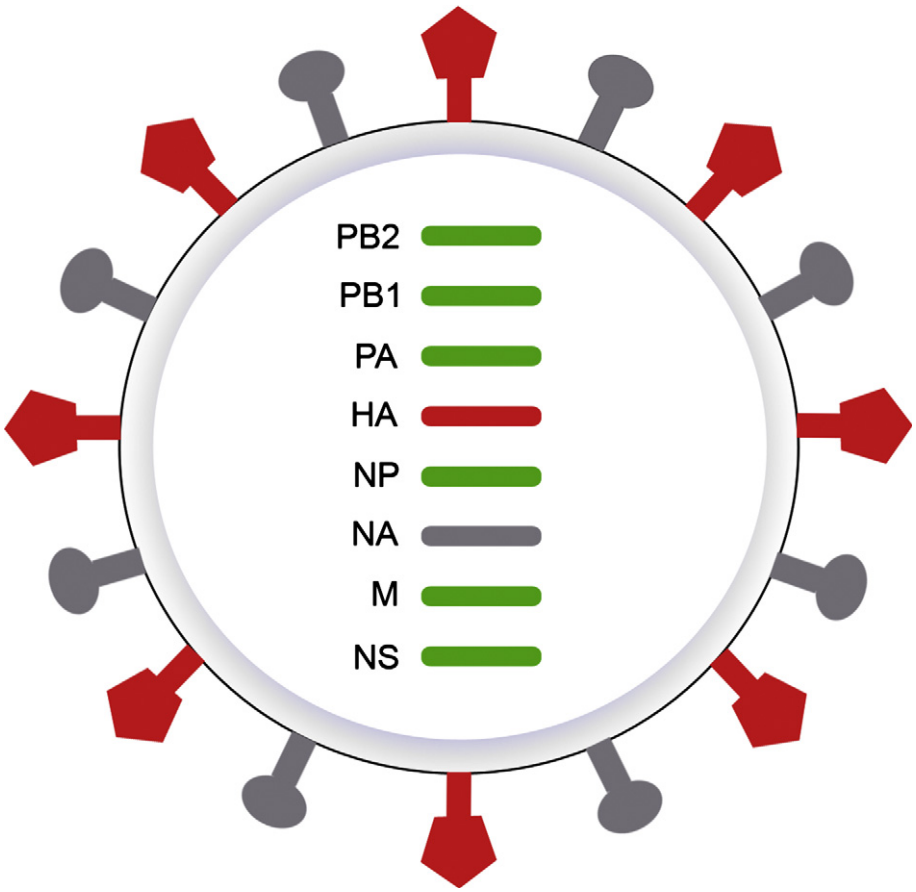
1. Introduction

Influenza viruses are classified into types A, B and C, in which, type A viruses are identified as most pathogenic to human [1–3]. Based on antigenic properties of external antigens Hemagglutinin (HA) and Neuraminidase (NA), influenza A viruses are further classified into subtypes H1–H16 and N1–N9, respectively [4–10]. Influenza A viruses possess a single stranded, negative sense 8 segmented RNA genome in an enveloped virion [11]. The genomic segments of influenza A viruses naturally allow for re-assortment that leads to the generation of new viruses, for instance, the 2009 pandemic influenza A/H1N1 virus. Remarkably, pathogenic avian influenza A subtypes H7N1, H7N2, H7N3, H7N7 and H9N2 or H10N7 viruses can cause a mild to moderate level of lower respiratory tract infections in humans. The transmission of the H7 subtype virus to mammals in Asia has rarely been reported. There has also been insufficient reports on human infections with the N9 subtype [12,13]. In March of 2013, a novel re-assorted strain of avian-origin influenza A/H7N9 subtype virus was identified in urban residents of Shanghai, China. These affected residents presented with a rapid progressive lower respiratory tract infection [13]. Evolution of the novel A/H7N9 virus likely occurred through the re-assortment of three subtypes of avian influenza viruses that co-infected a single host e.g. live birds in the market. The A/H7N9 virus likely gained its external HA and NA genes from domestic ducks (H7N3 virus) and wild birds (H7N9 virus), respectively, and the remaining six internal genes from domestic poultry (multiple H9N2 viruses) [13–15] (Fig. 1). Influenza viruses with these combinations of genes have not previously been identified among viruses isolated from birds, humans or any other species. This is the first time that the low pathogenic avian influenza A/H7N9 virus has caused a brutal outbreak in humans [16].

Typically, a periodic human infection by most of the avian influenza A viruses occur by animal to human transmission, but lack the ability of human-to-human transmission [8,9,17,18]. Unlike the 2009 pandemic influenza A/H1N1 virus, there is insufficient evidence of efficient human-to-human transmission of novel avian-origin influenza A/H7N9 virus [19]. It is predicted that the failure to mount an immune response, lack of previous infection and the constant changes in the viral genome, are critical reasons for the A/H7N9 viral outbreak in humans [10,20]. Few laboratories are currently working with the World Health Organization (WHO) and its Global Influenza Surveillance and Response System, to select the candidate strains for the development of a vaccine, which will then be shared with vaccine manufacturers worldwide. Currently, the existing H7 viral vaccine candidates provide some cross-protection against the novel influenza A/H7N9 virus [21]. It is evident that new vaccine candidates are required. Noteworthy, the novel influenza A/H7N9 virus grows well in eggs [13,16]. Utilizing this

E-mail address: arunachalamphd@gmail.com.
[1] Present address.

http://dx.doi.org/10.1016/j.ygeno.2014.10.012
0888-7543/



Note:

The A/H7N9 virus likely gained its external HA (Red) and NA (Grey) genes from domestic ducks (H7N3 virus) and wild birds (H7N9 virus), respectively, and remaining six internal genes (Green) from domestic poultry (multiple H9N2 viruses).

Fig. 1. Diagram of novel human influenza A/H7N9 virus. Note: The A/H7N9 virus likely gained its external HA (red) and NA (gray) genes from domestic ducks (H7N3 virus) and wild birds (H7N9 virus), respectively, and the remaining six internal genes (green) from domestic poultry (multiple H9N2 viruses).

Table 1
Summary of positive selection pressure on the proteins of the novel human influenza A/H7N9 viruses.

Protein	Total no. of sites	Total no. of codons	Model	Datamonkey						ω > 1 (total no. of sites)[a]	No. of positively selected sites with $p \leq 0.05$ (or) BF ≥ 20 (or) posterior prob ≥ 0.9 [b]	No. of statistically reliable positively selected sites identified by more than one method[c]
				Codon-based maximum likelihood methods				Hierarchical Bayesian method	Branch-site method			
				SLAC	FEL	REL	IFEL	FUBAR	MEME			
PB2	2280	759	001101	–	–	12	–	1	–	12	12	1
PB1	2274	757	HKY85	–	–	–	–	–	–	–	–	–
PB1-F2	273	90	F81	–	–	–	–	–	–	–	–	–
PA	2151	716	HKY85	–	–	–	–	–	1	1	1	–
HA	1683	560	F81	–	–	–	1	–	–	1	–	–
NP	1497	498	HKY85	–	–	–	–	–	1	1	1	–
NA	1398	465	010000	–	–	–	–	–	–	–	–	–
M1	759	252	HKY85	–	–	–	–	–	–	–	–	–
M2	294	97	000010	–	–	–	–	–	–	–	–	–
NS1	654	217	F81	–	–	159	–	–	–	159	159	–
NS2	366	121	F81	–	–	–	–	–	–	–	–	–

[a] The total number of positively selected sites (ω > 1) are observed from different methods for each gene/protein.
[b] Among the total number of positively selected sites, amino acid sites with statistically significant levels ($p \leq 0.05$ (SLAC, FEL, IFEL and MEME) or Bayes factors ≥20 (REL) or posterior probability ≥ 0.9 (FUBAR)) are indicated.
[c] The total number of statistically reliable positively selected sites for each gene that was observed by more than one method is indicated. Because, the significance of one method alone does not always infer that the sites are subjected to either positive or negative selection pressure thereby, only positions that have been detected by more than one method are finally considered as positively selected.

technique to identify new vaccine candidates along with utilizing published genomic sequences of influenza A/H7N9 viruses can be an effective method. The core objective of this investigation was to determine the natural selection pressures on each amino acid site of the novel avian-origin influenza A/H7N9 virus, which will be important in the efficient design of vaccines and drugs that are less susceptible to the mutant viral generations. Additional effort has been taken to annotate the functions of amino acid sites under strong natural selection pressures.

2. Results and discussion

2.1. Diversifying/positive selection

The strength of natural selection in each amino acid site of novel A/H7N9 viral proteins was inferred by calculating '*Omega*' (ω) using various methods under suitable substitution models. Tables 1–4 illustrate the results on diversifying and purifying selection pressures of each amino acid site of the entire proteins encoded by the A/H7N9 viral genomes. Comprehensive analysis showed 174 amino acid sites in the entire proteins of novel A/H7N9 viruses that underwent positive Darwinian selections. Among the 174 positively selected sites, 173 sites are identified with a statistically significant level of $p \leq 0.05$ for single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), internal fixed effects likelihood (IFEL), mixed effects model of episodic (MEME) selection or as posterior probability ≥ 0.9 for Fast Unconstrained Bayesian AppRoximation (FUBAR) or as BF ≥ 20 for random effects likelihood (REL) [22], when the CML, HB and BS methods were employed under best fit models (Tables 1, 3). Typically, in natural selection analysis, the significance of one method alone is not sufficient to infer that the amino acid site is subjected to either diversifying or purifying selection pressure. Hence, only amino acid sites that have been detected by more than one method, are considered positively or negatively selected [22,23]. One out of 173 statistically reliable amino acid sites is considered as a positively selected site with position 627 ((E) Glutamic acid (highlighted with red color)) of PB2 (Tables 1, 3; Supplementary Fig. 1) since it has been detected by more than one method. Markedly, no site was inferred as positively selected for PB1, PB1-F2, NA, M1, M2 and NS2 proteins, whereas, a single positively selected site of HA was inferred with no statistical significance (Table 1). In total, 11, 1, 1 and 159 reliable positively selected sites of PB2, PA, NP and NS1 proteins, respectively, were inferred just by a single method. As a result, there were no identical positively selected sites shown by other methods used in this study (Tables 1,3). The presence of the single reliable positively selected site 627 in PB2 of A/H7N9 virus indicates that this is an effect of molecular adaptation of the avian virus to human, which confers an evolutionary advantage to the virus. It validates the hypothesis that the positive selection is rare as it occurs in a few amino acid sites during the short time and is hardly detected effectively when compared to a large number of sites under neutral and purifying selection pressures [24,25]. Usually in the early stages of viral pandemic outbreaks, the virus adapting into a new host will be faced with higher selection pressures; particularly, codons/amino acid sites of the virus under positive selection pressures are likely to be more [26]. However, the presence of a single positively selected site suggests that the novel A/H7N9 virus with genomic segments from different subtypes/lineages has not undergone dramatic changes in order to adapt to humans. Typically, influenza pandemic outbreaks caused by different influenza viral strains may be of distinct severity [27]. The 2013 outbreak of low pathogenicity influenza A/H7N9 virus in China is unlike the recent 2009 pandemic outbreak of the A/H1N1 virus. The pandemic of the A/H1N1 virus is relatively mild, as prior to this outbreak, the seasonal H1N1 influenza viruses have been widely circulated in humans for decades. Hence, low levels of acquired immunity exist in humans against the new pandemic A/H1N1 virus, particularly in the older population. On the other hand, all age groups in humans are likely vulnerable to the A/H7N9 viral infections as this subtype has never been extensively circulated within the human population, which indicates that there is no acquired immunity that exists against the novel A/H7N9 virus. Hence, possibly in the near future, the A/H7N9 pandemic will likely be more severe than the 2009 A/H1N1 pandemic outbreak. With the driving force of natural selection, the low pathogenic A/H7N9 virus will most likely evolve into a highly pathogenic strain in birds and will probably then involve human-to-human transmission after its cross-species transmission [28]. Thus, the A/H7N9 viral outbreak has been assumed to be severe, and it is anticipated to become tremendously brutal in the near future, highlighting the critical need for scientific measures to develop highly protective vaccine and drugs.

To widen the horizon of analyses, it is essential to study the previous reports [3,10,26,29–35] on positive selection of influenza subtypes H1N1, H3N2 and H5N1 viruses. For an effective interpretation, I have compared the present result to the findings of some of the aforementioned studies. Suzuki [3] found that a total of 4 amino acid sites of the human influenza A/H3N2 viruses were under strong positive selection using parsimony method. Campitelli et al. [30] reported 16 positively selected amino acid sites in the entire proteins of the human and avian influenza A/H5N1 viruses using Bayesian method. Chen and Sun

Table 2
Summary of negative selection pressure on the proteins of the novel human influenza A/H7N9 viruses.

Protein	Datamonkey						ω > 1 (total no. of sites)[a]	No. of negatively selected sites with $p \leq 0.05$ (or) BF ≥ 20 (or) posterior prob ≥ 0.9 [b]	No. of statistically reliable negatively selected sites identified by more than one method[c]
	Codon-based maximum likelihood methods				Hierarchical Bayesian method	Branch-site method			
	SLAC	FEL	REL	IFEL	FUBAR	MEME			
PB2	–	12	–	–	6	–	12	6	4
PB1	–	8	–	–	6	–	9	7	2
PB1-F2	–	1	–	–	1	–	1	1	–
PA	1	21	–	–	9	–	23	12	4
HA	1	8	–	–	11	–	11	11	4
NP	–	5	–	–	6	–	8	6	3
NA	2	4	–	2	4	–	4	4	2
M1	–	7	20	5	5	–	20	20	6
M2	–	–	–	–	–	–	–	–	–
NS1	–	–	–	–	–	–	–	–	–
NS2	1	1	–	–	1	–	1	1	1

[a] The total number of negatively selected sites (ω < 1) are observed from different methods for each gene/protein.
[b] Among the total number of negatively selected sites, amino acid sites with statistically significant levels ($p \leq 0.05$ (SLAC, FEL, IFEL and MEME) or Bayes factors ≥20 (REL) or posterior probability ≥ 0.9 (FUBAR)) are indicated.
[c] The total number of statistically reliable negatively selected sites for each gene that was observed by more than one method is indicated. Because, the significance of one method alone does not always infer that the sites are subjected to either positive or negative selection pressure thereby, only positions that have been detected by more than one method are finally considered as negatively selected.

Table 3
List of positively and negatively selected sites of the novel human A/H7N9 influenza virus.

Gene	Datamonkey						
	Method	Total no. of statistically reliable sites		Positively selected sites	Total no. of statistically reliable sites		Purifying selected sites
PB2	REL	12	12	195, 197, 224, 292, 354, 395, 534, 559, 591,	–	6	6, 153, **185**, **490**, **499**, **652**
	FUBAR	1		**627**, 701, 740	6		
	FEL	–			4		
PB1	FEL	–	–	–	3	7	21, 208, 257, 662, 689, **738**, **740**
	FUBAR	–			6		
PB1-F2	FUBAR	–	–	–	1	1	24
PA	MEME	1	1	618	-	12	**62**, 203, 260, 558, 582, 590, 597, 599,
	FEL	–			7		**608**, **609**, 616, **681**
	FUBAR	–			9		
HA	FEL	–	–	–	4	11	**43**, 79, 80, **113**, 119, 194, 226, 337,
	FUBAR	–			11		**365**, **366**, 405
NP	MEME	1	1	15	–	6	**129**, 216, 221, 270, **286**, **480**
	FEL	–			3		
	FUBAR	–			6		
NA	FEL	–	–	–	2	4	137, **179**, **303**, 426
	IFEL	–			2		
	FUBAR	–			4		
M1	FEL	–	–	–	4	20	4, 12, **24**, **25**, 96, **109**, 134, **137**, 138,
	REL	–			20		139, **146**, 166, 169, 183, **205**, 207,
	IFEL	–			1		219, 230, 239, 248
	FUBAR	–			5		
NS1	REL	159	159	3, 5–13, 15, 18, 19, 21, 23, 24, 28, 30, 31, 33–38, 40–59, 61, 62, 64–70, 72, 73, 76, 81–88, 90–92, 94, 95, 99, 100, 103, 105–107, 109, 111–118, 120, 122–125, 127–130, 132, 134, 136–138, 140, 141, 143–149, 151, 154–158, 160–168, 170–172, 174, 177–184, 190–195, 197–202, 204–206, 208, 210–217	–	–	–
NS2	FEL	–	–		1	1	**92**
	FUBAR	–			1		

No statistically reliable positively and negatively selected site is found for M2. Positively and negatively selected common amino acid sites identified by more than one method in the Datamonkey are denoted by bold with underline. For instance, site **627** (bold with underline) of PB2 is notified as a common site that could be identified by both REL and FUBAR.

[35] reported a total of 43 positively selected sites in HA of the human H3N2 influenza viruses using Bayes and Naive Empirical Bayes analyses. Markedly, Furuse et al. [26] reported a total of 8, 4 and 2 positively selected amino acid sites in the entire HA gene of seasonal H1N1, swine H1 and 2009 H1N1 viruses, respectively using only FEL method [22]. Li et al. [34] reported 9 and 2 positively selected sites in HA and NA of pandemic influenza H1N1 viruses, respectively, using both SLAC and FEL methods; however, only 4 and 2 sites of HA and NA, respectively, could be identified by FEL method, but not by SLAC method. In my previous study on the 2009 pandemic human influenza A/H1N1 virus, I have reported a total of 9 amino acid sites including a single site from PB2 and M2, 2 sites from PB1 and NS1 and 3 sites from HA that were under strong positive selection pressures [10]. In the present study, just a single amino acid site 627 (E) of PB2 of A/H7N9 virus was considered as a highly variable (strong) site (Tables 1, 3; Supplementary Fig. 1). Nonetheless, no amino acid site was inferred as positively selected for all other proteins including PB1, PB1-F2, PA, HA, NP, NA, M1, M2, NS1 and NS2. Previous reports have to date not identified the same positively selected site highlighted in this paper [3,10,26,30,34,35]. This may be due to different factors, for instance, different viral strain sequences and statistical methods used for the natural selection analyses. It is noted that all previous analyses, except for Arunachalam [10], were carried out using a single statistical method. Consequently, the positively selected sites identified in such studies, might remain tentative. In contrast, the positively selected site identified in the present study is based on multiple statistical methods and therefore predicted to be more reliable.

2.2. Purifying/negative selection

Individual amino acid sites of novel A/H7N9 virus under purifying selection pressures are comparatively high. Tables 2–4 and Supplementary Fig. 1 illustrate the results of the above fact. As mentioned previously, the statistically reliable amino acid sites that have been detected by more than one method are finally considered as negatively selected [22,23]. The present results show that a total of 68 out of 89 negatively selected amino acid sites, were inferred with statistical significance. Among these 68 statistically reliable sites, only 26 sites of PB2, PB1, PA, HA, NP, NA, M1 and NS2 were identified under strong purifying selection pressures, since they have been detected by more than one method. The number of amino acid sites under purifying selection was comparatively more for M1 (6 sites) protein than other proteins. Typically, amino acid sites undergo negative selection more when a few positions can evolve by means of positive selection. For example, in this study just a single site (627) of PB2 was inferred as positively selected, whereas, a total of 26 sites were inferred as negatively selected. These findings evidently indicate that the amino acid sites of the A/H7N9 virus, subject to negative selections, evolve faster than those under positive selection. Also, it obviously suggests that the number of amino acid sites under negative selection pressure is extremely higher than those sites under positive selection pressure. This observation confirms the hypotheses of Kimura [36]. The present observation is partially in agreement with the findings of Sant'Anna et al. [37] who performed the natural selection analysis on the pandemic A/H1N1 viruses isolated in the state of Rio Grande do Sul, Brazil. Their finding shows that no gene is under positive selection, whereas, PB2, PB1, PA, HA, NP, NA and M1 genes are under purifying selection. Espinola (2012) reported that the external genes HA and NA of pandemic A/H1N1 viruses evolved through purifying selection pressures [38]. The findings of Sant'Anna et al. [37] and Espinola [38] might not be reliable because of two reasons, both studies i) are carried out just by using a single statistical method and ii) inferred the selection pressures at the entire gene level, but not at the individual codon/amino acid site level. Prominently, Barrero et al. [39] have employed multiple methods to detect the selection pressures on pandemic A/H1N1 viruses isolated from Argentina. They

Table 4
Functional annotations of positively and negatively selected sites of the novel human influenza A/H7N9 virus.

Protein	Statistically reliable amino acid sites	Method(s)	dN–dS/ω[a]	*p*-Value/posterior prob. /Bayes factor[b]	Functionally known epitope length (from–to)	IEDB-Epitope ID[c]	Function(s)	Host organism(s)	Reference(s)
Positively selected amino acid site									
PB2	627	REL;	1.22;	4059.28;	619–633	1035951[d]	TCE	*Mus musculus*	[47]
		FUBAR	8.19	0.94	618–634	1766537[d]	TCE	*Homo sapiens*	[48]
Negatively selected amino acid sites									
PB2	185	FUBAR;	−6.92;	0.92;	176–190	27386[d]	TCE	*Mus musculus*	[47]
		FEL	−1.10E+02	0.03					
					181–195	60501[d]	TCE	*Mus musculus*	[47]
					173–189	128680[d]	TCE	*Homo sapiens*	[48]
					185–201	129003[d]	TCE	*Homo sapiens*	[48]
	490	FUBAR;	−6.07;	0.91;	487–495	15092[d]	TCE	*Mus musculus*	[70]
		FEL	−7.23E+01	0.05					
	490; 499				489–503	61379[d]	TCE	*Mus musculus*	[47]
	499	FUBAR;	−6.49;	0.91;	494–510	130316[d]	TCE	*Homo sapiens*	[48]
		FEL	−1.31E+02	0.04					
	652	FUBAR;	−6.19;	0.91;	637–655	173541	TCE	*Homo sapiens*	[71]
		FEL	−7.98E+01	0.05					
PB1	738; 740	FEL; FUBAR	−1.85E+02;	0.05;	728–744	128898[d]	TCE	*Homo sapiens*	[48]
			−6.18 &	0.91 &					
			−6.98E+02;	0.03;					
			−6.05	0.91					
					734–750	129759[d]	TCE	*Homo sapiens*	[48]
PA	62	FEL; FUBAR	−1.67E+02;	0.01;	–	–	–	–	–
			−6.35	0.93					
	608; 609	FEL; FUBAR	−1.63E+02;	0.03;	587–636	97592	**BCE**	*Homo sapiens*	[72]
			−6.17 &	0.92 &					
			−7.53E+01;	0.03;					
			−8.53	0.97					
					604–618	12648[d]	TCE	*Mus musculus*	[47]
					599–613	14282[d]	TCE	*Mus musculus*	[47]
	681	FEL; FUBAR	−8.59E+01;	0.02;	674–688	9500[d]	TCE	*Mus musculus*	[47]
			−8.13	0.97					
HA	43	FEL; FUBAR	−2.62E+02;	0.02;	–	–	–	–	–
			−6.39	0.93					
	113	FEL; FUBAR	−2.02E+02;	0.03;	–	–	–	–	–
			−6.27	0.93					
	365, 366	FEL; FUBAR	−1.03E+02;	0.04;	356–372	128557[d]	TCE	*Homo sapiens*	[48]
			−5.96 &	0.92 &					
			−8.07E+01;	0.05;					
			−5.43	0.90					
NP	129	FEL; FUBAR	−8.35E+01;	0.05;	113–132	30347	TCE (HLA-DRB1*11:01)	*Homo sapiens*	[73]
			−5.49	0.91					
						143419	TCE (HLA-DRB1*03:01)	*Homo sapiens*	[74]
					110–129	164290	TCE	*Homo sapiens*	[75]
	286	FEL; FUBAR	−8.00E+01;	0.04;	282–300	21577	TCE	*Homo sapiens*	[76]
			−6.27	0.93					
					276–290	38391	TCE	*Mus musculus*	[47]
					276–292	129288	TCE (H-2-IAs)	*Mus musculus*	[77]
	480	FEL; FUBAR	−1.42E+02;	0.04;	473–481	60089	TCE (HLA-B*07:02)	*Mus musculus*	[78]
			−5.49	0.90					
					470–487	97410	TCE	*Homo sapiens*	[79]
					442–480	97680	**BCE**	*Homo sapiens*	[72]
NA	179	FEL;	−5.66E+02;	0.006;	176–192	1765922	TCE	*Homo sapiens*	[48]
		IFEL;	−5.66E+02;	0.03;					
		FUBAR	−6.53	0.94					
	303	FEL;	−3.26E+02;	0.01;	291–307	1765863[d]	TCE	*Homo sapiens*	[48]
		IFEL;	−3.26E+02;	0.05;					
		FUBAR	−6.28	0.93					
M1	24, 25	FEL;	−1.68E+02;	0.04;	18–29	21686	TCE (HLA-DRA*01:01/ DRB1*01:01; HLA-DR1)	*Homo sapiens*	[80,81]
		REL;	−2.39;	17541.5;					
		IFEL &	−1.68E+02 &	0.05 &					
		FEL;	−3.31E+01;	0.05;					
		REL;	−2.39;	20025;					
		FUBAR	−5.39	0.94					
					19–31	48376	TCE (HLA-DRB1*01:01; HLA-DRB1*01:02)	*Homo sapiens*	[82]
					8–36	97289	**BCE**	*Homo sapiens*	[72]
					1–51	97507	**BCE**	*Homo sapiens*	[72]
	109	FEL;	−5.37E+01;	0.04;	97–116	1271748	TCE (HLA-DRB1*01:03)	*Homo sapiens*	[83]
		REL;	−2.39;	19064.4;					
		FUBAR	−4.22	0.91					
						1478209	TCE (HLA-DRB1*07:01)	*Homo sapiens*	[73]

(continued on next page)

Table 4 *(continued)*

Protein	Statistically reliable amino acid sites	Method(s)	dN–dS/ω[a]	*p*-Value/posterior prob. /Bayes factor[b]	Functionally known epitope length (from–to)	IEDB-Epitope ID[c]	Function(s)	Host organism(s)	Reference(s)
					101–113	1512752	TCE (HLA-DRA*01:01/ DRB1*04:01)	*Homo sapiens*	[84]
					89–119	1003682	**BCE**	*Mus musculus*	[85]
	137	REL; FUBAR	−2.39; −3.98	19246.7; 0.91	121–140	1478196	TCE (HLA-DRB1*04:04)	*Homo sapiens*	[73]
					121–137	1765990	TCE (HLA-B*35:01)	*Homo sapiens*	[48]
	146	REL; FUBAR	−2.39; −4.20	1.44E+08; 0.96	146–160	1036074[d]	TCE	*Mus musculus*	[47]
	205	FEL; REL; FUBAR	−3.26E+01; −2.39; −4.79	0.04; 24927.7; 0.94	201–220	1764869	TCE (HLA-DRB1*11:01)	*Homo sapiens*	[86]
						1764871	TCE (HLA-DRB1*13:01)	*Homo sapiens*	[86]
					196–205	1828514	TCE (HLA-B*18:01)	*Homo sapiens*	[87]
					188–226	1005256	**BCE**	*Mus musculus*	[88]
NS2	92	FEL; FUBAR	−3.42E+02; −4.96	0.02; 0.92	91–105	1007685	TCE	*Mus musculus*	[47]

It is noted that no purifying selected site is inferred for M2 and among the 5 purifying selected sites, no statistically significant site is observed for NS2.

[a] It is appropriately scaled dN–dS from SLAC/FEL/IFEL/REL of Datamonkey.

[b] It shows *p*-values of the SLAC/FEL/IFEL (or) the posterior probability of FUBAR method (or) the Bayes factor values of the REL method (the posterior probabilities are included just for reference).

[c] Epitope identification codes of functionally known epitopes were obtained from the Immune Epitope Database and Analysis Resource (IEDB) (www.immuneepitope.org).

[d] These epitopes are categorized as negative in the qualitative measurement. Experimentally derived epitopes that are available in IEDB are quantitatively categorized either as positive or negative based on the B cell and/or T cell assays (ELISA/FACs/bioassay, etc.) 'immunogen epitope relation' reactions, where the epitopes of influenza are recognized by antisera of influenza infected host human/mouse; the immunogen that the host was exposed to was the influenza virus [52]. The present results show few quantitatively negative epitopes that should also be important to verify experimentally, whether it will produce positive response against the novel A/H7N9 virus.Experimentally identified epitopes, epitope identification codes, host organisms (which were used to expose the immunogen) and the literature information interpreted in this study, were collected from the IEDB.

have inferred a total of 11 and 17 amino acid sites of HA and NA respectively, under strong purifying selections. In my previous study on pandemic A/H1N1 viruses, I have identified 15 sites of PB2, PB1, PA, HA, NP, NA, M1 and NS1 proteins, as negatively selected using multiple statistical methods (unpublished report). In order to confirm, whether both the pandemic A/H1N1 and novel A/H7N9 viruses evolved through identical negatively selected amino acid sites, I have compared the present results with the observations of Barrero et al. [39] and my unpublished work, and found no identical negatively selected sites. It could be attributed to the use of different subtype sequences and different statistical methods for the adaptive selection analyses. Notably, negatively selected sites identified in the present study are based on the multiple methods obviously indicating that the strength of the amino acid sites under strong negative selection pressures will be more reliable. Although, there was no purifying selected site inferred from the M2 and NS1 proteins, a single statistically reliable amino acid site 24 of PB1-F2 has not been finally considered as negatively selected, because this site was just inferred by a single method.

2.3. Amino acid sites and their functional annotations

The human immune response against viruses highly depends on the antibodies' ability to recognize the conserved epitopes of the viruses. As a part of the immune response, T-cells provide a form of protection by identifying T-cell epitopes (TCEs) of the viruses [3,40,41]. On the other hand, B-cells are neutralizing the viral infectivity by recognizing the respective B-cell epitopes (BCEs) [42]. In addition to these natural immune responses, vaccines are promoting acquired immunity against viral infections. A distinctive method of producing a vaccine is to cultivate the influenza viruses in embryonated chicken eggs [43]. At this point, amino acid sites that facilitate the virus to adapt and grow up in an embryonated chicken egg can be identified by means of comparing the genomic sequences of the control virus and isolates from the embryonated chicken egg [44]. Apart from the natural and acquired immune responses, specific drugs are also available to successfully eliminate the viruses. A few amino acid sites of viruses are found to be involved in the resistance to drugs like Amantadine and Oseltamivir [3,45,46]. Hence, in the present study, there it is also important to understand the functions of amino acid sites of the novel A/H7N9 virus that are under strong natural selection pressures (Table 4).

The biological function of the single positively selected amino acid site 627 (E) of PB2 is identified as TCE [47,48]. This positively selected site in the T-cell antigenic region is possibly caused by vaccination and continuous use of anti-viral drugs. This site might help the A/H7N9 virus to better adapt to the human host and can reduce the efficiency of existing vaccines [34]. In addition, this site is known to enhance the replication of avian influenza viruses in mammals and is also identified as a marker of mammalian adaptation [13,49]. Further experimental analyses are critical to understand whether any potential changes that will affect the viral transmissibility between humans and their pathogenicity in birds. Notably, no positively selected site is found to be involved in the BCE, antiviral resistance and growth in eggs. Interestingly, the function of negatively selected amino acid positions 608 (T) and 609 (K) of PA, 480 (D) of NP, and 24 (I), 25 (A), 109 (F) and 205 (V) of M1 were found to be involved in both BCEs and TCEs. However, a total of sixteen sites, which included 185 (I), 490 (S), 499 (D) and 652 (N) of PB2, 738 (E) and 740 (F) of PB1, 681 (F) of PA, 365 (H) and 366 (Q) of HA, 129 (A) and 286 (A) of NP, 179 (S) and 303 (D) of NA, 137 (T) and 146 (L) of M1, and 92 (N) of NS2, were found to be involved only in TCEs (Table 4; Supplementary Fig. 1). The function of positions 62 (I) of PA, 43 (E) and 113 (E) of HA are unknown; however, the adjacent site 44 of HA is involved in BCE (epitope C) [3,50], suggesting that site 43 might also be involved in this function. No adjacent site with a known function was found for other functionally unknown sites such as 62 of PA and 113 of HA. Overall, functions for 23 out of 26 negatively selected amino acid sites were exactly inferred with known immune epitopes. Amino acid site 226 of HA was inferred with statistical significance (posterior probability 0.91) by single method FUBAR. It is imperative to discuss here as its substitution has been associated with reduced binding to avian-like receptors, with sialic acids linked to galactose by α-2,3 linkages that are found in the human lower respiratory tract. This is also associated with enhanced ability to bind to mammalian-like receptors bearing sialic acids linked to galactose by α-2,6 linkages, which are located in the upper airways of humans and other mammals. Experimentally, the Q226L of HA of A/H5N1 avian influenza virus was found to be linked with increased

evidence of respiratory transmission between ferrets [12,16,51]. Experimentally derived epitopes available in the Immune Epitope Database (IEDB) and Analysis Resource are quantitatively categorized either as positive or negative, based on the B-cell and/or T-cell assays (ELISA/FACs/bioassay, etc.) 'immunogen epitope relation' reactions, where the epitopes of influenza could be recognized by antisera of infected host like human, mouse, etc. [52]. The present results also show that a few quantitatively negative epitopes (Table 4), which will also be important to verify experimentally, will produce a positive response against the novel A/H7N9 virus. It would be interesting to examine the functions of both positive and negatively selected sites of the A/H7N9 virus, experimentally, using site directed mutagenesis. Remarkably, no negatively selected site is found to be involved in anti-viral drug resistance and growth in eggs. It should be noted that at present no vaccine is available for the prevention of the novel A/H7N9 virus. It is unclear whether the current candidate H7 vaccine viruses including three of North American viruses and the other three are avian viruses (e.g. A/Mallard/Netherlands/12/2000 (H7N3)) from the Netherlands might be effective. Recently, the influenza H7N9 A/Shanghai/2/2013 and wild type A/Anhui/1/2013 strains have been proposed to be the appropriate vaccine candidates; since, these have passed relevant safety testing and two-way HA inhibition tests [53]. In addition, the A/Anhui/1/2013 strain is found to be growing well in eggs [13]. The anti-viral drug NA inhibitors including Oseltamivir and Zanamivir are fairly effective in treating an early infection of the novel A/H7N9 virus [54].

The low pathogenic avian-origin A/H7N9 virus can more possibly become highly pathogenic in humans [28]. A single positively selected site of the A/H7N9 virus involved in TCE is supported by a statistical BF value of 4059.3 and a posterior probability of 0.94 (Tables 3, 4). The success rate of human antibody protection against viruses highly depends on how the antibodies can recognize the conserved BCEs and TCEs of viruses [41]. Amino acid mutations in many TCEs are expected to be advantageous, since, the haplotype of HLA restricts T-cells to recognize TCEs whereas it fails to restrict B-cells to recognize BCEs [3,55]. But, in the present study, one positively selected site and 23 negatively selected sites are identified to be involved in TCEs indicating that the majority of the amino acid mutations in A/H7N9 virus are deleterious. In support of this issue, a majority of the sites under negative selection are found to be involved in TCEs than in BCEs. Markedly, negatively selected sites are involved only in TCEs (p values 0.006–0.05; BFs 19246.7–1.44E+08) are efficient when compared to the sites involved in both BCEs and TCEs (p values 0.03–0.05; BFs 17541.5–24927.7) (Table 4). It apparently indicates that both humoral and cellular immune responses are engaged in the elimination of the A/H7N9 viruses in humans [56]. The present results reported are consistent with earlier reports [10,56,57], where they also observed both types of immune responses against hepatitis C viruses and influenza A viruses. Vaccines are available to induce immune responses against BCEs of influenza viruses [58]. Almost all the negatively selected sites of polio viruses were apparently involved in BCEs, and the vaccine based on these sites was known to be tremendously successful [59]. In the present study, 7 amino acid sites, which include a single site of NP, two sites of PA and four sites of M1 (p 0.03–0.05) under negative selection are found to be involved in both BCEs and TCEs (Table 4). Conspicuously, these negatively selected sites most likely have no potential to generate immune-escape variants, mainly due to the strong functional constraints operating them. Purifying selected sites under strong functional constraints will be considered as suitable targets for developing candidate vaccines and drugs, as most of the substitutions at these amino acid sites are probably intolerable [59–61]. Developing vaccines and drugs for multiple targets would be preferable due to its less susceptibility to make escape mutants [3,62–64]. Thus, amino acid sites engaged in both BCEs and TCEs are measured as highly suitable targets, as these sites would play a principal role in inducing strong humoral and cellular immune responses against targets. In the context of these findings, developing a successful vaccine against A/H7N9 viruses would require further experimental study to validate the accurate role of these sites in developing candidate vaccine.

3. Materials and methods

3.1. Sequence selection

The entire protein-coding regions of the avian-origin human influenza A/H7N9 viruses, which caused the recent outbreak in China, were extracted from the NCBI Influenza Virus Resource [65]. Among the eight gene segments, the five segments PB2, PA, HA, NP and NA encode a single protein whereas the rest of the three segments PB1, MP and NS encode two proteins, namely, PB1 and PB1-F2, M1 and M2, and, NS1 and NS2, respectively. Nucleotide positions 95–367 of PB1 and the entire region (positions 1–273) of PB1-F2, positions 1–27 of both M1 and M2 and positions 1–30 of both NS1 and NS2 overlapped in different reading frames. I have retrieved (dated 31-05-2013) all the available PB2, PB1, PB1-F2, PA, HA, NP, NA, M1, M2, NS1 and NS2 gene sequences of novel A/H7N9 viruses. In order to finalize the sequences for natural selection analysis, the following circumstances were taken into consideration for elimination, i) sequences which are bemused (different virus species and/or segments selected), ii) sequences derived from the same strains as others, iii) laboratory and vaccine strains, and iv) shorter sequences. The A/H7N9 viral strains names and their databank accession numbers employed in the present study are listed in Supplementary Table 1.

3.2. Natural selection analysis

Natural selection pressures operating on individual codon sites of A/H7N9 viruses were detected by computing 'ω', the ratio between non-synonymous (dN) and synonymous (dS) substitutions. In general, the ratio ($\omega = dN / dS$) has a straightforward measurement of selective pressure at each codon of protein-coding genes with 'ω' values of >1, 1 and <1 indicating diversifying, neutral and purifying selection, respectively [10,35,66]. For selection analysis, the CML, HB and BS methods under different substitution models were employed from Datamonkey [23], to estimate the 'ω' value for each codon of the A/H7N9 virus protein-coding genes. The methods are well described in detail [10,22,67–69]. In brief, the selection analyses were made with four independent CML methods namely SLAC, FEL, REL [67] and IFEL [22], one HB method namely FUBAR [69], and one BS evolutionary method namely MEME [68]. Interestingly, for detecting the selection pressures, the SLAC method estimates the number of non-synonymous and synonymous changes at each codon based on the evolutionary history of the sequences. The FEL method estimates the non-synonymous and synonymous substitution rates at each site directly, whereas, IFEL estimates the selection pressure at each codon of the sequences sampled from a population (i.e. along internal branches) [22]. The REL method works based on the predefined distribution in terms of selection pressure at each amino acid site inferred by using empirical Bayes approach. The FUBAR method uses a MCMC approach, but it ensures sturdiness against model misspecification by averaging over a large number of predefined site classes. This leaves the distribution of selection parameters basically unconstrained [69]. The MEME method used to find out the variation over branches (not just sites) is described in the REL framework [68]. Apart from the diverse or slightly diverse principles, the aforementioned methods are working based on the principle of underlying phylogeny and codon substitution model that allow for a reasonable comparison between different approaches. In this way, I have employed these six independent approaches on novel A/H7N9 virus coding genes to perform a series of simulations to evaluate the statistical properties of each approach.

The coding regions of A/H7N9 viruses were used as input in Datamonkey; consequently this server automatically translated the protein-coding genes into respective protein sequences. The best

model was identified among the 203 nucleotide substitution models for each method for each gene/protein via a model selection test (Table 1). The differential test significance levels were given as 0.1 for SLAC, FEL, MEME and IFEL analyses, 0.9 for FUBAR and the empirical Bayes factor (BF) was given as 50 for REL. The 'ω' ratio of amino acid sites under selection pressures either with p-values ≤ 0.05 or posterior probability ≥ 0.9 or Bayes factors ≥ 20 were considered statistically significant. The location of the amino acid sites under the positive and negative selection pressures in the entire protein sequences of the novel A/H7N9 virus were highlighted with red and blue colors, respectively, using BioEdit v7.1.3 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) (Supplementary Fig. 1). Moreover, in order to understand the functions of amino acid sites under strong positive and purifying selection pressures, I have annotated these sites with functionally known amino acid sites, which are involved in immune epitopes, growth in eggs and anti-viral drug resistance. The IEDB [52] was used as a source to retrieve the functionally verified immune epitopes, whereas, for growth in eggs and anti-viral drug resistance, the findings of earlier reports [3,44–46] were used as a source for annotations.

Acknowledgments

The author is grateful to Annapurna Vyakarnam (King's College London, UK and the Indian Institute of Science (IISc), India), Vaithi Arumugaswami and Deisy Contreras (Cedars-Sinai Medical Center and the University of California, Los Angeles, USA) for their valuable comments. The author also thanks G. Annadurai (Manonmaniam Sundaranar University, India) for providing partial facilities to carry out this work and Bharath K Sundararaj (IISc, India) for his help with the viral diagram. The author is grateful to the two anonymous reviewers for their fruitful comments. No funders had a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2014.10.012.

References

[1] W. Smith, C.H. Andrewes, P.P. Laidlaw, A virus obtained from influenza patients, Lancet 225 (1933) 66–68.
[2] Y. Suzuki, M. Nei, Origin and evolution of influenza virus hemagglutinin genes, Mol. Biol. Evol. 19 (2002) 501–509.
[3] Y. Suzuki, Natural selection on the influenza virus genome, Mol. Biol. Evol. 23 (2006) 1902–1911.
[4] WHO (World Health Organisation), A revision of the system of nomenclature for influenza viruses: a WHO memorandum, Bull. World Health Organ. 58 (1980) 585–591.
[5] M. Hatta, Y. Hatta, J.H. Kim, S. Watanabe, K. Shinya, T. Nguyen, P.S. Lien, Q.M. Le, Y. Kawaoka, Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice, PLoS Pathog. 3 (2007) e133.
[6] A. Gambotto, S.M. Barratt-Boyes, M.D. Jong, G. Neumann, Y. Kawaoka, Human infection with highly pathogenic H5N1 influenza virus, Lancet 371 (2008) 1464–1475.
[7] A.U. Tamuri, D.M. Reis, A.J. Hay, R.A. Goldstein, Identifying changes in selective constraints: host shifts in influenza, PLoS Comput. Biol. 5 (2009) e1000564.
[8] R. Arunachalam, K. Paulkumar, G. Annadurai, Phylogenetic analysis of pandemic influenza A/H1N1 virus, Biologia 67 (2012) 14–31.
[9] R. Arunachalam, K. Paulkumar, G. Annadurai, Genetic ancestor of external antigens of pandemic influenza A/H1N1 virus, Interdisc. Sci. Comput. Life Sci. 4 (2012) 282–290.
[10] R. Arunachalam, Detection of site-specific positive Darwinian selection on pandemic influenza A/H1N1 virus genome: integrative approaches, Genetica 141 (2013) 143–155.
[11] T. Noda, H. Sagara, A. Yen, A. Takada, H. Kida, H. Cheng, Y. Kawaoka, Architecture of ribonucleoprotein complexes in influenza A virus particles, Nature 439 (2006) 490–492.
[12] D. Liu, W. Shi, Y. Shi, D. Wang, H. Xiao, W. Li, Y. Bi, Y. Wu, X. Li, J. Yan, W. Liu, G. Zhao, W. Yang, Y. Wang, J. Ma, Y. Shu, F. Lei, G.F. Gao, Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses, Lancet 381 (2013) 1926–1932.
[13] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, X. Xu, H. Lu, W. Zhu, Z. Gao, N. Xiang, Y. Shen, Z. He, Y. Gu, Z. Zhang, Y. Yang, X. Zhao, L. Zhou, X. Li, S. Zou, Y. Zhang, X. Li, L. Yang, J. Guo, J. Dong, Q. Li, L. Dong, Y. Zhu, T. Bai, S. Wang, P. Hao, W. Yang, Y. Zhang, J. Han, H. Yu, D. Li, G.F. Gao, G. Wu, Y. Wang, Z. Yuan, Y. Shu, Human infection with a novel avian-origin influenza A (H7N9) virus, N. Engl. J. Med. 368 (2013) 1888–1897.
[14] CDC (Centers for Disease Control and Prevention), Genetic evolution of H7N9 virus in China, www.cdc.gov/flu/pdf/avianflu/h7n9-reassortment-diagram.pdf 2013 [Retrieved on November 10, 2013, Available from].
[15] G.-W. Chen, M.M.C. Lai, S.-C. Wu, S.-C. Chang, L.-M. Huang, S.-R. Shih, Is avian influenza A (H7N9) virus staggering its way to humans? J. Formos. Med. Assoc. 112 (2013) 312–318.
[16] URRA (Updated Rapid Risk Assessment), Human infection with a novel avian influenza virus, A(H7N9) — China, http://ecdc.europa.eu/en/publications/Publications/influenza-A(H7N9)-China-rapid-risk-assessment-8-may-2013.pdf 2013 [Released on May 08, 2013, Available from].
[17] R.J. Webby, R.G. Webster, Emergence of influenza A viruses, Philos. Trans. R. Soc. B 356 (2001) 1817–1828.
[18] N. Ding, N. Wu, Q. Xu, K. Chen, C. Zhang, Molecular evolution of novel swine-origin A/H1N1 influenza viruses among and before human, Virus Genes 39 (2009) 293–300.
[19] WHO (World Health Organisation), Western Pacific Region. Transcript of press conference in Beijing — The China–WHO joint mission on H7N9 assessment 2013, http://www.wpro.who.int/china/topics/h7n9_influenza/transcript_20130424/en/index.html 2013.
[20] R.J. Webby, R.G. Webster, Are we ready for pandemic influenza? Science 302 (2003) 1519–1522.
[21] WHO (World Health Organisation), Public health relevant virological features of Influenza A(H7N9) causing human infection in China, http://www.euro.who.int/__data/assets/pdf_file/0008/186677/050413-H7N9-influenza-viruses-Virologic-features_update.pdf 2013.
[22] S.L.K. Pond, S.D.W. Frost, Z. Grossman, M.B. Gravenor, D.D. Richman, A.J. Brown, Adaptation to different human populations by HIV-1 revealed by codon-based analyses, PLoS Comput. Biol. 2 (2006) e62.
[23] W. Delport, A.-F.Y. Poon, S.D.W. Frost, S.L.K. Pond, Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology, Bioinformatics 26 (2010) 2455–2457.
[24] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford University Press, New York, 2000.
[25] L. Duret, Neutral theory: the null hypothesis of molecular evolution, Nat. Educ. 1 (1) (2008) [Available from http://www.nature.com/scitable/topicpage/neutral-theory-the-null-hypothesis-of-molecular-839].
[26] Y. Furuse, K. Shimabukuro, T. Odagiri, R. Sawayama, T. Okada, I. Khandaker, A. Suzuki, H. Oshitani, Comparison of selection pressures on the HA gene of pandemic (2009) and seasonal human and swine influenza A H1 subtype viruses, Virology 405 (2010) 314–321.
[27] G. Neumann, H. Chen, G.F. Gao, Y. Shu, Y. Kawaoka, H5N1 influenza viruses: outbreaks and biological properties, Cell Res. 20 (2010) 51–61.
[28] Z. QingYe, W. SuChun, W. MeiLi, L. Shuo, J. WenMing, H. GuangYu, L. JinPing, W. KaiCheng, Y. JianMin, C. JiMing, C. JiWang, Epidemiological and risk analysis of the H7N9 subtype influenza outbreak in China at its early stage, Chin. Sci. Bull. 58 (2013) 3183–3187.
[29] M. Matrosovich, A. Tuzikov, N. Bovin, A. Gambaryan, A. Klimov, M.R. Castrucci, I. Donatelli, Y. Kawaoka, Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals, J. Virol. 74 (2000) 8502–8512.
[30] L. Campitelli, M. Ciccozzi, M. Salemi, F. Taglia, S. Boros, I. Donatelli, G. Rezza, H5N1 influenza virus evolution: a comparison of different epidemics in birds and humans (1997–2004), J. Gen. Virol. 87 (2006) 955–960.
[31] Y.I. Wolf, C. Viboud, E.C. Holmes, E.V. Koonin, D.J. Lipman, Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus, Biol. Direct 1 (2006) 34.
[32] J. Shen, J. Ma, Q. Wang, Evolutionary trends of A(H1N1) influenza virus hemagglutinin since 1918, PLoS ONE 4 (2009) e7789.
[33] D.A. Janies, I.O. Voronkin, J. Studer, J. Hardman, B.B. Alexandrov, T.W. Treseder, C. Valson, Selection for resistance to oseltamivir in seasonal and pandemic H1N1 influenza and widespread co-circulation of the lineages, Int. J. Health Geogr. 9 (2010) 13.
[34] W. Li, W. Shi, H. Qiao, S.Y.M. Ho, A. Luo, Y. Zhang, C. Zhu, Positive selection on hemagglutinin and neuraminidase genes of H1N1 influenza viruses, Virol. J. 8 (2011) 183.
[35] J. Chen, Y. Sun, Variation in the analysis of positively selected sites using nonsynonymous/synonymous rate ratios: an example using influenza virus, PLoS ONE 6 (2011) e19996.
[36] M. Kimura, The neutral theory of molecular evolution: a review of recent evidence, Jpn. J. Genet. 66 (1991) 367–386.
[37] F.H. Sant'Anna, L.G.A. Borges, P.R.V. Fallavena, T.S. Gregianini, F. Matias, R.A. Halpin, D. Wentworth, P.A. d'Azevedo, A.B.G. Veiga, Genome analysis of pandemic and post-pandemic influenza A pH1N1 viruses isolated in Rio Grande do Sul, Brazil, Arch. Virol. 159 (2014) 621–630.
[38] E.E. Espinola, Genome stability of pandemic influenza A(H1N1) 2009 based on analysis of hemagglutinin and neuraminidase genes, Open Virol. J. 6 (2012) 59–63.
[39] P.R. Barrero, M. Viegas, L.E. Valinotto, A.S. Mistchenko, Genetic and phylogenetic analyses of influenza A H1N1pdm virus in Buenos Aires, Argentina, J. Virol. 85 (2011) 1058–1066.
[40] A.L. Hughes, M. Nei, Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, Nature 335 (1988) 167–170.
[41] M.F. Cusick, S. Wang, D.D. Eckels, In vitro responses to avian influenza H5 by human CD4 T cells, J. Immunol. 183 (2009) 6432–6441.
[42] J. Klein, V. Horejsi, 2nd edn, ImmunologyBlackwell Science, Oxford, 1997.
[43] K.A. Smith, C.J. Colvin, P.S.D. Weber, S.J. Spatz, P.M. Coussens, High titer growth of human and avian influenza viruses in an immortalized chick embryo cell line without the need for exogenous proteases, Vaccine 26 (2008) 3778–3782.

[44] C.T. Hardy, S.A. Young, R.G. Webster, C.W. Naeve, R.J. Owens, Egg fluids and cells of the chorioallantoic membrane of embryonated chicken eggs can select different variants of influenza A (H3N2) viruses, Virology 211 (1995) 302–306.
[45] A.J. Hay, A.J. Wolstenholme, J.J. Skehel, M.H. Smith, The molecular basis of the specific anti-influenza action of amantadine, EMBO J. 4 (1985) 3021–3024.
[46] L.V. Gubareva, L. Kaiser, F.G. Hayden, Influenza virus neuraminidase inhibitors, Lancet 355 (2000) 827–835.
[47] S.R. Crowe, S.C. Miller, D.M. Brown, P.S. Adams, R.W. Dutton, A.G. Harmsen, F.E. Lund, T.D. Randall, S.L. Swain, D.L. Woodland, Uneven distribution of MHC class II epitopes within the influenza virus, Vaccine 24 (2006) 457–467.
[48] J.A.B. Babon, J. Cruz, L. Orphin, P. Pazoles, M.D.T. Co, F.A. Ennis, M. Terajima, Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of CD4+ T-cell responses to internal proteins, hemagglutinins, and neuraminidases, Hum. Immunol. 70 (2009) 711–721.
[49] T. Kageyama, S. Fujisaki, E. Takashita, H. Xu, S. Yamada, Y. Uchida, G. Neumann, T. Saito, Y. Kawaoka, M. Tashiro, Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013, Euro. Surveill. 18 (2013) [pii:20453].
[50] D.C. Wiley, I.A. Wilson, J.J. Skehel, Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation, Nature 289 (1981) 373–378.
[51] M. Jonges, A. Meijer, R. Fouchier, G. Koch, J. Li, J.C. Pan, H. Chen, Y.L. Shu, M.P. Koopmans, Guiding outbreak management by the use of influenza A(H7N9) virus sequence analysis, Euro. Surveill. 18 (2013) [pii:20460].
[52] R. Vita, L. Zarebski, J.A. Greenbaum, H. Emami, I. Hoof, N. Salimi, R. Damle, A. Sette, B. Peters, The immune epitope database 2, Nucleic Acids Res. 38 (2010) D854–D862.
[53] WHO, Summary of Status of Development and Availability of Avian Influenza A(H7N9) Candidate Vaccine Viruses and Potency Testing Reagents (Dated 26 May 2014), 2014. 1–2 [Retrieved on July 13, 2014].
[54] WHO, Avian Influenza A(H7N9) Virus, 2014. [Retrieved on July 13, 2014].
[55] E.G.M. Berkhoff, E. de-Wit, M.M. Geelhoed-Mieras, A.C.M. Boon, J. Symons, R.A. Fouchier, A.D. Osterhaus, G.F. Rimmelzwaan, Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes, J. Virol. 79 (2005) 11239–11246.
[56] P.G. Thomas, R. Keating, D.J. Hulse-Post, P.C. Doherty, Cell mediated protection in influenza infection, Emerg. Infect. Dis. 12 (2006) 48–54.
[57] Y. Suzuki, T. Gojobori, Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1B, Gene 276 (2001) 83–87.
[58] R.J. Cox, K.A. Brokstad, P. Ogra, Influenza virus: immunity and vaccination strategies: comparison of the immune response to inactivated and live, attenuated influenza vaccines, Scand. J. Immunol. 59 (2004) 1–15.
[59] Y. Suzuki, Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization, Gene 328 (2004) 127–133.
[60] R.M. Bush, W.M. Fitch, C.A. Bender, N.J. Cox, Positive selection on the H3 hemagglutinin gene of human influenza virus A, Mol. Biol. Evol. 16 (1999) 1457–1465.
[61] D.J. Smith, A.S. Lapedes, J.C. de-Jong, T.M. Bestebroer, G.F. Rimmelzwaan, A.D.M.E. Osterhaus, R.A.M. Fouchier, Mapping the antigenic and genetic evolution of influenza virus, Science 305 (2004) 371–376.
[62] E.D. Kilbourne, C. Smith, I. Brett, B.A. Pokorny, B. Johansson, N. Cox, 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. U. S. A. 99 (2002) 10748–10752.
[63] J.R. Gog, G.F. Rimmelzwaan, A.D. Osterhaus, B.T. Grenfell, Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 11143–11147.
[64] H. Jin, H. Zhou, H. Liu, W. Chan, L. Adhikary, K. Mahmood, M.S. Lee, G. Kemble, 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 (2005) 113–119.
[65] Y. Bao, P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, D. Lipman, The influenza virus resource at the National Center for Biotechnology Information, J. Virol. 82 (2008) 596–601.
[66] Z. Yang, J.P. Bielawski, Statistical methods for detecting molecular adaptation, Trends Ecol. Evol. 15 (2000) 496–503.
[67] S.L.K. Pond, S.D.W. Frost, Not so different after all: a comparison of methods for detecting amino acid sited under selection, Mol. Biol. Evol. 22 (2005) 1208–1222.
[68] S.L.K. Pond, B. Murrell, M. Fourment, S.D.W. Frost, W. Delport, K. Scheffler, A random effects branch-site model for detecting episodic diversifying selection, Mol. Biol. Evol. 28 (2011) 3033–3043.
[69] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S.L.K. Pond, K. Scheffler, FUBAR: a Fast, Unconstrained Bayesian AppRoximation for inferring selection, Mol. Biol. Evol. 30 (2013) 1196–1205.
[70] Y. Deng, J.W. Yewdell, L.C. Eisenlohr, J.R. Bennink, MHC affinity, peptide liberation, T cell repertoire, and immunodominance all contribute to the paucity of MHC class I-restricted peptides recognized by antiviral CTL, J. Immunol. 158 (1997) 1507–1515.
[71] T.J. Powell, A. Fox, Y. Peng, L.T.Q. Mai, V.T.K. Lien, N.L.K. Hang, L. Wang, L.Y.-H. Lee, C.P. Simmons, A.J. McMichael, J.J. Farrar, B.A. Askonas, T.N. Duong, P.Q. Thai, N.T.T. Yen, S.L. Rowland-Jones, N.T. Hien, P. Horby, T. Dong, Identification of H5N1-specific T-cell responses in a high-risk cohort in Vietnam indicates the existence of potential asymptomatic infections, J. Infect. Dis. 205 (2012) 20–27.
[72] S. Khurana, Amorsolo L. Suguitan Jr., Y. Rivera, C.P. Simmons, A. Lanzavecchia, F. Sallusto, J. Manischewitz, L.R. King, K. Subbarao, H. Golding, Antigenic fingerprinting of H5N1 avian influenza using convalescent sera and monoclonal antibodies reveals potential vaccine and diagnostic targets, PLoS Med. 6 (2009) [e1000049].
[73] M. Roti, J. Yang, D. Berger, L. Huston, E.A. James, W.W. Kwok, Healthy human subjects have CD4+ T cells directed against H5N1 influenza virus, J. Immunol. 180 (2008) 1758–1768.
[74] J. Yang, E. James, T. Gates, D. Berger, W.W. Kwok, Distinct influenza-Specific CD4+ T cells epitopes restricted to DRB1*1501, DRB1*1502, DRB5*0101, and DRB1*0301, 2009. [(dated 17/08/2013). Available from http://www.iedb.org/refId/1013404].
[75] T.M. Wilkinson, K.F. Li, C.S.C. Chui Chris, A.K.Y. Huang, M. Perkins, J.C. Liebner, R. Lambkin-Williams, A. Gilbert, J. Oxford, B. Nicholas, K.J. Staples, T. Dong, D.C. Douek, A.J. McMichael, X.-N. Xu, Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans, Nat. Med. 18 (2012) 274–280.
[76] S.J. Brett, J. Blau, C.M. Hughes-Jenkins, J. Rhodes, F.Y. Liew, J.P. Tite, Human T cell recognition of influenza A nucleoprotein: specificity and genetic restriction of immunodominant T helper cell epitopes, J. Immunol. 147 (1991) 984–991.
[77] J.L. Nayak, K.A. Richards, F.A. Chaves, A.J. Sant, Analyses of the specificity of CD4 T cells during the primary immune response to influenza virus reveals dramatic MHC-linked asymmetries in reactivity to individual viral proteins, Viral Immunol. 23 (2010) 169–180.
[78] A. Wahl, F. Schafer, W. Bardet, R. Buchli, G.M. Air, W.H. Hildebrand, HLA class I molecules consistently present internal influenza epitopes, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 540–545.
[79] L.Y.-H. Lee, D.L.A. Ha, C. Simmons, M.D. de Jong, N.V.V. Chau, R. Schumacher, Y.C. Peng, A.J. McMichael, J.J. Farrar, G.L. Smith, A.R.M. Townsend, B.A. Askonas, S. Rowland-Jones, T. Dong, Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals, J. Clin. Invest. 118 (2008) 3478–3490.
[80] N.A. Danke, W.W. Kwok, HLA class II-restricted CD4+ T cell responses directed against influenza viral antigens post influenza vaccination, J. Immunol. 171 (2003) 3163–3169.
[81] J.B. Rothbard, R.I. Lechler, K. Howland, V. Bal, D.D. Eckels, R. Sekaly, E.O. Long, W.R. Taylor, J.R. Lamb, Structural model of HLA-DR1 restricted T cell antigen recognition, Cell 52 (1988) 515–523.
[82] D.K. Newton-Nash, D.D. Eckels, Differential effect of polymorphism at HLA-DR1 beta-chain positions 85 and 86 on binding and recognition of DR1-restricted antigenic peptides, J. Immunol. 150 (1993) 1813–1821.
[83] J. Yang, E.A. James, L. Huston, N.A. Danke, A.W. Liu, W.W. Kwok, Multiplex mapping of CD4 T cell epitopes using class II tetramers, Clin. Immunol. 120 (2006) 21–32.
[84] W.W. Kwok, J. Yang, E. James, J. Bui, L. Huston, M. Roti, Identification of 13mer epitopes for DRB1*0101, DRB1*0401, DRB1*0404, and DRB1*0701 restricted CD4+ T Cell Epitopes for Tetanus Toxoid and several Influenza proteins, http://www.iedb.org/refId/10133602009 [Retrieved on 21 Nov, 2013. Available from].
[85] Z.P. Ye, N.W. Baylor, R.R. Wagner, Transcription–inhibition and RNA-binding domains of influenza A virus matrix protein mapped with anti-idiotypic antibodies and synthetic peptides, J. Virol. 63 (1989) 3586–3594.
[86] R. Faner, E. James, L. Huston, R. Pujol-Borrel, W.W. Kwok, M. Juan, Reassessing the role of HLA-DRB3 T-cell responses: evidence for significant expression and complementary antigen presentation, Eur. J. Immunol. 40 (2010) 91–102.
[87] J. Alexander, P. Bilsel, M.-F. del Guercio, A. Marinkovic-Petrovic, S. Southwood, S. Stewart, G. Ishioka, M.F. Kotturi, J. Botten, J. Sidney, M. Newman, A. Sette, Identification of broad binding class I HLA supertype epitopes to provide universal coverage of influenza A virus, Hum. Immunol. 71 (2010) 468–474.
[88] R.W. Hankins, K. Nagata, D.J. Bucher, S. Popple, A. Ishihama, Monoclonal antibody analysis of influenza virus matrix protein epitopes involved in transcription inhibition, Virus Genes 3 (1989) 111–126.