

Evolutionary Trends of A(H1N1) Influenza Virus Hemagglutinin Since 1918

Jun Shen¹, Jianpeng Ma^{1,2*}, Qinghua Wang^{2*}

1 Department of Bioengineering, Rice University, Houston, Texas, United States of America, **2** Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas, United States of America

Abstract

The Pandemic (H1N1) 2009 is spreading to numerous countries and causing many human deaths. Although the symptoms in humans are mild at present, fears are that further mutations in the virus could lead to a potentially more dangerous outbreak in subsequent months. As the primary immunity-eliciting antigen, hemagglutinin (HA) is the major agent for host-driven antigenic drift in A(H3N2) virus. However, whether and how the evolution of HA is influenced by existing immunity is poorly understood for A(H1N1). Here, by analyzing hundreds of A(H1N1) HA sequences since 1918, we show the first evidence that host selections are indeed present in A(H1N1) HAs. Among a subgroup of human A(H1N1) HAs between 1918~2008, we found strong diversifying (positive) selection at HA₁ 156 and 190. We also analyzed the evolutionary trends at HA₁ 190 and 225 that are critical determinants for receptor-binding specificity of A(H1N1) HA. Different A(H1N1) viruses appeared to favor one of these two sites in host-driven antigenic drift: epidemic A(H1N1) HAs favor HA₁ 190 while the 1918 pandemic and swine HAs favor HA₁ 225. Thus, our results highlight the urgency to understand the interplay between antigenic drift and receptor binding in HA evolution, and provide molecular signatures for monitoring future antigenically drifted 2009 pandemic and seasonal A(H1N1) influenza viruses.

Citation: Shen J, Ma J, Wang Q (2009) Evolutionary Trends of A(H1N1) Influenza Virus Hemagglutinin Since 1918. PLoS ONE 4(11): e7789. doi:10.1371/journal.pone.0007789

Editor: Darren P. Martin, Institute of Infectious Disease and Molecular Medicine, South Africa

Received: August 7, 2009; **Accepted:** October 15, 2009; **Published:** November 17, 2009

Copyright: © 2009 Shen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JM acknowledges support from National Institutes of Health (R01-GM067801), National Science Foundation (MCB-0818353), The Welch Foundation (Q-1512), the Welch Chemistry and Biology Collaborative Grant from John S. Dunn Gulf Coast Consortium for Chemical Genomics and the Faculty Initiatives Fund from Rice University. QW acknowledges support from a Beginning Grant-in-Aid award from American Heart Association (0865186F) and of a grant from THE National Institutes of Health (R01-AI067839). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: E-mail: qinghuaw@bcm.tmc.edu (QW); jpma@bcm.tmc.edu (JM)

Introduction

Since April 2009, a global outbreak caused by the swine-origin 2009 A(H1N1) influenza virus has spread to numerous countries [1,2,3,4,5,6,7,8,9,10], which warranted the declaration of “Pandemic (H1N1) 2009” by the World Health Organization on June 11, 2009. As of September 6, there had been over 277,607 infected individuals and at least 3,205 confirmed human deaths worldwide.

The Pandemic (H1N1) 2009 is not the first human pandemic caused by A(H1N1) influenza virus. During 1918~1919, the “Spanish” A(H1N1) influenza virus swept across the globe, infected ~25% of the entire population and claimed at least 50 million human lives worldwide [11]. In subsequent years, A(H1N1) influenza virus continued to circulate among humans and caused a number of severe outbreaks between 1920s and 1950s [12,13,14,15,16,17,18,19,20], in particular the A(H1N1) epidemic in 1950~1951 with mortality exceeding those of the 1957 “Asian” and 1968 “Hong Kong” pandemics [18,19,20]. In 1957, A(H1N1) influenza virus disappeared, replaced by a reassorted A(H2N2) influenza virus [21]. However, the A(H1N1) influenza virus reappeared in 1977, with a close genetic and antigenic similarity to those A(H1N1) viruses isolated in 1950 [22,23,24], and has co-circulated with A(H3N2) and type B influenza virus to cause seasonal human epidemics ever since.

The same 1918 pandemic A(H1N1) influenza virus was also spread to swine during 1918~1919, and became the so-called “classical” swine influenza [8,25,26,27], first isolated in North American in 1930 [26] and in Europe in 1976 [28,29]. In 1979, a novel lineage of avian-like A(H1N1) influenza virus, believed to have derived from closely related Eurasian avian influenza viruses, emerged in swine in Europe [30] and replaced the classical swine A(H1N1) virus in this region [31,32,33]. These two classes of swine A(H1N1) viruses displayed different evolutionary trajectories [34]. In 1998, a new triple-reassortant A(H3N2) virus, derived from North American avian, classical swine A(H1N1) and human A(H3N2) viruses, caused outbreaks in North American swine [35,36]. Mixing of the triple-reassortant H3N2 with established swine lineages gave rise to H1N1 and H1N2 reassortant swine viruses [37,38]. Since 2007, human infection caused by A(H1N1) swine virus has become a health concern in the United States [7].

The 2009 A(H1N1) influenza virus has its origin as a reassortant from a Eurasian avian-like swine A(H1N1) virus and a triple-reassortant virus circulating in North American swine [1,2,3,4,5,6,7,8]. As such, the 2009 A(H1N1) virus contains NA and M from Eurasian avian-like swine A(H1N1) virus, and the remaining genes from the triple-reassortant virus - PB2 and PA (avian virus), PB1 (human A(H3N2)), and HA, NP and NS (classical swine A(H1N1)) [1,2,3,4,5,6,7,8]. In a sense, we are continuingly living in a pandemic that started in 1918 [27]. Thus,

it is not surprising for the similarly mild first waves of the 1918 and 2009 pandemics. Notably, the second wave of the “Spanish” influenza in the fall of 1918 became much more lethal, peaked within one month of the initial introductions in many communities [11]. This makes influenza virologists and healthcare officials fear that further mutations in the 2009 A(H1N1) virus could also lead to a potentially more dangerous second wave in subsequent months. Thus, in-depth studies on the 1918 pandemic strains as well as their post-pandemic decedents should provide critical new insights into the evolution of A(H1N1) in general, and the pandemic potential of the 2009 A(H1N1) in particular.

HA is one of the two major glycoproteins on the surface of influenza virus. It is the primary antigen that elicits host immune response, and is also responsible for binding to sialic-acid receptors and for mediating viral entry into host cells [39]. The hallmarks of highly pathogenic influenza viruses among human population include easy human-to-human transmission as a result of high affinity of HA for human-like $\alpha(2,6)$ receptors, and significant difference in sequence and antigenicity of HA with existing seasonal and vaccine strains [1,2,39]. It has been demonstrated on 1918 A(H1N1) HA that HA₁ D190 and D225 are key determinants for effective binding to human-like $\alpha(2,6)$ receptors and consequently high infectivity of the virus among human population [11,40,41,42]. A single mutation D225G reduced the binding affinity for $\alpha(2,6)$ receptors [41,42] and the infectivity of the virus [40], while a double variant D190E/D225G rendered the HA non-binding to $\alpha(2,6)$ receptors [41,42] and the virus non-infectious [40].

In A(H3N2) virus, HA is the major agent for host-driven antigenic drift [43,44]. However, it is unclear whether or not and, if yes, how human immunity imposes selection on A(H1N1) HA. In order to address this critical issue, we undertook a systematic computational analysis of the evolution of H1 HA in the region of HA₁, which is the primary target for host immunity selection [43].

Recent years have witnessed an explosive expansion of available computational methods for phylogenetic analysis of selective pressure, including a variety of methods that look for different types of positive selection such as diversifying selection, toggling selection and directional selection [45,46,47,48,49,50,51,52,53,54] implemented in software packages such as HyPhy [55], MrBayes [56,57] and PAML [45]. Here we used PAML 4.0 [45] for calculation of heterogeneous selection pressure at each codon and HyPhy [55] for directional selection in 335 non-egg-adapted and 32 egg-adapted human A(H1N1) HA sequences. These sequences were from A(H1N1) viruses isolated all around the globe between 1918~2009. In addition, we also analyzed 42 classical swine A(H1N1) HA sequences for their close relationship to the 2009 A(H1N1) HA.

In PAML 4.0 [45], a number of models are available: the branch models allow the ω ratio to vary among branches in the phylogenetic tree and can be used to detect positive selection on particular branches [46,58]; the site models allow the ω ratio to vary among sites and can be used to detect positive selection at particular sites [59,60]; the branch-site models allow the ω ratio to vary both among sites and among branches [61] and can be used to detect positive selection that affects only a few sites in a few branches.

In this analysis, a large dataset composed of over 300 sequences was used to ensure sufficient representative sequences for the total time span of 91 years, which made it impractical for the use of branch-site models in our calculations. However, by separating the sequences into distinct subgroups based on their phylogenetic relationship and applying the site models in PAML 4.0 [45], we successfully detected the branch and the specific sites therein that

were under host-driven positive selection. Our study revealed differential evolutionary trends of A(H1N1) HA since 1918, which provided molecular signatures for monitoring future antigenically drifted 2009 pandemic and seasonal A(H1N1) influenza viruses.

Results and Discussion

Phylogenetic Analysis of Human A(H1N1) HA Sequences Since 1918

It is known that egg-adapted influenza viruses tend to have non-natural host-associated modifications at certain sites of HA sequences [62,63,64]. To eliminate the effects of such modifications in our analysis, we selected only 333 HA sequences of A(H1N1) viruses between 1918~2009 (as of July 10, 2009) with a well-documented record that they had never been passaged in chicken eggs at any stage. Furthermore, intragenic recombination may give rise to false positives in subsequent detection of positively selected codons [65], thus the Recombination Detection Program (RDP3) [66] was used to make sure that all HA sequences used in this study were free of recombination, agreeing with previous observations that intragenic recombination is rare for HA [67]. The nucleotide sequences of 333 A(H1N1) HAs in the region of HA₁ including the signal peptide, were analyzed by the ClustalW method [68]. The phylogeny tree suggested that these HA sequences belong to two major groups: the majority of HA sequences from 1918 to 2008 formed group I, and those of the 2009 A(H1N1) together with a strain isolated in 2007 formed group II (Fig. S1). The separation of the 2009 A(H1N1) HAs from HAs of established human A(H1N1) viruses between 1918~2008, including the 1918 pandemic and the seasonal A(H1N1) viruses, was consistent with the proposed swine origin of HAs in these viruses [1,2,3,4,5,6,7,8]. The low sequence identity (~73%) between the 2009 A(H1N1) HA with seasonal and vaccine A(H1N1) HAs might explain why people were in general immunologically naive to the former [8,69]. In fact, there did not exist cross-reactivity between the 2009 and seasonal A(H1N1) viruses [8], nor did the vaccination with recent (2005~2009) annual vaccines provide immune protection against the 2009 A(H1N1) virus [69].

Evidence for Host-Driven Antigenic Drift in Human A(H1N1) HAs

In order to understand whether host-driven antigenic drift is imposed on the evolution of HA₁ of A(H1N1) virus, we used likelihood ratio tests (LRT) in the software package PAML 4.0 [45] to identify the presence or absence of *positive selection*. In this context, positive selection referred to a significant excess of amino-acid altering (non-synonymous) substitutions over silent (synonymous) substitutions in nucleotide sequences. Large LRT values (or small *p*-values) between alternative models and null models, such as M2a vs. M1a, M8 vs. M7, or M8 vs. M8a, led to the rejection of the null models.

Since HA sequences of group I was further divided into five subgroups (Fig. S1), the PAML calculation was carried out on each of these five subgroups and on group II (Table 1). For group I-i that included three 1918 pandemic A(H1N1) HAs, in order to increase the sample size, we also included two partial sequences, A/London/1/1918 and A/London/1/1919 [11]. Except for the subgroup I-v, all other subgroups of group I had very low LRT values and large *p*-values (Table 1, 2), indicating predominantly neutral or purifying selection. These results were consistent with the overall low prevalence of A(H1N1) virus during the period of 1979~2006 [70], and agreed well with a previous study that focused on 1995~2005 A(H1N1) isolates where no positive

Table 1. The values of log-likelihood (l), d_N/d_S , and parameter estimates in CODEML analysis of human A(H1N1) HAs.

Model	l	d_N/d_S	Parameters estimates
I-i 1918~1919 (5 strains)¹			
M0 (one-ratio)	-806.78	0.516	$\omega = 0.516$
M1a (nearly neutral)	-805.91	0.323	$p_0 = 0.677$ ($p_1 = 0.323$), $\omega_0 = 0$ ($\omega_1 = 1$)
M2a (positive selection)	-804.19	0.564	$p_0 = 0.963$, $p_1 = 0$ ($p_2 = 0.037$), $\omega_0 = 0$ ($\omega_1 = 1$), $\omega_2 = 15.421$
M7 (beta)	-805.92	0.300	$p = 0.005$, $q = 0.012$
M8a (beta& $\omega = 1$)	-805.91	0.323	$p_0 = 0.846$ ($p_1 = 0.154$), $p = 0.005$, $q = 0.020$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-804.19	0.561	$p_0 = 0.963$ ($p_1 = 0.037$), $p = 0.005$, $q = 7.228$, $\omega_3 = 15.242$
I-ii 1979~2000 (45 strains)²			
M0 (one-ratio)	-2489.23	0.223	$\omega = 0.223$
M1a (nearly neutral)	-2486.25	0.242	$p_0 = 0.855$ ($p_1 = 0.145$), $\omega_0 = 0.113$ ($\omega_1 = 1$)
M2a (positive selection)	-2486.25	0.242	$p_0 = 0.855$, $p_1 = 0.056$ ($p_2 = 0.089$), $\omega_0 = 0.113$ ($\omega_1 = 1$), $\omega_2 = 1$
M7 (beta)	-2485.63	0.232	$p = 0.327$, $q = 1.074$
M8a (beta& $\omega = 1$)	-2485.63	0.232	$p_0 = 1$ ($p_1 = 0$), $p = 0.327$, $q = 1.074$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-2485.63	0.232	$p_0 = 1$ ($p_1 = 0$), $p = 0.327$, $q = 1.074$, $\omega_3 = 1$
I-iii 2000~2001 (22 strains)²			
M0 (one-ratio)	-1686.21	0.279	$\omega = 0.279$
M1a (nearly neutral)	-1684.58	0.261	$p_0 = 0.793$ ($p_1 = 0.207$), $\omega_0 = 0.068$ ($\omega_1 = 1$)
M2a (positive selection)	-1683.09	0.287	$p_0 = 0.994$, $p_1 = 0$ ($p_2 = 0.006$), $\omega_0 = 0.225$ ($\omega_1 = 1$), $\omega_2 = 10.636$
M7 (beta)	-1684.52	0.265	$p = 0.046$, $q = 0.127$
M8a (beta& $\omega = 1$)	-1684.58	0.261	$p_0 = 0.793$ ($p_1 = 0.207$), $p = 7.211$, $q = 98.93$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-1683.10	0.287	$p_0 = 0.994$ ($p_1 = 0.006$), $p = 28.774$, $q = 99$, $\omega_3 = 10.638$
I-iv 2001~2007 (89 strains)²			
M0 (one-ratio)	-2720.83	0.187	$\omega = 0.187$
M1a (nearly neutral)	-2709.20	0.181	$p_0 = 0.883$ ($p_1 = 0.117$), $\omega_0 = 0.072$ ($\omega_1 = 1$)
M2a (positive selection)	-2708.27	0.187	$p_0 = 0.906$, $p_1 = 0.076$ ($p_2 = 0.018$), $\omega_0 = 0.085$ ($\omega_1 = 1$), $\omega_2 = 1.915$
M7 (beta)	-2708.99	0.183	$p = 0.137$, $q = 0.612$
M8a (beta& $\omega = 1$)	-2709.40	0.180	$p_0 = 0.885$ ($p_1 = 0.115$), $p = 7.871$, $q = 98.995$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-2708.82	0.186	$p_0 = 0.973$ ($p_1 = 0.027$), $p = 0.375$, $q = 2.298$, $\omega_3 = 1.936$
I-v 2006~2008 (100 strains)²			
M0 (one-ratio)	-3871.33	0.303	$\omega = 0.303$
M1a (nearly neutral)	-3813.28	0.241	$p_0 = 0.828$ ($p_1 = 0.172$), $\omega_0 = 0.082$ ($\omega_1 = 1$)
M2a (positive selection)	-3782.07	0.313	$p_0 = 0.807$, $p_1 = 0.188$ ($p_2 = 0.005$), $\omega_0 = 0.083$ ($\omega_1 = 1$), $\omega_2 = 11.142$
M7 (beta)	-3814.21	0.246	$p = 0.139$, $q = 0.427$
M8a (beta& $\omega = 1$)	-3812.40	0.231	$p_0 = 0.862$ ($p_1 = 0.138$), $p = 0.497$, $q = 3.969$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-3781.55	0.305	$p_0 = 0.994$ ($p_1 = 0.006$), $p = 0.180$, $q = 0.546$, $\omega_3 = 10.554$
II 2007~2009 (74 strains)²			
M0 (one-ratio)	-2374.98	0.277	$\omega = 0.277$
M1a (nearly neutral)	-2373.20	0.282	$p_0 = 0.922$ ($p_1 = 0.078$), $\omega_0 = 0.222$ ($\omega_1 = 1$)
M2a (positive selection)	-2372.81	0.282	$p_0 = 0.922$, $p_1 = 0.035$ ($p_2 = 0.042$), $\omega_0 = 0.222$ ($\omega_1 = 1$), $\omega_2 = 1$
M7 (beta)	-2373.21	0.281	$p = 1.218$, $q = 3.079$
M8a (beta& $\omega = 1$)	-2374.03	0.282	$p_0 = 0.945$ ($p_1 = 0.055$), $p = 3.243$, $q = 10.201$, $\omega_3 = 1$
M8 (beta& $\omega > 1$)	-2372.81	0.282	$p_0 = 0.946$ ($p_1 = 0.054$), $p = 3.158$, $q = 9.900$, $\omega_3 = 1$

¹Due to the inclusion of two partial sequences of A/London/1/1918 and A/London/1/1919 in this subgroup, the analysis was performed on a total of 187 amino-acid residues that covered the antigenic and receptor-binding sites in the region of HA₁ (51~237) [11]. ²The analysis was performed on the first 340 residues of HA₁ including the signal peptide.

doi:10.1371/journal.pone.0007789.t001

selection was detected [71]. In sharp contrast, group I-v 2006~2008 had $\omega > 10$ and $LRT > 60$, which provided strong evidence for positive selection (Table 1, 2) and agreed with the necessity to update the A(H1N1) vaccine strain using A/Brisbane/59/07 for the 2008~2009 season. Group II including 73 HAs of

2009 A(H1N1) and one of 2007 A(H1N1) also had a very low LRT rate ratio (Table 1, 2). Given the largely nonexistence of human immunity against the 2009 A(H1N1), the lack of positive selection among group II was expected. However, with more mild infections rapidly propagating among human population in the

Table 2. LRT tests for HA₁ sequences of human A(H1N1) influenza viruses.

	LRT (M2a – M1a) (2Δl) (p-values) ¹	LRT (M8 – M7) (2Δl) (p-values) ¹	LRT (M8 – M8a) (2Δl) (p-values) ²
I-i 1918~1919 (5 strains)	3.44 (0.1791)	3.46 (0.1773)	3.44 (0.0318)
I-ii 1979~2000 (45 strains)	0	0	0
I-iii 2000~2001 (22 strains)	2.98 (0.2254)	2.84 (0.2417)	2.96 (0.0427)
I-iv 2001~2007 (89 strains)	1.86 (0.3946)	0.34 (0.8437)	1.16 (0.1407)
I-v 2006~2008 (100 strains)	62.42 (0.0000)	65.32 (0.0000)	61.70 (0.0000)
II 2007~2009 (74 strains)	0.78 (0.6771)	0.80 (0.6703)	2.44 (0.0591)

¹We used the degree of freedom of 2 for these LRT tests that is expected to be too conservative. ²The p-values were calculated from χ^2 distribution using degree of freedom of 1 that was then divided by a factor of 2 for the mixture distribution, as suggested by the author of PAML 4.0.
doi:10.1371/journal.pone.0007789.t002

first wave, the gradually established human immunity might drive positive selection in future isolates of 2009 A(H1N1) strains.

Identification of Positively Selected Codons in Human A(H1N1) HAs

In order to understand how H1 HA sequences were positively selected by human existing immunity, the CODEML [72] program in PAML 4.0 was used on subgroup I-v in which about 0.6% codons were found to be under positive selection (Table 1, 2). Both M2a and M8 models identified HA₁ 156 and 190 with greater than 95% posterior probabilities to be under positive selection (Table 3). In previous studies, the antigenic structure of H1 HA (A/PuertoRico/8/1934) had been determined to include five distinct antigenic sites on the globular domain: Sa, Sb, Ca1, Ca2 and Cb [73,74] (Fig. 1a). Both of these positively selected codons were located on the site Sb (Fig. 1b). The focus of positive selection on the Sb antigenic site was consistent with a cross-reactivity analysis of various epidemic H1N1 strains using monoclonal antibodies that it was under much higher pressure for mutations [74].

HA₁ 138, 186, 190, 194, 225, 226 and 228 had been previously shown to affect receptor binding to H1 HA [75,76]. Among them, two residues, HA₁ 190 and 225, play predominant roles in determining the receptor-binding specificity of H1 HA: D190/D225 for $\alpha(2,6)$ receptors in humans, D190/G225 for $\alpha(2,6)$ and $\alpha(2,3)$ receptors in swine, and E190/G225 for $\alpha(2,3)$ receptors in avian [11,39,40,41,42,76]. Although changes at these two sites had been previously reported to cause antigenic drift in A(H1N1) epidemic strains [77], it was a somewhat common belief that key determinants of receptor-binding specificity are in general not subject to selection. Thus, the strong positive selection at HA₁ 190 within subgroup I-v is quite unexpected.

Table 3. Codons under positive selection in HA₁ of human A(H1N1) influenza viruses.

Positively selected sites ¹		
I-v 2006~2008 (100 strains)	M2a	156 (97.5%), 190 (100%)
	M8	156 (99.7%), 190 (100%)

¹Positively selected sites from PAML 4.0 [45] using Bayes Empirical Bayes analysis [72]. Only codons with greater than 95% posterior probabilities to be under positive selection were listed with the corresponding posterior probabilities shown in parentheses.
doi:10.1371/journal.pone.0007789.t003

Positive Selection of Egg-Adapted Human A(H1N1) HAs during 1933~1979

To compensate for the lack of non-egg-adapted human A(H1N1) HAs for the period of 1933~1978, we separately collected a total of 32 different egg-adapted A(H1N1) HA sequences between 1933~1979 that were free of sequence ambiguity (Fig. S2). These sequences as a group were analyzed by PAML 4.0, as well as two subgroups that covered the periods of 1947~1957 (12 sequences) and 1948~1979 (17 sequences) (Table 4, 5), keeping in mind of the egg-adapted mutations at HA₁ 138, 144, 163, 189, 190, 225, and 226 [62,63,64]. The two subgroups 1947~1957 and 1948~1979 represented A(H1N1) viruses circulating in the 1950s and in the 1970s upon its reemergence in 1977, respectively. Given the close genetic and antigenic similarity of the reappeared A(H1N1) influenza virus in 1977 with the A(H1N1) viruses isolated in 1950 [22,23,24], it was of particular interest to investigate whether different evolutionary trends were adopted by the 1947~1957 and 1948~1979 subgroups.

For both the entire group 1933~1979 and the subgroup 1947~1957, comparisons of M2a-M1a, M8-M7, or M8-M8a yielded large LRT values and very small p-values, suggesting the presence of positive selection at about 5% and 3% codons, respectively (Table 4, 5). However, it is noteworthy that the subgroup 1948~1979 had much smaller LRT values, suggesting that the positive pressure of the entire group 1933~1979 be mostly from the contribution of the subgroup 1947~1957.

We further employed the CODEML in PAML 4.0 to analyze the positively selected codons in each group. The results were shown in Table 5 where highlighted in bold were the codons not known to be possible egg-adapted mutations (HA₁ 138, 144, 163, 189, 190, 225, and 226) [62,63,64]. For the entire group 1933~1979, HA₁ 77, 225 and 227 were found to be under positive selection with greater than 95% posterior probability in model M8 (Fig. 1b). They were located in the antigenic sites Cb (HA₁ 77) and Ca2 (HA₁ 225 and 227), respectively (Fig. 1b). In addition, for the subgroup 1948~1979, HA₁ 225 was found to be under positive selection with greater than 99% posterior probability in both models M2a and M8. However, given the fact that these HAs were from egg-adapted A(H1N1) viruses in which HA₁ 225 was one of the most frequently changed site [62,63,64], and the predominant residue at this site (Table 6), G225, was commonly found in swine and avian A(H1N1) HAs, it was possible that the changes at HA₁ 225 was due to positive selection imposed by adaptation in eggs. At posterior probability of 90%, HA₁ 138 and 189 were positively selected as well, however, both sites were involved in egg-adapted substitutions [62,63,64]. In sharp contrast, however, HA₁ 143, 166 and 264 in the subgroup

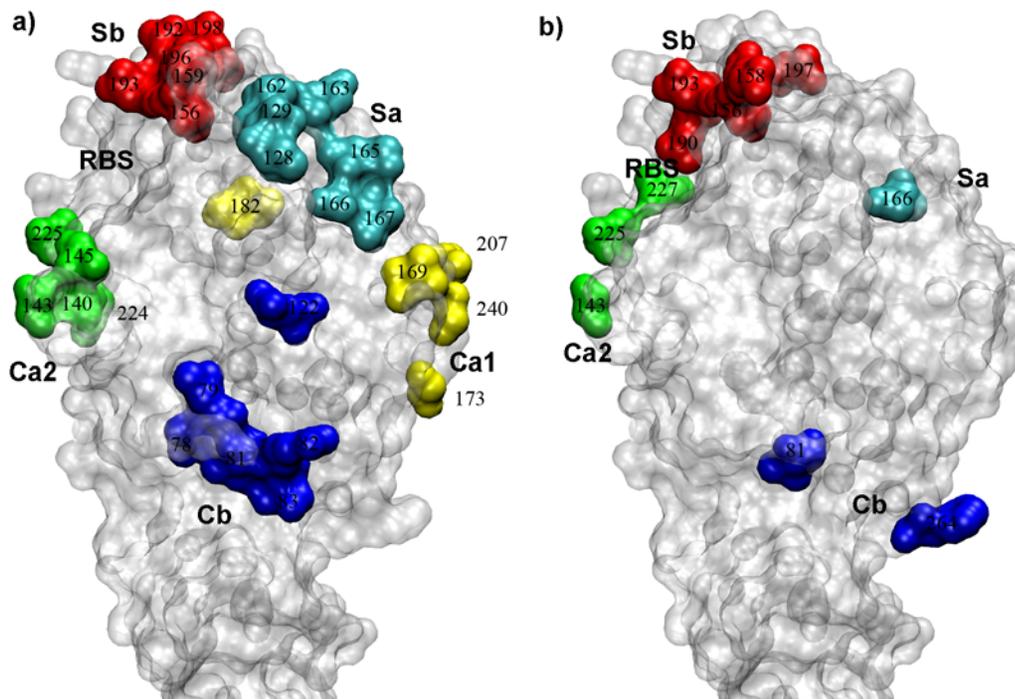


Figure 1. Antigenic structure and positive selection of A(H1N1) HA. a) Antigenic structure of A/PR/8/34 (H1N1) HA (PDB accession code 1RU7 [78]). Five antigenic sites were identified by using a large number of monoclonal antibodies [73,74]: Sa (cyan), Sb (red), Ca1 (yellow), Ca2 (green), Cb (blue), using H3 HA numbering. The receptor-binding site (RBS) was labeled for reference. b) Codons on A(H1N1) HA that were identified to be under various selection in PAML and HyPhy analysis.
doi:10.1371/journal.pone.0007789.g001

Table 4. The values of log-likelihood (l), d_N/d_S , and parameter estimates in CODEML analysis of egg-adapted human A(H1N1) HAs between 1933–1979.

Model	l	d_N/d_S	Parameters estimates
1933~1979 (32 strains)¹			
M0 (one-ratio)	-3336.01	0.411	$\omega = 0.411$
M1a (nearly neutral)	-3283.75	0.336	$p_0 = 0.705$ ($p_1 = 0.295$), $\omega_0 = 0.057$ ($\omega_1 = 1$)
M2a (positive selection)	-3275.24	0.454	$p_0 = 0.721$, $p_1 = 0.234$ ($p_2 = 0.046$), $\omega_0 = 0.079$ ($\omega_1 = 1$), $\omega_2 = 3.571$
M7 (beta)	-3285.56	0.345	$p = 0.068$, $q = 0.129$
M8a (beta& $\omega = 1$)	-3283.78	0.336	$p_0 = 0.705$ ($p_1 = 0.295$), $p = 6.100$, $q = 99$, $\omega_s = 1$
M8 (beta& $\omega > 1$)	-3275.41	0.452	$p_0 = 0.942$ ($p_1 = 0.058$), $p = 0.206$, $q = 0.534$, $\omega_s = 3.283$
1947~1957 (12 strains)¹			
M0 (one-ratio)	-2068.01	0.435	$\omega = 0.435$
M1a (nearly neutral)	-2056.27	0.337	$p_0 = 0.689$ ($p_1 = 0.311$), $\omega_0 = 0.038$ ($\omega_1 = 1$)
M2a (positive selection)	-2046.27	0.501	$p_0 = 0.967$, $p_1 = 0$ ($p_2 = 0.033$), $\omega_0 = 0.256$ ($\omega_1 = 1$), $\omega_2 = 7.651$
M7 (beta)	-2056.44	0.324	$p = 0.012$, $q = 0.023$
M8a (beta& $\omega = 1$)	-2056.28	0.337	$p_0 = 0.688$ ($p_1 = 0.312$), $p = 3.886$, $q = 99$, $\omega_s = 1$
M8 (beta& $\omega > 1$)	-2046.29	0.501	$p_0 = 0.967$ ($p_1 = 0.033$), $p = 34.141$, $q = 99$, $\omega_s = 7.667$
1948~1979 (17 strains)¹			
M0 (one-ratio)	-1759.58	0.385	$\omega = 0.385$
M1a (nearly neutral)	-1751.46	0.260	$p_0 = 0.740$ ($p_1 = 0.260$), $\omega_0 = 0$ ($\omega_1 = 1$)
M2a (positive selection)	-1747.17	0.408	$p_0 = 0.794$, $p_1 = 0.168$ ($p_2 = 0.039$), $\omega_0 = 0$ ($\omega_1 = 1$), $\omega_2 = 6.226$
M7 (beta)	-1751.61	0.300	$p = 0.005$, $q = 0.012$
M8a (beta& $\omega = 1$)	-1751.46	0.260	$p_0 = 0.740$ ($p_1 = 0.260$), $p = 0.005$, $q = 2.350$, $\omega_s = 1$
M8 (beta& $\omega > 1$)	-1747.19	0.411	$p_0 = 0.969$ ($p_1 = 0.031$), $p = 0.006$, $q = 0.025$, $\omega_s = 6.964$

¹The analysis was performed on the first 337 residues of HA₁ including the signal peptide.

doi:10.1371/journal.pone.0007789.t004

Table 5. LRT tests and codons under positive selection for HA₁ sequences of egg-adapted human A(H1N1) influenza viruses between 1933-1979.

			LRT (2Δl) (p-values)	Positively selected sites ¹
1933~1979 (32 strains)	M2a	M2a-M1a	17.02 (0.0002)	77 (95.8%) , 225 (98.8%)
	M8	M8-M7	20.30 (0.0000)	77 (98.7%) , 225 (99.6%), 227 (97.7%)
		M8-M8a	16.74 (0.0000)	
1947~1957 (12 strains)	M2a	M2a-M1a	20.0 (0.0000)	143 (99.3%) , 264 (99.6%)
	M8	M8-M7	20.30 (0.0000)	143 (99.6%) , 166 (95.1%) , 264 (99.7%)
		M8-M8a	19.98 (0.0000)	
1948~1979 (17 strains)	M2a	M2a-M1a	8.58 (0.0137)	225 (99.1%)
	M8	M8-M7	8.84 (0.0120)	225 (99.8%)
		M8-M8a	8.54 (0.0017)	

¹Positively selected sites from PAML 4.0 [45] using Bayes Empirical Bayes analysis [72]. Only codons with greater than 95% posterior probabilities to be under positive selection were listed with the corresponding posterior probabilities shown in parentheses. Highlighted in bold were codons that were not associated with egg-adapted substitutions [62,63,64].

doi:10.1371/journal.pone.0007789.t005

1947~1957 were found to be under positive selection (Table 5), none of which was among the previously identified egg-adapted mutations. Antigenically, these codons were located in the antigenic sites Ca2, Sa and Cb, respectively (Fig. 1b). For their relatively distant location from the receptor-binding site, HA₁ 143, 166 and 264 are probably mutations driven by existing human immunity for antibody escape.

Thus, there appeared to have different evolutionary patterns for the subgroup 1947~1957 circulating in the 1950s and the subgroup 1948~1979 circulating mostly in the 1970s. The former subgroup was subjected to positive selection pressure at HA₁ 143, 166 and 264 (Table 5), and had a much larger variability at HA₁ 190, with 25% being non-D190 (Table 6). In marked contrast, the latter subgroup was probably not under host-driven positive selection in humans and had highly conserved HA₁ 190 (94.1% being D190).

Evolution of Swine A(H1N1) HAs during 1990~2009

Given the swine origin of the 2009 pandemic A(H1N1) HA, we also analyzed 42 non-redundant, non-ambiguous swine A(H1N1) HA sequences during 1990~2009 available from GISAID/Epifludb (Table 7, Fig. S3). The reason that we focused on this period was mainly for the antigenic stasis of swine A(H1N1) until 1998 [8] since the introduction of the 1918 “Spanish” A(H1N1) virus into swine [1,2,3,4,5,6,7,8]. Overall, the alternative models

M2a and M8 fitted the data only marginally better than the null models M1a, M7 and M8a, respectively (Table 7). Thus, it seemed that swine A(H1N1) HAs during 1990~2009 were not subjected to strong host-driven positive selection.

Directional Evolution of Human A(H1N1) HAs

In order to test whether directional evolution of protein sequences existed in the evolution of human A(H1N1) HAs, we employed a maximum likelihood method developed by Kosakovsky Pond and colleagues [49]. In each subgroup, we used the oldest HA sequence as the root. In agreement with CODEML analysis reported in previous sections, among all non-egg adapted human A(H1N1) HAs, directional evolution was only identified in the subgroup I-v, at sites HA₁ 143, 156, 158, 190, 193 and 197 (Table 8, 9). HA₁ 143 belonged to the antigenic site Ca2 of A(H1N1) HA, whilst all other sites were located in the antigenic site Sb (Fig. 1b). Among these sites, HA₁ 156, 190 and 193 were identified by CODEML in PAML 4.0 to be under positive selection with 99.7%, 100%, and 80.9% posterior probability in model M8, respectively (Table 3). In previous structural studies, residue HA₁ 190 in 1934 human A(H1N1) HA and HA₁ 190 and 193 in 1930 swine A(H1N1) HA were found to directly interact with bound human-like α(2,6)-receptors [78]. Thus, it remains to be investigated the impacts of directional evolution at HA₁ 190 and 193 on receptor binding and antigenic drift.

Table 6. Codons at HA₁ 190 and 225 in human and swine A(H1N1) influenza viruses.

	D190	Non-D190	D225	G225	Non-D225/G225
Human 1979~2008 Epidemic (575 sequences)	477 (83.0%) [*]	98 (17.0%)	565 (98.2%)	2 (0.4%)	8 (1.4%)
Human 1918 Pandemic (5 sequences)	5 (100%)	0	3 (60%)	2 (40%)	0
Human 2009 Pandemic (73 sequences)	73 (100%)	0	69 (94.5%)	1 (1.4%)	3 (4.1%)
Human 1947~1957 (12 strains) (egg-adapted)	9 (75%)	3 (25%)	2 (16.7%)	10 (83.3%)	0
Human 1948~1979 (17 strains) (egg-adapted)	16 (94.1%)	1 (5.9%)	4 (23.5%)	13 (76.5%)	0
Swine 1990~2009 (42 sequences)	41 (97.6%)	1 (2.4%)	28 (66.6%)	12 (28.6%)	2 (4.8%)

^{*}The number of cases that a particular type of residues occurs at each site. Shown in parenthesis was the occurrence in percentage.

doi:10.1371/journal.pone.0007789.t006

Table 7. The values of log-likelihood (l), d_N/d_S , and parameter estimates in CODEML analysis of swine A(H1N1) HAs between 1990–2009.

Model	l	d_N/d_S	Parameters estimates	LRT (2Δl) (p-values)
M0 (one-ratio)	−6021.08	0.158	$\omega = 0.158$	
M1a (nearly neutral)	−5949.30	0.217	$p_0 = 0.864$ ($p_1 = 0.136$), $\omega_0 = 0.094$ ($\omega_1 = 1$)	LRT (M2a-M1a) = 1.24 (0.5379)
M2a (positive selection)	−5948.68	0.224	$p_0 = 0.864$, $p_1 = 0.134$ ($p_2 = 0.002$), $\omega_0 = 0.095$ ($\omega_1 = 1$), $\omega_2 = 3.682$	
M7 (beta)	−5925.79	0.174	$P = 0.381$, $q = 1.768$	LRT (M8-M7) = 6.78 (0.0337) LRT (M8-M8a) = 3.40 (0.0326)
M8a (beta& $\omega = 1$)	−5924.10	0.171	$p_0 = 0.970$ ($p_1 = 0.030$), $p = 0.460$, $q = 2.617$, $\omega_3 = 1$	
M8 (beta& ω)	−5922.40	0.177	$p_0 = 0.995$ ($p_1 = 0.005$), $p = 0.413$, $q = 2.052$, $\omega_3 = 2.546$	

The analysis was performed on the first 338 residues of HA₁ including the signal peptide.

doi:10.1371/journal.pone.0007789.t007

We also performed directional evolution study on egg-adapted human A(H1N1) HA sequences, and found that in both the entire group 1933~1979 and the subgroup 1948~1979, multiple favored mutations of D225→G and G225→D were detected (Table 8, 9). Given its involvement in egg-adaptation, the directional evolution at HA₁ 225 may be the consequence of egg-adaptation. In contrast, no residues in the subgroup 1947~1957 were identified to be under directional selection.

Evolution of Human and Swine A(H1N1) HAs at HA₁ 190 and 225

For their predominant roles in determining receptor-binding specificity of A(H1N1) HA, and the positive selection on HA₁ 190 in the subgroup I-v, we further investigated the evolution of HA₁ 190 and 225 in A(H1N1) strains during 1918~1009. These included 653 non-egg-adapted HAs (five pandemic HAs from 1918~1919, 575 epidemic HAs from 1979~2008, and 73 pandemic HAs from 2009), and 42 swine HAs (Table 6). For the 575 epidemic HAs, HA₁ 190 was highly variable (17.0% sequences did not have D190), while HA₁ 225 was more conserved (only 1.8% sequences did not have D225) (Table 6). Among all the deviations (a total of 107 cases) from the ideal D190/D225 combination for human A(H1N1) viruses, two predominant ones were N190/D225 (69.2%) and V190/D225 (19.6%). At present, we don't know the exact effects of these mutations, or in combination with other concurring mutations at or around the receptor-binding site, on binding to human receptors. Further experiments are needed to clarify these issues. However, in previous studies, a single mutation D190N of A(H1N1) HA was shown to result in a lower binding affinity for human-like $\alpha(2,6)$ receptors, and a higher binding affinity for avian-like $\alpha(2,3)$ receptors [64].

The five HA sequences retrieved from victims of 1918 “Spanish” A(H1N1) influenza virus shared 98.9% to 99.8% sequence identity [11]. Among them, there were two non-synonymous substitutions of D225G, one in A/New York/1/1918 and the other one in A/London/1/1919 (Table 6). The HAs harboring the mutation D225G had reduced binding affinity for human receptors [11,40,41].

In the 73 HA sequences from the 2009 pandemic A(H1N1), D190 was strictly conserved, while D225 was 94.5% conserved (Table 6). At HA₁ 225, the deviations were 1.1% for G225 and 3.3% for E225. Thus, the complete conservation at HA₁ 190 and the nearly complete conservation at HA₁ 225 were consistent to the importance of these residues in allowing for binding to human-like $\alpha(2,6)$ receptors [40,41], supporting the substantially higher human-to-human transmissibility of the 2009 A(H1N1) virus than seasonal A(H1N1) viruses [5,8].

Therefore, there were two distinct evolutionary trends in host-driven antigenic drift of human A(H1N1) HAs at residues in the receptor-binding site: the 1918 pandemic HAs underwent antigenic drift at HA₁ 225, while the epidemic HAs undertook antigenic drift at HA₁ 190. In the absence of selection, the 2009 A(H1N1) viruses were highly conserved at both HA₁ 190 and 225, which was distinct from those two host-selected evolutionary trends (Table 6). With gradually established immunity among human population, we wondered how the 2009 A(H1N1) virus would undergo antigenic drift in the months to come. Thus, we also looked at the conservation at HA₁ 190 and 225 in 42 swine A(H1N1) HA sequences (Table 6). Surprisingly, among these sequences, D190 was conserved at 97.6%, while D225 and G225 were observed at 66.6% and 28.6%, respectively. The similarly high variability of HA₁ 225 in swine A(H1N1) HAs with that of 1918 pandemic HAs was consistent with the relative antigenic

Table 8. Directional selection analysis on human A(H1N1) HAs.

	Tree L	Residue	P-Value	Bias	Proportion (%)	No. of Sites
Human I-v (100 sequences)	0.474	T	0.0002	32.995	5.4	2
		R	0.0002	12.583	11.8	1
		V	0.0004	34.478	4.0	2
		K	0.0023	29.301	7.2	1
Human 1933~1979 (32 strains) (egg-adapted)	0.613	D	0.0000	78.148	3.4	1
Human 1948~1979 (17 strains) (egg-adapted)	0.091	D	0.0007	133.611	5.1	1

doi:10.1371/journal.pone.0007789.t008

Table 9. Sites found to be under directional selection in human A(H1N1) HAs.

	Sites	Composition	Root	Preferred	Inferred Substitutions
Human I-v (100 sequences)	143	V ₉₉ T ₁	V	V	T→ ₁ V
	156	G ₉₀ R ₉ E ₁	R	R	G→ ₁ E, G→ ₇ R
	158	N ₉₆ K ₄	N	K	N→ ₄ K
	190	D ₆₇ N ₂₅ V ₈	D	V	D→ ₁₆ N, D→ ₄ V
	193	A ₅₂ T ₄₈	A	T	A→ ₆ T
	197	T ₈₂ K ₁₈	T	T	K→ ₂ T
Human 1933~1979 (32 strains) (egg-adapted)	225	G ₂₃ D ₉	D	D	D→ ₁ G, G→ ₆ D
Human 1948~1979 (17 strains) (egg-adapted)	225	G ₁₃ D ₄	G	D	G→ ₄ D

doi:10.1371/journal.pone.0007789.t009

stasis of swine A(H1N1) until 1998 [8] and agreed well with the suggestion that the introduction of the 2009 pandemic A(H1N1) virus into humans be a single event or multiple events of similar viruses [1,2,3,4,5,6,7,8].

The deviations from the ideal D190/D225 combination in A(H1N1) HAs might result in reduced binding to human receptors [11,41,42,64,79]. However, two possibilities, which are not mutually exclusive, may explain the fact that mutations are frequently observed at these two sites: one is that other concurring mutations at or around the receptor-binding site may sufficiently maintain the receptor binding affinity so that the overall binding affinity is largely unaffected; the second is that the gain in evading antibody neutralization far outweighs the reduction in receptor binding. Due to the overlapping locations of the ever-changing antigenic sites and the more-conserved receptor-binding site of HA, there is a constant dilemma of whether or not a residue at the receptor-binding site should change. Although the involvement of residues in antigenic drift that are critical for receptor binding was also observed in HAs of other types and subtypes including influenza B virus HA [80], H3 [43,44] and H5 HA [44,81], the interplay between these two opposing forces in HA evolution is still very poorly understood. Although previous studies on A(H3N2) HAs suggested covariation of antigenicity and receptor-binding specificity as a possible mechanism for the antigenic differences observed in viruses propagated in different cells [82], questions such as how residues involved in receptor binding are actively utilized for antigenic drift in influenza evolution in the same hosts need to be urgently addressed in order for us to comprehend the powerful strategies that the virus employs for recurring influenza infections.

Implications for the 2009 Pandemic

By analyzing hundreds of A(H1N1) HA sequences between 1918~2009, our study revealed positive selection in the subgroup I-v of A(H1N1) HAs. The positively selected codons were located at HA₁ 156 and 190 in the Sb antigenic site [83]. It was surprising that HA₁ 190, which is critical for receptor-binding specificity of A(H1N1) HAs, was also under positive selection. Through further analysis of HA₁ 190, together with HA₁ 225, the other critical determinant for receptor-binding specificity of A(H1N1), we found that the epidemic HAs and the 1918 pandemic and swine HAs favored one of these two sites for antigenic drift. Whether the 2009 pandemic A(H1N1) HA will adopt any of these two trends, or use a novel mechanism that does not involve HA₁ 190 and 225, will unfold in the coming months. If the latter is to be used, the 2009 A(H1N1) viruses may maintain their intrinsic high transmissibility, which, together with mutations in other genes such as NS1 and

PB1-F2 with signatures of elevated pathogenicity [1,2], may suffice a new disastrous pandemic in the near future.

Materials and Methods

Phylogenetic Analysis of A(H1N1) HAs

We obtained all available HA sequences (over 1,000) of non-egg-adapted A(H1N1) viruses for the period of 1918~2009 (as of July 10, 2009) from GISAID/Epifludb. We then removed the sequences with one or more ambiguous nucleotide sequences within the HA₁ region and deleted identical sequences. This gave us a dataset of 652 HA sequences that included three 1918 pandemic HAs, 575 epidemic HAs from 1979~2008 that collectively formed group I, and 73 pandemic HAs from 2009 and one HA from 2007 that belonged to group II. To facilitate the speed of computing, we further removed closely related sequences and obtained a dataset of 333 HA sequences. The program RDP3 (<http://darwin.uvigo.es/rdp/rdp.html>) [66] was used to make sure that no recombination was present in any of these HA sequences. The ClustalW method [68] with the MEGALIGN program of DNASTAR package (www.dnastar.com) was used for phylogenetic analysis of H1 HA sequences in the region of HA₁ (Fig. S1).

Due to the historic use of eggs for amplification of influenza viruses before sequencing, there presented a vacuum in sequence for non-egg-adapted A(H1N1) viruses between 1919 and 1979. In order to gain insights into the evolution of A(H1N1) viruses for this period, we separately collected a total of 32 different egg-adapted A(H1N1) HA sequences between 1933~1979 that were free of sequence ambiguity (Fig. S2). These sequences were similarly analyzed while keeping in mind of the possible egg-adapted mutations at HA₁ 138, 144, 163, 189, 190, 225, and 226 [62,63,64].

In order to compare the evolution of swine A(H1N1) HA sequences, we also retrieved 42 unique swine H1 HA sequences for the period of 1990~2009 that were free of ambiguous nucleotide sequences (Fig. S3). The reason that we focused on 1990~2009 was that previous studies suggested that swine A(H1N1) viruses be antigenically stable for the period of 1930 to 1990s [84].

Analysis of Positive Selection by PAML 4.0

The site-specific models implemented in the CODEML program in PAML 4.0 [45] was used to calculate heterogeneous selection pressure at amino-acid positions [45,54,72,85]. The models used in this study were M0, M1a, M2a, M7 and M8. M1a (nearly neutral), M7 (beta) and M8a (beta and $\omega = 1$) were null models that did not support $\omega > 1$. In contrast, the alternative

models M2a (positive selection) and M8 (beta and ω), compared to M1a and M7 respectively, each had an additional class that allowed $\omega > 1$. Likelihood ratio tests (LRT) comparing M2a versus M1a, M8 versus M7, and M8 versus M8a provided test for the existence of positive selection. In the test, twice the log likelihood difference, $2\Delta l = 2(l_1 - l_0)$, was calculated where l_1 and l_0 were the log likelihoods for the alternative model and null model, respectively. A larger value of LRT over those of χ^2 distribution led to rejection of the null models [72]. In order to calculate the codon-substitution models for heterogeneous selection pressure at each codon, the Bayes Empirical Bayes (BEB) analysis implemented in CODEML [72] was used, which has been shown to yield robust results even for small datasets. For all calculations, multiple runs, each with different initial parameter values, were performed to ensure optimization and convergence.

Directional Evolution of Protein Sequences Using HyPhy

Each group of A(H1N1) HA sequences aligned by the ClustalW method (Fig. S1, S2, S3) was input to the PhyML program [86] to generate an unrooted phylogenetic tree, which was then rooted using the Treeview software [87] by selecting the oldest sequence in each group as the root/ancestor. This rooted phylogenetic tree was used for directional evolution of protein sequences [49] implemented in the HyPhy [55] software package.

References

- Wang TT, Palese P (2009) Unraveling the mystery of swine influenza virus. *Cell* 137: 983–985.
- Neumann G, Noda T, Kawaoka Y (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 459: 931–939.
- Peiris JS, Poon LL, Guan Y (2009) Emergence of a novel swine-origin influenza A virus (S-OIV) H1N1 virus in humans. *J Clin Virol* 45: 169–173.
- Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, et al. (2009) Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 360: 2605–2615.
- Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, et al. (2009) Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science* 324: 1557–1561.
- Soloviyov A, Palacios G, Briese T, Lipkin WI, Rabadan R (2009) Cluster analysis of the origins of the new influenza A(H1N1) virus. *Euro Surveill* 14.
- Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, et al. (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459: 1122–1125.
- Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, et al. (2009) Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans. *Science*.
- Munster VJ, de Wit E, van den Brand JM, Herfst S, Schrauwen EJ, et al. (2009) Pathogenesis and Transmission of Swine-Origin 2009 A(H1N1) Influenza Virus in Ferrets. *Science*.
- Maines TR, Jayaraman A, Belser JA, Wadford DA, Pappas C, et al. (2009) Transmission and Pathogenesis of Swine-Origin 2009 A(H1N1) Influenza Viruses in Ferrets and Mice. *Science* 325(5939): 484–487.
- Reid AH, Janczewski TA, Lourens RM, Elliot AJ, Daniels RS, et al. (2003) 1918 influenza pandemic caused by highly conserved viruses with two receptor-binding variants. *Emerg Infect Dis* 9: 1249–1253.
- Logan WP, Mac KD (1951) Development of influenza epidemics. *Lancet* 1: 264–265.
- Collins SD, Lehmann J (1951) Trends and epidemics of influenza and pneumonia: 1918–1951. *Public Health Rep* 66: 1487–1516.
- Rasmussen AF Jr, Stokes JC, Smadel JE (1948) The Army experience with influenza, 1946–1947; laboratory aspects. *Am J Hyg* 47: 142–149.
- Sartwell PE, Long AP (1948) The Army experience with influenza, 1946–1947; epidemiological aspects. *Am J Hyg* 47: 135–141.
- Salk JE, Suriano PC (1949) Importance of antigenic composition of influenza virus vaccine in protecting against the natural disease; observations during the winter of 1947–1948. *Am J Public Health Nations Health* 39: 345–355.
- Kilbourne ED, Smith C, Brett I, Pokorny BA, Johansson B, et al. (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 U S A* 99: 10748–10752.
- Isaacs A, Gledhill AW, Andrewes CH (1952) Influenza A viruses; laboratory studies, with special reference to European outbreak of 1950–1. *Bull World Health Organ* 6: 287–315.

Supporting Information

Figure S1 Phylogenetic tree of 333 HA sequences of A (H1N1) influenza viruses isolated between 1918~2009 without egg-adaptation.

Found at: doi:10.1371/journal.pone.0007789.s001 (41.64 MB TIF)

Figure S2 Phylogenetic tree of 32 HA sequences of egg-adapted human A(H1N1) influenza viruses isolated between 1933~1979.

Found at: doi:10.1371/journal.pone.0007789.s002 (1.02 MB TIF)

Figure S3 Phylogenetic tree of 42 HA sequences of swine A(H1N1) influenza viruses isolated between 1990~2009.

Found at: doi:10.1371/journal.pone.0007789.s003 (1.39 MB TIF)

Acknowledgments

We gratefully thank Dr. Robert Couch for insightful comments on the manuscript, Drs. Alexander Klimov, Xiyan Xu and Rebecca Garten for help with sequence retrieval, Dr. Sergei Kosakovsky Pond for help with the HyPhy software, Mingyang Lu for Figure 1 and Athanasios Dousis for computational help.

Author Contributions

Conceived and designed the experiments: JM QW. Performed the experiments: JS QW. Analyzed the data: JM QW. Contributed reagents/materials/analysis tools: JM QW. Wrote the paper: JS JM QW.

- Viboud C, Tam T, Fleming D, Miller MA, Simonsen L (2006) 1951 influenza epidemic, England and Wales, Canada, and the United States. *Emerg Infect Dis* 12: 661–668.
- Viboud C, Tam T, Fleming D, Handel A, Miller MA, et al. (2006) Transmissibility and mortality impact of epidemic and pandemic influenza, with emphasis on the unusually deadly 1951 epidemic. *Vaccine* 24: 6701–6707.
- Scholtissek C, Rohde W, Von Hoyningen V, Rott R (1978) On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* 87: 13–20.
- Nakajima K, Desselberger U, Palese P (1978) Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* 274: 334–339.
- Kendal AP, Noble GR, Skehel JJ, Dowdle WR (1978) Antigenic similarity of influenza A (H1N1) viruses from epidemics in 1977–1978 to “Scandinavian” strains isolated in epidemics of 1950–1951. *Virology* 89: 632–636.
- Scholtissek C, von Hoyningen V, Rott R (1978) Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). *Virology* 89: 613–617.
- Taubenberger JK, Reid AH, Janczewski TA, Fanning TG (2001) Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos Trans R Soc Lond B Biol Sci* 356: 1829–1839.
- Shope RE (1931) The Etiology of Swine Influenza. *Science* 73: 214–215.
- Morens DM, Taubenberger JK, Fauci AS (2009) The Persistent Legacy of the 1918 Influenza Virus. *N Engl J Med* 361(3): 225–229.
- Brown IH (2000) The epidemiology and evolution of influenza viruses in pigs. *Vet Microbiol* 74: 29–46.
- Olsen CW (2002) The emergence of novel swine influenza viruses in North America. *Virus Res* 85: 199–210.
- Pensaert M, Ottis K, Vandeputte J, Kaplan MM, Bachmann PA (1981) Evidence for the natural transmission of influenza A virus from wild ducts to swine and its potential importance for man. *Bull World Health Organ* 59: 75–78.
- Brown IH, Ludwig S, Olsen CW, Hannoun C, Scholtissek C, et al. (1997) Antigenic and genetic analyses of H1N1 influenza A viruses from European pigs. *J Gen Virol* 78 (Pt 3): 553–562.
- Donatelli I, Campitelli L, Castrucci MR, Ruggieri A, Sidoli L, et al. (1991) Detection of two antigenic subpopulations of A(H1N1) influenza viruses from pigs: antigenic drift or interspecies transmission? *J Med Virol* 34: 248–257.
- Reid AH, Fanning TG, Janczewski TA, Lourens RM, Taubenberger JK (2004) Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. *J Virol* 78: 12462–12470.
- Dunham EJ, Dugan VG, Kaser EK, Perkins SE, Brown IH, et al. (2009) Different evolutionary trajectories of European avian-like and classical swine H1N1 influenza A viruses. *J Virol* 83: 5485–5494.
- Brown IH, Harris PA, McCauley JW, Alexander DJ (1998) Multiple genetic reassortment of avian and human influenza A viruses in European pigs, resulting

- in the emergence of an H1N2 virus of novel genotype. *J Gen Virol* 79 (Pt 12): 2947–2955.
36. Webby RJ, Swenson SL, Krauss SL, Gerrish PJ, Goyal SM, et al. (2000) Evolution of swine H3N2 influenza viruses in the United States. *J Virol* 74: 8243–8251.
 37. Newman AP, Reisdorf E, Beinemann J, Uyeki TM, Balish A, et al. (2008) Human case of swine influenza A (H1N1) triple reassortant virus infection, Wisconsin. *Emerg Infect Dis* 14: 1470–1472.
 38. Shinde V, Bridges CB, Uyeki TM, Shu B, Balish A, et al. (2009) Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N Engl J Med* 360: 2616–2625.
 39. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69: 531–569.
 40. Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solorzano A, et al. (2007) A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* 315: 655–659.
 41. Srinivasan A, Viswanathan K, Raman R, Chandrasekaran A, Raguram S, et al. (2008) Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses. *Proc Natl Acad Sci U S A* 105: 2800–2805.
 42. Stevens J, Blixt O, Glaser L, Taubenberger JK, Palese P, et al. (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol* 355: 1143–1155.
 43. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM (1999) Predicting the evolution of human influenza A. *Science* 286: 1921–1925.
 44. 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: 1457–1465.
 45. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
 46. Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* 46: 409–418.
 47. Delport W, Scheffler K, Seoighe C (2008) Frequent toggling between alternative amino acids is driven by selection in HIV-1. *PLoS Pathog* 4: e1000242.
 48. Seoighe C, Ketwaroo F, Pillay V, Scheffler K, Wood N, et al. (2007) A model of directional selection applied to the evolution of drug resistance in HIV-1. *Mol Biol Evol* 24: 1025–1031.
 49. Kosakovsky Pond SL, Poon AF, Leigh Brown AJ, Frost SD (2008) A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol Biol Evol* 25: 1809–1824.
 50. Anisimova M, Kosiol C (2009) Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol Biol Evol* 26: 255–271.
 51. Delport W, Scheffler K, Seoighe C (2009) Models of coding sequence evolution. *Brief Bioinform* 10: 97–109.
 52. Lemey P, MS, Vandamme A (2009) *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge: Cambridge University Press. 430 p.
 53. Yang Z (2006) *Computational Molecular Evolution*. Oxford: Oxford University Press. 376 p.
 54. Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.
 55. Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
 56. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
 57. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
 58. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15: 568–573.
 59. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929–936.
 60. Yang Z (2000) Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol* 51: 423–432.
 61. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19: 908–917.
 62. Robertson JS, Bootman JS, Newman R, Oxford JS, Daniels RS, et al. (1987) Structural changes in the haemagglutinin which accompany egg adaptation of an influenza A(H1N1) virus. *Virology* 160: 31–37.
 63. Xu X, Rocha EP, Regenerly HL, Kendal AP, Cox NJ (1993) Genetic and antigenic analyses of influenza A (H1N1) viruses, 1986–1991. *Virus Res* 28: 37–55.
 64. Gambaryan AS, Robertson JS, Matrosovich MN (1999) Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. *Virology* 258: 232–239.
 65. Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164: 1229–1236.
 66. Heath L, van der Walt E, Varsani A, Martin DP (2006) Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 80: 11827–11832.
 67. Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. *Nat Rev Genet* 8: 196–205.
 68. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
 69. Serum cross-reactive antibody response to a novel influenza A (H1N1) virus after vaccination with seasonal influenza vaccine. *MMWR Morb Mortal Wkly Rep* 58: 521–524.
 70. Finkelmann BS, Viboud C, Koelle K, Ferrari MJ, Bharti N, et al. (2007) Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One* 2: e1296.
 71. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ (2006) Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol Direct* 1: 34.
 72. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118.
 73. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W (1982) The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* 31: 417–427.
 74. Gerhard W, Yewdell J, Frankel ME, Webster R (1981) Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* 290: 713–717.
 75. Rogers GN, D'Souza BL (1989) Receptor binding properties of human and animal H1 influenza virus isolates. *Virology* 173: 317–322.
 76. Matrosovich MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, et al. (1997) Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* 233: 224–234.
 77. Nobusawa E, Nakajima K, Nakajima S (1987) Determination of the epitope 264 on the hemagglutinin molecule of influenza H1N1 virus by site-specific mutagenesis. *Virology* 159: 10–19.
 78. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, et al. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 303: 1838–1842.
 79. Gambaryan AS, Tuzikov AB, Piskarev VE, Yamnikova SS, Lvov DK, et al. (1997) Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyl)lactosamine. *Virology* 232: 345–350.
 80. Shen J, Kirk BD, Ma J, Wang Q (2009) Diversifying selective pressure on influenza B virus hemagglutinin. *J Med Virol* 81: 114–124.
 81. Shi W, Gibbs MJ, Zhang Y, Zhuang D, Dun A, et al. (2008) The variable codons of H5N1 avian influenza A virus haemagglutinin genes. *Sci China C Life Sci* 51: 987–993.
 82. Daniels RS, Douglas AR, Skehel JJ, Wiley DC, Naeve CW, et al. (1984) Antigenic analyses of influenza virus haemagglutinins with different receptor-binding specificities. *Virology* 138: 174–177.
 83. 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: 373–378.
 84. Sheerar MG, Easterday BC, Hinshaw VS (1989) Antigenic conservation of H1N1 swine influenza viruses. *J Gen Virol* 70 (Pt 12): 3297–3303.
 85. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–556.
 86. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
 87. Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357–358.