

Evolutionary Dynamics and Emergence of Panzootic H5N1 Influenza Viruses

Dhanasekaran Vijaykrishna^{1,2}, Justin Bahl^{1,2}, Steven Riley³, Lian Duan^{1,2}, Jin Xia Zhang^{1,2}, Honglin Chen^{1,2}, J. S. Malik Peiris¹, Gavin J. D. Smith^{1,2*}, Yi Guan^{1,2*}

1 State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Special Administrative Region, China, **2** International Institute of Infection and Immunity, Shantou University Medical College, Shantou, Guangdong, China, **3** Department of Community Medicine and School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China

Abstract

The highly pathogenic avian influenza (HPAI) H5N1 virus lineage has undergone extensive genetic reassortment with viruses from different sources to produce numerous H5N1 genotypes, and also developed into multiple genetically distinct sublineages in China. From there, the virus has spread to over 60 countries. The ecological success of this virus in diverse species of both poultry and wild birds with frequent introduction to humans suggests that it is a likely source of the next human pandemic. Therefore, the evolutionary and ecological characteristics of its emergence from wild birds into poultry are of considerable interest. Here, we apply the latest analytical techniques to infer the early evolutionary dynamics of H5N1 virus in the population from which it emerged (wild birds and domestic poultry). By estimating the time of most recent common ancestors of each gene segment, we show that the H5N1 prototype virus was likely introduced from wild birds into poultry as a non-reassortant low pathogenic avian influenza H5N1 virus and was not generated by reassortment in poultry. In contrast, more recent H5N1 genotypes were generated locally in aquatic poultry after the prototype virus (A/goose/Guangdong/1/96) introduction occurred, i.e., they were not a result of additional emergence from wild birds. We show that the H5N1 virus was introduced into Indonesia and Vietnam 3–6 months prior to detection of the first outbreaks in those countries. Population dynamics analyses revealed a rapid increase in the genetic diversity of A/goose/Guangdong/1/96 lineage viruses from mid-1999 to early 2000. Our results suggest that the transmission of reassortant viruses through the mixed poultry population in farms and markets in China has selected HPAI H5N1 viruses that are well adapted to multiple hosts and reduced the interspecies transmission barrier of those viruses.

Citation: Vijaykrishna D, Bahl J, Riley S, Duan L, Zhang JX, et al. (2008) Evolutionary Dynamics and Emergence of Panzootic H5N1 Influenza Viruses. *PLoS Pathog* 4(9): e1000161. doi:10.1371/journal.ppat.1000161

Editor: Ron A. M. Fouchier, Erasmus Medical Center, The Netherlands

Received: June 17, 2008; **Accepted:** August 22, 2008; **Published:** September 26, 2008

Copyright: © 2008 Vijaykrishna 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: This study was supported by the Li Ka Shing Foundation, the National Institutes of Health (NIAID contract HHSN266200700005C), the Area of Excellence Scheme of the University Grants Committee (Grant AoE/M-12/06) and Research Grants Council (HKU 7512/06 M) of the Hong Kong SAR Government. GJDS is supported by a career development award under NIAID contract HHSN266200700005C.

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

* E-mail: gjsmith@hku.hk (GJDS); yguan@hku.hk (YG)

Introduction

Outbreaks of highly pathogenic avian influenza (HPAI) H5N1 virus were first recorded in Guangdong, China in 1996 [1]. Since its emergence, the A/goose/Guangdong/1/96 (Gs/GD) virus lineage has become the longest recorded HPAI virus to remain endemic in poultry [2]. The ecological success of this virus in diverse avian and mammalian species [3] with frequent introduction to humans suggests this virus is the most likely candidate of the next human pandemic [4]. Therefore, the evolutionary and ecological characteristics of its emergence from wild birds into poultry are of considerable interest.

The virus gradually became endemic in poultry in different regions of China, developing into genetically and antigenically distinct sublineages [5,6]. The geographic spread of these sublineages outside China is also unprecedented, with two sublineages spreading to Southeast Asia in late 2003, and another westwards to Central Asia, Europe, Africa, the Middle East and the Indian subcontinent in mid-2005 [1,5–9]. During mid-2005, one

sublineage (Fujian-like or clade 2.3.4) became dominant in China and subsequently spread to Laos, Thailand and Vietnam [2,10].

Influenza surveillance in southern China has revealed that the Gs/GD virus lineage underwent extensive genetic reassortment to generate many different reassortant viruses (or genotypes) between 1997 and 2006 [5,6]. The non-reassortant Gs/GD-like viruses were prevalent only from 1996 to 2000 [11]. Afterwards, all H5N1 viruses detected were reassortant genotypes. Amongst all recognized reassortants, only genotypes B, X0, W, Z, G and V, were persistent for more than two years or predominant at different time points, while many genotypes were only detected occasionally [5,6].

While the genetic and antigenic evolution and geographic spread of the HPAI H5N1 panzootic viruses are well documented after the initiation of systematic surveillance in 2000 [5], little is known about the source and early evolutionary dynamics of H5N1 virus. In particular, it is still unknown whether the Gs/GD virus itself was a reassortant virus or introduced wholly from migratory waterfowl. Even though the internal gene sources for most genotypes have been identified from aquatic birds [12], the order

Author Summary

H5N1 influenza virus has been responsible for poultry outbreaks over the last 12 years—the longest recorded example of highly pathogenic avian influenza (HPAI) circulation in poultry. The ecological success of this virus in diverse species of both poultry and wild birds with sporadic introduction to humans suggests that it is a likely source of the next human pandemic. Genome sequences of H5N1 viruses reveal extensive genetic reassortment (mixing) with other influenza subtypes to produce many H5N1 genotypes that have developed into multiple genetically distinct clades, some of which have spread to affect over 60 countries. Here, we analyze all available sequence data of avian influenza viruses from Eurasia and show that the original HPAI H5N1 virus (referred to as A/goose/Guangdong/1/96) was likely introduced directly into poultry as an intact virus particle from wild aquatic birds. In contrast, H5N1 genotypes were generated in aquatic poultry populations after the introduction of A/goose/Guangdong/1/96 virus. Our results suggest that the transmission of reassortant viruses through the diverse poultry populations in farms and markets in China has selected H5N1 viruses that are well-adapted to multiple hosts and reduced the interspecies transmission barrier of those viruses.

of emergence and reassortment events that generated these genotypes remains to be answered.

Recently, Dugan et al. [13] described an evolutionary dynamics of avian influenza virus (AIV) in wild bird populations. They found that the viruses exist as a large pool of functionally equivalent genes, under strong purifying selection, and reassort without overall loss of virus fitness resulting in many transient genotypes. Other studies have shown that all eight gene segments of the HPAI H5N1 viruses exhibit rapid nucleotide and amino acid substitution rates in comparison to other subtypes and that reassortment with the wild bird AIV gene pool has increased H5N1 genotype diversity, producing both persistent and transient genotypes [5,12,14]. However, the mechanism behind the population behavior of H5N1 viruses has still not been investigated. Here we aim to address these gaps in our knowledge of the evolutionary dynamics of HPAI H5N1 viruses through the estimation of divergence times of gene segments of major reassortants and population dynamics analyses of H5N1 viruses in poultry.

Results

Study design

In the present study, all H5N1 influenza viruses with a hemagglutinin gene derived from A/goose/Guangdong/1/96 were designated as belonging to the “Gs/GD lineage”. The H5N1 reassortants from the Gs/GD lineage were referred to by their genotype designation, such as B, X0, Z, V, W and G, as described previously [5,6]. Non-reassortant viruses with all eight gene segments closely related to A/goose/Guangdong/1/96 were referred to as “Gs/GD-like”.

Our results are presented in four parts: 1) estimation of the time of emergence of the Gs/GD-like virus; 2) estimation of the dates of emergence of genotypes; 3) estimation of the dates of emergence of H5N1 clades; and 4) description of the evolutionary dynamics of HPAI H5N1 viruses in poultry in China. Phylogenetic analysis was conducted using relaxed clock and Bayesian skyline coalescent models [15,16] as implemented in the BEAST software package version 1.4.7 [17].

First, dates of divergence of the H5-hemagglutinin (HA) and N1-neuraminidase (NA) genes of representative Eurasian viruses were estimated. Due to the large number of Gs/GD-like H5N1 virus sequences available, a representative subset of H5N1 sequences were aligned with all publicly available avian H5-HA and N1-NA sequences. The final datasets consisted of 116 H5-HA and 89 N1-NA genes of 1,014 and 1,356 nucleotides in length, respectively, and are referred to as “Eurasian” datasets. Because the exact isolation dates of many viruses were not available, the mid-years of virus isolation were used as calibration points.

For dating the emergence of the internal genes of the Gs/GD-like virus and of the H5N1 genotypes, we used internal gene datasets (referred to as “Asian” datasets) that included representative viruses of all identified H5N1 genotypes, plus viruses of other subtypes from poultry and wild aquatic birds. The mid-years of virus isolation were used as calibration points to estimate the time of incorporation of novel gene segments into Gs/GD-like genotypes. Because the non-structural (NS) genes of Gs/GD-like viruses (allele A) and most of its reassortants (allele B) are highly divergent [1], the time of most recent common ancestor (TMRCA) for each allele were calculated separately.

To estimate the dates of emergence of different Gs/GD-like H5N1 clades, H5-HA datasets were recruited that contained representative viruses of known H5N1 variants. The final dataset consisted of the complete HA gene (1,733 nucleotides in length) of 192 viruses, and is also referred to as the “Asian” H5-HA dataset. The exact dates of sampling were known for most of these samples and were used as calibration time points. For those sequences for which exact virus isolation dates were not available, the mid-month date (either the 14th or 15th) was used as the calibration point. However, in some instances, only the date of initial detection of an H5N1 outbreak was available and this was used for calibration [18].

Finally, to examine changes in genetic diversity during the evolution of the Gs/GD lineage in China, we constructed Bayesian skyline plots of the virus population described by a modified Asian H5-HA dataset [19]. This dataset consisted of HA genes of viruses isolated from chicken ($n = 54$), duck ($n = 52$), goose ($n = 15$), pheasant (1) and Guinea fowl (1) in China.

Rates of nucleotide substitution

Our analyses showed that the mean substitution rates for the H5 and N1 datasets were 4.77×10^{-3} and 5.19×10^{-3} substitutions per site, per year (subst/site/year), respectively, which were significantly higher than the rates of the internal gene segments (1.84 – 2.62×10^{-3} subst/site/year) (Table 1). These rates are similar to those previously described for avian, human, equine and swine influenza viruses [14,20,21].

Emergence of the Gs/GD virus

Analysis of the Eurasian H5-HA dataset revealed four distinct H5 lineages currently circulating in Eurasia: the Gs/GD lineage plus two lineages exclusively of European virus isolates (groups A and B) and one containing isolates from throughout Eurasia (group C) (Figure 1A, S1A). The time of divergence of these four lineages (node I) was Jul 1989 (95% highest posterior density Jul 1986–Jun 1992) (Figure 1A; Table 2). The TMRCA for the Gs/GD lineage (node IV) was estimated to be Jan 1994 (95% highest posterior density (HPD) Apr 1992–Nov 1995) (Figure 1A; Table 2). The TMRCA of the remaining groups are presented in Table 2.

Similarly, multiple N1-NA lineages were also recognized in Eurasia (Figure 1B, S1B). The TMRCA for the Gs/GD lineage NA (node V) was estimated to be Oct 1994 (HPD Mar 1993–Jan 1996) (Figure 1B; Table 2). It was noted that the N1 gene of the

Table 1. Best-fit relaxed clock model and mean nucleotide substitution rates

Dataset	Gene	Uncorrelated relaxed clock model	Mean substitution rate ($\times 10^{-3}$)	Substitution rate HPD ($\times 10^{-3}$)
Eurasian	H5	exponential	4.77	3.88–5.74
	N1	exponential	5.19	4.17–6.17
Asian	PB2	lognormal	2.41	1.92–2.90
	PB1	lognormal	2.59	2.13–3.04
	PA	exponential	2.62	1.9–3.2
	NP	exponential	2.47	1.9–3.03
	M	exponential	1.84	1.3–2.30
	NS	exponential	2.51	1.77–3.44
	N1	exponential	4.27	3.27–5.25

doi:10.1371/journal.ppat.1000161.t001

HK/97-like H5N1 virus clustered along with the NA genes of H6N1 viruses from poultry (node VII), with an estimated TMRCA of Jan 1994 (HPD Aug 1990–Oct 1996) (Figure 1B; Table 2). The TMRCA for each of the internal gene segments of the Gs/GD-like viruses were estimated using the Asian datasets and ranged from Jul 1993 for the polymerase acidic (PA) gene to May 1995 for the NS gene (Figure 2 and Table 3).

Bayes factor (BF) tests showed no significant difference between TMRCA of all eight gene segments of the Gs/GD viruses (Table S1). These results indicate that the common ancestor of the Gs/GD-like virus was generated, probably in wild waterfowl, approximately 2 years before it was first detected in poultry in 1996.

Generation of H5N1 reassortant viruses

The TMRCA of the internal gene segments for each major H5N1 genotype (B, X0, W, Z, G and V) were estimated using the Asian datasets from which we inferred the dates of emergence of these genotypes (Figures 2 and 3; Table 3). The mean TMRCA for the internal genes of genotype B virus ranged from Apr 1995 to Nov 1997 (Table 3). Bayes factor tests showed that the internal

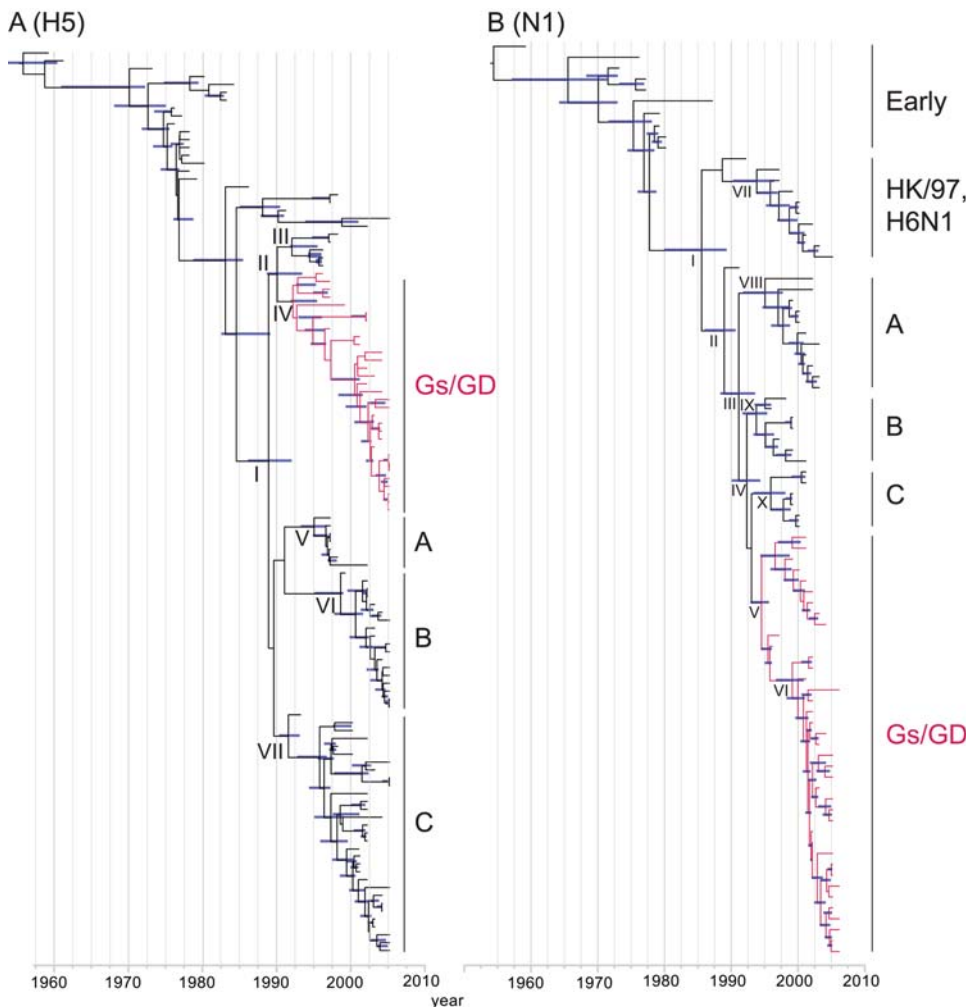


Figure 1. Dated phylogeny of the surface genes of H5N1 viruses isolated in Eurasia. The HA gene (A) and N1 gene (B) trees scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. Red branches indicate Gs/GD lineage H5N1 viruses. Identical phylogenetic trees with virus names are shown in Figure S1. TMRCA and HPDs for each of the nodes marked with Roman numerals are given in Table 2.
doi:10.1371/journal.ppat.1000161.g001

Table 2. Estimated TMRCAs for the Eurasian H5-HA and N1-NA datasets

Dataset	Node ¹	Description	Mean TMRCA (95% HPD) ²
H5 HA dataset	I	Current Eurasia	1989.51 (1986.52–1992.42)
	II	Gs/GD+out-group	1991.69 (1989.06–1993.83)
	III	Hokkaido/Singapore	1994.21 (1992.19–1995.91)
	IV	Gs/GD H5N1	1994.04 (1992.27–1995.84)
	V	Europe A	1995.64 (1993.64–1997.06)
	VI	Europe B	1997.88 (1995.47–1999.39)
	VII	Eurasia C	1992.48 (1992.71–1993.49)
N1 NA dataset	I	Current Eurasia	1985.83 (1980.32–1989.67)
	II	Eurasia	1989.21 (1986.30–1991.01)
	III	Eurasia	1991.40 (1988.73–1993.94)
	IV	Eurasia	1992.59 (1990.45–1994.71)
	V	GsGD	1994.81 (1993.23–1996.02)
	VI	GsGD with 20 AA deletion	1999.42 (1997.05–2001.23)
	VII	HK97	1994.08 (1990.61–1996.79)
	VIII	Eurasia A	1995.33 (1992.10–1998.08)
	IX	Eurasia B	1994.03 (1992.04–1995.75)
	X	Eurasia C	1996.19 (1993.68–1998.49)

¹Nodes indicate TMRCAs of major H5 and N1 lineages, as shown in Figures 1A, B.
²The dates are presented as year followed by proportion of days.

doi:10.1371/journal.ppat.1000161.t002

genes of genotype B were incorporated from three different sources (Table S1). This suggests that the three internal gene segments of genotype B virus (polymerase basic 2 (PB2), polymerase basic 1 (PB1), polymerase acidic (PA)) were derived from the same virus (mean TMRCA range Nov 1995–Jul 1997), while the NP and NS genes (mean TMRCA range Apr 1995–Nov 1996) was from a different source (Table 3), while the M gene (mean TMRCA Nov 1997) was from an independent sources. Therefore, genotype B virus was generated from viruses of four independent sources, i.e. it contains surface genes from Gs/GD and internal genes from three other sources. The mean TMRCA of the last incorporated sources (M gene) was Nov 1997. This date represents the earliest possible time when all genotype B internal gene segments co-circulated in the poultry gene pool in China. Therefore, genotype B virus was generated after mid-1997 (Figure 3).

Both genotypes B and W appear to have emerged during the same period, i.e. after mid-1997 (Figure 3). Genotype W shares four internal gene segments (PB1, PA, M and NS) with genotype B, while its PB2 and NP segments are from two different aquatic sources (Figures 2 and S2). The TMRCAs for the PB2 and NP genes were estimated at Dec 1996 (HPD Feb 1993–Mar 2000) and Jan 1995 (HPD Mar 1992–Jan 1997), respectively, and these dates were significantly different (Tables 3 and S1). These results suggest that genotype W has been generated through a reassortment of viruses from six different sources.

However, the estimated number of gene sources for genotypes B and W should be treated with caution. As the mid-year of isolation (e.g. 2000.5) was used to calibrate the internal gene datasets, the extension of confidence intervals to account for this error indicates that the three internal gene sources for genotype B may not be significantly different. The availability of the date of isolation for

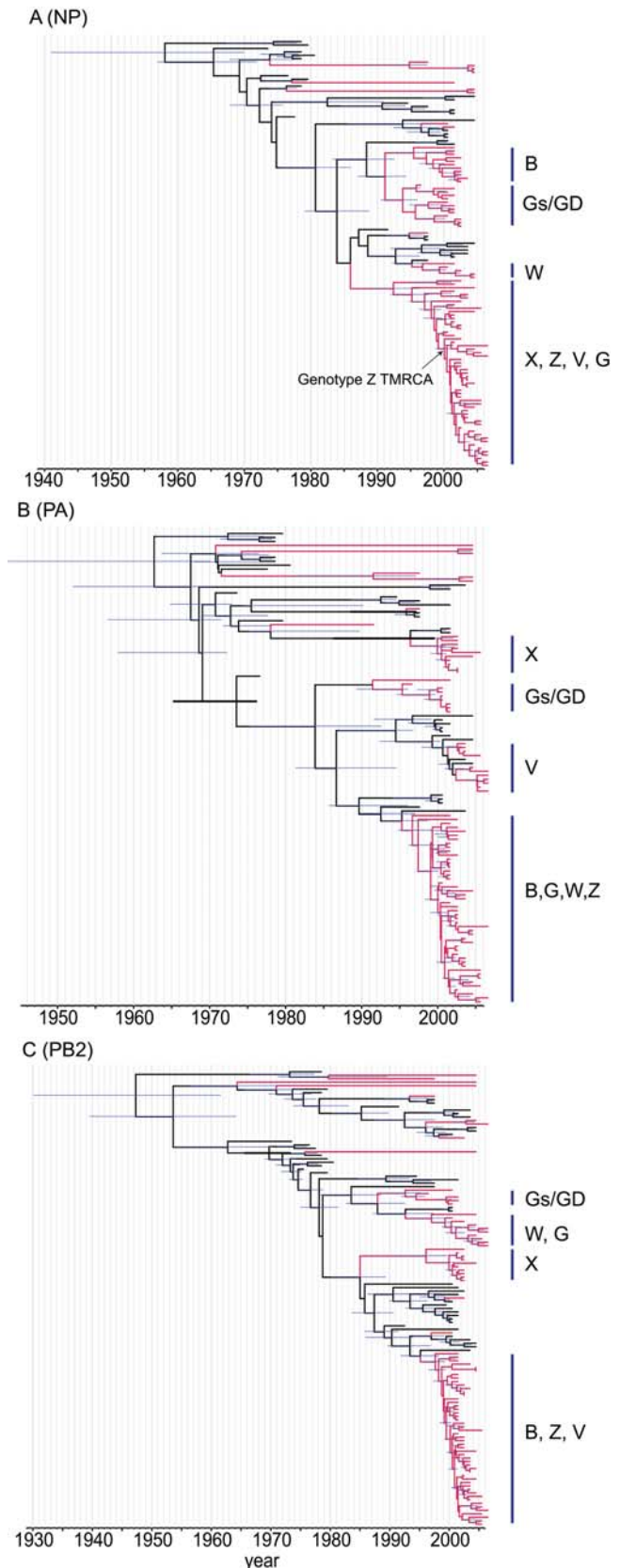


Figure 2. Dated phylogeny of the internal genes of viruses isolated in Asia. The NP (A), PA (B) and PB2 (C) gene trees scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCAs

and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. Red branches indicate Gs/GD lineage H5N1 viruses. Identical phylogenetic trees with virus names are shown in Figure S2. TMRCA and HPDs for each of the major H5N1 genotype internal genes are given in Table 3.

doi:10.1371/journal.ppat.1000161.g002

viruses would be needed to estimate the age of gene sources for genotypes B and W with greater confidence.

Genotype X0 does not share any internal genes with genotypes B and W, and has incorporated internal genes from two different sources, i.e. the PB1, PA, M and NS genes were from one source, and while PB2 and NP genes are from another (Figures 2 and S2). For genotype X0 virus, the mean TMRCA range of the internal genes was Apr 1992 to May 2000 (Table 3). The TMRCA of the PB1, PA, M and NS genes were not significantly different (Table S1). However, the TMRCA of the NP and PB2 genes were significantly earlier than the other genes, but without significant difference between them, suggesting a common source. The averaged TMRCA means of the five most recent internal genes (Jul 1999) indicates that genotype X0 was generated after mid 1999 (Table 3 and Figure 3).

The predominant genotype Z virus was derived from genotype B, with which it shares five internal genes (PB2, PB1, PA, M and NS). The NP gene has a common source with genotype X0 viruses. The TMRCA of the NP gene (Figure 2A) was estimated at Dec 1999 (HPD Sep 1998–Jan 2001), indicating that genotype Z emerged late 1999/early 2000 (Figure 3 and Table 3). Genotype V virus shares all internal genes with genotype Z except for the PA gene. The PA gene of genotype V viruses are most closely related to viruses isolated from the aquatic gene pool and therefore is most likely derived from an aquatic source (Figure 2B). The TMRCA for the PA gene was estimated as Mar 2002 (HPD Jan 2001–Jan 2003), suggesting genotype V was generated after early-2002 (Table 3). Genotype G viruses have five internal genes (PB2, PB1, PA, M and NS) in common with genotype W, while the NP gene groups with those from genotype Z and V viruses (Figures 2 and 3). Therefore, it was not possible to estimate the date of emergence of genotype G virus.

Geographical expansion of H5N1

The introduction to Indonesia occurred approximately 3–4 months before the introduction to Vietnam and Thailand (Figure 4 and Table 4). The TMRCA for viruses isolated from Vietnam, Thailand and Malaysia (clade 1, node VI) was estimated at Mar 2003 (HPD Oct 2002–Aug 2003), while the TMRCA for the Indonesia viruses (clade 2.1, node VIII) was estimated at Nov 2002 (HPD Jul 2002–Feb 2003) (Figure 4). A BF test indicated these dates were significantly different. Both of these TMRCA are approximately 3–6 months before the first observed H5N1 outbreaks in the field in these countries.

The TMRCA of the Qinghai-like HA variant (clade 2.2, node IX) was estimated as Dec 2004 (HPD Aug 2004–Feb 2005), while the Fujian-like HA TMRCA (clade 2.3.4, node XIII) was estimated as Sep 2004 (Mar 2004–Feb 2005) (Figure 4). These TMRCA were not significantly different (Table S1), indicating that these two variants arose during the same period.

Estimation of genetic diversity and selection pressure

Coalescent reconstruction using a HA-H5 dataset of poultry isolated in China revealed a rapid increase in the relative genetic diversity of Gs/GD lineage viruses from mid-1999 to 2000 (Figure 5A). It was during this period, preceding the current epizootic, that each of the major HA lineages were generated and

Table 3. TMRCA of internal gene segments of H5N1 viruses

	GsGD (Gs/GD/1/1996)	X (Dk/ST/4912/2001)	B (Ck/HK/YU562/01)	W (Dk/Zhejiang/52/00)	Z (Ck/HK/YU22/02)	V (B.bird/Hunan/1/04)	G (Dk/Guangxi/13/04)
PB2	1994.34 (1992.01–1995.95)	1996.04 (1991.11–1999.75)	1997.58 (1995.63–1999.39)	1996.96 (1993.12–2000.19)	B*	B	W
PB1	1994.77 (1993.62–1995.71)	2000.41 (1999.37–2001.26)	1995.84 (1993.1–1998.27)	B	B	B	B
PA	1993.52 (1989.21–1996.05)	1999.16 (1996.56–2000.88)	1997.29 (1994.73–1999.29)	B	B	2002.18 (2001.03–2003.06)	B
NP	1993.67 (1990.46–1995.94)	1992.29 (1988.19–1996.43)	1995.32 (1991.69–1998.41)	1995.08 (1992.21–1997.06)	1999.94 (1998.67–2001.06)	X0	X0
M	1994.23 (1990.7–1996.36)	1999.49 (1997.42–2001.08)	1997.83 (1995.58–1999.7)	B	B	B	B
NS	1995.35 (1992.33–1997.29)	1999.31 (1997.25–2001.14)	1996.88 (1993.7–1999.46)	B	B	B	B

*Refers to the gene being derived from an earlier H5N1 genotype. doi:10.1371/journal.ppat.1000161.t003

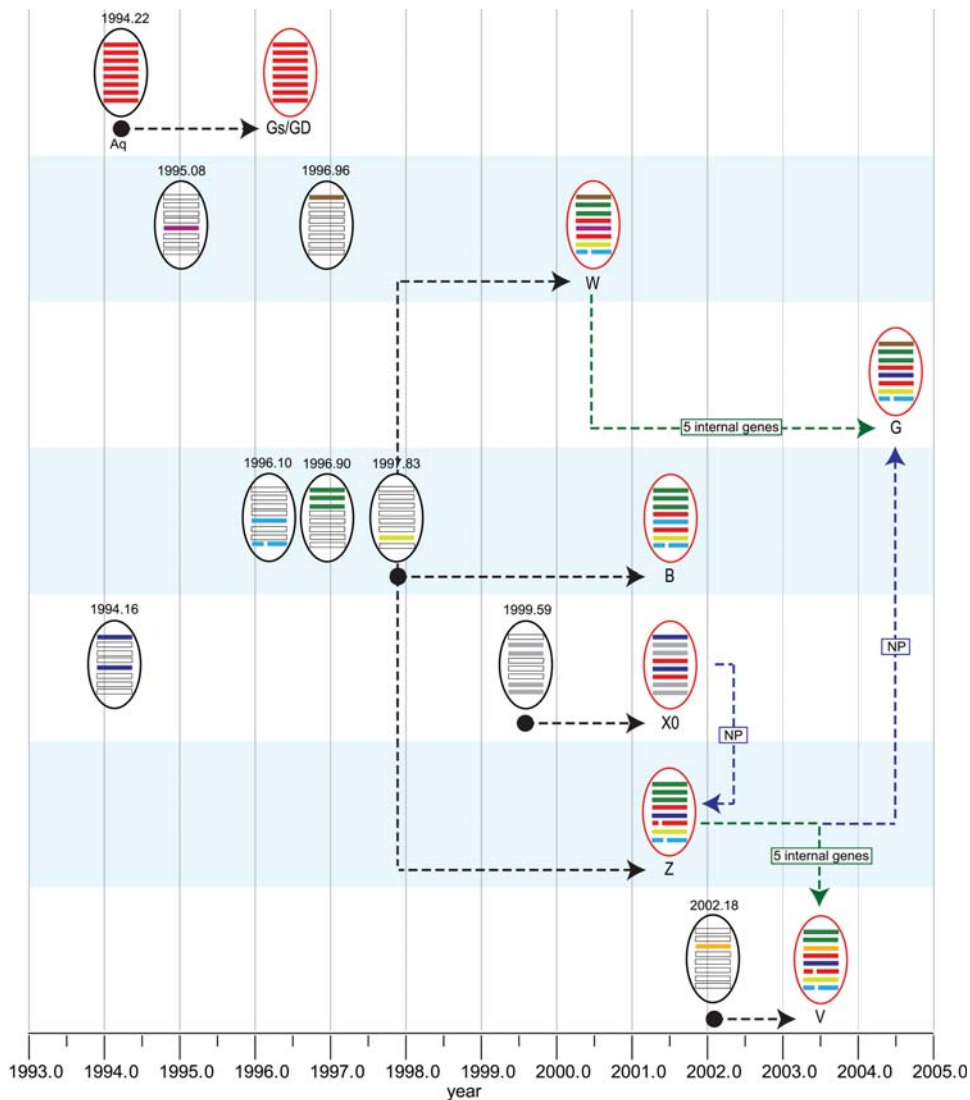


Figure 3. Diagram representing the emergence of major H5N1 reassortant viruses. Virus particles outlined in black represent donor viruses (with mean TMRCAs above the particle) and those outlined in red represent characterized H5N1 genotypes placed at the year of first detection. Gene segments are ordered PB2, PB1, PA, HA, NP, NA, M and NS from top to bottom within the virus particle diagram. Arrows represent possible reassortment pathways of genotype development. The start of the black arrows (filled circles) indicate the earliest possible time of corresponding genotype generation. Colored arrows represent reassortment between existing H5N1 genotypes.
doi:10.1371/journal.ppat.1000161.g003

subsequently became widespread in poultry throughout China (Figure 5B).

Analysis of selection pressures show that the HA of all the major H5N1 sublineages (or clades) were subject to strong purifying selection (mean ratio of rates of nonsynonymous to synonymous substitutions per site, $d_N/d_S = 0.272$). However, four amino acid sites (positions 138, 140, 141 and 156) were under positive selection ($P < 0.05$), consistent with previous results for AIV [14,22,23].

Discussion

We used extensive sequence data and relaxed clock models to date each of the eight gene segments of the Gs/GD-like H5N1 virus. Previously, phylogenetic analyses have shown that the gene segments of Gs/GD-like viruses were most closely related to those viruses from migratory birds [12]. However, the time of introduction into poultry was not established. We have shown

that the prototype virus was likely introduced into poultry as a non-reassortant low pathogenic avian influenza H5N1 virus, rather than being generated by reassortment within poultry. We have also shown that the reassortment events that generated these H5N1 genotypes occurred locally in domestic duck after the Gs/GD-like virus introduction (Figure 3).

Our data help provide insights into the sequence of events that led to each of the three H5N1 transmission waves. Previously, the earliest records of H5N1 wave 1 outbreaks in Vietnam and Indonesia were August and October 2003, respectively [22]. However, estimation of the TMRCAs of these two lineages indicates that H5N1 virus introduction to Indonesia occurred approximately 3–4 months before the introduction to Vietnam. Furthermore, both of these estimated dates of introduction are approximately 3–6 months before the first observed H5N1 outbreaks in these countries. This may represent the time for disease development and under reporting once outbreaks

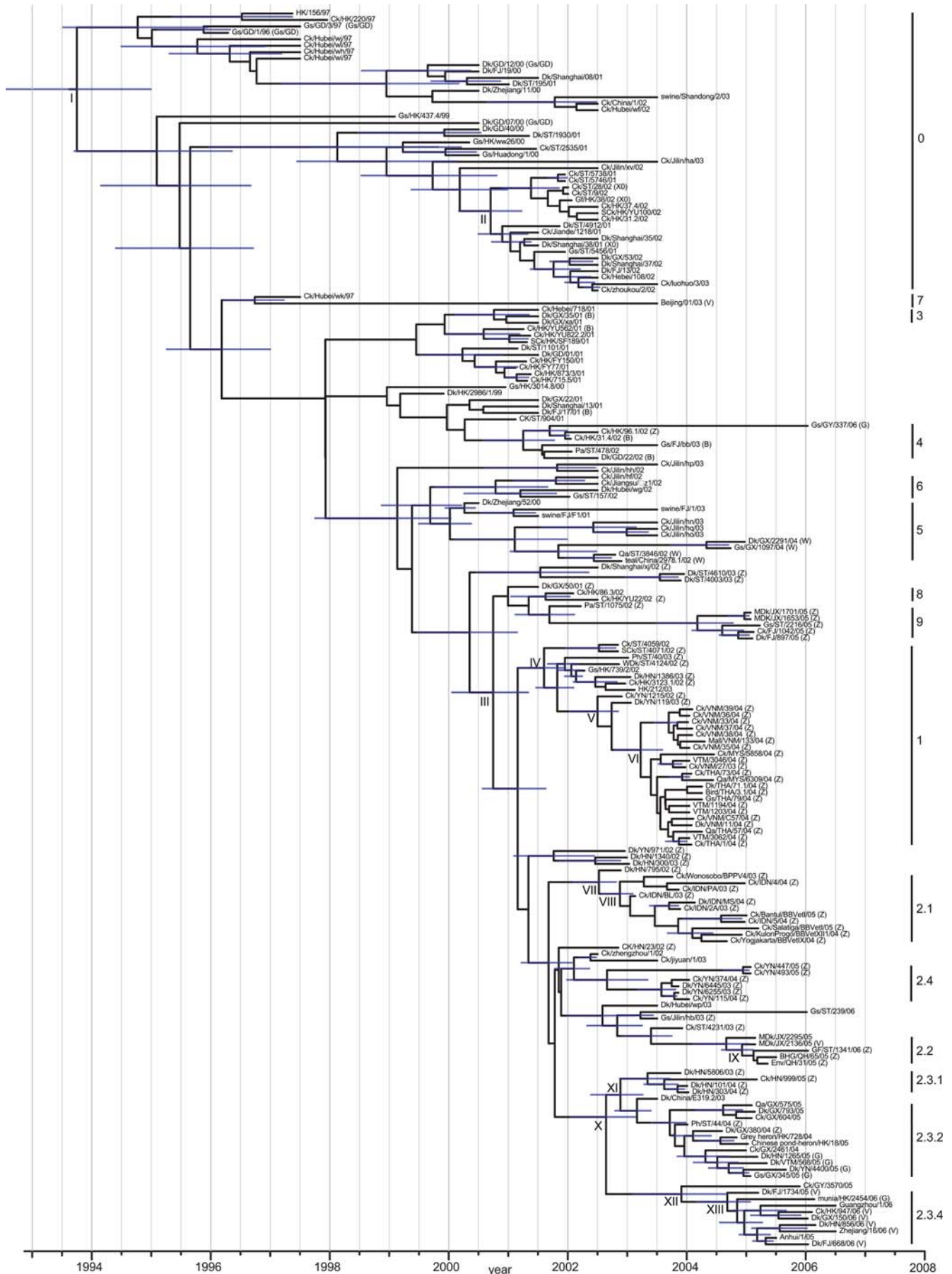


Figure 4. Dated phylogeny of the HA gene of H5N1 viruses isolated in Asia. The tree is scaled to time (horizontal axis) and was generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. TMRCA and HPDs for each of the major H5N1 lineage are given in Table 4. doi:10.1371/journal.ppat.1000161.g004

occurred. In contrast, the TMRCA of the wave 2 and 3 H5N1 viruses indicated that these HA variants arose during the same period in late 2004.

Analysis of virus population dynamics revealed a rapid increase in the genetic diversity of Gs/GD lineage in poultry in China from mid-1999 to early 2000. This corresponds with the period when each of the major Gs/GD-like H5N1 variants or sublineages diverged and subsequently became widespread in poultry throughout China [6]. It is likely that combined strong ecological and evolutionary factors led to this rapid increase in diversity, namely, the spread of the virus through large, immunologically naive poultry populations consisting of diverse species coupled with relatively high rates of nucleotide substitution and selection pressure in the HA. Interestingly, the first detection of current H5N1 reassortants occurred in early 2000, some time after the 1997 Hong Kong ‘Bird Flu’ outbreak in poultry occurred [11]. This leads us to hypothesize that these reassortant viruses, particularly genotypes B, X0, W and Z, were also generated during this period. This hypothesis is supported by our field data in which different H5N1 genotype viruses have been isolated from the same market on the same sampling occasion [24,25].

We also explored a possible mechanism to explain the population behavior of this virus, particularly the generation and maintenance of multiple H5N1 reassortants. The frequent reassortment of the polymerase complex (PB2, PB1, PA and NP) and matrix genes observed in H5N1 viruses indicates that the fitness landscape is similar to that observed for AIV in their natural gene pools, wherein little or no change in fitness is associated with frequent reassortment of functionally equivalent gene segments [5,6,13]. The presence of most of these internal genes in domestic duck before their detection in H5N1 genotypes suggests that reassortment occurred in these hosts [11].

We propose a specific mechanism to explain observed patterns of genetic drift and reassortment in H5N1. First, AIV (of different subtypes) from the natural gene pool in wild birds are introduced

into domestic duck. In domestic duck, these viruses undergo regular reassortment with endemic H5N1 viruses. Subsequently, transmission of these reassortant viruses within large highly connected populations of duck and other poultry species results in frequent interspecies transmission and genetic drift. Therefore, it is likely that this process selects for relatively fit viruses with a broad host range which are subsequently exported to other geographical regions. It is interesting to note that further reassortment has not been observed once those H5N1 viruses

Table 4. Estimated TMRCA for H5N1 HA clades

Nodes	Description	TMRCA (95% HPD)
I	GsGD	1993.75 (1992.50–1994.96)
II	X-series	2000.70 (2000.11–2001.18)
III	Clades 1, 2, 8, 9	2000.74 (2000.06–2001.36)
IV	Clade 1	2001.16 (2000.57–2001.66)
V	Vietnam, Thailand, Malaysia (VTM)+precursor	2000.74 (2000.06–2001.36)
VI	VTM	2003.22 (2002.80–2003.62)
VII	Indonesia+precursor	2002.52 (2002.09–2002.84)
VIII	Clade 2.1 (Indonesia)	2002.87 (2002.58–2003.12)
IX	Clade 2.2 (Qinghai lineage)	2004.93 (2004.59–2005.14)
X	Clade 2.3	2002.64 (2002.06–2003.16)
XI	Clades 2.3.1, 2.3.2	2002.88 (2002.39–2003.29)
XII	Clades 2.3.3, 2.3.4	2003.91 (2003.11–2004.70)
XIII	Clade 2.3.4 (Fujian-like)	2004.68 (2004.17–2005.09)

doi:10.1371/journal.ppat.1000161.t004

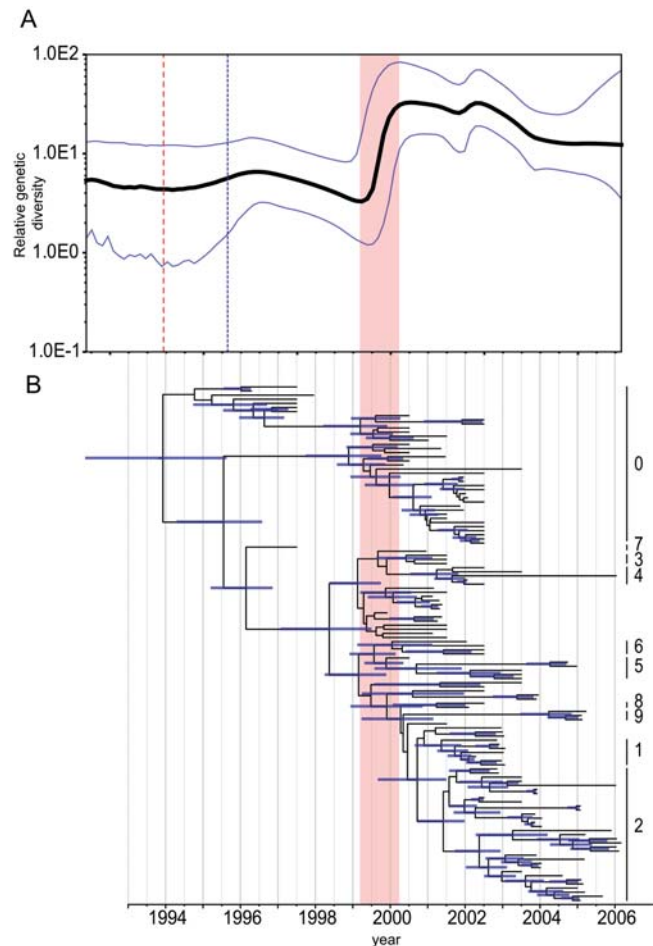


Figure 5. Population dynamics of genetic diversity of H5N1 viruses isolated from poultry in China. Bayesian skyline plot of the HA gene (A) showing changes in genetic diversity of H5N1 viruses. A measure of genetic diversity is given on the y-axis with the 95% HPD shown in blue. The red dashed line indicates the mean TMRCA of the Gs/GD lineage; the blue dashed line represents the time of the first detected H5N1 outbreak in China. HA gene tree (B) scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. Numbers to the right of the HA tree indicate H5N1 clades based on the World Health Organization nomenclature system [36]. An identical phylogenetic tree with virus names is shown in Figure S3. The red vertical bar in both panels indicates the period of divergence of major H5N1 lineages in poultry. doi:10.1371/journal.ppat.1000161.g005

were transmitted out of China. We suggest that host population structures elsewhere may not result in the same intense multi-species transmission we observe in southern China.

Methods

Virus and viral sequence data

We sequenced the full genome of 31 low pathogenic avian influenza H5 viruses plus 135 HPAI H5N1 viruses isolated from our surveillance program in southern China or otherwise available in our repository (GenBank accession numbers CY028924–CY029507, CY030878–CY031006, EF597247–EF597498). A further 114 partial gene sequences from 29 H5N1 and other subtype viruses isolated from 2000–2005 were also resequenced to provide full-length genomes. These sequences were analyzed together with all publicly available sequence data, including genes from 93 H6 and 279 H9 viruses that were also recently sequenced in our laboratories. Viral genome sequencing was conducted as previously described [2,24].

Phylogenetic reconstruction and Bayesian skyline plots

To estimate the times of divergence, a total of nine full-length datasets were analyzed: two Eurasian datasets for the H5-HA and N1-NA, and seven Asian datasets for each of the influenza gene segments, except the NA gene. To examine changes in genetic diversity during the evolution of the Gs/GD lineage, we constructed Bayesian skyline plots using a modified Asian HA dataset. This modified dataset consisted of only HA genes of viruses isolated from chicken ($n = 54$), duck ($n = 52$), goose ($n = 15$), pheasant (1) and Guinea fowl (1) in China.

To estimate the rates of nucleotide substitution and TMRCA, we used a Bayesian Markov chain Monte Carlo (MCMC) method as implemented in the program BEAST v1.4.7 [17,26]. Each gene was analyzed with the codon based SRD06 nucleotide substitution model [27]. For each analysis the Bayesian skyline coalescent model was used [16]. Three clock models were compared statistically for each dataset using a Bayes factor test in Tracer v1.4 [28,29]: the strict clock that assumes a single evolutionary rate along all branches, and the uncorrelated lognormal relaxed (UCLD) clock and uncorrelated exponential relaxed (UCED) clock that allow evolutionary rates to vary along branches within lognormal and exponential distributions, respectively [15]. A Bayes factor test of clock models showed that the UCED clock was most appropriate for datasets other than PB2 and PB1, for which the UCLD clock most appropriately described the data. For each dataset, three to five independent Bayesian MCMC runs were conducted for 10–20 million generations sampled to produce 10,000 trees. Convergence of the runs was confirmed using Tracer v1.4 [29] and effective sample size values of >500 indicated a sufficient level of sampling. The results of the multiple runs were then combined using LogCombiner v1.4.7 [17]. Mean evolutionary rates and divergence times were calculated using TreeAnnotator v1.4.7 and TreeStat v1.1 after the removal of an appropriate burnin, 10–20% in most cases, and phylogenetic trees were visualized with FigTree v1.1.2 [17,30,31].

Finally, to evaluate if the TMRCA of each of the gene segments of a given genotype were significantly different or not, the TMRCA of each gene segment was compared to the remaining genes of the genotype using a Bayes factor test [28]. This test was calculated in as follows; given a genotype, the

probability of any gene (e.g. PB2) being older than any other segment (e.g. PB1) divided by the probability of PB1 being older than PB2 given the data (tree estimates of TMRCA) multiplied by the inverse estimation for the priors (PB1 being older than PB2 divided by PB2 being older than PB1 of the priors) was calculated for each Bayesian MCMC run [19].

Detection of selection pressure

To determine selection pressures on the HA of Gs/GD-like H5N1 viruses in poultry the modified Asian H5-HA dataset was analyzed using the single-likelihood ancestor counting (SLAC) [32] and genetic algorithm (GA) methods [33] available in DataMonkey [34] and HYPHY [35]. The SLAC method calculates global and site-specific nonsynonymous (d_N) and synonymous (d_S) nucleotide substitution rate ratios ($\omega = d_N/d_S$) based on the BEAST generated phylogenetic tree and the best-fit nucleotide substitution model. The GA method assigns four ω classes to each lineage in search of the model of lineage-specific evolution that best fits the data [33]. The probability ($\geq 95\%$) of ω being >1 along a specific lineage was calculated from the averaged model probability of all models rather than by inference from the single best-fitting model [33]. This approach does not require any *a priori* hypothesis of lineage-specific evolution.

Supporting Information

Figure S1 The HA gene (A) and N1 gene (B) trees scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. Red branches indicate Gs/GD lineage H5N1 viruses.

Found at: doi:10.1371/journal.ppat.1000161.s001 (1.33 MB PDF)

Figure S2 The PB2 (A), PB1 (B), PA (C), NP (D), M (E) and NS (F) gene trees scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA. Red branches indicate Gs/GD lineage H5N1 viruses.

Found at: doi:10.1371/journal.ppat.1000161.s002 (1.65 MB PDF)

Figure S3 HA gene tree of H5N1 viruses isolated from poultry in China, scaled to time (horizontal axis) generated using the SRD06 codon model and uncorrelated relaxed clock model. Nodes correspond to mean TMRCA and blue horizontal bars at nodes represent the 95% HPDs of TMRCA.

Found at: doi:10.1371/journal.ppat.1000161.s003 (0.43 MB PDF)

Table S1

Found at: doi:10.1371/journal.ppat.1000161.s004 (0.07 MB PDF)

Acknowledgments

We thank CL Cheung, Huang Kai and Wang Jia, State Key Laboratory of Emerging Infectious Diseases, for collection and processing of samples.

Author Contributions

Conceived and designed the experiments: DV GJS YG. Performed the experiments: DV JB LD JXZ. Analyzed the data: DV JB LD YG. Wrote the paper: DV JB SR HC JSMP GJS YG.

References

- Xu X, Subbarao K, Cox NJ, Guo Y (1999) Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its

hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology* 261: 15–19.

2. Smith GJD, Fan XH, Wang J, Li KS, Qin K, et al. (2006) Emergence and predominance of an H5N1 influenza variant in China. *Proc Natl Acad Sci U S A* 103: 16936–16941.
3. Thiry E, Zicola D, Addie H, Egberink K, Hartmann H, et al. (2007) Highly pathogenic avian influenza in cats and other carnivores. *Vet Microbiol* 122: 25–31.
4. World Health Organization (2008) Cumulative number of confirmed human cases of Avian Influenza A (H5N1) reported to WHO (WHO, Geneva). Available: http://www.who.int/csr/disease/avian_influenza/country/en/. Accessed 11 August 2008.
5. Li KS, Guan Y, Wang J, Smith GJD, Xu KM, et al. (2004) Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* 430: 209–213.
6. Chen H, Smith GJD, Li KS, Wang J, Fan XH, et al. (2006) Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *Proc Natl Acad Sci U S A* 103: 2845–2850.
7. Chen H, Smith GJD, Zhang SY, Qin K, Wang J, et al. (2005) H5N1 virus outbreak in migratory waterfowl. *Nature* 436: 191–192.
8. Bragstad K, Jørgensen PH, Handberg K, Hammer AS, Kabell S, et al. (2007) First introduction of highly pathogenic H5N1 avian influenza A viruses in wild and domestic birds in Denmark, Northern Europe. *Virology* 350: 258–268.
9. Ducatez MF, Olinger CM, Owoade AA, DeLandsheer S, Ammerlaan W, et al. (2006) Introductions of H5N1 in Nigeria. *Nature* 442: 37.
10. Nguyen TD, Nguyen TV, Vijaykrishna D, Guan Y, Peiris JSM, et al. (2008) Multiple lineages of influenza A (H5N1) viruses in Vietnam (2006–2007). *Emerg Infect Dis* 14: 632–636.
11. Guan Y, Peiris JSM, Kong KF, Dyrting KC, Ellis TM, et al. (2002) H5N1 influenza viruses isolated from geese in Southeastern China: evidence for genetic reassortment and interspecies transmission to ducks. *Virology* 292: 16–23.
12. Duan L, Campitelli L, Fan XH, Leung YHC, Vijaykrishna D, et al. (2007) Characterization of low pathogenic H5 subtype influenza viruses from Eurasia: Implications for the origin of highly pathogenic H5N1 viruses. *J Virol* 81: 7529–7539.
13. Dugan VG, Chen R, Spiro DJ, Sengamalay N, Zaborsky J, et al. (2008) The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog* 4: e1000076. doi:10.1371/journal.ppat.1000076.
14. Chen R, Holmes EC (2006) Avian Influenza exhibits rapid evolutionary dynamics. *Mol Biol Evol* 23: 2336–2341.
15. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88. doi:10.1371/journal.pbio.0040088.
16. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185–1192.
17. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
18. Food and Agriculture Organization (2008) Avian influenza disease emergency news. Available: <http://www.fao.org/avianflu/en/AIDNews.html>. Accessed 11 August 2008.
19. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, et al. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453: 615–619.
20. Fitch WM, Bush RM, Bender CA, Cox NJ (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc Natl Acad Sci U S A* 94: 7712–7718.
21. Lindstrom S, Endo A, Sugita S, Pecoraro M, Hiromoto Y, et al. (1998) Phylogenetic analyses of the matrix and non-structural genes of equine influenza viruses. *Arch Virol* 143: 1585–1598.
22. Smith GJD, Naipospos TSP, Nguyen TD, de Jong MD, Vijaykrishna D, et al. (2006) Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology* 350: 258–268.
23. Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, et al. (2006) Large-scale sequence analysis of avian influenza isolates. *Science* 311: 1576–1580.
24. Guan Y, Peiris JSM, Lipatov AS, Ellis TM, Dyrting KC, et al. (2002) Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc Natl Acad Sci U S A* 99: 8950–8955.
25. Guan Y, Poon LLM, Cheung CY, Ellis TM, Lim W, et al. (2004) H5N1 influenza: a protean pandemic threat. *Proc Natl Acad Sci U S A* 101: 8156–8161.
26. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307–1320.
27. Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* 23: 7–9.
28. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18: 1001–1013.
29. Rambaut A, Drummond AJ (2007) Tracer v1.4: MCMC trace analyses tool. Available: <http://beast.bio.ed.ac.uk/Tracer>. Accessed 20 June 2008.
30. Rambaut A, Drummond AJ (2007) TreeStat v1.1: Tree statistic calculation tool. Available: <http://tree.bio.ed.ac.uk/software/treestat/>. Accessed 20 June 2008.
31. Rambaut A (2008) FigTree v1.1.1: Tree figure drawing tool. Available: <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 20 June 2008.
32. Kosakovsky Pond SL, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22: 1208–1222.
33. Kosakovsky Pond SL, Frost SDW (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol* 22: 478–485.
34. Kosakovsky Pond SL, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21: 2531–2533.
35. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
36. WHO/OIE/FAO H5N1 Evolution Working Group (2008) Towards a unified nomenclature system for the highly pathogenic avian influenza virus (H5N1). *Emerg Infect Dis* 14: e1. doi:10.3201/eid1407.071681.