

# Spatial, temporal and genetic dynamics of highly pathogenic avian influenza A (H5N1) virus in China

Tian *et al.*

RESEARCH ARTICLE

Open Access

# Spatial, temporal and genetic dynamics of highly pathogenic avian influenza A (H5N1) virus in China

Huaiyu Tian<sup>1†</sup>, Yujun Cui<sup>2†</sup>, Lu Dong<sup>3†</sup>, Sen Zhou<sup>4</sup>, Xiaowen Li<sup>1,4</sup>, Shanqian Huang<sup>1</sup>, Ruifu Yang<sup>2</sup> and Bing Xu<sup>1,4,5\*</sup>

## Abstract

**Background:** The spatial spread of H5N1 avian influenza, significant ongoing mutations, and long-term persistence of the virus in some geographic regions has had an enormous impact on the poultry industry and presents a serious threat to human health.

**Methods:** We applied phylogenetic analysis, geospatial techniques, and time series models to investigate the spatiotemporal pattern of H5N1 outbreaks in China and the effect of vaccination on virus evolution.

**Results:** Results showed obvious spatial and temporal clusters of H5N1 outbreaks on different scales, which may have been associated with poultry and wild-bird transmission modes of H5N1 viruses. Lead-lag relationships were found among poultry and wild-bird outbreaks and human cases. Human cases were preceded by poultry outbreaks, and wild-bird outbreaks were led by human cases. Each clade has gained its own unique spatiotemporal and genetic dominance. Genetic diversity of the H5N1 virus decreased significantly between 1996 and 2011; presumably under strong selective pressure of vaccination. Mean evolutionary rates of H5N1 virus increased after vaccination was adopted in China. A clear signature of positively selected sites in the clade 2.3.2 virus was discovered and this may have resulted in the emergence of clade 2.3.2.1.

**Conclusions:** Our study revealed two different transmission modes of H5N1 viruses in China, and indicated a significant role of poultry in virus dissemination. Furthermore, selective pressure posed by vaccination was found in virus evolution in the country.

## Background

A particular subtype of influenza A virus, highly pathogenic avian influenza (HPAI) virus H5N1, is transmitted by contact with infected birds [1]. It is epizootic in many bird populations, especially in Southeast Asia. Clade 2.2 of the virus has spread globally, including Europe, the Middle East and Africa after first appearing in Asia in 2005 [2]. The spatial spread of H5N1 avian influenza and long-term persistence of the virus in some regions has had an enormous impact on the poultry industry and presents a serious threat to the health of humans and migratory birds [3,4]. It has been 17 years since the

first case in geese of H5N1 avian influenza was discovered in Hong Kong in 1996 [5]. As of 19 May 2013, H5N1 has caused 628 human cases of influenza in 15 different countries, with 374 deaths.

The high lethality and virulence of H5N1, its epizootic presence, its increasingly large host reservoir, its significant ongoing mutations, and its potential transmissibility between humans, make it one of the greatest current pandemic threats [6]. Substantial progress has been made in researching various aspects of the virus and preparing for a potential influenza pandemic [7,8]. Several studies on the global spread of avian flu using phylogenetic relationships of virus isolates have indicated that migratory bird movements, and trade in poultry and wild birds could explain the pathway for introduction events into different countries [9,10]. However, the underlying mechanism of the long-term persistence of the virus and its various

\* Correspondence: bingxu@tsinghua.edu.cn

<sup>†</sup>Equal contributors

<sup>1</sup>State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

<sup>4</sup>Ministry of Education Key Laboratory for Earth System Modelling, Center for Earth System Science, Tsinghua University, Beijing 100084, China

Full list of author information is available at the end of the article

spatiotemporal transmission pathways, with their corresponding genetic footprints, remain poorly understood.

China is one of the world primary producers of poultry products [11], and is among the regions most affected by H5N1 [12]. Poultry production generates 16 million tons of meat and 27 million tons of eggs annually in China [11]. In 2004, the total loss caused by HPAI virus H5N1 was 4.5 billion US dollars in China. To control H5N1 infection in poultry, many countries have implemented a vaccination policy, including China, Vietnam, Indonesia and Egypt. In Mainland China, a poultry vaccine was first used at the end of 2004. Over 55 billion doses of vaccines were applied to control the outbreaks between 2004 and 2008 [13]. Antigenic variants of H5N1 avian influenza virus have occurred along with its spatiotemporal transmission. Furthermore, vaccination may change the evolutionary dynamics of H5N1 virus [14,15]. Vaccine strains can be selected from the seeding region of Southeast Asia where its genetic and antigenic characteristics can be determined earlier for human or avian influenza [2,16]. However, the effect of vaccination in China has rarely been studied and is discussed in the present study.

Here, we applied phylogenetic analysis, geospatial techniques, and time series models to investigate the spatiotemporal pattern of H5N1 outbreaks in China and the effect of vaccination on the dynamics of virus evolution. These data, combined with spatiotemporal information on the H5N1 outbreaks and viruses, were used to understand viral evolutionary dynamics from the beginning to its circulation in China. This study aimed to illustrate the spatiotemporal pattern of H5N1 outbreaks in China, understand the role of migratory birds and poultry in contributing such a pattern, and the effect of vaccination on the dynamics of virus evolution.

## Methods

### Virus sequences and outbreak data

Full-length hemagglutinin (HA) sequences of HPAI virus H5N1 were obtained from the GenBank database (<http://www.flu.lanl.gov>) hosted by the National Center for Biotechnology Information (NCBI). A total of 674 non-repetitive sequences were used in the phylogenetic and selection analysis at provincial level, from 1996 to 2011 (Additional file 1: Table S1). Hosts were classified into different groups, according to the list of species affected by H5N1 avian influenza from the US Geological Survey National Wildlife Health Center. We also collected H5N1 outbreak data in China from the World Organisation for Animal Health (OIE, <http://www.oie.int/>) and the United Nations Food and Agriculture Organization (FAO, <http://www.fao.org/home/en/>), from January 2004 to December 2011. We removed redundant records and geocoded them according to their spatial information. Mainland China had 323 outbreaks in total, of which 217 were in poultry,

85 in wild birds, and 21 in humans (Figure 1). The HA sequences of HPAI H5N1 are accurate to the year, while outbreak records are accurate to the day. The HA sequences and outbreak records were not linked, and they were collected from two separate data sets.

### Spatial point pattern analysis

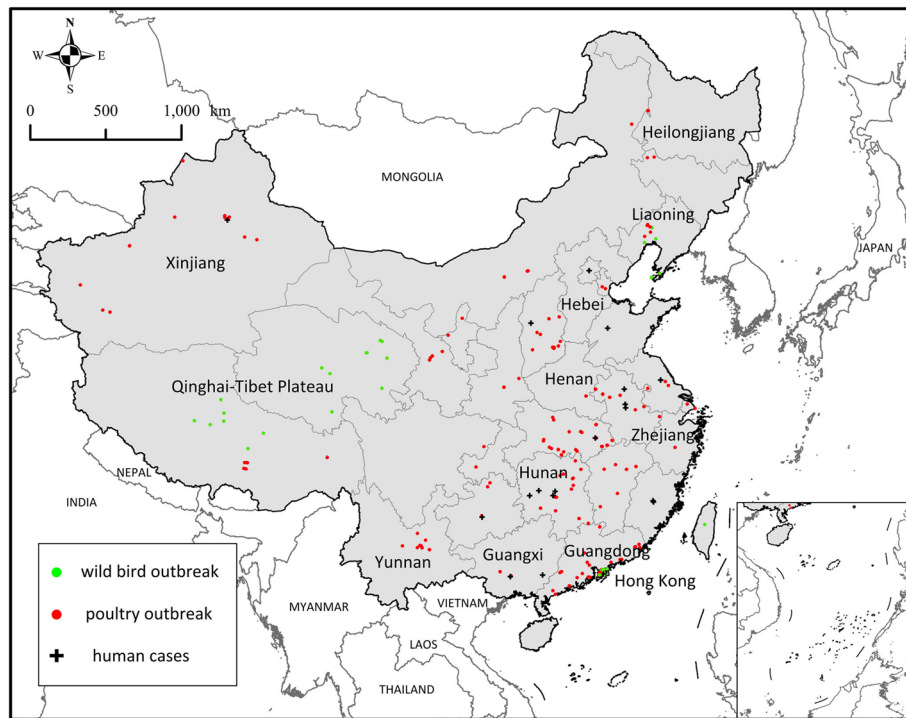
A point pattern analysis of H5N1 outbreaks was used to determine where the cases were spread, as well as to determine which spatial scales were optimal for disease clustering [10,17]. Choosing the right scale was critical for subsequent analyses. We used exploratory spatial statistical techniques to examine the patterns of outbreaks. Ripley's K function describes how the expected value of a point process changes over different spatial and temporal lags [18,19]. A peak value in Ripley's K function indicates a clustering at the scale of the corresponding lag. An estimate of spatial and temporal K function can be calculated by [20]:

$$K(d) = \frac{R}{n^2} \sum_{i=1}^n \sum_{j \neq i} \frac{I_d(i, j)}{W_{ij}} \quad (1)$$

$$K(t) = \frac{T}{n^2} \sum_{i=1}^n \sum_{j \neq i} I_t(i, j) \quad (2)$$

where  $R$  is the total area of the study,  $n$  is the number of observed events,  $I_d(i, j)$  is an indicator function that takes a value of 1 when the spherical distance between points  $i$  and  $j$  is less than  $d$ .  $W_{ij}$  is the adjustment factor of the edge effect. In clustering,  $K(d)$  would be greater than  $A(d)$ , which is the area of the spherical circle of arc-radius  $d$ ; and less than  $A(d)$  under regularity. We then apply a transformation to  $K(d)$  to have  $L(d) = R * \arccos\left(1 - \frac{K(d)}{2\pi R^2}\right) - d$ , and to plot  $L(d)$  against  $d$ . Deviation,  $L(d)$ , above the zero line suggest clustering (more events within distance  $d$  than expected), while deviations below zero suggest regularity (fewer events within distance  $d$  than expected) [21].  $T$  is the time span from the earliest to latest outbreak in the dataset; we changed Gregorian date to Julian date.  $I_t(i, j)$  is an indicator function that takes a value of 1 when the time interval between points  $i$  and  $j$  is less than  $t$ . All the calculations used the same study area – the extent of Mainland China.

The upper and lower bounds were determined by undertaking Monte Carlo simulations 999 times. For each simulation, we generated the same number of random points as the cases, and then calculated their K functions. To each lag, the upper and lower bounds were the minimum and maximum K values among the set of 999 simulations.



**Figure 1 H5N1 virus outbreaks, avian and human cases.** Outbreak records in China began in January 2004, and outbreaks were mainly concentrated in four periods: early 2004, late 2005 to early 2006, late 2007 to early 2008, and early 2009.

### Phylogenetic analysis

Sequences were aligned using the ClustalW algorithm [22] implemented in BioEdit (version 7.0) [23]. The optimal nucleotide substitution model was selected using the Akaike Information Criterion (AIC) [24] and a hierarchical likelihood ratio test in ModelTest [25]. Phylogenetic relationships of the 674 H5N1 HA sequences were constructed using the neighbor-joining approach and the GTR + I +  $\Gamma_4$  model of nucleotide substitution by PAUP 4.0 [26]. Tree topology reliability was tested with 1000 bootstrap replicates. The tree was rooted by A/goose/Guangdong/1/96 and structured according to the World Health Organization system of avian influenza cladistics ([http://www.who.int/influenza/gisrs\\_laboratory/h5n1\\_nomenclature/en/](http://www.who.int/influenza/gisrs_laboratory/h5n1_nomenclature/en/)).

Rates of nucleotide substitution per site and year were estimated by the BEAST program [27], which uses a Bayesian Markov chain Monte Carlo approach. For each analysis the constant size, exponential growth, Bayesian skyline and Bayesian skyline coalescent prior were used with the codon-based SRD06 nucleotide substitution model [28]. The strict and uncorrelated lognormal (UCLN) relaxed molecular clocks were compared by analyzing values of the coefficient of variation in the tracer. The statistical uncertainty in all analyses was reflected by the 95% highest probability density (HPD) values for parameter estimate. Additionally, the chain lengths in each analysis were run

for sufficient time to achieve coverage (up to 10 million generations). To estimate the variations of evolutionary rates resulting from vaccination, the HA sequences were categorized by sampling time. We inferred the evolutionary rates in a window of 2 years with a step of 1 year forward.

To visualize the evolutionary process of the virus, we calculated the genetic distances between any pair of sequences and reprojected the distances to a genetic map using multi-dimensional scaling (MDS) [29,30]. MDS created a matrix of inter-point distances. These points could be reconstructed using Euclidean distance to replace genetic distance. MDS minimized the loss function as follows:

$$L = \min \sum_{i,j=1}^{i,j=M,N} (D_{ij} - \delta_{ij})^2, \quad (3)$$

where  $D_{ij}$  is the genetic distance from strain  $i$  to  $j$ , and  $\delta_{ij}$  is the estimated Euclidean distance. The MDS method was conducted in Matlab (version R2009b) (MathWorks, Natick, MA, USA).

### Detection of positively-selected sites

The selective pressure can be detected by measuring nonsynonymous-synonymous rate ratio (dN/dS) [31]. A ratio >1 overall represents positive selection of genes [32]. To illustrate the variation of raw mutation rate over sites, we compared the virus with the HA genes of reassortant

avian influenza virus vaccine Re-1, Re-4 and Re-5 by the LPB93 method [33], with introduction of a  $\gamma$  distribution using the KaKs Calculator (version 2.0) [34]. We aligned the pair of sequences in a window of 57 bp with a step of 3 bp forward. Fisher's exact test was applied for justification of small sample size. The dN/dS ratio among codons was calculated using the maximum likelihood method with the codeml program in PAML 4.5 [35,36], as well as the fixed effect likelihood (FEL) method implemented in HyPhy (version 0.99b) [37]. Positive selection was tested using a likelihood ratio test (LRT) comparing a null model (M7) that did not allow dN/dS >1 with an alternative model (M8) that did. M8 had two more parameters than M7, so that  $\chi^2_2$  was used to conduct the LRT [38].

### Time series analysis

A cross-correlation function was used for identification of the relationship between outbreaks in different hosts. We obtained cross-correlation coefficient time series based on poultry and wild-bird outbreaks and human cases. The cross-correlation coefficient time series was calculated as:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad k = 0, \pm 1, \pm 2, \quad (4)$$

where  $\rho_{xy}(k)$  is the cross-correlation coefficient at lag  $k$ ;  $\sigma_x$  and  $\sigma_y$  are standard deviations of X and Y, respectively; and  $\gamma$  is the covariance function.

The time series analyzed here was characterized by strong autocorrelation. Thus, an autoregressive integrated moving average (ARIMA) model [39] was applied to account for this autocorrelation. An ARIMA model is notated as  $(p, d, q)$ , where  $p$  is the autoregressive (AR) order,  $d$  the differencing order, and  $q$  the moving average (MA) order. The autocorrelation function and partial autocorrelation function were applied to determine the AR and MA order. Models and coefficients were examined through calculated AIC and the coefficient of determination ( $R^2$ ), and the optimal model in each relationship was selected: poultry outbreaks lead human cases; human cases lead poultry outbreaks, wild-bird outbreaks led human cases; human cases lead wild-bird outbreaks; wild-bird outbreaks led poultry outbreaks; and poultry outbreaks lead wild-bird outbreaks. All ARIMA modeling and corresponding statistical tests were performed using R (version 2.13.2).

## Results

### Spatial and temporal pattern of H5N1 outbreak

Figure 2a shows values of the spatial K function. There were three peaks located at approximately 50, 700 and 1200 km for the entire dataset. The poultry outbreaks exhibited the same three peaks as the overall trend (Figure 2b), whereas wild-bird outbreaks showed a single peak at 50 km

(Figure 2c). The 50-km lag corresponded to the average distance between neighboring counties, within a city or at a specific locale. The 700- and 1200-km lags may fit the distance between provinces. Outbreaks in wild birds were concentrated in the Qinghai-Tibetan Plateau, Hong Kong and Liaoning Province. We suspect that the lag for wild birds was related to their migration distances; the 50-km lag corresponded to an outbreak within a city, or at breeding, wintering and stopover sites.

Figure 2d shows values of the temporal K function. There were two peaks located at approximately 20 and 115 days. Poultry outbreaks exhibited a single peak at 20 days (Figure 2e), while wild-bird outbreaks had a single peak at 115 days (Figure 2f). The 20-day peak may correspond to poultry trade within a province or between provinces. The 115-day peak may fit the migration cycle of migratory birds.

### Correlations between poultry and wild-bird outbreaks and human cases

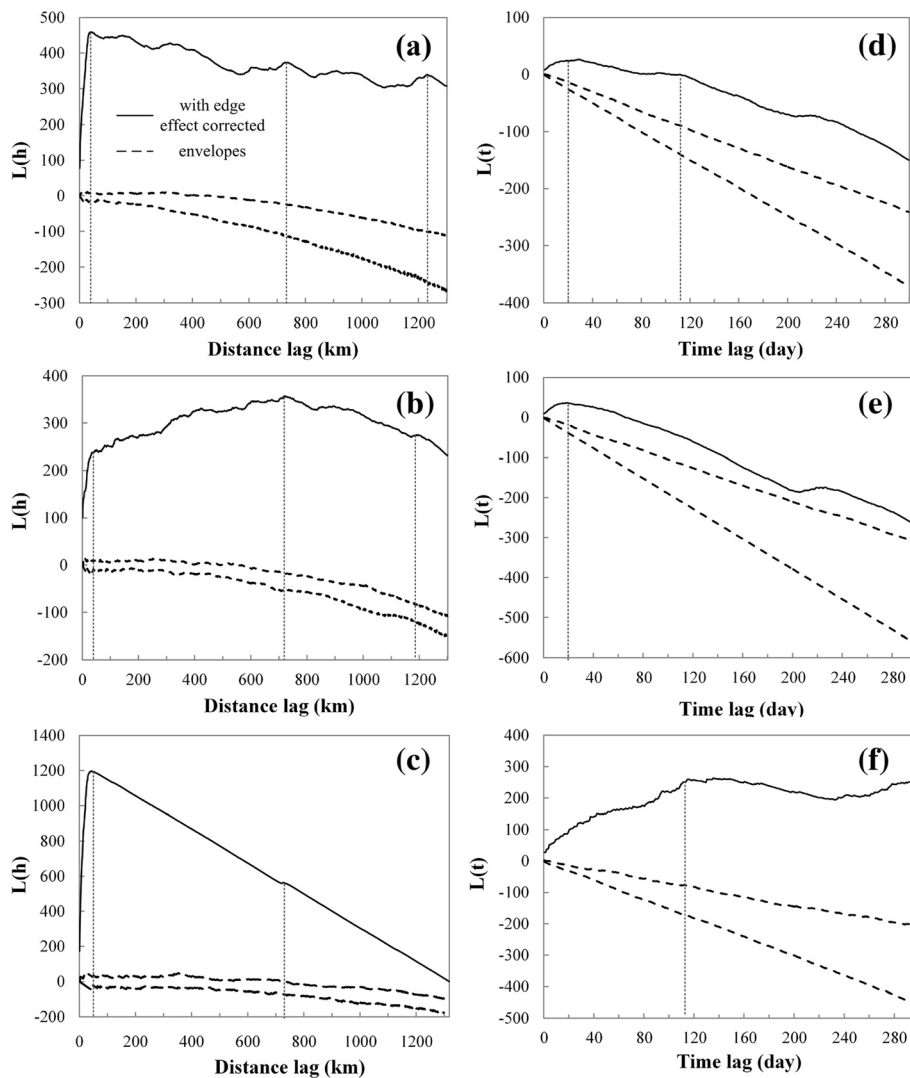
Outbreak data began in 2004, and there were epidemic waves during four periods in China: January–February 2004, October 2005–June 2006, January–June 2005, and December 2008–February 2009. The outbreaks revealed significant seasonal patterns. Wild-bird outbreaks tended to concentrate in January, February, April and May; poultry outbreaks in January, February, June and November; and human cases in January (Additional file 2: Figure S1).

An ARIMA model was obtained for each of the six relationships (Figure 3), and a total of 384 models ( $4 \times 4 \times 4 \times 6$ ,  $p: 0-4$ ,  $d: 0-4$ ,  $q: 0-4$ ) were tested (Additional file 3: Table S2). The results indicated that human cases were preceded by poultry outbreak with 1–4 months (Figure 3a), and wild-bird outbreaks were led by human cases with 1–3 months (Figure 3d). These lead–lag relationships were considered to be stable, with approximately 12 and 11 months periodicity, respectively.

### Circulation of H5N1 virus in China, 1996–2011

Virus strains in China were classified into different clades according to the supplement from WHO (Figures 4a and Additional file 4: FigureS2). A distribution map of different virus clades based on spatiotemporal information of isolates was drawn (Figure 4b). Clade 0 virus was the circulating strain at the early stage; after 2004, it was replaced and several new clades emerged and began to circulate, such as 2.3.2, 2.3.4, and 7. The virus began to spread westward and has been detected in almost all of China since then.

Clade 0 virus circulated in poultry south of the Yangtze River before 2000. The Yangtze River has been historically considered as a geographic middle line dividing the country into north and south. With time, the virus spread north and crossed the Yangtze River in 2001. After 2004,



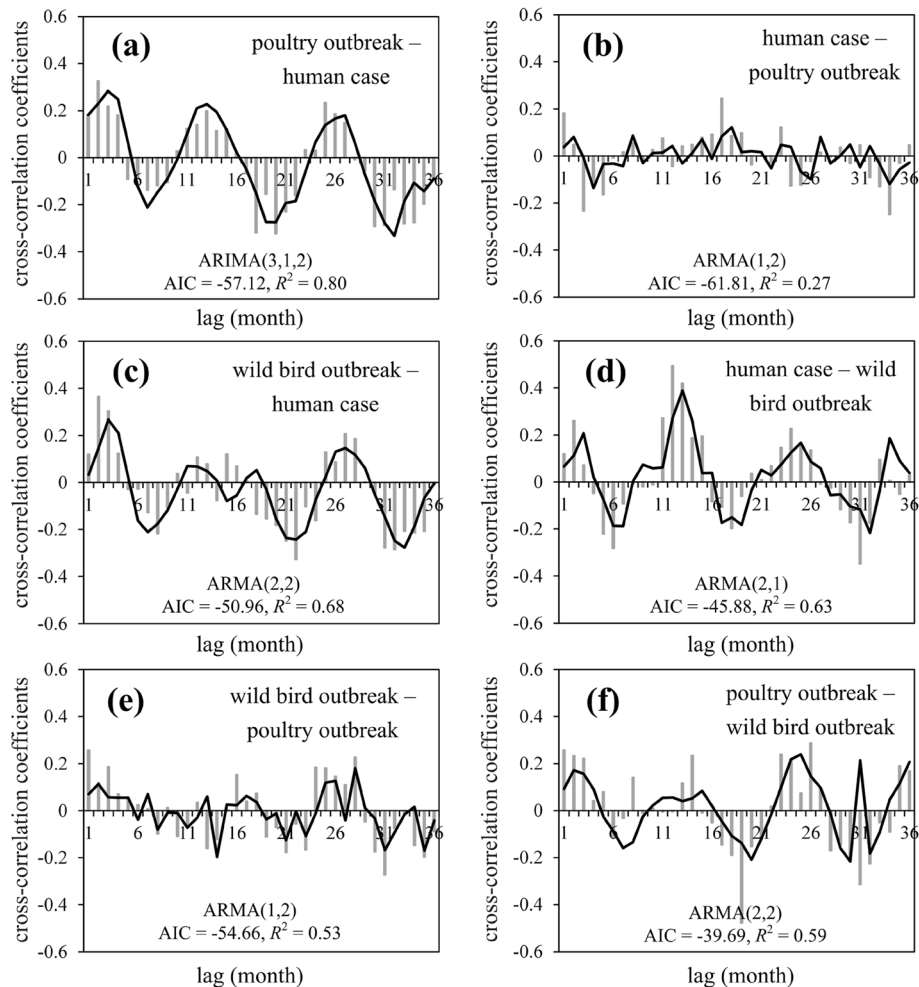
**Figure 2 Temporal and spatial K-function results.** Spatial K-function calculated for: (a) all infections, (b) poultry and (c) wild birds. Temporal K-function calculated for: (d) all infections, (e) poultry and (f) wild birds. Clusters are shown by thin dashed lines. The peak value in spatial K-function for wild birds represents the outbreaks in Hong Kong, because of data bias.

there was no clade 0 virus isolated from birds, indicating that it may have been replaced by other clades (Figure 5a). In 2004, clade 2.3.2 emerged in Southern China and clade 7 in Northern China. These may have taken over the dominant role of clade 0 in those parts of the country.

Clade 2.3.2 virus appeared in 2004 and circulated south of the Yangtze River, except for one wild-bird isolate found in Qinghai in 2009 [40]. After 2006, a variant of this virus, clade 2.3.2.1, began circulating in China and Vietnam, and the regularly-used vaccines did not provide full protection (Figure 5b). Clade 2.3.4 emerged in 2005 and spread from Southern to Northern China. It reached Henan Province, the geographic centroid of the country, in 2006, and appeared in wild birds in Heilongjiang, the northernmost province, in 2007 (Figure 5c).

Clade 7 virus appeared in wild birds in Henan Province in 2004. In the following year the virus spread south, reaching Hunan and Yunnan Provinces; however, it disappeared in these areas in 2006. The virus also spread north, reaching Xinjiang and Hebei Provinces, and became the primary circulating strain in poultry in Northern China (Figure 5d).

After an outbreak in migratory waterfowl at Qinghai Lake in 2005 [41], clade 2.2 virus persisted in the Qinghai–Tibet Plateau region during 2005–2007, and began to decrease in 2008. Surprisingly, clade 2.2 was seldom isolated from poultry in China, and it was not the strain circulating in birds throughout China, although it had a large and long-lasting global impact on various host species.



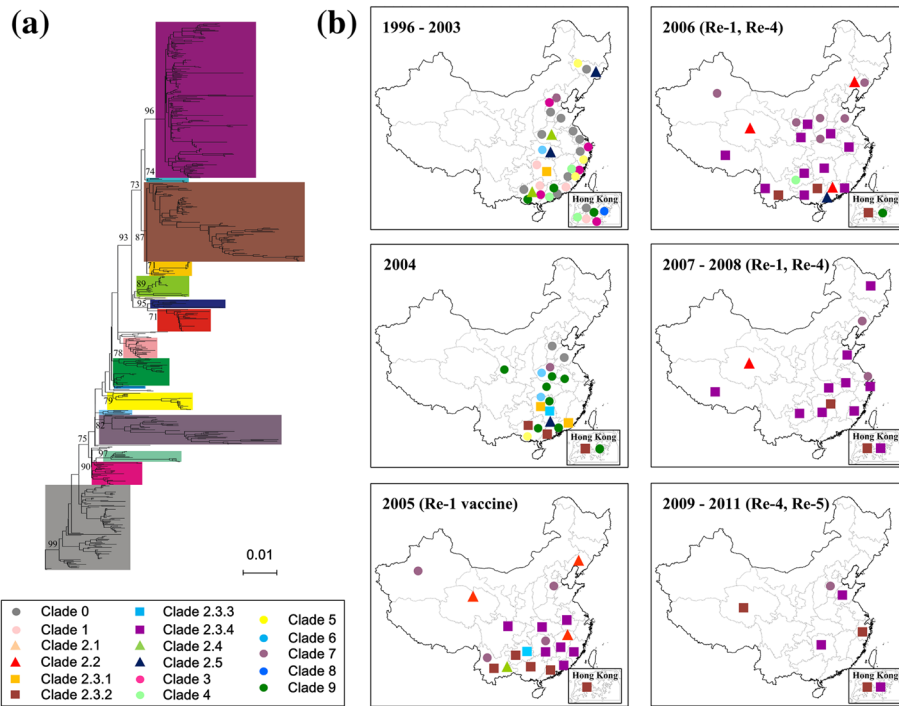
**Figure 3** Cross-correlation coefficient time series and ARIMA model-fitted curves. (a) Poultry outbreak – human case; (b) human case – poultry outbreak; (c) wild-bird outbreak – human case; (d) human case – wild-bird outbreak; (e) wild-bird outbreak – poultry outbreak; and (f) poultry outbreak – wild-bird outbreak. Thick lines are best-fit results of cross-correlation coefficient time series by ARIMA modeling.

### Effect of vaccination on evolution of H5N1

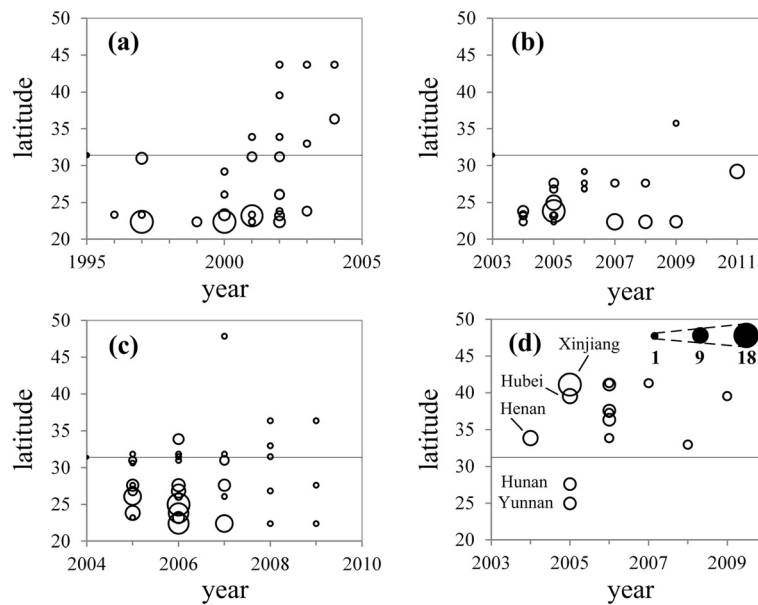
Three H5N1 avian influenza vaccines, Re-1, Re-4 and Re-5, were widely adopted in China (Table 1). The Re-1 vaccine was first used at the end of 2004, and for almost 10 years after clade 0 appeared. The Re-4 vaccine was first used in August 2006, 2 years after clade 7 appeared. The Re-5 vaccine was first used in May 2008, 4 years after clade 2.3.4 emerged. After compulsory vaccination began in 2004, several specific sub-clades appeared, such as 2.3.2 and 2.3.4. The mean evolutionary rates of H5N1 virus increased after vaccination was implemented in the country. The mean nucleotide substitution rate for all H5N1 viruses collected from 1996 to 2004 was  $3.77 \times 10^{-3}$  substitutions/site/year (95% HPD from  $3.33 \times 10^{-3}$  to  $4.22 \times 10^{-3}$  substitutions/site/year), and  $4.39 \times 10^{-3}$  substitutions/site/year (95% HPD from  $4.03 \times 10^{-3}$  to  $4.76 \times 10^{-3}$  substitutions/site/year) from 2005 to 2012.

The results were consistent with those from other countries where vaccination was adopted [15]. We also conducted a sliding separation window analysis to examine the variation of evolutionary rates over time. The results indicated that evolutionary rates of H5N1 virus increased in the year when the vaccine was used in China (Additional file 5: Figure S3, Table 2).

To examine further the effect of vaccination on the genomic dynamics of the virus, we focused on clade 2.3.2 that circulated in Southern China, and a total of 97 strains were used for analysis (Additional file 6: Figure S4). The strain appeared in Guangdong and Guangxi Provinces and Hong Kong in 2004. Then it spread to several neighboring provinces of Southern China (Figure 6a). Based on the unrooted tree of clade 2.3.2 in the country, it was clear that the virus was divided into two groups: classical clade 2.3.2 isolated in 2003–2006, and its



**Figure 4** Spatiotemporal distribution of different virus clades isolated from birds in China, 1996–2011. **(a)** Phylogenetic tree based on H5N1 HA gene sequences. **(b)** Distribution of H5N1 virus isolated from birds in China, 1996–2011. The definition criteria of clades were developed by the WHO/OIE/FAO H5N1 Evolution Working Group: sharing of a common (clade-defining) node in the phylogenetic tree; monophyletic grouping with a bootstrap value  $\geq 60$  at the clade-defining node; and average percentage pairwise nucleotide distances between and within clades of  $>1.5\%$  and  $<1.5\%$ , respectively.



**Figure 5** Distribution of prevailing H5N1 virus clades in China. **(a)** Clade 0; **(b)** clade 2.3.2; **(c)** clade 2.3.4; **(d)** clade 7. Thin black line represents latitude of Yangtze River. Circle size is proportional to sample size, with smallest circle representing one sample, and largest 18 samples.



**Table 1 Major vaccine types used for control of H5N1 avian influenza in China from 2004 to 2011**

Vaccine	HA gene donor virus	First use
Re-1	A/goose/Guangdong/1/1996 (Clade 0)	End of 2004
Re-4	A/chicken/Shanxi/2/2006 (Clade 7)	Aug. 2006
Re-5	A/duck/Anhui/1/2006 (Clade 2.3.4)	May 2008

descendent, clade 2.3.2.1, isolated in 2007–2011. Various host species including poultry, wild birds and humans were infected with these two sub-clades of virus (Figure 6b).

The genetic map (Figure 6c) revealed that there were mainly two distinct clusters with one cluster dominated by poultry stains (81.6% among a total of 49) between 2003 and 2006, and the other cluster dominated by wild-bird strains (40.7% among a total of 27) between 2007 and 2008. The two clusters displayed in the genetic map represented the two groups of sub-clades constructed from the rooted phylogenetic tree. They clearly indicated that the virus might have been transmitted from domestic poultry to wild birds during 2006–2007. Although wild-bird strains have been constantly isolated in Hong Kong and Qinghai between 2007 and 2008, the virus transmission started to return to poultry in 2008 and even the new host of humans in 2009. In 2009–2011, clade 2.3.2.1 developed in two evolutionary directions. The first was in Hunan, Qinghai and Hubei Provinces, whereas the second was in Hong Kong and Zhejiang Province.

The major population divergence of clade 2.3.2 virus was closely related to the beginning of vaccine use in China (Figure 6b, Table 1). The first instance of divergence occurred in 2004, corresponding to the use of Re-1 vaccine; the second in 2006 with Re-4; and the third in 2008 with Re-5 (Figure 6b). The dN/dS ratio comparing clade 2.3.2 virus with the vaccine antigen was calculated. It was found that positions of 130–160 in the HA1 protein were under positive selections during 2005–2011 but not during 2003–2004, suggesting an effect of Re-1 vaccine, which the government started to implement in 2004 (Figure 7a). Similar phenomena occurred again in 2006. Positions 118–137 and 161–179 in the HA1 protein exhibited strong signals of positive selection during 2007–2011 in comparison with those during 2003–2006, suggesting again an effect of Re-4 vaccine, which the government started to implement in 2006 (Figure 7b).

To identify positively-selected sites among codons in the HA1 protein and avoid chance mutations, we divided clade 2.3.2 into three groups based on the timing of three adopted vaccines: unvaccinated group I in 2003–2004; early stage vaccinated group II in 2005–2006 and 2005–2007; and post-vaccination group III in

2007–2008, 2007–2011 and 2009–2011. The results showed no positively-selected sites in group I; the dN/dS ratio over all sites was 0.13–0.19 among models. In group II (2005–2006), model M8 suggested a proportion of sites (2.7%) under positive selection, with dN/dS ratio = 2.56 (Tables 3 and 4). The LRT statistics for comparing M7 and M8 were  $2 \ln L = 2 \times (\ln L_{M8} - \ln L_{M7}) = 6.38 > \chi^2_{5\%} = 5.99$  with d.f. = 2. For 2005–2007, model M8 suggested a proportion of sites (7.4%) under positive selection, with dN/dS ratio = 1.61 ( $2 \ln L = 2 \times (\ln L_{M8} - \ln L_{M7}) = 8.50 > \chi^2_{5\%} = 5.99$  with d.f. = 2). However, in group III, the model did not detect any positive selection in the dataset. The results indicated that the selected sites occurred after vaccination policy implementation, and the regularly-used vaccines in China did not provide full protection against clade 2.3.2.1 virus after 2007.

Some sites in the HA gene (156 and 172 in group II) were consistently detected as positively selected by both the M7 and M8 models in codeml and by the FEL method. We compared the variable sites between clade 2.3.2 (2003–2006) and clade 2.3.2.1 (2007–2011), summarizing the key sites of variation (Table 5). Based on these results (Figure 7a,b, Tables 3 and 5), we believe that the amino acid site 156 was under selective pressure from the vaccine used in China. These results provide strong support for adaptive evolution of the clade 2.3.2 virus, and that it evolved into clade 2.3.2.1 to escape the vaccine pressure since 2006.

## Discussion

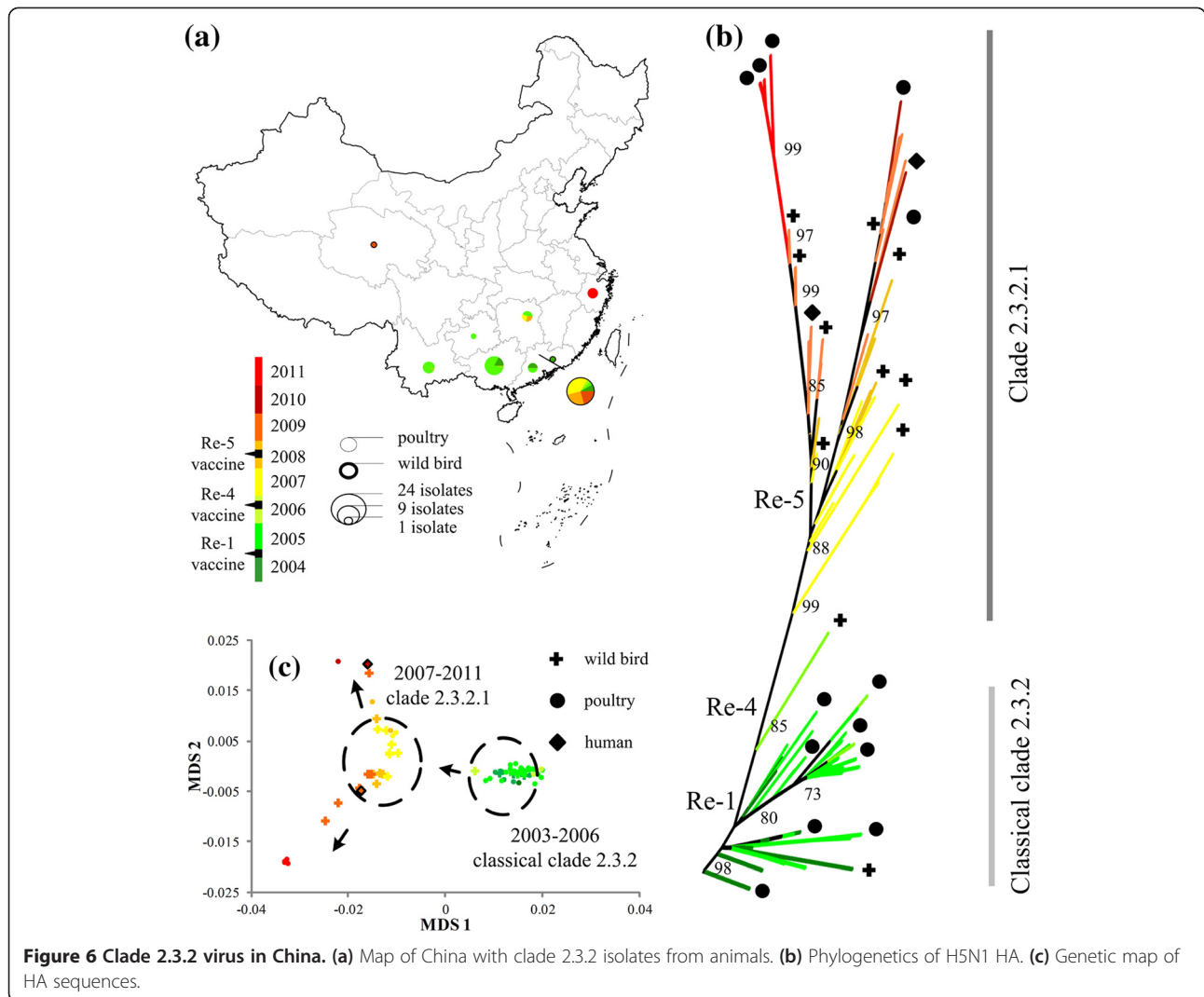
In 1996, A/goose/Guangdong/1/1996 (H5N1), the precursor of currently circulating HPAI H5N1 virus was identified in China [42]. The virus has circulated for 17 years, and presented an imminent threat to humans, poultry production, and wild animals in China [43,44]. Our study provides insight into the spatiotemporal pattern of H5N1 outbreaks in China and the dynamics of virus evolution. The results showed obvious spatiotemporal clusters of H5N1 outbreaks on different scales associated with two transmission modes of H5N1 viruses. Viral evolutionary dynamics were analyzed, and the effect of vaccination on virus circulation in China was identified.

Our time series analysis indicated a significant temporal relationship between poultry outbreaks, human cases, and wild-bird outbreaks. The current transmission chain of H7N9 virus in the country coincides with this H5N1 path. Results showed that human cases correlated with poultry outbreaks with a 1–4-month lag, which indicated that the infection may have started in poultry from live-bird markets and then transmitted to humans, even though the timing of wild-bird infection is still uncertain [45,46]. It is postulated that the virus might have

**Table 2 Evolutionary profiles of H5N1 HPAI viruses: comparisons between strict and relaxed (uncorrelated lognormal) molecular clocks using the constant, exponential growth, Bayesian skyride and Bayesian skyline coalescent prior**

		Evolutionary rates (sub/site/year) $\times 10^{-3}$ (95% HPD*)							
Year	Best fit clock model	Strict clock				Uncorrelated lognormal molecular clock (ULCN)			
		Constant size	Exponential growth	Bayesian skyride	Bayesian skyline	Constant size	Exponential growth	Bayesian skyride	Bayesian skyline
2000-2001	Strict-Skyride	4.15 (2.97-5.44)	3.49 (2.34-4.76)	3.85 (2.98-4.80)	4.12 (3.07-5.39)	4.36 (2.96-5.98)	4.36 (2.93-5.82)	4.61 (3.25-6.08)	4.65 (2.93-6.37)
2001-2002	Strict-Constant	4.06 (3.03-5.17)	3.24 (2.11-4.48)	3.83 (2.92-4.78)	3.72 (2.73-4.72)	5.17 (3.63-6.74)	4.86 (3.21-6.59)	5.16 (3.58-6.63)	5.48 (3.80-7.18)
2002-2003	Strict-Exponential	2.51 (1.87-3.26)	2.34 (1.68-3.09)	2.29 (1.68-2.93)	2.47 (1.81-3.13)	2.52 (1.78-3.29)	2.97 (2.01-3.94)	2.79 (1.85-3.87)	2.61 (1.77-3.49)
2003-2004	Strict-Constant	4.31 (3.38-5.23)	4.15 (3.31-5.00)	4.19 (3.28-5.12)	4.21 (3.24-5.13)	4.83 (3.76-5.93)	4.84 (3.7-6.22)	1.04 (0.01-5.02)	4.86 (3.38-6.60)
2004-2005	Strict-Skyride	6.81 (5.69-7.95)	6.88 (5.66-8.03)	6.63 (5.73-7.56)	6.70 (5.69-7.67)	8.99 (6.69-11.08)	7.83 (6.31-10)	7.15 (5.55-8.46)	7.37 (4.20-9.98)
2005-2006	Strict-Exponential	3.55 (2.94-4.17)	3.62 (3.09-4.19)	3.85 (2.5-5.14)	3.14 (2.03-3.92)	3.52 (2.73-4.44)	4.31 (3.27-5.38)	4.03 (1.25-5.69)	3.44 (1.29-4.53)
2006-2007	UCLN-Exponential	4.18 (3.28-5.17)	3.76 (2.59-4.68)	4.07 (2.8-5.45)	3.41 (2.28-4.47)	5.72 (3.78-7.44)	5.48 (3.99-6.92)	4.83 (1.25-5.69)	5.61 (2.20-8.40)
2007-2008	Strict-Skyline	4.37 (3.29-5.56)	4.19 (3.06-5.37)	4.52 (3.4-5.62)	4.66 (3.63-5.81)	4.08 (2.76-5.55)	4.58 (3.04-6.27)	4.96 (3.63-6.44)	5.01 (3.22-7.13)
2008-2009	Strict-Exponential	4.91 (3.64-6.12)	4.84 (3.58-6.07)	5.03 (3.85-6.26)	4.74 (3.57-6.02)	5.98 (4.04-8.06)	6.33 (3.94-8.79)	7.31 (4.72-10.08)	6.44 (4.12-8.96)
1996-2004	UCLN-Exponential	3.24 (2.92-3.60)	3.19 (2.86-3.51)	4.03 (3.58-4.52)	3.63 (3.13-4.10)	3.54 (3.08-3.97)	3.77 (3.33-4.22)	4.03 (3.57-4.51)	3.63 (3.13-4.10)
2005-2011	Strict-Constant	4.39 (4.03-4.76)	4.44 (4.06-4.80)	4.35 (3.88-4.81)	4.45 (4.05-4.87)	4.28 (3.67-4.84)	4.26 (3.15-5.11)	4.30 (3.82-4.78)	3.95 (3.58-4.28)

\*HPD, highest posterior density.

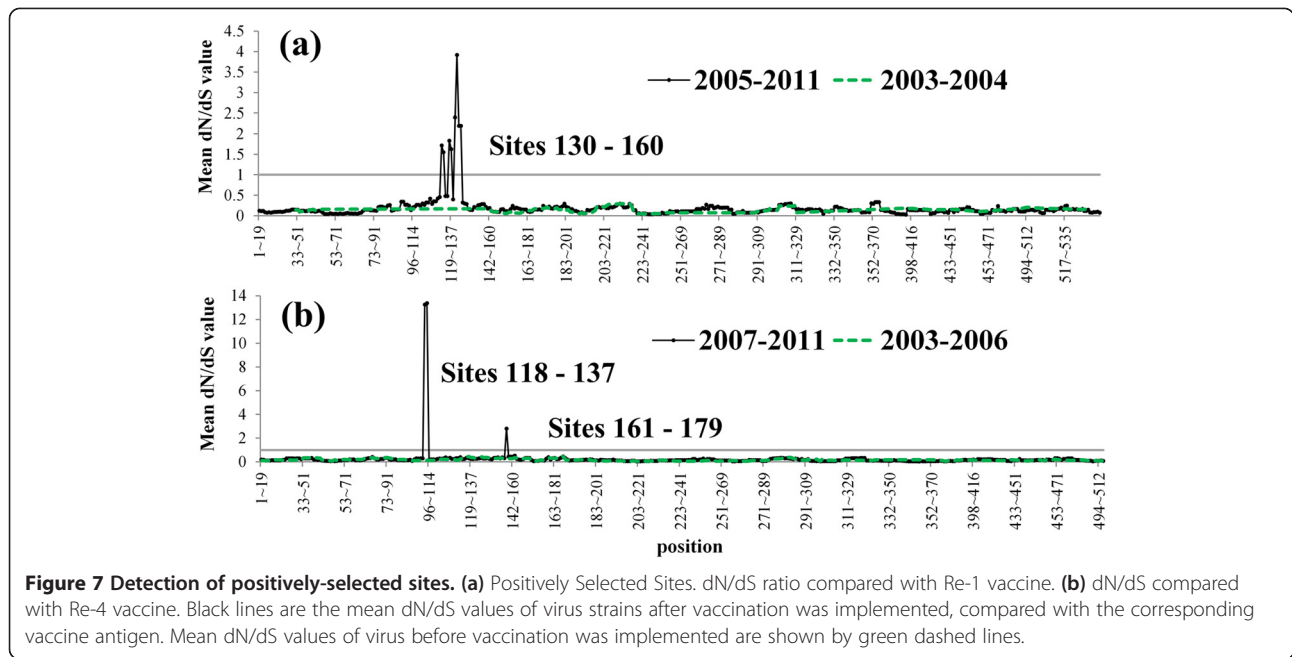


undergone a period of time in poultry and the environment without being detected [47,48]. With an increase in imported poultry or active poultry production to meet the demand prior to Lunar New Year activities, most poultry outbreaks occur during these festivals in China [49]. Human infections occurred 1–4 months thereafter because of dense human population, increased exposure, and possibly evolved high-affinity binding of the virus to human receptors. Wild-bird infections might occur last (correlated with human cases with a 1–3-month lag) after exposure and sufficient population of wild species. Many wild bird species winter in Southern China, and start spring migration in early April, and during these periods, wild birds may be infected directly through contact with infected poultry, or the environment [50].

The transmission mode of poultry was faster, with a shorter cycle (Figure 2e). Additionally, the evolving capability for sustained transmission across species barriers

represents a major adaptive challenge, because the number of required mutations is often large [51]. However, a vast number of poultry hosts serve as a reservoir, providing a sufficient population basis for accumulating such mutations [52]. Poultry may therefore play a crucial role in the avian influenza epidemics in the country.

Study of the spatiotemporal pattern of outbreaks is important in the prevention and control of epidemics over large regions [53]. The seasonal characteristics of outbreaks in different hosts means that as the season alternates, epidemic areas shift significantly [17]. The epidemic regions will naturally be affected by future climate change [54,55]. Poultry outbreaks concentrate in January, February, June and November. Wild-bird outbreaks concentrate in January and February in Hong Kong, and April and May in Northeastern China and Qinghai–Tibet Plateau. Through spatiotemporal pattern analysis, we can identify possible areas of subsequent outbreaks around



epidemic areas. We can calculate an appropriate radius for prevention and culling, and establish early warning systems for regions potentially affected by outbreaks. The results of this study will help to develop appropriate prevention and control policies toward various host outbreaks in different time periods, to avoid the possibility of continuous large outbreaks.

A culling plus vaccination mixed strategy was initiated for control of HPAI H5N1 outbreaks in Mainland China [56]. The actual effect of vaccination in different years and epidemic regions is debated. Although direct association between vaccination and H5N1 virus evolution is difficult to establish, we found that both evolutionary rates and positively selected sites were affected by vaccination for H5N1 in China. OIE has recognized that the emergence of clade 2.3.2.1 virus was one of the genetic mutations occurring as part of neutral virus

evolution [57]. However, we believe that the mutation possibly resulted from vaccination, and some mutation sites in clade 2.3.2.1 virus exhibited a strong signature of positive selection, which was against the signature of neutral evolution.

When an outbreak occurs in many regions, simple culling may not be effective. This is also true in some other infectious diseases [12]. In developing countries, controlling virus spread by only long-term, large-scale culling is an immeasurable encumbrance on human living standards [58]. Therefore, vaccination in combination with culling may be the most appropriate way for the country to control infections. The evolutionary patterns and spatiotemporal distribution of the virus are important in making targeted vaccination policy and developing appropriate prevention measures [59]. An appropriate vaccination strategy that includes immunity planning

**Table 3** Positively-selected sites in the HA genes of clade 2.3.2 virus

Group	Year	Vaccine	M7 lnL	M8 lnL	Positively selected sites by BEB (>50%) <sup>†</sup>	Positively selected sites by FEL ( <i>P</i> <0.05) <sup>†</sup>
I	2003-2004	None	-2649.7	-2649.6	None	
II	2005-2006	Re-1	-4227.7	-4224.5	145,154, <b>156</b> ,157,171, <b>172</b> ,194,204,205	156,171,172,199,205,216
II	2005-2007	Re-1, Re-4	-4872.9	-4868.6	61,136,145,154, <b>156*</b> ,157,170,171,172,178,199,200,204,205	56,61,69,131,136,154,156,172,199,204,205
III	2007-2008	Re-1, Re-4	-3807.5	-3806.1	None	
III	2007-2011	Re-1, Re-4, Re-5	-5008.7	-5007.1	None	
III	2009-2011	Re-4, Re-5	-4035.2	-4033.8	None	

Note. Sites inferred under selection at the 90% level are listed in bold, and those at the 95% level are marked with an asterisk.

<sup>†</sup>The posterior probability and *P* value in LRT with null hypothesis of the site are shown in parenthesis for the BEB and FEL analyses, respectively.

**Table 4 Parameter estimates for PAML models**

Group	Model	Parameter estimates	Sites* with $\omega > 1$ , $\omega$ (SE)
II 2005-2006	M7 ( $\beta$ )	$p = 0.19, q = 0.64$	145, 1.379 (0.763)
	M8 ( $\beta$ & $\omega$ )	$p = 0.63, q = 2.96, f_0 = 0.97$ $\omega = 2.56, f_1 = 0.03$	154, 1.34 (0.765) <b>156</b> , 1.748 (0.571) 157, 1.313 (0.765) 171, 1.409 (0.714) <b>172</b> , 1.749 (0.571) 178, 1.653 (0.636) 204, 1.429 (0.712) 205, 1.517 (0.702)
II 2006-2007	M7 ( $\beta$ )	$p = 0.16, q = 0.60$	61, 1.229 (0.634)
	M8 ( $\beta$ & $\omega$ )	$p = 1.81, q = 13.43, f_0 = 0.93$ $\omega = 1.61, f_1 = 0.07$	136, 1.506 (0.584) 145, 1.288 (0.687) 154, 1.252 (0.688) <b>156*</b> , 1.701 (0.454) 157, 1.211 (0.688) 170, 1.507 (0.585) 171, 1.496 (0.588) 172, 1.632 (0.516) 178, 1.504 (0.585) 199, 1.202 (0.63) 200, 1.358 (0.637) 204, 1.268 (0.637) 205, 1.371 (0.636)

$p$  and  $q$  are parameters of the  $\beta$  distribution.  $f$  is the proportion of sites assigned to an individual  $\omega$  ( $= dN/dS$ ) category or to a  $\beta$  distribution with shape parameters  $p$  and  $q, f_1 = 1 - f_0$ . Sites inferred under selection at the 90% level are listed in bold, and those at the 95% level are marked with an asterisk.

directed at the spatiotemporal distribution of the circulating virus and its possible evolutionary pattern is more important than efficacy of the vaccine itself [60].

The limitations of this study should be mentioned. First, the spatiotemporal information of H5N1 virus depended on sampling. Virus samples may have been concentrated in supposed high-risk areas, which could have led to data bias. Second, H5N1 outbreak information was collected from a passive surveillance system, and data quality varied significantly across provinces.

## Conclusions

In this study, we highlighted two transmission modes on a regional scale through a spatiotemporal analysis, and revealed lead-lag relationships among poultry outbreaks, human cases and wild-bird outbreaks. We analyzed the circulation of H5N1 virus, and established an effect of vaccination on virus evolution in China. In conclusion, the H5N1 virus has become endemic in China, and variant strains continue to emerge [42,49]. For infection control, we need to develop effective vaccination and culling strategies, based on analysis of outbreaks and evolutionary patterns of H5N1 virus.

**Table 5 Key modified sites between the classical clade 2.3.2 and clade 2.3.2.1 virus, China**

Clade	Codon position								
	18	61	69	136	156	178	196	242	279
Classical 2.3.2	Gln (100%)	Asp (97.96%)	Arg (97.96%)	Ser (97.96%)	Ser (91.84%)	Arg (95.92%)	Asn (53.06%)	Met (95.92%)	Ala (100%)
2.3.2.1	His (100%)	Asn (91.67%)	Lys (100%)	Asp (95.83%)	Asn (93.75%)	Lys (100%)	Asp (95.83%)	Ile (100%)	Thr (91.67%)

Note. Based on the results of Figure 7a and b, we conclude that the virus may have escaped from the vaccine in sites 118–137, 130–160 and 161–179. Table 3 indicates that site 156 received selective pressure with 95% probability. The table summarizes the key sites that varied from the classical clade 2.3.2 virus to clade 2.3.2.1 virus.

## Additional files

**Additional file 1: Table S1.** Sequences used in this study.

**Additional file 2: Figure S1.** The cobweb chart plot of seasonal distribution of avian influenza A (H5N1). Reported monthly average occurrence of avian influenza A (H5N1) in China, January 2004–December 2011. **(a)** Poultry outbreaks; **(b)** wild-bird outbreaks.

**Additional file 3: Table S2.** Summary of model performance and the estimated coefficients.

**Additional file 4: Figure S2.** Phylogenetic tree based on H5N1 HA gene sequences in China.

**Additional file 5: Figure S3.** Mean rate of nucleotide substitution of all the H5N1 viruses collected from China using the best fit clock model.

**Additional file 6: Figure S4.** Neighbor-joining (NJ) tree of clade 2.3.2 H5N1 HA sequences in China.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

HT and BX designed the overall scope of the research. HT, YC, LD, SZ, XL and SH analyzed the data. XL, RY and BX contributed reagents/materials/analysis tools. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

### Funding sources

This research was supported by the Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407), and the National Natural Science Foundation of China (41271099). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Data accessibility section

DNA sequences: Genbank accessions in Additional file 1: Table S1 (Supporting information).

Final DNA sequence assembly uploaded as online supplemental material. H5N1 outbreak data (FAO): <http://empres-iafao.org/eipws3g/#h=0>. H5N1 outbreak data (OIE): <http://www.oie.int/?id=1682>.

### The sources of funding for each author:

H.Y.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407), and the National Natural Science Foundation of China (41271099).

Y.C.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407).

L.D.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407).

S.Z.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407), and the National Natural Science Foundation of China (41271099).

X.L.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407), and the National Natural Science Foundation of China (41271099).

R.Y.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407).

B.X.: Ministry of Science and Technology, China, National Research Program (2010CB530300, 2012CB955501, 2012AA12A407), and the National Natural Science Foundation of China (41271099).

### Author details

<sup>1</sup>State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China.

<sup>2</sup>State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China. <sup>3</sup>Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China.

<sup>4</sup>Ministry of Education Key Laboratory for Earth System Modelling, Center for

Earth System Science, Tsinghua University, Beijing 100084, China.

<sup>5</sup>Department of Geography, University of Utah, Salt Lake City, UT 84112, USA.

Received: 5 July 2014 Accepted: 19 January 2015

Published online: 13 February 2015

### References

- Tian H, Zhou S, Dong L, Van Boeckel TP, Pei Y, Wu Q et al. Climate change suggests a shift of H5N1 risk in migratory birds. *Ecol Model*. 2014, doi:10.1016/j.ecolmodel.2014.08.005.
- Webster RG, Govorkova EA. H5N1 influenza—continuing evolution and spread. *N Engl J Med*. 2006;355(21):2174–7.
- Carrel M, Wan X-F, Nguyen T, Emch M. Highly pathogenic H5N1 avian influenza viruses exhibit few barriers to gene flow in Vietnam. *Ecohealth*. 2012;9(1):60–9.
- Carrel M, Wan X-F, Nguyen T, Emch M. Genetic Variation of highly pathogenic H5N1 avian influenza viruses in vietnam shows both species-specific and spatiotemporal associations. *Avian Dis*. 2011;55(4):659–66.
- Xu X, Subbarao, Cox NJ, Guo Y. 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*. 1999;261(1):15–9.
- Smallman-Raynor M, Cliff AD. The geographical spread of avian influenza A (H5N1): panzootic transmission (December 2003–May 2006), pandemic potential, and implications. *Ann Assoc Am Geogr*. 2008;98(3):553–82.
- Carrel MA, Emch M, Jobe RT, Moody A, Wan X-F. Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza viruses in Vietnam. *PLoS ONE*. 2010;5(1):e8631.
- Carrel MA, Emch M, Nguyen T, Todd Jobe R, Wan X-F. Population-environment drivers of H5N1 avian influenza molecular change in Vietnam. *Health Place*. 2012;18(5):1122–31.
- Kilpatrick AM, Chmura AA, Gibbons DW, Fleischer RC, Marra PP, Daszak P. Predicting the global spread of H5N1 avian influenza. *Proc Natl Acad Sci U S A*. 2006;103(51):19368–73.
- Liang L, Xu B, Chen YL, Liu Y, Cao WC, Fang LQ, et al. Combining spatial-temporal and phylogenetic analysis approaches for improved understanding on global H5N1 transmission. *PLoS ONE*. 2010;5(10):e13575.
- China Animal Industry Yearbook Editorial Committee. *China Animal Industry Yearbook*. Beijing, China: Agriculture Press; 2011.
- Wang Y, Jiang Z, Jin Z, Tan H, Xu B. Risk factors for infectious diseases in backyard poultry farms in the poyang lake area, china. *PLoS ONE*. 2013;8(6):e67366.
- Chen H. H5N1 avian influenza in China. *Sci China C Life Sci*. 2009;52(5):419–27.
- Lee CW, Senne DA, Suarez DL. Effect of vaccine use in the evolution of Mexican lineage H5N2 avian influenza virus. *J Virol*. 2004;78(15):8372–81.
- Cattoli G, Fusaro A, Monne I, Coven F, Joannis T, El-Hamid HS, et al. Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. *Vaccine*. 2011;29(50):9368–75.
- Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global circulation of seasonal influenza A (H3N2) viruses. *Science*. 2008;320(5874):340–6.
- Xu B, Jin Z, Jiang Z, Guo J, Timberlake MXM. Climatological and Geographical Impacts on Global Pandemic of Influenza A (H1N1). In: Weng Q, editor. *Global Urban Monitoring and Assessment through Earth Observation*. Boca Raton, Florida, U.S.A: Taylor & Francis/CRC Press; 2013.
- Bailey TC, Gatrell AC. *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific & Technical Essex; 1995.
- Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. *Trans Inst Br Geogr*. 1996;21(1):256–74.
- Ripley BD. *Spatial Statistics*. New York: John Wiley; 1981.
- Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. New Jersey: John Wiley & Sons; 2004.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997;25(24):4876.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp*. 1999;41:95–8.

24. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;19:716–23.
25. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics.* 1998;14(9):817–8.
26. Swofford DL. PAUP\*: Phylogenetic Analysis Using Parsimony, Version 4.0 b10. Sunderland, MA: Sinauer; 2002.
27. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7(1):214.
28. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol.* 2006;23(1):7–9.
29. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika.* 1964;29(1):1–27.
30. Smith DJ, Lapedes AS, De Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* 2004;305(5682):371–6.
31. Miyata T, Miyazawa S, Yasunaga T. Two types of amino acid substitutions in protein evolution. *J Mol Evol.* 1979;12(3):219–36.
32. Yang Z, Nielsen R, Goldman N, Pedersen AMK. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* 2000;155(1):431–49.
33. Li WH. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol.* 1993;36(1):96–9.
34. Wang D, Zhang S, He F, Zhu J, Hu S, Yu J. How do variable substitution rates influence Ka and Ks calculations? *Genomics Proteomics Bioinformatics.* 2009;7(3):116–27.
35. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
36. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 1998;15(5):568–73.
37. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;21(5):125–81.
38. Yang Z. *Computational Molecular Evolution.* Oxford: Oxford University Press; 2006.
39. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis.* San Francisco: Holden-day; 1976.
40. Li Y, Liu L, Zhang Y, Duan Z, Tian G, Zeng X, et al. New avian influenza virus (H5N1) in wild birds, Qinghai. *China Emerg Infect Dis.* 2011;17(2):265–7.
41. Chen H, Smith GJ, Zhang SY, Qin K, Wang J, Li KS, et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature.* 2005;436(7048):191–2.
42. Li K, Guan Y, Wang J, Smith G, Xu K, Duan L, et al. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature.* 2004;430(6996):209–13.
43. Chen H, Deng G, Li Z, Tian G, Li Y, Jiao P, et al. The evolution of H5N1 influenza viruses in ducks in southern China. *Proc Natl Acad Sci U S A.* 2004;101(28):10452–7.
44. Smith GJ, Vijaykrishna D, Ellis TM, Dyrting KC, Leung YH, Bahl J, et al. Characterization of avian influenza viruses A (H5N1) from wild birds, Hong Kong, 2004–2008. *Emerg Infect Dis.* 2009;15(3):402–7.
45. Shi J, Deng G, Liu P, Zhou J, Guan L, Li W, et al. Isolation and characterization of H7N9 viruses from live poultry markets—implication of the source of current H7N9 infection in humans. *Chinese Sci Bull.* 2013;58:1857–63.
46. Liu D, Shi W, Shi Y, Wang D, Xiao H, Li W, et al. Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet.* 2013;381(9881):1926–32.
47. Zhang H, Xu B, Chen Q, Chen Z. Characterization of H9N2 influenza viruses isolated from Dongting Lake wetland in 2007. *Arch Virol.* 2011;156(1):95–105.
48. Yao Y, Wang H, Chen Q, Zhang H, Zhang T, Chen J, et al. Characterization of low-pathogenic H6N6 avian influenza viruses in central China. *Arch Virol.* 2013;158(2):367–77.
49. Smith GJ, Fan XH, Wang J, Li KS, Qin K, Zhang JX, et al. Emergence and predominance of an H5N1 influenza variant in China. *Proc Natl Acad Sci U S A.* 2006;103(45):16936–41.
50. Si Y, Skidmore AK, Wang T, de Boer WF, Debba P, Toxopeus AG, et al. Spatiotemporal dynamics of global outbreaks match bird migration patterns. *Geospat Health.* 2009;4:65–78.
51. Kuiken T, Holmes EC, McCauley J, Rimmelzwaan GF, Williams CS, Grenfell BT. Host species barriers to influenza virus infections. *Science.* 2006;312(5772):394–7.
52. Boyce WM, Sandrock C, Kreuder-Johnson C, Kelly T, Cardona C. Avian influenza viruses in wild birds: a moving target. *Comp Immunol Microbiol Infect Dis.* 2009;32(4):275–86.
53. Xu B, Gong P, Seto E, Liang S, Yang CHSW, Qiu DC, et al. A spatial temporal model for assessing the effects of inter-village connectivity in schistosomiasis transmission. *Ann Assoc Am Geogr.* 2006;96(1):31–46.
54. Wu X, Tian H, Zhou S, Chen L, Xu B. Impact of global change on transmission of human infectious diseases. *Sci China Earth Sci.* 2013.
55. Yang J, Gong P, Fu R, Zhang M, Chen J, Liang S, et al. The role of satellite remote sensing in climate change studies. *Nature Clim Change.* 2013. forthcoming.
56. Li Y, Shi J, Zhong G, Deng G, Tian G, Ge J, et al. Continued evolution of H5N1 influenza viruses in wild birds, domestic poultry, and humans in China from 2004 to 2009. *J Virol.* 2010;84(17):8389–97.
57. OIE: Avian influenza H5N1 clade 2.3.2.1. Paris; 2011: <http://www.oie.int/for-the-media/press-releases/detail/article/avian-influenza-h5n1-clade-2321/>
58. Pimentel D, Cooperstein S, Randell H, Filiberto D, Sorrentino S, Kaye B, et al. Ecology of increasing diseases: population growth and environmental degradation. *Hum Ecol.* 2007;35(6):653–68.
59. Xu B, Gong P. Spatial temporal modeling of endemic diseases: Schistosomiasis transmission and control as an example. In Madden M, editor. *Manual of geographic information systems.* ASPRS Pubns; 2009. p. 413–26.
60. Kim JK, Kayali G, Walker D, Forrest HL, Ellebedy AH, Griffin YS, et al. Puzzling inefficiency of H5N1 influenza vaccines in Egyptian poultry. *Proc Natl Acad Sci U S A.* 2010;107(24):11044–9.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

