



H5Nx Viruses Emerged during the Suppression of H5N1 Virus Populations in Poultry

Yao-Tsun Li,^a  Yvonne C. F. Su,^a  Gavin J. D. Smith^{a,b,c}

^aProgramme in Emerging Infectious Diseases, Duke-NUS Medical School, Singapore

^bSingHealth Duke-NUS Global Health Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore

^cDuke Global Health Institute, Duke University, Durham, North Carolina, USA

ABSTRACT Highly pathogenic avian influenza (HPAI) H5 viruses have posed a substantial pandemic threat through repeated human infection since their emergence in China in 1996. Nationwide control measures, including vaccination of poultry, were implemented in 2005, leading to a sharp reduction in H5N1 virus outbreaks. In 2008, novel non-N1 subtype (H5Nx) viruses emerged, gradually replacing the dominant H5N1 subtype and causing global outbreaks. The cause of this major shift in the ecology of HPAI H5 viruses remains unknown. Here, we show that major H5N1 virus lineages underwent population bottlenecks in 2006, followed by a recovery in virus populations between 2007 and 2009. Our analyses indicate that control measures, not competition from H5Nx viruses, were responsible for the H5N1 decline, with an H5N1 lineage capable of infecting poultry and wild birds experiencing a less severe population bottleneck due to circulation in unaffected wild birds. We show that H5Nx viruses emerged during the successful suppression of H5N1 virus populations in poultry, providing an opportunity for antigenically distinct H5Nx viruses to propagate. Avian influenza vaccination programs would benefit from universal vaccines targeting a wider diversity of influenza viruses to prevent the emergence of novel subtypes.

IMPORTANCE A major shift in the ecology of highly pathogenic avian influenza (HPAI) H5 viruses occurred from 2008 to 2014, when viruses with non-N1 neuraminidase genes (termed H5Nx viruses) emerged and caused global H5 virus outbreaks. Here, we demonstrate that nationwide control measures, including vaccination in China, successfully suppressed H5N1 populations in poultry, providing an opportunity for antigenically distinct H5Nx viruses to emerge. In particular, we show that the widespread use of H5N1 vaccines likely conferred a fitness advantage to H5Nx viruses due to the antigenic mismatch of the neuraminidase genes. These results indicate that avian influenza vaccination programs would benefit from universal vaccines that target a wider diversity of influenza viruses to prevent potential emergence of novel subtypes.

KEYWORDS evolution, pandemic, zoonotic, influenza, zoonotic infections

Influenza A virus has a wide host range and is able to infect and transmit between diverse animals, including both birds and mammals (1). Highly pathogenic avian influenza (HPAI) A/Goose/Guangdong/96 (Gs/GD) H5N1 viruses have caused recurrent outbreaks in poultry and wild birds since they were first recognized in southern China in 1996 (2). After the successful establishment of Gs/GD viruses in China's enormous domestic poultry population, the virus spread throughout Eastern and Southeast Asia, resulting in devastating economic losses as a consequence of infection or culling of infected poultry (3). H5N1 viruses subsequently underwent diversification (Fig. 1a) and are classified into multiple clades that are often not antigenically cross-reactive, greatly complicating vaccination strategies (3, 4). Reverse spillover of H5N1 into migratory birds

Citation Li Y-T, Su YCF, Smith GJD. 2021. H5Nx viruses emerged during the suppression of H5N1 virus populations in poultry. *Microbiol Spectr* 9:e01309-21. <https://doi.org/10.1128/Spectrum.01309-21>.

Editor Daniel R. Perez, University of Georgia

Copyright © 2021 Li et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Gavin J. D. Smith, gavin.smith@duke-nus.edu.sg.

Received 28 August 2021

Accepted 30 August 2021

Published 29 September 2021

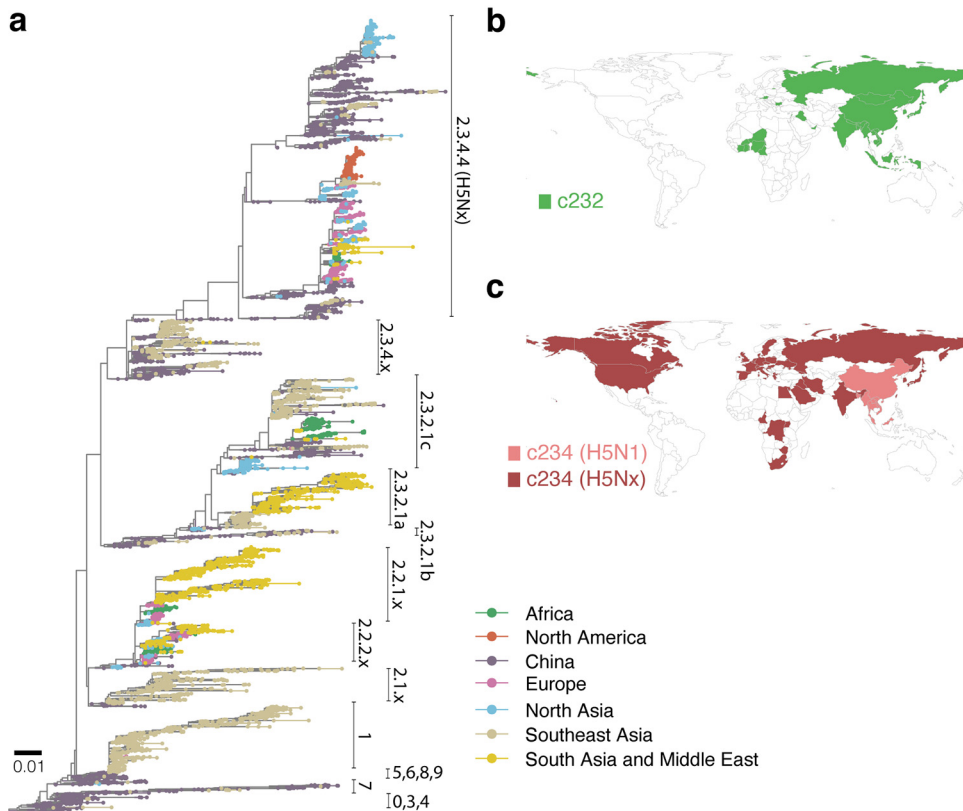


FIG 1 Geographical distribution of *A/goose/Guangdong/1/96*-like (Gs/GD) viruses based on available HA sequences. (a) Maximum likelihood tree was inferred with HA sequences of Gs/GD viruses with tips colored based on the location of isolation ($n = 8,487$). Different sublineages classified by the H5 nomenclature system are labeled next to the tips. Scale bar represents 0.01 substitutions per nucleotide site. Global distributions of clade 2.3.2 (b) and H5N1 and H5Nx clade 2.3.4 (c) viruses are shown in pink and dark red, respectively.

occurred at Qinghai Lake in 2005 (5), leading to intercontinental spread of the virus to the Middle East and Europe (6). Since then, many studies have focused on the distinct roles played by avian species in different ecological niches in virus dispersal (7–9). As of May 2020, 861 human H5N1 cases and 455 fatalities have been recorded (10).

Until 2014, outbreaks in poultry and human infections caused by Gs/GD viruses were exclusively H5N1 subtype (11, 12). H5N1 underwent frequent reassortments with low-pathogenic avian influenza (LPAI) viruses carrying different neuraminidase (NA) subtypes but only acquired internal genes to form distinct genotypes (13, 14). In contrast, Gs/GD isolates possessing an NA gene other than N1 were infrequently detected (11, 12), with H5 genes generally coevolving with N1 genes within the Gs/GD sublineages (13). However, since 2014, a significant change in the NA subtype occurred with the emergence of reassortant H5Nx virus outbreaks in poultry (15, 16). These H5Nx viruses include H5N2, H5N6, and H5N8 subtypes initially identified in China and Northern Asia that subsequently spread to different regions across Asia, Europe, and North America, far exceeding the distribution of H5N1 viruses (Fig. 1b and c). These H5Nx viruses demonstrated unprecedented compatibility with different NA subtypes, rapidly generating novel H5-NA combinations with endemic strains when introduced to other countries (15, 17). It is unclear what factors may have driven the transition from a single avian N1 subtype to multiple non-N1 NA subtypes.

RESULTS

Evolution and coevolution of diverse H5 clades in China. To better understand the mechanisms underlying the emergence of H5Nx viruses, we first characterized overall H5 virus diversity in China. By the early 2000s, H5N1 viruses were highly prevalent

in poultry in China with spillover to wild birds (18), and three major monophyletic clades (clades 2.3.2, 2.3.4, and 7) had become predominant (Fig. 2a), constituting >50% of all available H5-hemagglutinin (HA) sequences in GenBank and GISAID databases from 2005 to 2019 (see Fig. S1a in the supplemental material). Other H5N1 clades were undetected from 2008 onward. With temporal phylogenies of H5-HA genes, we inferred the time of the most recent common ancestors (tMRCAs) of clades 2.3.2, 2.3.4, and 7 as 2002, 2003, and 2001 (Fig. 2a), indicating that they cocirculated for more than 10 years. The first detected clade 2.3.4.4 H5Nx virus was H5N5 isolated from a domestic duck in China in 2008 (19), which was >1 year divergent from the most closely related clade 2.3.4 H5N1 virus (Fig. 2a), indicating H5Nx viruses circulated undetected in that period. H5Nx viruses subsequently became predominant in China, eventually replacing their ancestral H5N1 viruses (Fig. 2a; see also Fig. S1b). The tMRCAs of clades 2.3.4.4 and 7 H5Nx viruses are both estimated as early 2007 (March 2007) with an uncorrelated rate model (Table 1), suggesting that they emerged during the same period. The concordant emergence of H5Nx viruses in multiple clades represents a major change in the evolutionary landscape of the Gs/GD H5N1 lineage. H5Nx viruses subsequently spread from China to multiple countries, including the first incursion of Gs/GD viruses to North America (Fig. 1a and c). These results show the dynamic lineage replacement occurring in China that corresponded with H5Nx emergence.

Comparative population dynamics of H5 viruses. To understand virus population behavior prior to the emergence of H5Nx viruses, we compared temporal changes in population sizes and growth rates of clades 2.3.2, 2.3.4, and 7. We first used the HA genes to infer effective population size using the Skygrid model implemented in a Bayesian phylogenetic framework (20, 21). The population size of all three clades (2.3.2, 2.3.4, and 7) increased exponentially to peak during 2005 to 2006, followed by an approximate 0.5 times contraction in population size, which recovered to their original levels after 4 to 5 years (Fig. 2b; see also Fig. S2a and S3 in the supplemental material). The growth rate estimates reflect the same population dynamics, with positive rates occurring until 2005 when clade 2.3.2 rates dropped below zero, followed by negative growth rates for clades 2.3.4 and 7 in 2006, indicating a contraction in population size. These negative growth rates matched the reduction in population size and remained low for subsequent years until 2010 (Fig. 2b). The patterns of population dynamics by the NA-N1 gene (Fig. 2c; Fig. S2b) demonstrated some distinguishing characteristics. For instance, the NA-N1 of clade 2.3.2 viruses had a relatively constant population size, but fluctuations were observed for clades 2.3.4 and 7. In clade 7, the HA and NA both showed a population decline following a peak in early 2006. Although the population contraction observed in the clade 2.3.4 HA gene was less obvious for the NA gene, growth rates of clades 2.3.4 and 7 both dropped from mid-2005 to 2007, the same pattern shared by HA estimates for all three clades (Fig. 2b). The decline of clade 2.3.4 after 2009 and the dramatic reduction in clade 7 inferred by NA-N1 are associated with the emergence of non-N1 reassortant viruses (Fig. 2c; Fig. S1b). Consistent patterns in population dynamics were present across the three replicates of these analyses (Fig. 2b and c; Fig. S2) regardless of subsampling (Fig. S3). We observed discordant patterns of population size and growth rates between HA and NA genes, which are likely due to the effect of interclade reassortment, as frequent reassortment events are common among clades, especially clade 2.3.2 (see Fig. S4 in the supplemental material). Notably, the emergence of clade 2.3.4.4 and 7 H5Nx viruses in early 2007 (Table 1) occurred after the observed population contraction and shows that they survived the bottleneck.

Comparative analysis of a contemporaneous H6 virus lineage (A/wild duck/Shantou/2853/2003-like) (22, 23), which is highly prevalent in poultry in China, showed no comparable bottleneck during 2005 to 2011 (see Fig. S3 and S5 in the supplemental material), suggesting that these population declines were not uniform across all avian influenza subtypes in China. To further test the effect of sampling, we arbitrarily created a sampling scheme in which no sequences were isolated in 2007 plus only half of 2008 sequences were selected compared to that in other sampling years. Results from

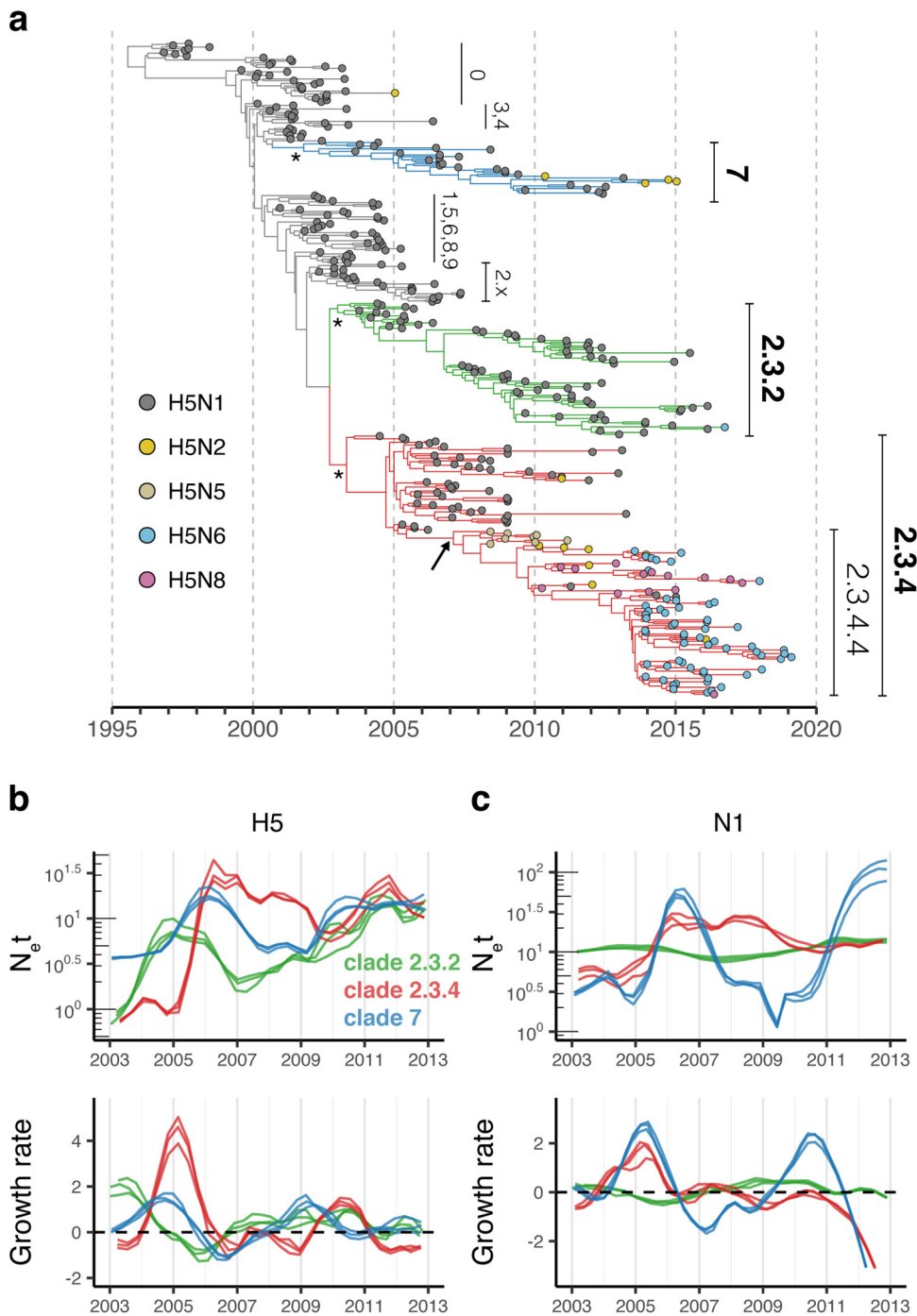


FIG 2 Phylogeny and population dynamics of Gs/GD viruses in China. (a) Maximum clade credibility (MCC) tree was reconstructed using HA genes from viruses isolated in China ($n = 369$). Tips are colored according to their subtypes. Different sublineages classified by the H5 nomenclature system are labeled next to the tips. Three major clades (2.3.2, 2.3.4, and 7) are labeled in bold, with their times of most recent common ancestors (tMRCA) highlighted by asterisks. tMRCA of clade 2.3.4.4 (H5Nx) is indicated by an arrow. Bayesian Skygrid method was used to infer effective population sizes ($N_e t$) using H5 ($n = 122$ to 123, 201, and 155 for clades 2.3.2, 2.3.4, and 7, respectively) (b) and N1 genes ($n = 135$, 88, and 35 for clades 2.3.2, 2.3.4, and 7) (c). Corresponding growth rates of population size were estimated by skygrowth method using annotated MCC trees generated in the same Bayesian run. Results of three randomly sampled data sets are shown. N1 sequences belonging to clade 2.3.4.4 viruses were not included in the analyses. Horizontal dashed line in the growth rate panels indicates zero.

TABLE 1 Estimated dates of nodes within the H5-HA phylogeny as shown in Fig. 2a

Model combinations	tMRCA (95% HPD ^c)		
	Clade 234 H5Nx (internal node ^a)	Clade 234 H5Nx (root ^b)	Clade 7 H5N2
Strict/Skygrid	2007.28 (2006.91–2007.61)	2006.91 (2006.33–2007.41)	2006.17 (2005.50–2006.85)
UCLN/Skygrid	2007.18 (2006.66–2007.63)	2006.91 (2005.89–2007.66)	2007.18 (2006.42–2007.93)

^atMRCAs were identified from MCC trees built by clade 2.3.4-N1 (i.e., clade 2.3.4.1 to 2.3.4.3) and clade 2.3.4.4.

^btMRCA were identified from MCC trees built using only clade 2.3.4.4 virus sequences.

^cHPD, highest posterior density.

this artificial scheme show a drop in population size lasting for less than 2 years with a magnitude smaller than 0.2 times (Fig. S3 and S5), suggesting the observed H5 clade population dynamics in China were not easily shaped by temporal absence of sampling effort or sequence availability.

We then performed neutrality tests with the HA nucleotide sequences to investigate population dynamics independent of Bayesian coalescent-based methods. Tajima's D is a scaled comparison of the genetic diversity and the number of segregation sites in the sequences (24). When more rare alleles are found in the samples, which may result from population expansion or selection sweep, Tajima's D would be negative. Conversely, when fewer low-frequency alleles are found, which may result from ongoing population bottleneck or balancing selection, Tajima's D would be positive (24, 25). Our results indicate that Tajima's D values of clades 2.3.2, 2.3.4, and 7 viruses all showed an upward trajectory from 2008 to 2010 and were generally negative outside of that window (see Fig. S6 in the supplemental material). These results suggest a dramatic reduction in excess low-frequency alleles occurred in H5N1 during 2008 to 2010, supporting the bottleneck inferred by the Bayesian coalescent-based methods (Fig. 2b). Furthermore, the Tajima's D values of all three clades were generally consistent with Fu and Li's D (Fig. S6b) and Fu and Li's F (Fig. S6c), two derivatives of Tajima's D following similar principles but taking into account singleton sites instead of segregation sites (26). These results suggest a demographic change in H5N1 viruses in China prior to the emergence of H5Nx viruses.

Ecosystem interactions and transmission dynamics of H5 virus. To understand whether agricultural activities influenced virus population size, we investigated H5N1 virus transmission between domestic and wild ecosystems in China. The rationale here is that, for example, increasing poultry production would boost the virus population, while control measures such as culling or vaccination would have the opposite effect, and virus populations in wild birds would be largely unaffected by human activities. To achieve this, we integrated the ecosystem type (i.e., domestic versus wild host populations) into an ancestral state reconstruction analysis and estimated the rate of viral transmission between domestic and wild ecosystems using a Bayesian phylogenetic diffusion model (27). Our results indicate that domestic-to-domestic transmissions dominated during the early diversification of clade 2.3.2, 2.3.4, and 7 viruses in China (Fig. 3a and b; see also Fig. S7 in the supplemental material). However, some marked differences between these virus clades were observed. From 2005 to 2007, clade 2.3.2 viruses showed noticeable transition from circulation in domestic to wild hosts, with the wild ecosystem reaching 100% of the trunk proportion in 2006, although subsequent transmissions were predominantly mediated by domestic-domestic spread (Fig. 3a). In contrast, clade 2.3.4 and 7 viruses showed limited cross-ecosystem transmission, and the domestic ecosystem occupied the trunk across their phylogenies (Fig. 3b; Fig. S7). Distribution of the ecological source of virus sequences in public databases also reflected the inferred trunk proportions, notably with over half of clade 2.3.2 viruses isolated from wild animals during 2007 to 2009 (Fig. S1c). Interestingly, we found that the transition rates (Fig. 3c) of clade 2.3.2 viruses were significantly greater than those of clade 2.3.4 and clade 7 viruses, reflecting more frequent interecosystem virus flow within clade 2.3.2. The total number of state transitions were higher in clade 2.3.2 and clade 2.3.4 but lower in clade 7 virus (Fig. 3d); however, only clade 2.3.2 had

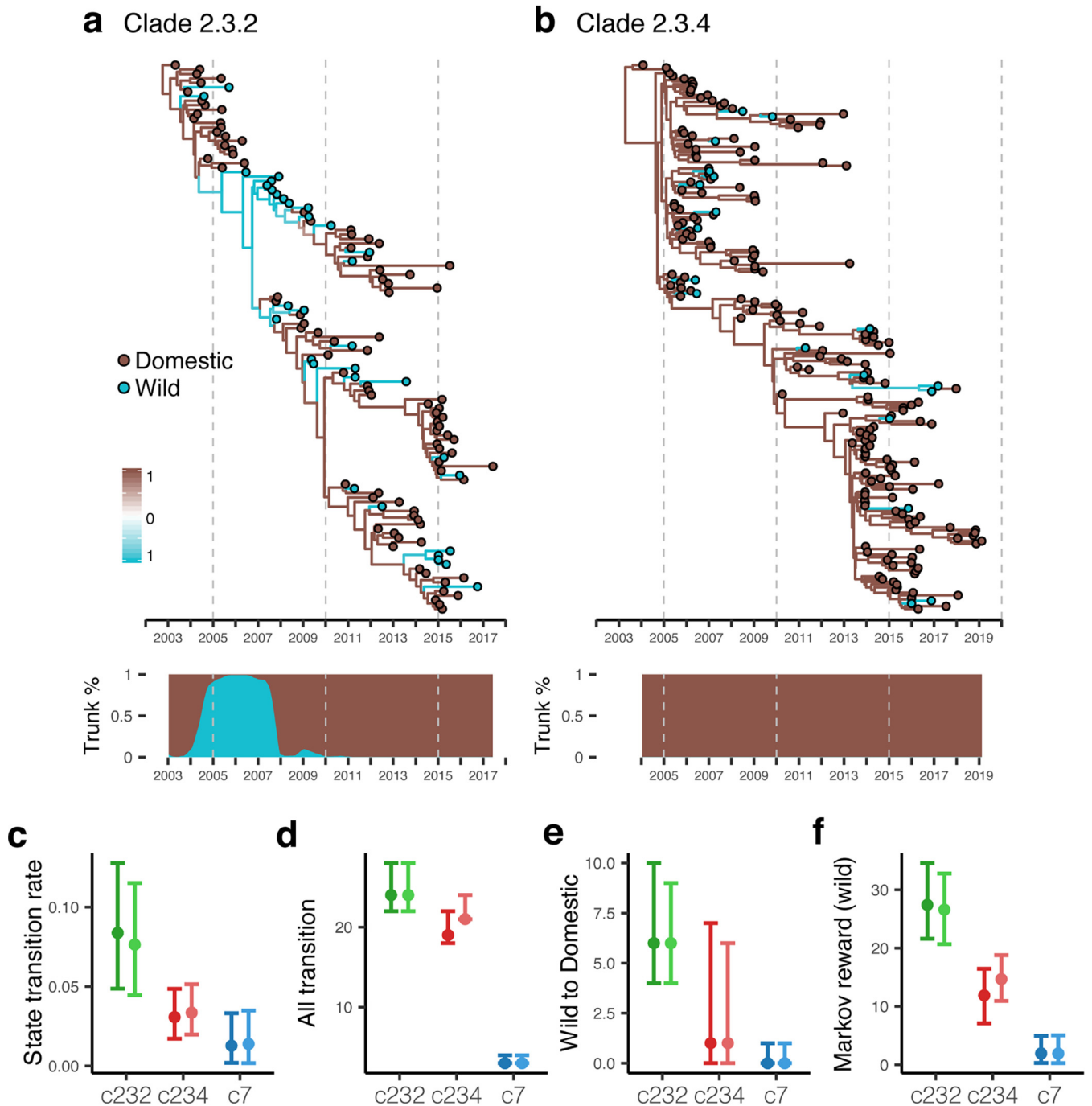


FIG 3 Gene flow of Gs/GD clades between ecological systems in China. MCC trees were reconstructed using HA genes of clade 2.3.2 ($n = 123$) (a) and 2.3.4 ($n = 201$) (b) viruses. Tips are colored according to their assigned ecological states, while internal branches are colored according to ancestral states inferred using a Bayesian phylogenetic framework. The shaded boxes indicate posterior probability values for both ecological states on the phylogenetic trees. Trunk proportions of phylogenetic trees occupied by the two ecological states summarized by PACT are shown below the trees. Overall transition rates (c), total counts of Markov jump (d), Markov jump from wild to domestic states (e), and Markov reward in the wild states (f) were estimated from the same Bayesian analyses conducted in panels a and b and also for clade 7 (see Fig. S7 in the supplemental material). Results using another random sampled data set allowing for greater sample sizes for each clade are shown in lighter shades. Error bars represent 95% highest posterior density (HPD) intervals.

a greater number of state transitions from wild to domestic ecosystems (Fig. 3e). This is also reflected by the Markov rewards (Fig. 3f), which indicate that clade 2.3.2 viruses have circulated in wild birds significantly longer (~ 27 years) than clades 2.3.4 and 7 viruses (~ 12 and ~ 2 years, respectively). Analysis of a larger H5-HA data set indicated that these results were not significantly affected by sample size. As such, clade 2.3.2

viruses demonstrated more extensive circulation in wild birds and more readily transmitted from wild birds to poultry. Overall, our analyses show that the wider ecological distribution of clade 2.3.2 viruses shielded them from the H5N1 virus population declines in poultry that led to the replacement of N1 viruses in clade 2.3.4 and 7 with non-N1 viruses (Fig. 2a; Fig. S1b).

DISCUSSION

In this study, we analyzed the population dynamics of the three major HPAI H5N1 virus sublineages in China. We show that the first non-N1 Gs/GD H5 viruses, i.e., H5Nx viruses, emerged after population bottlenecks were observed in H5N1 viruses, indicating that H5N1 population declines did not result from competition with H5Nx viruses. A scenario involving competitive exclusion entails limited host population and cocirculation for an extended duration (28), which is not the case considering the abrupt change in population size that we observed (Fig. 2) and the tremendous poultry population in China (see Fig. S8b in the supplemental material). We further show that clade 2.3.2 viruses were subject to a less severe population decline that is attributable to their circulation in wild bird populations during the bottleneck period, and that H6 viruses in poultry were not subject to any population bottleneck. Together, this data suggests that the implementation of comprehensive H5N1 control strategies in China in 2005 (11, 29) successfully suppressed H5N1 viruses in poultry directly before the emergence of H5Nx viruses.

Vaccination and aggressive culling (11, 29, 30) successfully reduced the number of H5 outbreaks in China until the emergence of H5Nx viruses (Fig. S8a) despite stable poultry production numbers (Fig. S8b) (31). Chinese national surveillance data (11) subsequently showed dramatic changes in the proportion of H5N1 virus clades detected in poultry that matches the results of our analysis. Specifically, in 2007, clade 2.3.4 viruses accounted for 100% of H5N1 viruses but fell to <35% from 2008 to 2013 before rebounding to over 92% from 2014 to 2018 (11). In contrast, the proportion of clade 2.3.2 viruses went from zero in 2007 to >51% from 2008 to 2013 before falling to <7% from 2014 to 2018. These dynamics are further supported by the absence of H5N1 in poultry markets in Hong Kong from 2004 to 2008, although viruses were isolated in wild birds (32). We therefore hypothesize that the widespread suppression of the H5N1 virus population in poultry provided susceptible hosts in which antigenically distinct H5Nx viruses (33, 34) could propagate and emerge. Recombinant H5N1 vaccines produce potent anti-HA and anti-NA antibody responses in chicken and other poultry (35, 36) but do not provide sterilizing immunity (37–39). The widespread use of H5N1 vaccines would therefore confer a fitness advantage to H5Nx viruses due to the antigenic mismatch of the NA. However, we cannot discount the possibility that H5Nx viruses were fixed stochastically in the H5 virus population due to the strong genetic drift caused by the observed bottlenecks (40).

Our study stresses the importance of intensified monitoring of pathogen populations during the application of control measures, particularly widespread vaccination programs that have the potential to lead to the emergence of antigenic escape (41, 42). This includes continued surveillance of related viruses in wildlife that may act as a reservoir for the reemergence of those pathogens (5, 43, 44). Vaccination programs against avian influenza viruses would benefit from universal vaccines that target a wider diversity of viruses, preventing the potential emergence of novel subtypes residing in poultry or wild bird populations.

MATERIALS AND METHODS

Data source and preparation. H5 and H6 hemagglutinin (HA) and associated neuraminidase (NA) sequences of avian influenza A viruses were downloaded from NCBI Influenza Virus Resource (<https://www.ncbi.nlm.nih.gov/genomes/FLU/>) and GISAID (<https://www.gisaid.org>) on 15 January 2020. Sequences downloaded were subjected to the following search criteria: (i) each viral strain required both HA and NA sequences, (ii) sequences had known host information, and (iii) lab-derived strains and mixed subtypes were excluded. Short sequences (0.95× the lengths of coding region) or sequences containing more than 5% ambiguous nucleotides were then removed. Curated data sets were aligned using MAFFT v7.407 (45) and trimmed to coding regions and additional multiple basic amino acids at the H5 cleavage site also removed.

Preliminary phylogenetic analyses and data set design. A maximum likelihood (ML) tree was inferred for the HA and NA alignments using IQ-TREE v1.6.12 (46) with a general time reversible (GTR) substitution model plus a gamma-distributed rate. Outliers generated by mislabeling or sequencing errors were detected using TempEst v1.5.3 (47) and discarded. Viral sublineages defined by HA genes, or clades, were classified according to World Health Organization Gs/GD H5N1 nomenclature (12) or previous studies of H6 viruses (22, 23). We define clades to include viruses in all descendant subclades, e.g., clade 7 includes viruses in clades 7.1 and 7.2. Due to their monophyletic relationship and small sample sizes, clade 2.3.1 and clade 2.3.3 viruses were combined with clade 2.3.2 and clade 2.3.4, respectively (12).

Identical sequences were removed from the analyses prior to any subsampling. To obtain an amenable number of sequences for Bayesian analyses and to avoid biases made by oversampled sequences presented by distinct outbreaks (48), we subsampled both the whole Gs/GD lineage and each clade uniformly with at most 25 sequences per year. Sequences isolated in the same animal or that clustered spatially were grouped, and a single representative sequence was randomly chosen in each year. HA sequences of H5 reference strains were included to maintain the topology of phylogenetic trees (12). From the total available clade 2.3.2 ($n = 276$), clade 2.3.4 ($n = 1,085$), and clade 7 ($n = 74$) nonredundant HA sequences in China, we produced subsampled data sets that contained 122 to 123, 201, and 55 sequences for the three clades, respectively. Likewise, a total of 250, 174, and 51 nonredundant N1 sequences in China for clades 2.3.2, 2.3.4, and 7, respectively, were subsampled to data sets of 135, 88, and 35 sequences. Larger data sets, containing up to 30 HA sequences per year, were used for the cross-ecosystem diffusion analyses; this sampling scheme increased the size of the clade 2.3.4 data set ($n = 223$) but not the sizes of data sets for clades 2.3.2 and 7. To examine the possible effect of uneven sampling efforts that may be present in surveillance data, we generated an H6 data set using an artificial sampling scheme in which no sequences were sampled in 2007 and only 15 sequences were selected in 2008 compared to up to 30 sequences being sampled in other years. The discontinuously sampled data sets contain 184 H6 sequences sampled from 535 nonredundant A/wild duck/Shantou/2853/2003-like sequences in China; the size is comparable to the typical data sets of H6 sequence generated by the methods applied to H5/N1 genes ($n = 182$).

Population dynamics. Changes in viral population size over time were estimated using the Skygrid demographic model (21) with an SRD06 nucleotide substitution model (49) and uncorrelated lognormal relaxed molecular clock (50) as implemented in BEAST v1.10.4 (20). For taxa without an associated day or month, date samplings were applied from uniform distributions within the known temporal bounds. Markov chain Monte Carlo (MCMC) analysis was run for 120 million steps and sampled every 10,000 steps. Tracer v1.7.1 (51) was used to inspect convergence of parameters (ESS values of >200) after discarding the first 20 million steps per MCMC chain. We tested the fit of the Skygrid model against a constant coalescent prior to our data sets by comparing the marginal likelihood using both path-sampling and stepping-stone approaches (52, 53) and found the Skygrid prior provided a better fit for all data sets (see Table S1 in the supplemental material). Maximum clade credibility (MCC) trees were summarized using TreeAnnotator (20). The annotated MCC trees were also used with the skygrowth model to calculate growth rates (54). To test the robustness of the estimates, time of the most recent common ancestor (tMRCA) of H5Nx was acquired in the internal node of the MCC tree containing all HA genes of clade 2.3.4 and from the root height of the MCC tree containing only HA genes of clade 2.3.4.4 (H5Nx). To obtain the tMRCA of clade 7 viruses, an arbitrary taxa group was set to include all H5N2 viruses in each data set. Combinations of either a strict or a relaxed molecular clock model were employed to further test the robustness of tMRCA estimates.

Discrete trait transition analyses. For analyses on the distribution of virus ecology, we classified each sequence as either wild or domestic according to strain name and information provided in the original database entry or publication. The sequences isolated from poultry (e.g., duck, chicken, goose) and human infections were classified as domestic state. For other animal species and environmental samples, their ecological states were designated based on the location of sample collection (see Fig. S1 in the supplemental material or source data at <https://github.com/yaotli/n1tonx> for these ecological states). Diffusion patterns between the two ecosystems were inferred using a discrete-state continuous time Markov chain (CTMC) model implemented in BEAST (27). We employed an asymmetric substitution model with Bayesian stochastic search variable selection (BSSVS) for transition parameters (27). Equal probability of the two states were assigned to sequences that were unable to be classified based on available information as described in a previous study (55). Number of transitions (Markov jumps) between the two states and the time spent in each state (Markov rewards) along the branches were recorded during MCMC processes (56, 57). Priors used in the diffusion analyses were identical to those in the Skygrid analyses. A subset of 1,000 posterior trees from each MCMC process was resampled by LogCombiner (20), and PACT v0.9.4 (<https://github.com/trvrb/PACT>) was utilized to summarize trunk proportion occupied by the two states. Trunks were defined as ancestral branches shared by sequences isolated within 0.1 year of the most recent samples. All trees were visualized with ggtree (58).

Population genetics statics. Tajima's D (24) and the derivative F_u and F_L 's D and F (26) were calculated using HA sequences isolated in every two consecutive years for each clade. We applied F_u and F_L 's D and F methods that are estimated without an outgroup in this study (i.e., D^* , F^*) (26). The 2.5% and 97.5% percentiles were calculated from 500 bootstrapped samples.

Data and code availability. Custom codes used in the data set preparation, details of state designation, and XML files required for BEAST are available at <https://github.com/yaotli/n1tonx>. Accession numbers of sequences analyzed in our study are listed in Table S2 in the supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 3.1 MB.

ACKNOWLEDGMENTS

The study was supported by the Duke-NUS Signature Research Program funded by the Ministry of Health, Singapore, and by contracts HHSN272201400006C and 75N93021C00016 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, USA.

The computational work for this article was partially performed using resources of the National Supercomputing Centre, Singapore (<https://www.nscg.sg>).

Y.-T.L. and G.J.D.S. conceived and designed the experiments; Y.-T.L. performed analysis and contributed analysis tools; Y.-T.L., Y.C.F.S., and G.J.D.S. wrote the manuscript. All authors reviewed and edited the paper.

We declare no conflicts of interest.

REFERENCES

- Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol Rev* 56:152–179. <https://doi.org/10.1128/mr.56.1.152-179.1992>.
- 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. <https://doi.org/10.1006/viro.1999.9820>.
- Guan Y, Smith GJ. 2013. The emergence and diversification of panzootic H5N1 influenza viruses. *Virus Res* 178:35–43. <https://doi.org/10.1016/j.virusres.2013.05.012>.
- World Health Organization. 2020. Antigenic and genetic characteristics of zoonotic influenza A viruses and development of candidate vaccine viruses for pandemic preparedness (February, 2020). World Health Organization, Geneva, Switzerland.
- Chen H, Smith GJ, Zhang SY, Qin K, Wang J, Li KS, Webster RG, Peiris JS, Guan Y. 2005. Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* 436:191–192. <https://doi.org/10.1038/nature03974>.
- Sonnberg S, Webby RJ, Webster RG. 2013. Natural history of highly pathogenic avian influenza H5N1. *Virus Res* 178:63–77. <https://doi.org/10.1016/j.virusres.2013.05.009>.
- Bahl J, Pham TT, Hill NJ, Hussein ITM, Ma EJ, Easterday BC, Halpin RA, Stockwell TB, Wentworth DE, Kayali G, Krauss S, Schultz-Cherry S, Webster RG, Webby RJ, Swartz MD, Smith GJD, Rungstadler JA. 2016. Ecosystem interactions underlie the spread of avian influenza A viruses with pandemic potential. *PLoS Pathog* 12:e1005620. <https://doi.org/10.1371/journal.ppat.1005620>.
- Trovao NS, Suchard MA, Baele G, Gilbert M, Lemey P. 2015. Bayesian inference reveals host-specific contributions to the epidemic expansion of influenza A H5N1. *Mol Biol Evol* 32:3264–3275. <https://doi.org/10.1093/molbev/msv185>.
- Lu L, Brown AJL, Lycett SJ. 2017. Quantifying predictors for the spatial diffusion of avian influenza virus in China. *BMC Evol Biol* 17:16. <https://doi.org/10.1186/s12862-016-0845-3>.
- World Health Organization. 2020. (Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003–2020 (8 May 2020). World Health Organization, Geneva, Switzerland.
- Liu S, Zhuang Q, Wang S, Jiang W, Jin J, Peng C, Hou G, Li J, Yu J, Yu X, Liu H, Sun S, Yuan L, Chen J. 2020. Control of avian influenza in China: strategies and lessons. *Transbound Emerg Dis* 67:1463–1471. <https://doi.org/10.1111/tbed.13515>.
- Smith GJ, Donis RO, World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization (WHO/OIE/FAO) H5 Evolution Working Group. 2015. Nomenclature updates resulting from the evolution of avian influenza A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013–2014. *Influenza Other Respir Viruses* 9:271–276. <https://doi.org/10.1111/irv.12324>.
- Vijaykrishna D, Bahl J, Riley S, Duan L, Zhang JX, Chen H, Peiris JSM, Smith GJD, Guan Y. 2008. Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. *PLoS Pathog* 4:e1000161. <https://doi.org/10.1371/journal.ppat.1000161>.
- Bi Y, Chen Q, Wang Q, Chen J, Jin T, Wong G, Quan C, Liu J, Wu J, Yin R, Zhao L, Li M, Ding Z, Zou R, Xu W, Li H, Wang H, Tian K, Fu G, Huang Y, Shestopalov A, Li S, Xu B, Yu H, Luo T, Lu L, Xu X, Luo Y, Liu Y, Shi W, Liu D, Gao GF. 2016. Genesis, evolution and prevalence of H5N6 avian influenza viruses in China. *Cell Host Microbe* 20:810–821. <https://doi.org/10.1016/j.chom.2016.10.022>.
- Saito T, Tanikawa T, Uchida Y, Takemae N, Kanehira K, Tsunekuni R. 2015. Intracontinental and intercontinental dissemination of Asian H5 highly pathogenic avian influenza virus (clade 2.3.4.4) in the winter of 2014–2015. *Rev Med Virol* 25:388–405. <https://doi.org/10.1002/rmv.1857>.
- Global Consortium for H5N8 and Related Influenza Viruses. 2016. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science* 354:213–217. <https://doi.org/10.1126/science.aaf8852>.
- Taubenberger JK, Morens DM. 2017. H5Nx panzootic bird flu—influenza's newest worldwide evolutionary tour. *Emerg Infect Dis* 23:340–342. <https://doi.org/10.3201/eid2302.161963>.
- Chen H, Smith GJD, Li KS, Wang J, Fan XH, Rayner JM, Vijaykrishna D, Zhang JX, Zhang LJ, Guo CT, Cheung CL, Xu KM, Duan L, Huang K, Qin K, Leung YHC, Wu WL, Lu HR, Chen Y, Xia NS, Naipospos TSP, Yuen KY, Hassan SS, Bahri S, Nguyen TD, Webster RG, Peiris JSM, Guan Y. 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. <https://doi.org/10.1073/pnas.0511120103>.
- Liu C-G, Liu M, Liu F, Lv R, Liu D-F, Qu L-D, Zhang Y. 2013. Emerging multiple reassortant H5N5 avian influenza viruses in ducks, China, 2008. *Vet Microbiol* 167:296–306. <https://doi.org/10.1016/j.vetmic.2013.09.004>.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. <https://doi.org/10.1093/ve/vey016>.
- Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. 2013. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol* 30:713–724. <https://doi.org/10.1093/molbev/mss265>.
- Huang K, Zhu H, Fan X, Wang J, Cheung C-L, Duan L, Hong W, Liu Y, Li L, Smith DK, Chen H, Webster RG, Webby RJ, Peiris M, Guan Y. 2012. Establishment and lineage replacement of H6 influenza viruses in domestic ducks in southern China. *J Virol* 86:6075–6083. <https://doi.org/10.1128/JVI.06389-11>.
- Li J, Quan C, Xie Y, Ke C, Nie Y, Chen Q, Hu T, Chen J, Wong G, Wang Q, Feng L, Yu H, Liu Y, Liu W, Gao GF, Liu WJ, Shi W, Bi Y. 2019. Continued reassortment of avian H6 influenza viruses from Southern China, 2014–2016. *Transbound Emerg Dis* 66:592–598. <https://doi.org/10.1111/tbed.13037>.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709. <https://doi.org/10.1093/genetics/133.3.693>.

27. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5:e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>.
28. Bahl J, Vijaykrishna D, Holmes EC, Smith GJ, Guan Y. 2009. Gene flow and competitive exclusion of avian influenza A in natural reservoir hosts. *Virology* 390:289–297. <https://doi.org/10.1016/j.virol.2009.05.002>.
29. Chen H. 2009. H5N1 avian influenza in China. *Sci China C Life Sci* 52: 419–427. <https://doi.org/10.1007/s11427-009-0068-6>.
30. Wan X-F, Dong L, Lan Y, Long L-P, Xu C, Zou S, Li Z, Wen L, Cai Z, Wang W, Li X, Yuan F, Sui H, Zhang Y, Dong J, Sun S, Gao Y, Wang M, Bai T, Yang L, Li D, Yang W, Yu H, Wang S, Feng Z, Wang Y, Guo Y, Webby RJ, Shu Y. 2011. Indications that live poultry markets are a major source of human H5N1 influenza virus infection in China. *J Virol* 85:13432–13438. <https://doi.org/10.1128/JVI.05266-11>.
31. Food and Agriculture Organization of the United Nations. 2020. FAOSTAT (April, 2020). Food and Agriculture Organization of the United Nations, Rome, Italy.
32. Smith GJD, Vijaykrishna D, Ellis TM, Dyrting KC, Leung YHC, Bahl J, Wong CW, Kai H, Chow MKW, Duan L, Chan ASL, Zhang LJ, Chen H, Luk GSM, Peiris JSM, Guan Y. 2009. Characterization of avian influenza viruses A (H5N1) from wild birds, Hong Kong, 2004–2008. *Emerg Infect Dis* 15: 402–407. <https://doi.org/10.3201/eid1503.081190>.
33. World Health Organization. 2015. Antigenic and genetic characteristics of zoonotic influenza viruses and development of candidate vaccine viruses for pandemic preparedness (February, 2015). World Health Organization, Geneva, Switzerland.
34. World Health Organization. 2016. Antigenic and genetic characteristics of zoonotic influenza viruses and development of candidate vaccine viruses for pandemic preparedness (February, 2016). World Health Organization, Geneva, Switzerland.
35. Sylte MJ, Hubby B, Suarez DL. 2007. Influenza neuraminidase antibodies provide partial protection for chickens against high pathogenic avian influenza infection. *Vaccine* 25:3763–3772. <https://doi.org/10.1016/j.vaccine.2007.02.011>.
36. Nayak B, Kumar S, DiNapoli JM, Paldurai A, Perez DR, Collins PL, Samal SK. 2010. Contributions of the avian influenza virus HA, NA, and M2 surface proteins to the induction of neutralizing antibodies and protective immunity. *J Virol* 84:2408–2420. <https://doi.org/10.1128/JVI.02135-09>.
37. Swayne DE, Pavade G, Hamilton K, Vallat B, Miyagishima K. 2011. Assessment of national strategies for control of high-pathogenicity avian influenza and low-pathogenicity notifiable avian influenza in poultry, with emphasis on vaccines and vaccination. *Rev Sci Tech* 30:839–870. <https://doi.org/10.20506/rst.30.3.2081>.
38. Sims LD. 2012. Progress in control of H5N1 highly pathogenic avian influenza and the future for eradication. *Avian Dis* 56:829–835. <https://doi.org/10.1637/10198-041012-Review.1>.
39. Capua I. 2007. Vaccination for notifiable avian influenza in poultry. *Rev Sci Tech* 26:217–227. <https://doi.org/10.20506/rst.26.1.1741>.
40. Gutierrez S, Michalakos Y, Blanc S. 2012. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol* 2:546–555. <https://doi.org/10.1016/j.coviro.2012.08.001>.
41. Read AF, Baigent SJ, Powers C, Kgosana LB, Blackwell L, Smith LP, Kennedy DA, Walkden-Brown SW, Nair VK. 2015. Imperfect vaccination can enhance the transmission of highly virulent pathogens. *PLoS Biol* 13: e1002198. <https://doi.org/10.1371/journal.pbio.1002198>.
42. Sjaak de Wit JJ, Cook JK, van der Heijden HM. 2011. Infectious bronchitis virus variants: a review of the history, current situation and control measures. *Avian Pathol* 40:223–235. <https://doi.org/10.1080/03079457.2011.566260>.
43. Dudas G, Rambaut A. 2014. Phylogenetic analysis of Guinea 2014 EBOV ebolavirus outbreak. *PLoS Curr* 6:currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d. <https://doi.org/10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d>.
44. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, Liu R, He X, Shuai L, Sun Z, Zhao Y, Liu P, Liang L, Cui P, Wang J, Zhang X, Guan Y, Tan W, Wu G, Chen H, Bu Z. 2020. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 368:1016–1020. <https://doi.org/10.1126/science.abb7015>.
45. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
46. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
47. Rambaut A, Lam TT, Carvalho LM, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2:vev007. <https://doi.org/10.1093/ve/vev007>.
48. Hall MD, Woolhouse ME, Rambaut A. 2016. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: a simulation study. *Virus Evol* 2:vev003. <https://doi.org/10.1093/ve/vev003>.
49. 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. <https://doi.org/10.1093/molbev/msj021>.
50. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88. <https://doi.org/10.1371/journal.pbio.0040088>.
51. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67: 901–904. <https://doi.org/10.1093/sysbio/syy032>.
52. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29:2157–2167. <https://doi.org/10.1093/molbev/mss084>.
53. Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 30:239–243. <https://doi.org/10.1093/molbev/mss243>.
54. Volz EM, Didelot X. 2018. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Syst Biol* 67:719–728. <https://doi.org/10.1093/sysbio/syy007>.
55. Scotch M, Tahsin T, Weissenbacher D, O'Connor K, Magge A, Viente M, Suchard MA, Gonzalez-Hernandez G. 2019. Incorporating sampling uncertainty in the geospatial assignment of taxa for virus phylogeography. *Virus Evol* 5:vey043. <https://doi.org/10.1093/ve/vey043>.
56. Minin VN, Suchard MA. 2008. Counting labeled transitions in continuous-time Markov models of evolution. *J Math Biol* 56:391–412. <https://doi.org/10.1007/s00285-007-0120-8>.
57. Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci* 363:3985–3995. <https://doi.org/10.1098/rstb.2008.0176>.
58. Yu G, Lam TT, Zhu H, Guan Y. 2018. Two methods for mapping and visualizing associated data on phylogeny using Ggtree. *Mol Biol Evol* 35: 3041–3043. <https://doi.org/10.1093/molbev/msy194>.