Spatiotemporal Genotype Replacement of H5N8 Avian Influenza Viruses Contributed to H5N1 Emergence in 2021/2022 Panzootic

Spatio-temporal Genotype Replacement of H5N8 Avian Influenza Viruses Contributed to H5N1 Emergence in 2021/2022 Panzootic

Jinfeng Zeng,[a,b*] Fanshu Du,[c*] Linna Xiao,[d] Honglei Sun,[c] Lu Lu,[e] Weipan Lei,[d] Jialu Zheng,[a,b] Lu Wang,[c] Sicheng Shu,[c] Yudong Li,[c] Qiang Zhang,[f] Kang Tang,[a,b] Qianru Sun,[a,b] Chi Zhang,[a,b] Haoyu Long,[a,b] Zekai Qiu,[a,b] Ke Zhai,[a,b] Zhichao Li,[f] Geli Zhang,[g] Yipeng Sun,[c] Dayan Wang,[h] Zhengwang Zhang,[d] Samantha J. Lycett,[e] George F. Gao,[i] Yuelong Shu,[a,j] Jinhua Liu,[c] Xiangjun Du,[a,b,k]# Juan Pu[c]#

[a]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, P.R. China.

[b]School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, P.R. China.

[c]National Key Laboratory of Veterinary Public Health and Safety, Key Laboratory for Prevention and Control of Avian Influenza and Other Major Poultry Diseases, Ministry of Agriculture and Rural Affairs, College of Veterinary Medicine, China Agricultural University, Beijing 100193, P.R. China.

[d]Key Laboratory for Biodiversity Science and Ecological Engineering, Demonstration Center for Experimental Life Sciences & Biotechnology Education, College of Life Sciences, Beijing Normal University, Beijing 100875, P. R. China.

[e]The Roslin Institute, University of Edinburgh, EH25 9RG Edinburgh, United Kingdom.

[f]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, P. R. China.

[g]College of Land Science and Technology, China Agricultural University, Beijing 100193, P. R. China.

[h]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, P.R. China.

[i]CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, P. R. China.

[j]Institute of Pathogen Biology of Chinese Academy of Medical Science (CAMS)/Peking Union Medical College (PUMC), Beijing 100000, P.R. China.

[k]Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou 510000, P.R. China.

Running Head: Spatio-temporal spread of H5Ny in 20-22 pandemic

#Address correspondence to Juan Pu, pujuan@cau.edu.cn and Xiangjun Du, duxj9@mail.sysu.edu.cn.

*These authors contributed equally to this work. Author order was determined on the basis of seniority.

Abstract: 156 words. Main text: 6787 words.

ABSTRACT

Since 2020, clade 2.3.4.4b highly pathogenic avian influenza (HPAI) H5N8 and H5N1 viruses have swept through continents, posing serious threats to the world. Through comprehensive analyses of epidemiological, genetic, and bird migration data, we found that the dominant genotype replacement of the H5N8 viruses in 2020 contributed to the H5N1 outbreak in the 2021/2022 wave. The 2020 outbreak of H5N8 G1 genotype instead of G0 genotype produced reassortment opportunities and led to the emergence of a new H5N1 virus with G1's HA and MP genes. Despite extensive reassortments in the 2021/2022 wave, the H5N1 virus retained the HA and MP genes, causing a significant outbreak in Europe and North America. Furtherly, through the wild bird migration flyways investigation, we found that the temporal-spatial coincidence between the outbreak of H5N8 G1 virus and the bird autumn migration may have expanded H5 viral spread, which maybe one of the main drivers of the emergence of 2020-2022 H5 panzootic. ~~We further found that viral evolution in poultry of Egypt and surrounding area and bird migration from Russia-Kazakhstan region to Europe are important drivers in the emergence of the 2020-2022 H5 panzootic.~~

IMPORTANCE

Since 2020, highly pathogenic avian influenza (HPAI) H5 subtype variants of clade 2.3.4.4b have spread across continents, posing unprecedented threats globally. However, the factors promoting the genesis and spread of H5 HPAI viruses remain unclear. Here we found that the spatio-temporal genotype replacement of H5N8 HPAI viruses contributed to the emergence of H5N1 variant that caused the 2021/2022 panzootic, and viral evolution in poultry of Egypt and surrounding area and autumn bird migration from Russia-Kazakhstan region to Europe are important drivers of the

emergence of the 2020-2022 H5 panzootic. These findings provide important targets for early warning and could help control the current and future HPAI epidemics.

KEYWORDS Avian influenza virus (AIV), H5N1, H5N8, genesis, spread, bird migration

MAIN TEXT

The ongoing avian influenza virus (AIV) H5N1 outbreak is the largest ever recorded and is affecting wild birds, poultry, mammals, and humans across multiple continents. In Europe, the 2021-2022 epidemic has resulted in a total of 2,520 HPAI outbreaks in poultry, with 50 million birds dead or culled, 227 findings in captive birds, 3,867 HPAI virus detections in wild birds, and 1 human case(1). In the United States, between January 2022 and 15 March 2023, more than 58.6 million domestic birds were affected, along with 6,444 wild bird detections and 1 human case(2, 3). The origin of the virus and how to control its spread are urgent concerns in light of this unprecedented outbreak.

Since the first isolation of the HPAI virus [A/Goose/Guangdong/1/96 (H5N1)] belonging to GS/GD lineage was detected in China in 1996, H5 HPAI has undergone rapid genetic divergence and genetic reassortment, resulting in the emergence of numerous clades and subclades(4–6). Over the years, there have been six waves of intercontinental transmissions caused by the GS/GD lineage viruses through migratory birds(7–10). In 2005-2006, clade 2.2 H5N1 virus spread out from Qinghai Lake of China to other countries in Europe and Africa(11, 12). In 2009-2010, clade 2.3.2.1 H5N1 virus affected Asia and Europe(13). Subsequently, clade 2.3.4.4a and 2.3.4.4b of H5N8 viruses caused multiple waves of worldwide epidemics in 2014-2015 and 2016-2021, respectively(10, 14, 15). Following that, a novel H5N1 virus of clade 2.3.4.4b has emerged, circulating in Europe, North America, South America, Africa, and Asia(16, 17), causing the most lasting and devastating threats to the world.

The clade 2.3.4.4b H5N8 outbreaks were first identified in Qinghai Lake of China during the epidemic season of 2016/2017, and rapidly spread to almost all of Eurasia(15). However, in 2020-2021, the clade 2.3.4.4b H5N8 caused an unexpected epidemic peak(18, 19). In Europe alone, the

virus caused 1,298 outbreaks in poultry, affecting 229 million domestic birds, 85 detections in captive birds, and 2,294 HPAI events in wild birds(20). Moreover, during this wave, the World Health Organization (WHO) reported the first human infections with HPAI H5N8 viruses(21). In the following epidemic season of 2021/2022, the H5N1 subtype replaced the H5N8 and triggered a larger panzootic(16, 22), affecting more species. However, there are still some key questions related to these outbreaks that remain unclear. Specifically, it is unclear how the successive outbreaks of H5N8 and H5N1 are related, as well as what factors contribute to the two outbreaks. Addressing these questions is significant for the control and prevention of the potential pandemic.

In this study, we collected all available viral genome sequences of H5N8 and H5N1 that were sampled between October 2016 and September 2022, along with other related sequences before this time period. We then used Bayesian phylodynamic modeling to investigate the genetic evolution and spread of HPAI H5N8 and H5N1 viruses during the epidemic seasons of 2020/2021 and 2021/2022, and incorporated epidemiological data and bird migratory data to independently assess the consistency.

RESULTS

H5 HPAI viruses initiated the biggest outbreak in 2020-2022 after a quiet period

Our analysis of the epidemiological data from FAO Empres-i revealed four primary subtypes of H5N1, H5N2, H5N6, and H5N8 viruses prevalent in Europe, America, Asia, and Africa. H5N1 and H5N8 caused three epidemics since 2016 (Fig. 1A and F), with H5N8 causing the first wave in Europe, Africa, and Asia from October 2016 to September 2017 (2016/2017 Wave). In 2017/2018, 2018/2019 and 2019/2020, the period looks like a quiet stage as there weren’t massive reported outbreaks. After that, the H5N8 initiated the second wave in same continents in

2020/2021. However, another subtype of H5N1 replaced the H5N8 virus in the following 2021/2022 wave, causing a wider outbreak across Eurasia, Africa, and North America (Fig. 1B-E). Prior to the 2020/2021 wave, there was a low-level prevalence caused by H5N8 virus in 2019/2020 wave (Fig. 1A and C). Among the three major epidemics of 2016/2017, 2020/2021 and 2021/2022 waves, there were outbreaks occurred in domestic poultry and wild birds, with wild birds contributing to a greater percentage of the outbreaks in the later waves (Fig. 1F, Table S1). A more detailed case and death data were further summarized by waves, showing a growth trend across waves in wild birds (Fig. 1G). Human cases infected with HPAI H5Ny also peaked in 2020-2022 (Fig. 1H). The results indicated that the 2021/2022 wave of H5N1 was the largest and most devastating HPAI epidemic. Our analysis raises questions about why the H5N8 virus was able to re-initiate a big outbreak in 2020/2021 and whether this outbreak contributed to the subsequent H5N1 panzootic.

The genomes of H5N8 and H5N1 viruses in 2020-2022 are divergent from other clade 2.3.4.4b viruses

To explore the factors inducing the 2020/2021 H5N8 and 2021/2022 H5N1 waves, a phylogenetic analysis of HA gene in clade 2.3.4.4b H5Ny viruses was performed using a full genome dataset composed of 3781 HPAI H5Ny viruses collected around the world between 2016 and 2022 (Fig. 2A). We further calculated the pairwise identity of eight gene segments of all sequences and visualized the results using heatmap (Fig. 2B-I). Our analysis of the reconstructed phylogenetic tree of the HA gene revealed that the 2016/2017 wave of HPAI H5N8 virus had high gene homology (98.75%, Table S2) but different geographical distributions. Some were mainly found in Asia and Africa, while others mainly in Europe and Africa (Fig. 2A). The HA gene of 2019/2020 and 2020/2021 waves of H5N8 viruses formed from two distinct groups (Fig. 2A). The 2019/2020

viruses diverged from those isolated in Asia and Africa during 2016/2017 wave (with 97.2% identity, Table S2), and 2020/2021 viruses mainly diverged from those isolated in Europe and Africa in 2016/2017 wave (with 97.38% identity, Table S2). However, both two groups have a relatively long leading branch connecting the viruses of 2016/2017 wave. Apart from HA, the two groups of viruses also have different sequences from those of 2016/2017 wave in other seven segments (76.79% to 97.83% identity, Table S2, Fig. 2B-I). Therefore, a potential hypothesis is that the 2019/2020 and 2020/2021 waves were caused by two new H5N8 variants differently from the initial clade 2.3.4.4b viruses. However, the H5N1 virus of the 2021/2022 wave had high identities with those of H5N8 2020/2021 wave, in HA (98.29%, Table S2, Fig. 2E) and MP (98.77%, Table S2, Fig. 2H) genes, but low identities (68.85%-95.14%, Table S2) with those viruses in other genes (Fig. 2B-I). So, the consecutive emergent 2020/2021 H5N8 and 2021/2022 H5N1 viruses had divergent genomes from those of the previous waves, but they shared similar HA and MP genes with each other.

H5N1 virus in panzootic acquired the HA and MP genes from H5N8 HPAI viruses and other six genes of diverse LPAI viruses

To investigate the genetic origins of H5N1 virus in the 2021/2022 wave, we constructed phylogenetic trees using data from eight segments (Fig. S1-S9). The results showed the earliest H5N1 virus strain isolated in Europe and in North America (A/Eurasian wigeon/Netherlands/1/2020 and A/Fancy chicken/NL/FAV-0035/2021, respectively) were grouped together with the H5N8 virus of the 2020/2021 wave in Europe in the trees of HA and MP genes (Fig. S1, S5 and S8), but were separated from the H5N8 virus of the 2019/2020 wave and earlier ones (Data S1 and Fig. 2). In the trees of the other six genes, the two viruses exhibited the closest genetic relationship with diverse LPAI wild bird strains from Eurasia. Subsequently,

some H5N1 viruses isolated in the United States underwent further reassortment with the local LPAI strains. Specifically, all eight segments of the early H5N1 strains of North America clustered with European H5N1, while late North American H5N1 strains were rooted with North American LPAIVs, especially in PB2 and NP segments (Fig. S2-S9). North American-like PB2 and NP strains were predominantly found in the north-central United States, such as Minnesota, South Dakota, North Dakota, and Iowa in March and April of 2022, whereas other strains with European-like genes were distributed mostly in March of the eastern seaboard of the States, such as Maine and North Carolina (Fig. S2 and S6). Overall, the results suggested that European H5N1 strains first arrived on the east coast of the North American continent and then gradually speared westwards, acquiring local LPAIV genes to form more reassortants. Notably, regardless of whether they were European H5N1 or later North American H5N1 strains, they consistently retained the HA and MP genes originating from the 2020/2021 H5N8 virus, suggesting these two genes might have contributed to the emergence and further transmission of the H5N1 virus.

Dominant genotype change in H5N8 virus drives the 2020/2021 wave

To identify the specific H5N8 virus that contributed the HA and MP genes to the H5N1 virus, we analyzed the genotypes of the earlier H5N8 virus isolated during the 2019/2020 and 2020/2021 waves (Fig. 3A). The sequences were divided into 3 to 12 distinct phylogenetic monophyletic groups for each gene segment (Fig. S10 and S11) and 22 genotypes were identified for all HPAI H5N8 viruses, including two major genotypes (G0, G1, dominate circulate genotypes), nine minor genotypes (reassortant genotypes of major), and eleven other transient genotypes (with one sampled isolate, Fig. 3A, Table S3). Notably, we found that the strains providing the HA and MP genes for the H5N1 virus belonged to the H5N8 G1 genotype (Fig. S5 and S8, Data S1).

Temporal dynamic analysis revealed that G0 and G1 genotypes have successively dominated the HPAI H5N8 epidemic in the Eurasian continent since 2020 (Fig. 3B and C, Fig. S12). G0 dominated in the first phase, from January 2020 to June 2020, primarily affecting domestic poultry in the middle and eastern Europe (Fig. 3B and D, Fig. S12). However, during the low epidemic period from June to August 2020, the distinct genome set of G1 replaced G0 and dominated the following 2020/2021 wave (Fig. 3B and C). G1 affected both domestic poultry and wild birds for a longer time, from August 2020 to December 2021, and showed a higher epidemic scale in terms of increased number of outbreaks, species of wild birds and geographic distribution (Fig. 3B and E, Fig. S12). Furthermore, reassortment during this epidemic generated the diversified G1-like genotypes (Fig. 3B and C, Table S3), increasing the likelihood of G1 providing genes to the novel H5N1 virus. The dominant replacement of G0 by G1 was a prerequisite for the 2021/2022 H5N1 wave.

H5N8 virus circulating in Egypt and surrounding area since 2018 identified as the ancestor of the dominant G1 in 2020/2021 wave

To investigate how the H5N8 re-emerged and became dominant in the Eurasian continent after a long quiet period, we compared the evolution history of the two distinct genotypes of G0 and G1 using phylogeographic continuous diffusion model. Our findings for G0 revealed the most recent common ancestor (MRCA) of all gene segments emerged in Central Europe (Fig. S13A-H). However, the most recent common grand ancestor (MRCGA, first recent ancestor node of MRCA) had six segments originating from South Africa or Western Africa (PB2, PA, HA, NA, MP, NS, Fig. S13 A, C-D, F-H), and two from Eurasia (PB1, NP, Fig. S13B and E).

The inferred time of MRCA (TMRCA) for the genotype G0 was October 2019 [95% HPD: August 2019 to November 2019], while the inferred time of MRCGA (TMRCGA) was February 2019

[November 2018 to May 2019] (Fig. S13I). The likely hosts of MRCA and MRCGA for almost all gene segments were domestic *Anseriformes* or domestic *Galliformes*, except for PB1 and NP genes, which the MRCGA were inferred to have waterfowls as hosts (Fig. S13J). Therefore, it is possible that the HPAI H5N8 viruses circulating in Western African and South African poultry provided the backbone genes (6 of 8 gene segments) for the genotype G0 at the beginning of 2019. The ancestor of genotype G0 emerged by reassorting the PB1 and NP gene segments of LPAI viruses from Eurasian waterfowls at the wintering sites of middle Europe around October 2019 (Fig. S13K). Following that, the G0 strain A/turkey/Poland/23/2019 was firstly isolated in Poland and caused a peak infection from January to April in 2020.

The location of MRCA for all segments of G1 was traced to a larger area around the border of Russia-Kazakhstan, which acted as a crossroads between Europe and Asia (Fig. S14A-H). However, nearly all segments of MRCGA for G1 were estimated from Egypt and surrounding area, especially the PB1, PA, HA and NP segments are mainly located in Egypt (Fig. S14A-F, S14H). The only exception is MP segment, which was most likely located in northwest China (Fig. S14G). The TMRCA for G1 was estimated to be March 2020 [December 2019 to May 2020], similar to G0. But the TMRCGA for G1 was estimated to be much earlier, in May 2018 [May 2017 to November 2018] (Fig. S14I). Except for the MP gene, whose host is unclear (domestic *Anseriformes* or waterfowl), the other segments of MRCA and MRCGA are all presumed to be domestic *Anseriformes* or domestic *Galliformes* (Fig. S14J). Similarly, it is possible that the HPAI H5N8 virus endemic in poultry of Egypt and surrounding area provided the backbone genes (7 of 8 gene segments) for the genotype G1. The ancestor of genotype G1 emerged at the wintering sites along the Asia-Europe border around December 2019, by reassorting MP gene segment of HPAI

H5N8 viruses from Xinjiang, China (Fig. S14K). Then, the first G1 strain A/chicken/Iraq/1/2020 was isolated in Iraq in May 2020 and caused an outbreak peak from December 2020 to April 2021.

Thus, in the "quiet" period, H5N8 viruses kept evolving and spreading. Both G0 and G1 were probably generated by reassortment with genes from poultry and wild birds in the autumn of 2019. Their poultry-origin genes are both from Africa, but those of G0 are from West and South Africa, while those of G1 are from Egypt in North Africa and surrounding area. G0 emerged in middle Europe, while G1 emerged in the Asia-Europe border. These results suggest that Egypt and surrounding area is an important source for genetic diversification of avian influenza viruses, from where the ancestor G1 virus was generated as early as 2018, which then spread to neighboring continents of Asia and Europe through migratory birds.

Temporal-spatial coincidence between outbreaks and bird autumn migration expanded H5 HPAI viral spread

To investigate the factors influencing the spread of H5 HPAI virus, we examined the impact of migratory birds, given their geographical relationship to the viral gene source and epidemic scale. Using continuous phylogeographic analysis, we reconstructed the spatial-temporal spread routes of G0 and G1 H5N8 genotypes. We then collected 108 species of migratory birds which were reported to have been infected by HPAI H5N8 viruses, including 29 species with detailed migration data (Table S4). The migratory flyways, breeding sites, and wintering sites of these birds were collected and summarized in Fig. S15. By analyzing the collected spread routes, migration patterns, as well as sampling locations of viruses, we gained a unique perspective on the role of migratory birds in the spread dynamic of HPAI H5N8 viruses.

As shown in Fig. S15B, the breeding sites of these migratory birds cover most areas of Eurasian continent, while their wintering sites cover nearly all of Africa continent and southern part of the

Eurasian continent. Some areas in Europe, Middle East, and eastern Asia have both breeding and wintering sites (Fig. S15B). The birds migrate along eight major flyways that connect Europe, Asia, and Africa, including four flyways from southern to northern Asia, two from Africa to Europe, one from Africa to Asia, and one from Europe to Asia (Fig. S15B).

As shown in Fig. 4A and C, the spatially explicit phylogeographic analysis revealed the inferred transmission dynamic of genotype G0 and G1 since their MRCGA phase. By mapping the wild bird migration pattern onto the sampling locations of viruses (Fig. 4B and D), we found that the viral transmission matched well with the flyways of wild birds (Fig. 4). Specifically, for G0 viruses, ancestral viruses were transmitted from the wintering sites in South Africa and western Africa to the breeding sites in Europe through northward migration between March and June 2019 (Fig. 4A and 4B, the two green arrows below). Genotype G0 was produced in Europe in the winter of 2019 (Fig. 4A) and spread to northern breeding areas (Fig. 4B, the green arrows up). G0 viruses initiated local outbreaks in Europe during the spring of 2020 (Fig. 4A) and were soon spread to the breeding sites in the Arctic through bird autumn migration (Fig. 4B, the purple arrows). Some of the G0 viruses were transmitted to East Asia (Fig. 4B, the dark red dots) during the autumn migration in 2020 but were not isolated in Europe again (Fig. 4B, no dark red dots there).

For G1 viruses, the ancestors of genotype G1 viruses were spread via the routes of Egypt, Middle East, Kazakhstan, and Russia during the northward migration (Fig. 4C). Some G1 viruses circulated at the breeding sites near the border of Russia-Kazakhstan, where they spilled over to nearby poultry and amplified (Fig. 4D, the green arrows and the light blue dots). Other viruses might have continued north to the Arctic. The time and place of the outbreaks in Russia-Kazakhstan coincided with the ongoing autumn migration of 2020, which made G1 viruses easily diffuse into different directions across the Eurasian continent, including westward to Europe and

southward to eastern Asia, and caused expanding transmission (Fig. 4D, the purple arrows and the dark blue dots). From another perspective, discrete phylogeographic analysis of host traits further confirmed these findings. The first wave of HPAI H5N8 epidemic caused by G0 viruses in the spring of 2020 in Europe, was mainly restricted in domestic poultry. However, a shift of trunk probability from domestic poultry to wild birds was observed in the second wave (including G0 and G1 viruses), which was in coincidence with the widely spread of 2020-2021 H5N8 epidemic (Fig. S16).

Therefore, the spring migration of wild birds in 2020 may initiate the spread of G0 from Europe to the Arctic, while the autumn migration of 2020 may facilitate the spread of G1 from Russia-Kazakhstan region to Eurasia, resulting in contrasting transmission effects, with G0 contracting and G1 dispersing. Collectively, the spatial and temporal coincidence between autumn migration of wild birds and the G1 outbreak greatly promoted wide wider spread, particularly in Europe during the 2020/2021 wave, as observed in Fig. 5A and B. The extensive outbreak in Europe further created numerous opportunities for reassortment, leading to the emergence of the H5N1 virus (Fig. 5A and B). In October 2020, the first H5N1 reassortant carrying H5N8 original genes was detected in wild birds in the Netherlands (Fig. S1-S9), and subsequently, more viruses were isolated in other European countries, resulting in a larger outbreak (Fig. 5B). By late 2021, the H5N1 virus had spread further via wild birds from Northwestern Europe to the East coast of North America (Fig. 5B). Thus, the temporal-spatial coincidence between the outbreak and the bird autumn migration expanded H5N8 viral spread from Russia-Kazakhstan region to Eurasia and contributed to the emergence of the H5N1 virus in Europe, which was later introduced to North America.

DISCUSSION

Among multiple clades of H5Ny viruses causing intercontinental spread, the clade 2.3.4.4b H5 subtype HPAI virus caused the longest and most serious epidemic in wild birds and poultry. Here, we explained the genesis and spread of 2020-2022 H5 HPAI viruses, and revealed the important factors that contributed to their emergence.

During the "quiet" period of 2017/2018, 2018/2019, and 2019/2020, the H5N8 virus had ample time to evolve into a new variant of G1 with divergent HA and other segments, through interactions between poultry and wild birds in Egypt and surrounding area as well as Asia-Europe border. This new variant then spread widely during bird autumn migration. The widespread occurrence of H5N8 G1 virus, particularly in Europe, facilitated the reassortment of the H5N8 virus with other subtype viruses from wild birds, resulting in the emergence of the novel H5N1 virus that caused the 2021/2022 panzootic in Europe and North America. Therefore, the emergence of two new viruses, the H5N8 G1 virus and the H5N1 virus, led to widespread outbreaks in the two waves of 2020/2021 and 2021/2022. Viral evolution in north-east Africa and middle-east Asia, especially the poultry of Egypt and surrounding area, as well as bird autumn migration from Russia-Kazakhstan region to Europe were critical factors initiating the 2020-2022 H5 outbreak.

Endemic of AIVs in poultry populations provides a breeding ground for their evolution through reassortment and mutation. For instance, the prevalence of H9N2 AIVs in chickens in China enabled the virus to participate in reassortment events, leading to the emergence of the novel H7N9 virus in 2013(23). In this study, we revealed similar consequences resulting from the endemic epidemic of H5 subtype AIVs in north-east Africa and middle-east Asia, especially the poultry of Egypt and surrounding area, which served as a gene source for the H5N8 (7 of 8 segments) and H5N1 (mutated HA segment) viruses during the 2020-2022 outbreak. Similar findings about the

genetic origins of recently resurgent HPAI H5 epidemics were also reported by Xie et al.(24). Differently, we here found two genotypes of HPAI H5N8 (G0 and G1 viruses) before the emergence of H5N1, and the replacement from G0 to G1 is a critical event contributing to the later emergence of HPAI H5N1, and the spatial-temporal coincidence between the G1 outbreak and the bird autumn migration may have expanded H5 viral spread.

Egypt, being located at the intersection of the Black Sea-Mediterranean and East African-West Asian Flyways, is a key habitat for a wide variety of waterfowl that connect Africa and Eurasia(25, 26). The country also experiences an annual increase in poultry production, which, coupled with a relatively rough farming approach, creates fertile ground for the endemicity of highly pathogenic avian influenza, enabling the disease to persist in the area. The selection pressure brought by poultry vaccination further complicated the evolution of HPAI virus(27, 28). Phylogeographic research has revealed that during the 2016-2017 epidemic season, there were at least six independent introductions of H5N8 virus from the Eurasian continent to Africa, including three introductions to Egyptian poultry, two introductions to western African poultry, and one introduction to South African poultry(29, 30). Some of these introduced viruses have persisted in Egypt since then(31–35). The frequent two-way transmission of AIVs between poultry and wild birds has provided ample opportunities for the H5N8 HPAI virus to adapt to both poultry and wild birds through mutation and reassortment, as evidenced by the observed long branch of HA gene and distinct gene constellation observed in our study. Therefore, the maintenance of HPAI virus in Egyptian poultry acted as an important source for generating new successful viruses.

In addition, Russia-Kazakhstan border is a crucial intersection for various travel routes between Eastern Europe and China, as well as between European Russia-Western Siberia and Central Asia(36–38). The wetlands in the vicinity are important breeding grounds or stopping points for

birds' migratory across the region(36, 37). Both Kazakhstan and Russia also have a large density of poultry(38). Since 2014, Russia has recorded H5N8 HPAI outbreaks, initially in Russia's East, relatively far from Kazakhstan(39). The first recorded outbreak in Kazakhstan occurred in poultry during the fall of 2020, along the Kazakhstan-Russia border. By year-end, the outbreaks had been found in eleven provinces of Kazakhstan(37). Migrating birds from Russia may introduce the virus into Kazakhstan, as earlier outbreaks in Russia have shown(7). Our data also indicate that, between July and September 2020, the genotype G1 was frequently detected in domestic poultry and waterfowl near the border between Russia and Kazakhstan. This time and place coincided with the autumn migration routes of many waterfowl. During the autumn migration period, the wild birds can travel from their breeding grounds in arctic or temperate breeding sites to the wintering sites around the world, and disseminate the virus to wider regions, which is known as the primary mechanism for long-distance transmission of HPAI virus(14). In the 2016/2017 wave, the previously identified major reassortant AAAAA8AA also emerged during pre-migratory or early autumn migration in 2016 somewhere between Belarus and Kazakhstan, and spread westward to Europe(15), which highlights the ecological significance of the Russian-Kazakhstan border and autumn bird migration as factors driving viral evolution, and facilitate the spread to Europe.

Europe serves as a diversified gene pool for avian influenza viruses due to its location at the crossroads of various wild bird migration channels connecting Asia, Africa, North America, and the Arctic. This virus gene pool can provide a large number of genes of wild bird origin to generate multiple new AIVs, which in turn become the outbreak center of these new viruses(36, 40, 41). Despite Europe's long history of culling measures to control highly pathogenic avian influenza, since 2014, HPAI H5 clade 2.3.4.4 viruses have dominated outbreaks in the region, yielding various subtypes such as H5N1, H5N2, H5N3, H5N4, H5N5, H5N6 and H5N8 through genetic

reassortments(14, 15). The majority of HPAI H5 virus detections in wild and domestic birds within Europe coincide with southwest/westward fall migration and large local water bird aggregations during wintering(36). This study demonstrates that the H5N8 (G1) virus, which differs from those (G0) circulating earlier in 2020, was dispersed in the autumn from the Europe-Asia border through westward migration. This H5N8 virus subsequently spread to at least 19 European countries(36). Simultaneously, multiple HPAI H5 reassortant viruses were detected in the region(42). The vast majority of HPAI viruses in wild birds and poultry were H5N8 viruses, while H5N1, H5N3, H5N4, and H5N5 were mainly isolated from wild birds(20). During this period, a novel H5N1 virus emerged, which replaced the H5N8 virus and dominated the next wave of 2021/2022. In December 2021, the Europe-origin H5N1 virus was found in Newfoundland, Canada, and then in North Carolina and South Carolina, USA(43–45). It was suggested that the H5N1 introduction is through the Atlantic Flyway probably including wild bird migratory routes from northern Europe that overlap Arctic regions of North America, eventually dispersal farther south into Canada and the United States(43).

The emergence of new dominant genotypes or subtypes for influenza A viruses is usually an important signal associated with disease outbreaks in animals or humans. In this study, we observed two dominant replacement events: one is the replacement from G0 to G1 in 2020, and the other is the replacement from H5N8 to H5N1 in 2021. The two novel HPAI viruses of H5N8 and H5N1 subtype successively caused the most serious 2020/2021 and 2021/2022 outbreaks, including 11 confirmed human cases, with 7 and 4 described in H5N8 and H5N1 respectively. Although they both belonged to the clade 2.3.4.4b GS/96 lineage, the H5N8 and H5N1 have divergent genomes from earlier H5 HPAI viruses. But they shared the same original HA and MP genes each other. Even when they were widely spread in Europe, Asia, and North America, they

later experienced frequent reassortment, their HA and MP genes were always retained. HA gene, a membrane glycoprotein, mainly influences receptor binding and antigenicity, which makes it significant in viral pathogenicity, transmission, and cross-species infection(46). MP gene, encoding M1 and M2 proteins, plays a significant role in the assembly and budding of influenza virus, determining virus morphology, as well as affecting viral replication, and transmission(47, 48)This suggests that these two segments are critical for viral adaption in wild birds and poultry, and the acquirements of HA and MP gene may have critical functional effects in breaking host barriers among wild birds and poultry, leading to worldwide outbreaks. Although other genes may also have positive effects, such as the N1 gene, further research is needed to clarify its role in viral adaption.

The continuous HPAI outbreaks caused by H5N8 and H5N1 suggested that migratory birds are not only the natural reservoir, but also potential disease outbreak sources that generate and spread novel variants. To mitigate the damage caused by bird migration, we face a significant new challenge. Traditional measures such as killing or immunization are not practical for migratory birds, and therefore, early warning is the most effective prevention measure.

In all, our study on the H5N8 and H5N1 outbreaks emphasizes the importance of identifying risk factors, such as hot spots and high-risk periods, to control and prevent the global spread of HPAI viruses in both poultry and wild birds, and even in humans. Comprehensive and systematic surveillance is needed, with a focus on these risk factors to improve our understanding and management of HPAI outbreaks.

MATERIALS AND METHODS

Data collection

Epidemic data

Outbreak data were downloaded from FAO Empres-i (https://empres-i.apps.fao.org). Number of cases and death for poultry and wild birds were retrieved from WOAH six-monthly report (https://wahis.woah.org). Only records of H5Ny HPAI after 2016 were collected, and 7021 records of Empres-i and 12763 records of WOAH WAHIS were used. Epidemic wave was defined based on the epidemic curve of outbreak data. Reported Human cases in H5Ny HPAI data were summarized from WHO (https://www.who.int/).

Genetic data

A total of 6966 HPAI H5 influenza isolates were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID) for all available countries and hosts during the period October 2016 to December 2022(49). From this data set, strains with known sampling locations, dates, hosts, and all segments available were selected. ~~Only one isolate with the earliest sampling date was retained if identical viral genome sequences were found.~~ Sequences that were less than 75% of the overall length of the segment were also removed. Finally, 3781 strains with unique genomes were obtained. Among these, 2965 strains with full genome were collected between December 2019 to December 2022, including 1460 H5N1 strains, 1391 H5N8 strains, which will be refer to as dataset-N1, dataset-N8 hereafter. All sequence data were aligned using MAFFT v7.310(50) with default parameters and subsequently manually edited.

Given our interest in the genesis of H5Ny viruses, the major agents of 2020-2022 wave, two expanded dataset ex-dataset-N1 and ex-dataset-N8 were also created. The additional sequences

were chosen from Basic Local Alignment Search Tool (BLAST) analysis to be genetically close to different groups and are not restrict to any specific subtype(51). Groups were defined as a collection of sequences that are similar to each other in each segment (detailed below). Specifically, BLAST analysis was run on each group in each segment of dataset-N1 or dataset-N8, collecting up to 500 sequences from GISAID with a sequence identity no less than 97%. Then, Maximum Likelihood (ML) phylogenetic trees were estimated using FastTree v2.1.11 with default settings for all of the unique retrieved sequences and original sequences(52), and sequences within the larger clade subtending the second ancestor node with bootstrap support no less than 70% of original sequences were selected. Finally, duplicate sequences were eliminated from these selected and original sequences. For each included genome, the centroid geographic coordinates of its sampling location were retrieved at the secondary administrative level using Google Earth (earth.google.com).

Migration data

Based on the annotation of sampling location of all H5N8 viruses, we have systematically compiled a list of 108 avian hosts reported to have been infected by HPAI H5N8 viruses (Table S4). Using the residency data provided by Bow (https://birdsoftheworld.org), we have filtered out 78 migratory bird species. By conducting searches on Google Scholar (https://scholar.google.com) using the keywords "host name & migration route", we obtained a total of 228 relevant publications (those mentioning species migration route information were considered valid). After careful screening and verification, we excluded literature lacking specific data on migration routes, resulting in the final confirmation of 29 bird species supported by 48 articles (Fig. S15A, Tables S4 and S5). To ensure analytical accuracy, we only selected a single data route for each bird species in different migration directions to avoid bias caused by multiple individuals of the same species

sharing the same route data in hotspot analysis. Using the ArcGIS Pro platform, we integrated this data and employed spatial analysis and data reclassification techniques to extract several important migration corridors. In order to comprehensively showcase the migration patterns of these 29 host species, we downloaded corresponding partition vector maps from the IUCN (https://www.iucn.org) website. In ArcGIS Pro, we decomposed the layers of each species based on breeding and non-breeding periods and overlaid the breeding and non-breeding zones of all species separately, resulting in an overall distribution map of both breeding and non-breeding areas for the hosts.

Phylogenetic analysis

Maximum likelihood trees

A ML tree for all clade 2.3.4.4b HPAI H5Ny viral HA gene sequences was first constructed using IQTREE v1.6.12 with "GTR+F+I+G4" model. Node support was determined by 1000 ultrafast bootstrap replicates. Ancestor time was further determined by TreeTime v0.9.5. Sequence identity matrix for each gene was calculated using a customized Python v3.10 script and visualized using heatmap. ML trees were also built for all eight genes for HPAI H5N1 and HAPI H5N8 viruses isolated between 2020 and 2022 respectively. For each gene segment, only sequence with earliest sampling date was kept when identical sequences were found. Then, a ML tree was constructed using IQTREE with same parameters above. These trees were used for further genotypic assignment. All phylogenetic trees were visualized using a Python package "baltic".

Group and genotype delineating

Within these H5N8 trees, well-supported monophyletic distinct groups were delineated base on mean paired patristic distance (MPD) (53). For a given internal node, the MPD value is defined as:

$$MPD = \frac{\sum d_{ij}}{\binom{n}{2}}$$

where d_{ij} is the phylogenetic distance between sequence i and j, and n is the number of sequences under this internal node. First, the MPD value for each internal node was calculated. Then, a depth-first search algorithm was used to find well-supported monophyletic group(54). At each step of the depth-first visit, a subtree was identified as a distinct group if the MPD value of the subtree was below a t-percentile threshold of the whole-tree MPD value distribution and the posterior support of the subtree was no less than 70%. If this condition was met in a node, the search at that node was stopped, ignoring the children's nodes, passing to analyze other node siblings. The threshold t was evaluated and optimized over the range [0th, 100th] percentile of the whole-tree MPD value distribution, with a step of 0.1. The mean cluster size against the t-percentile was further plotted. Based on this plot, the last value of t was chosen as the t-percentile threshold when the mean cluster size reaches the first plateau, at which value means a relative stable cluster results with small number of group numbers(55). Above processes were repeated for each segment using custom script in https://github.com/DuLab-SYSU/reEmergenceH5Ny/blob/main/tree_processing.ipynb.

The index for each monophyletic group was named in alphabet order based on number of sequences. Hence, the genotype of each strain can be annotated by the group indexes of eight genes in the order of PB2, PB1, PA, HA, NP, NA, MP, NS, according to a previous study(15). For example, the genotype aaaaaaaa represents that all the group index of eight gene segments is a. We also applied the following rules to assign alias for each genotype. Based on the main epidemic time, genotype G0 and genotype G1 was assigned for viruses if all genes fell into the group bbbbbbbb or aaaaaaaa, respectively. Within each G0 and G1 series, genotypes were further

assigned (for example, G0R1) if any internal gene come from a different monophyletic group. The decimal number was sequentially assigned based on the number of different genes. Some transient genotype with only one strain were merged as G0other or G1other based on the group index of HA gene.

Ancestor status estimation

The extended H5N8 dataset was divided into groups based on the index of monophyletic group for each segment, resulting in multiple smaller datasets(15). Each group's ancestor status was estimated using a Bayesian statistical framework. To be more explicit, host status was estimated using a continuous time Markov chain (CTMC) process and location was estimated using a Brownian random walk process to model diffusion in continuous space. Despite the fact that Bayesian framework allows for a joint inference of time, location and host status for ancestor status, it is computationally challenging for the data set sizes we examined here(56). Because of this and the fact that both diffusion processes were modelled separately from the substitution process throughout evolutionary history, the inference problem was divided into two steps: first, only sequence evolution process was considered to generate an empirical distribution of trees, and then discrete or continuous trait diffusion processes mentioned above were fitted conditioned on this set of posterior trees(29, 57). All MCMC sampling analyses were performed using BEAST in conjunction with the Broad-platform Evolutionary Analysis General Likelihood Evaluator (BEAGLE) library to enhance computation(58).

Bayesian time-resolved phylogenetic trees were first estimated per group per segment using BEAST v1.10.4(59). Coding genes were partitioned into first + second and third codon positions for all segments apart from NS and MP segment and applied a separate Hasegawa-Kishino-Yano 85 (HKY85) substitution model with gamma-distributed rate variation among sites to both

partitions. An uncorrelated lognormal relaxed molecular clock was used to account for evolutionary rate variation among lineages and specified a constant population size coalescent tree prior. Ten independent Markov Monte Carlo (MCMC) chains were run for 200 million generations and sampled for every 40000 generation with 10% as burn-in. Stationarity and mixing were investigated using Tracer version 1.5, making sure that effective sample sizes for the continuous parameters were greater than 200, which is the accepted standard in BEAST analyses. A subset of 500 trees were randomly selected from the combined posterior tree distribution. Then, these trees were used as an empirical distribution in the subsequent spatial and host diffusion inference. This is achieved by incorporating a proposal mechanism that randomly draws a new tree from the empirical distribution. In the discrete host status estimation, the Bayesian stochastic search variable selection (BSSVS) approach with asymmetric rates was also used to identify best-supported lineage transitions events between hosts(60). In the continuous geography estimation, RRW diffusion model was used to perform continuous phylogeographic reconstructions along groups delineated in the previous step, and a Cauchy distribution was used to model the among-branch heterogeneity in diffusion velocity(61). Such Bayesian inference resulted in a posterior distribution of time-measured trees, each annotated with inferred ancestral locations and host status. The time of MRCA and MRCGA were reported as middle value of posterior distribution with 95% HPD to quantify the uncertainty. The host of MRCA and MRCGA were reported with the highest posterior probability. The location of MRCA and MRCGA were reported with the middle value of posterior distribution for latitude and longitude and using 2-dimension kernel density estimation to visualization the 95% HPD. The ancestor statuses for specific node were retrieved from posterior trees using a customized Python script. The trunk host through the time was determined from posterior phylogenies using PACT v.0.9.5 (https://github.com/trvrb/PACT). The trunk is

comprised of all branches ancestral to a virus that were sampled within a year of the most recent samples(62).

ACKNOWLEDGMENT

We acknowledge the authors for originating and submitting laboratories of the sequences from GISAID's EpiFlu database, on which this research is based. A full name is available as Data S1. This work was supported by the National Natural Science Foundation of China grant 81961128002 (Y.S.), National Key Research and Development Program of China grant 2022YFF0802403 (J.P.), National Natural Science Foundation of China grant 32192451 (J.L.), Natural Science Foundation of Hainan Province of China grant 323CXTD37 (J.P.), the National Waterfowl-Industry Technology Research System (CARS-42) (J.P.), National Key Research and Development Program of China grant 2022YFC2303800 (D.W.), Shenzhen Science and Technology Program grant KQTD20180411143323605 (X.D.), Guangdong Frontier and Key Tech Innovation Program grant 2019B020228001, 2019B111103001, 2021A111112007, 2022B1111020006 (X.D.), European Union's Horizon 2020 research and innovation programme grant No. 874735 (S.J.L. and L.L.), Biological Sciences Research Council (BBSRC) ecology and evolution of infectious diseases (EEID) project under Grant No. BB/V011286/1 (S.J.L. and L.L.), and BBSRC Biotechnology and programme grant to Roslin Institute No. BBS/E/D/20002173 (S.J.L.).

AUTHOR CONTRIBUTIONS

J.P., H.S., X.D., J.L., Y.S., G.F.G. and L.L. designed research; J.Z., F.D., L.X., S.S. and Y.L. contributed to the data collection and processing; J.Z., F.D. and J.P. performed bioinformatics analyses; J.Z., L.X., F.D. and J.P. performed migration analysis; J.Z. and F.D. wrote the initial manuscript; J.P., X.D., J.Z., F.D., L.L., J.L., G.F.G., S.J.L., Y.S., L.X., H.S., W.L., J.Z., L.W., S.S., Y.L., Q.Z., K.T, Q.S., C.Z., H.L., Z.Q., K.Z., Z.L., G.Z., Y.S., D.W. and Z.Z. discussed and revised the manuscript.

CODE AVAILABILITY

XML file used for BEAST software and code used for data analysis are available on GitHub at: https://github.com/DuLab-SYSU/reEmergenceH5Ny.

DATA AVAILABILITY

Raw sequence data used in this work are available on GISAID EpiFlu database under strain name provided in Data S1. Summarized epidemiology data and migration data are available in Supplemental material.

DECLATRATION OF INTERESTS

Authors declare that they have no competing interests.

Supplemental Material

Supplemental figures and tables (Supplemental Material.pdf): Figures S1 to S16; Tables S1 to S5.

Supplemental data (DataS1.xlsx): Data S1.

REFERENCES

1. European Food Safety Authority, European Centre for Disease Prevention and Control, European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Marangon S, Niqueux É, Staubach C, Terregino C, Aznar I, Guajardo IM, Baldinelli F. 2023. Avian influenza overview September – December 2022. Efsa J 21.

2. CDC. 2023. H5N1 Bird Flu Detections across the United States (Backyard and Commercial). Centers for Disease Control and Prevention. https://www.cdc.gov/flu/avianflu/data-map-commercial.html. Retrieved 18 March 2023.

583 3. CDC. 2023. H5N1 Bird Flu Detections across the United States (Wild Birds). Centers
584 for Disease Control and Prevention. https://www.cdc.gov/flu/avianflu/data-map-
585 wild-birds.html. Retrieved 18 March 2023.

586 4. Guan Y, Smith GJD. 2013. The emergence and diversification of panzootic H5N1
587 influenza viruses. Virus Res 178:35–43.

588 5. Group WHEW. 2012. Continued evolution of highly pathogenic avian influenza A
589 (H5N1): updated nomenclature. Influenza Other Resp 6:1–5.

590 6. Smith GJD, Donis RO, World Health Organization/World Organisation for Animal
591 Health/Food and Agriculture Organization (WHO/OIE/FAO) H5 Evolution Working
592 Group. 2015. Nomenclature updates resulting from the evolution of avian influenza
593 A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013–2014. Influenza Other Resp
594 9:271–276.

595 7. Lee D-H, Criado MF, Swayne DE. 2020. Pathobiological Origins and Evolutionary
596 History of Highly Pathogenic Avian Influenza Viruses. Csh Perspect Med a038679.

597 8. Lycett SJ, Duchatel F, Digard P. 2019. A brief history of bird flu. Philos T Roy Soc B
598 374:20180257.

599 9. Cui P, Shi J, Wang C, Zhang Y, Xing X, Kong H, Yan C, Zeng X, Liu L, Tian G, Li C, Deng G,
600 Chen H. 2022. Global dissemination of H5N1 influenza viruses bearing the clade
601 2.3.4.4b HA gene and biologic analysis of the ones detected in China. Emerg Microbes
602 Infec 11:1693–1704.

603 10. Shi W, Gao GF. 2021. Emerging H5N8 avian influenza viruses. Science 372:784–786.

11. Liu J, Xiao H, Lei F, Zhu Q, Qin K, Zhang X-W, Zhang X-L, Zhao D, Wang G, Feng Y, Ma J, Liu W, Wang J, Gao GF. 2005. Highly pathogenic H5N1 influenza virus infection in migratory birds. Science 309:1206.

12. Chen H, Smith GJD, Zhang SY, Qin K, Wang J, Li KS, Webster RG, Peiris JSM, Guan Y. 2005. H5N1 virus outbreak in migratory waterfowl. Nature 436:191–192.

13. Tian H, Zhou S, Dong L, Boeckel TPV, Cui Y, Newman SH, Takekawa JY, Prosser DJ, Xiao X, Wu Y, Cazelles B, Huang S, Yang R, Grenfell BT, Xu B. 2015. Avian influenza H5N1 viral and bird migration networks in Asia. P Natl Acad Sci Usa 112:172–177.

14. The 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.

15. Lycett SJ, Pohlmann A, Staubach C, Caliendo V, Woolhouse M, Beer M, Kuiken T, Global Consortium for H5N8 and Related Influenza Viruses. 2020. Genesis and spread of multiple reassortants during the 2016/2017 H5 avian influenza epidemic in Eurasia. P Natl Acad Sci Usa 117:20814–20825.

16. Wille M, Barr IG. 2022. Resurgence of avian influenza virus. Science 376:459–460.

17. Pan American Health Organization / World Health Organization. 2023. Epidemiological Alert: Outbreaks of avian influenza caused by influenza A(H5N1) in the Region of the Americas.

18. European Food Safety Authority, European Centre for Disease Prevention and Control and European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A,

Kuiken T, Niqueux E, Staubach C, Terregino C, Guajardo IM, Baldinelli F. 2020. Avian influenza overview November 2019– February2020. Efsa J 18.

19. European Food Safety Authority, European Centre for Disease Prevention and Control and European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Marangon S, Niqueux É, Staubach C, Terregino C, Muñoz Guajardo I, Lima E, Baldinelli F. 2021. Avian influenza overview December 2020 – February 2021. Efsa J 19.

20. European Food Safety Authority, European Centre for Disease Prevention, Control, European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Marangon S, Niqueux É, Staubach C, Terregino C, Aznar I, Muñoz Guajardo I, Baldinelli F. 2022. Avian influenza overview May – September 2021. Efsa J 20.

21. Pyankova OG, Susloparov IM, Moiseeva AA, Kolosova NP, Onkhonova GS, Danilenko AV, Vakalova EV, Shendo GL, Nekeshina NN, Noskova LN, Demina JV, Frolova NV, Gavrilova EV, Maksyutov RA, Ryzhikov AB. 2021. Isolation of clade 2.3.4.4b a(H5N8), a highly pathogenic avian influenza virus, from a worker during an outbreak on a poultry farm, russia, december 2020. Euro Surveill 26.

22. European Food Safety Authority, European Centre for Disease Prevention and Control, European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Marangon S, Niqueux É, Staubach C, Terregino C, Guajardo IM, Chuzhakina K, Baldinelli F. 2022. Avian influenza overview June – September 2022. Efsa J 20.

23. Pu J, Wang S, Yin Y, Zhang G, Carter RA, Wang J, Xu G, Sun H, Wang M, Wen C, Wei Y, Wang D, Zhu B, Lemmon G, Jiao Y, Duan S, Wang Q, Du Q, Sun M, Bao J, Sun Y, Zhao J, Zhang H, Wu G, Liu J, Webster RG. 2015. Evolution of the H9N2 influenza genotype that facilitated the genesis of the novel H7N9 virus. Proc Natl Acad Sci USA 112:548–553.

24. Xie R, Edwards KM, Wille M, Wei X, Wong S-S, Zanin M, El-Shesheny R, Ducatez M, Poon LLM, Kayali G, Webby RJ, Dhanasekaran V. 2023. The episodic resurgence of highly pathogenic avian influenza H5 virus. Nature 622:810–817.

25. Scott D, Rose P. 1996. Atlas of Anatidae Populations in Africa and Western Eurasia.

26. 2006. Waterbirds around the world: a global overview of the conservation, management and research of the world's waterbird flyways. The Stationery Office, Edinburgh.

27. Fasanmi OG, Odetokun IA, Balogun FA, Fasina FO. 2017. Public health concerns of highly pathogenic avian influenza H5N1 endemicity in Africa. Vet World 10:1194–1204.

28. El-Shesheny R, Kandeil A, Mostafa A, Ali MA, Webby RJ. 2021. H5 Influenza Viruses in Egypt. Csh Perspect Med 11:a038745.

29. Fusaro A, Zecchin B, Vrancken B, Abolnik C, Ademun R, Alassane A, Arafa A, Awuni JA, Couacy-Hymann E, Coulibaly M' B, Gaidet N, Go-Maro E, Joannis T, Jumbo SD, Minoungou G, Meseko C, Souley MM, Ndumu DB, Shittu I, Twabela A, Wade A, Wiersma L, Akpeli YP, Zamperin G, Milani A, Lemey P, Monne I. 2019. Disentangling

668 the role of Africa in the global spread of H5 highly pathogenic avian influenza. Nat
669 Commun 10:5310.

670 30. Laleye AT, Bianco A, Shittu I, Sulaiman L, Fusaro A, Inuwa B, Oyetunde J, Zecchin B,
671 Bakam J, Pastori A, Olawuyi K, Schivo A, Meseko C, Vakuru C, Fortin A, Monne I,
672 Joannis T. 2022. Genetic characterization of highly pathogenic avian Influenza H5Nx
673 clade 2.3.4.4b reveals independent introductions in nigeria. Transbound Emerg Dis
674 69:423–433.

675 31. Abolnik C, Pieterse R, Peyrot BM, Choma P, Phiri TP, Ebersohn K, Heerden CJ van,
676 Vorster AA, Zel G van der, Geertsma PJ, Laleye AT, Govindasamy K, Rauff DL. 2018.
677 The Incursion and Spread of Highly Pathogenic Avian Influenza H5N8 Clade 2.3.4.4
678 Within South Africa. Avian Dis 63:149–156.

679 32. Salaheldin AH, Elbestawy AR, Abdelkader AM, Sultan HA, Ibrahim AA, Abd El-Hamid
680 HS, Abdelwhab EM. 2022. Isolation of Genetically Diverse H5N8 Avian Influenza
681 Viruses in Poultry in Egypt, 2019-2021. Viruses-Basel 14:1431.

682 33. Kandeil A, Hicks JT, Young SG, El Taweel AN, Kayed AS, Moatasim Y, Kutkat O, Bagato
683 O, McKenzie PP, Cai Z, Badra R, Kutkat M, Bahl J, Webby RJ, Kayali G, Ali MA. 2019.
684 Active surveillance and genetic evolution of avian influenza viruses in Egypt, 2016–
685 2018. Emerg Microbes Infec 8:1370–1382.

686 34. Hagag NM, Yehia N, El-Husseiny MH, Adel A, Shalaby AG, Rabie N, Samy M, Mohamed
687 M, El-Oksh ASA, Selim A, Arafa A-S, Eid S, Shahein MA, Naguib MM. 2022. Molecular

688 Epidemiology and Evolutionary Analysis of Avian Influenza A(H5) Viruses Circulating
689 in Egypt, 2019-2021. Viruses-Basel 14:1758.

690 35. Naguib MM, Verhagen JH, Samy A, Eriksson P, Fife M, Lundkvist Å, Ellström P, Järhult
691 JD. 2019. Avian influenza viruses at the wild–domestic bird interface in Egypt. Infect
692 Ecol Epidemiol 9:1575687.

693 36. Verhagen JH, Fouchier RAM, Lewis N. 2021. Highly Pathogenic Avian Influenza
694 Viruses at the Wild–Domestic Bird Interface in Europe: Future Directions for
695 Research and Surveillance. 2. Viruses-Basel 13:212.

696 37. Amirgazin A, Shevtsov A, Karibayev T, Berdikulov M, Kozhakhmetova T, Syzdykova L,
697 Ramankulov Y, Shustov AV. 2022. Highly pathogenic avian influenza virus of the
698 A/H5N8 subtype, clade 2.3.4.4b, caused outbreaks in Kazakhstan in 2020. Peerj
699 10:e13038.

700 38. Hill NJ, Smith LM, Muzaffar SB, Nagel JL, Prosser DJ, Sullivan JD, Spragens KA,
701 DeMattos CA, DeMattos CC, El Sayed L, Erciyas-Yavuz K, Davis CT, Jones J, Kis Z, Donis
702 RO, Newman SH, Takekawa JY. 2021. Crossroads of highly pathogenic H5N1: overlap
703 between wild and domestic birds in the Black Sea-Mediterranean impacts global
704 transmission. Virus Evol 7:veaa093.

705 39. Marchenko VY, Susloparov IM, Kolosova NP, Goncharova NI, Shipovalov AV,
706 Durymanov AG, Ilyicheva TN, Budatsirenova LV, Ivanova VK, Ignatyev GA, Ershova SN,
707 Tulyahova VS, Mikheev VN, Ryzhikov AB. 2015. Influenza A(H5N8) virus isolation in
708 Russia, 2014. Arch Virol 160:2857–2860.

709 40. King J, Harder T, Conraths FJ, Beer M, Pohlmann A. 2021. The genetics of highly
710 pathogenic avian influenza viruses of subtype H5 in Germany, 2006–2020.
711 Transbound Emerg Dis 68:1136–1150.

712 41. King J, Schulze C, Engelhardt A, Hlinak A, Lennermann S-L, Rigbers K, Skuballa J,
713 Staubach C, Mettenleiter TC, Harder T, Beer M, Pohlmann A. 2020. Novel HPAIV H5N8
714 Reassortant (Clade 2.3.4.4b) Detected in Germany. Viruses-Basel 12:281.

715 42. European Food Safety Authority, European Centre for Disease Prevention Control and
716 European Union Reference Laboratory for Avian Influenza, Adlhoch C, Fusaro A,
717 Gonzales JL, Kuiken T, Marangon S, Niqueux É, Staubach C, Smietanka K, Terregino C,
718 Van der Stede Y, Aznar I, Baldinelli F. 2020. Avian influenza overview – update on 19
719 November 2020, EU/EEA and the UK. Efsa J 18.

720 43. Bevins SN, Shriner SA, Cumbee JC, Dilione KE, Douglass KE, Ellis JW, Killian ML,
721 Torchetti MK, Lenoch JB. 2022. Intercontinental Movement of Highly Pathogenic
722 Avian Influenza A(H5N1) Clade 2.3.4.4 Virus to the United States, 2021. Emerg Infect
723 Dis 28:1006–1011.

724 44. Caliendo V, Lewis NS, Pohlmann A, Baillie SR, Banyard AC, Beer M, Brown IH, Fouchier
725 R a. M, Hansen RDE, Lameris TK, Lang AS, Laurendeau S, Lung O, Robertson G, van der
726 Jeugd H, Alkie TN, Thorup K, van Toor ML, Waldenström J, Yason C, Kuiken T, Berhane
727 Y. 2022. Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds
728 from Europe to North America in 2021. 1. Sci Rep 12:11729.

729 45. Günther A, Krone O, Svansson V, Pohlmann A, King J, Hallgrimsson GT,
730 Skarphéðinsson KH, Sigurðardóttir H, Jónsson SR, Beer M, Brugger B, Harder T. 2022.
731 Iceland as stepping stone for spread of highly pathogenic avian influenza virus
732 between europe and north america. Emerg Infect Dis 28.

733 46. Gamblin SJ, Vachieri SG, Xiong X, Zhang J, Martin SR, Skehel JJ. 2020. Hemagglutinin
734 Structure and Activities. Csh Perspect Med a038638.

735 47. Rossman JS, Lamb RA. 2011. Influenza virus assembly and budding. Virology
736 411:229–236.

737 48. Cross TA, Dong H, Sharma M, Busath DD, Zhou H-X. 2012. M2 protein from influenza
738 A: from multiple structures to biophysical and functional insights. Curr Opin Virol
739 2:128–133.

740 49. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from
741 vision to reality. Euro Surveill 22:30494.

742 50. Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-
743 scale multiple sequence alignments. Bioinformatics 34:2490–2492.

744 51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
745 search tool. J Mol Biol 215:403–410.

746 52. Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution
747 Trees with Profiles instead of a Distance Matrix. Mol Biol Evol 26:1641–1650.

53. Gong X, Hu M, Chen W, Yang H, Wang B, Yue J, Jin Y, Liang L, Ren H. 2021. Reassortment network of influenza a virus. Front Microbiol 12:793500.

54. Prosperi MCF, Ciccozzi M, Fanti I, Saladini F, Pecorari M, Borghi V, Di Giambenedetto S, Bruzzone B, Capetti A, Vivarelli A, Rusconi S, Re MC, Gismondo MR, Sighinolfi L, Gray RR, Salemi M, Zazzi M, De Luca A, ARCA collaborative group. 2011. A novel methodology for large-scale phylogeny partition. Nat Commun 2:321.

55. Du X, Dong L, Lan Y, Peng Y, Wu A, Zhang Y, Huang W, Wang D, Wang M, Guo Y. 2012. Mapping of H3N2 influenza antigenic evolution in china reveals a strategy for vaccine strain recommendation. Nat Commun 3:709.

56. Trovão 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.

57. Dellicour S, Lequime S, Vrancken B, Gill MS, Bastide P, Gangavarapu K, Matteson NL, Tan Y, du Plessis L, Fisher AA, Nelson MI, Gilbert M, Suchard MA, Andersen KG, Grubaugh ND, Pybus OG, Lemey P. 2020. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. Nat Commun 11:5620.

58. Ayres DL, Cummings MP, Baele G, Darling AE, Lewis PO, Swofford DL, Huelsenbeck JP, Lemey P, Rambaut A, Suchard MA. 2019. BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics. Syst Biol 68:1052–1061.

59. 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.

60. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, Suchard MA. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog 10:e1003932.

61. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, Delwart EL. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. P Natl Acad Sci Usa 109:15066–15071.

62. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, Daniels RS, Gunasekaran CP, Hurt AC, Kelso A, Klimov A, Lewis NS, Li X, McCauley JW, Odagiri T, Potdar V, Rambaut A, Shu Y, Skepner E, Smith DJ, Suchard MA, Tashiro M, Wang D, Xu X, Lemey P, Russell CA. 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 523:217–220.

Figure legends

FIG 1 Epidemiological situation of global HPAI H5Ny during 2016-2022. (**A, F**) Epidemic curve of the confirmed global H5Ny HPAI outbreaks reported to FAO, colored by subtypes (A) and hosts (F). The temporal span for each epidemic wave is represented by grey square shadows with black text. (**B-E**) Counts of confirmed outbreaks in affected countries impacted in four waves of global HPAI H5Ny epidemics (from yellow for small counts to red color for large counts). Countries with a grey slash have no outbreaks or statistics. (**G**) Number of confirmed cases (filled bar) and deaths (no filled bar) reported to WOAH in four waves of HPAI H5Ny for domestic (blue bar) and wild birds (red bar) (in log scale). (**H**) Reported human cases with circle size proportion to case number. HPAI stands for high pathogenic avian influenza; FAO stands for the Food and Agriculture Department of the United Nations; and WOAH stands for the World Organization for Animal Health.

FIG 2 Evolution of clade 2.3.4.4b HPAI H5Ny viruses during 2016-2022. (**A**) Time-scaled phylogenetic tree of HPAI clade 2.3.4.4b H5Ny generated using 3781 HA gene sequences. Two main genetic groups have been donated and are denoted by arrows in black text. Major epidemic waves are denoted by grey or white bars, and donations are written in black writing. The subtypes of each strain colored the tips (red for H5N1, blue for H5N8, gold for H5N6 and grey for other H5Ny). Isolated regions are depicted on the right column (purple for Europe, light blue for Asia, pink for North America and green for Africa). (**B-I**) Pairwise genetic identity for each gene segment of 3781 HPAI H5Ny viruses. Sequences are sorted by the time of isolation. Black dashed lines separate major epidemic waves and the color from light purple to dark purple indicates genetic identity from low to high.

FIG 3 Genotype dynamic of HPAI H5N8 viruses during 2020-2021. (**A**) Gene constellations of H5N8 viruses are presented by the order of HA Maximum likelihood tree. Different colors of tips represent different genotypes. Colored bars on the right show the group classification of eight gene segments: PB2, PB1, PA, HA, NP, NA, MP and NS. (**B**) The number of isolated H5N8 viruses for every half month for different genotypes. (**C**) The proportion of different H5N8 viral genotypes. Gaussian kernel density estimation with bandwidth as 0.5 year was used to calculate the relative frequency at given time points. (**D**-**E**) The number of isolated H5N8 by hosts for G0 H5N8 and G1 H5N8. HPAI: high pathogenic avian influenza; Dom-ans: domestic *Anseriformes*; Dom-gal: domestic *Galliformes*; Wild-gal: Wild *Galliformes*.

FIG 4 Spread of genotype G0/G1 H5N8 and related migration flyways of wild birds in Eurasia and Africa continents. (**A**, **C**) Continuous spatiotemporal dispersal of genotype G0 and genotype G1 HPAI H5N8 viruses. The solid lines and dots represent the branches and nodes of the MCC tree. Contours represent statistical uncertainty of the estimated locations at the internal nodes (95% HPD based on 2-dimensional kernel density estimates). Dots, lines and contour are colored according to the time (from red for the earliest to the blue for the latest). (**B**, **D**) Migration pattern summarized from the existing literature. Yellow area represents breeding sites. Blue represents wintering sites. The purple arrows indicate major G0/G1 H5N8 spread related southward migration routes. The green arrows indicate major G0/G1 spread related northward migration routes. The red dots indicate sampling locations of genotype G0 viruses and blue dots indicate sampling locations of genotype G1 viruses (from light color for the earliest to dark color for the latest).

FIG 5 Schematics illustrating the generation and spread dynamic of the HPAI H5Ny in 2020-2022. (**A**) a schematic of key genetic genesis and reassortment history for G0/G1 H5N8

and H5N1. Viruses with different colors represent different genotypes or subtypes (gold for G0 H5N8, blue for G1 H5N8, red for H5N1). Orange lines and blue lines represent G0-like and G1-like genes. Dashed orange lines and dashed blue lines represent non-G0-like and non-G1-like genes. Red lines represent H5N1-like genes. Grey lines in clouds represent LPAIV pool in Eurasian and North American waterfowls. Dashed quadrilateral with color from orange to blue represent dominated genotype replacement from G0 H5N8 to G1 H5N8. (**B**) a schematic of 20/21 H5N8 and 21/22 H5N1 viruses spread between continents. Blue and red dots represent HPAI H5N8 and HPAI H5N1 outbreaks. Blue rhombuses represent hot spots for gene source/pool. Red rhombuses represent hot spots for transmission. Blue circles represent G1 H5N8 and red circles represent H5N1, lines in circles represent viral gene segments, and are colored by gene origin (blue for G1 H5N8, red for H5N1). The number shown in yellow hexagon indicates the steps of generation and spread of the successive H5N8 and H5N1 panzootic.