Proposal for a Global Classification and Nomenclature System for A/H9 Influenza Viruses

Alice Fusaro,[1] Juan Pu,[1] Yong Zhou,[1] Lu Lu, Luca Tassoni, Yu Lan, Tommy Tsan-Yuk Lam, Zoe Song, Justin Bahl, Jiani Chen, George F. Gao, Isabella Monne, Jinhua Liu, The International H9 Evolution Consortium[2]

Influenza A/H9 viruses circulate worldwide in wild and domestic avian species, continuing to evolve and posing a zoonotic risk. A substantial increase in human infections with A/H9N2 subtype avian influenza viruses (AIVs) and the emergence of novel reassortants carrying A/H9N2-origin internal genes has occurred in recent years. Different names have been used to describe the circulating and emerging A/H9 lineages. To address this issue, an international group of experts from animal and public health laboratories, endorsed by the WOAH/FAO Network of Expertise on Animal Influenza, has created a practical lineage classification and nomenclature system based on the analysis of 10,638 hemagglutinin sequences from A/H9 AIVs sampled worldwide. This system incorporates phylogenetic relationships and epidemiologic characteristics designed to trace emerging and circulating lineages and clades. To aid in lineage and clade assignment, an online tool has been created. This proposed classification enables rapid comprehension of the global spread and evolution of A/H9 AIVs.

Avian influenza viruses (AIVs), including type A of the H9 subtype (A/H9), have a worldwide distribution in wild birds and are endemic in poultry populations in several countries in Asia, the Middle East, and Africa. Despite being classified as low pathogenicity, AIV strains of the A/H9N2 subtype are becoming an increased threat to domestic birds and humans (*1*,*2*). A/H9N2 subtype viruses can cause major economic damage to the poultry industry because of decreased egg production and increased mortality, related to co-infection with bacteria and other viruses (*3*–*6*). The economic consequences from this subtype have prompted several countries to adopt poultry vaccination policies to try to control the virus spread (*7*,*8*). Those policies have further driven genetic diversification of A/H9 AIVs (*8*–*11*).

Since 1998, viruses of the A/H9 subtype have occasionally crossed over into mammal species, leading to detection in humans, pigs, dogs, horses, pikas, minks, and bats (*12*). As of December 1, 2023, A/H9N2 was associated with 128 human infections, 90% of which were reported in China (*13*), and more than half of cases (52%) were identified during 2020–2023. Of note, this number may be underestimated because A/H9N2 infected humans do not often show obvious symptoms (*14*). Furthermore, A/H9N2 AIVs were identified as a source of internal gene mutations, leading to the emergence of several reassortants in human infection reports, such as H3N8, H5N6, H7N9, and H10N8 (*15*–*21*).

To date, different names have been adopted to describe A/H9 influenza virus lineages or sublineages and clades based on the hemagglutinin (HA) gene, often with differing names for the same group (*5*,*22*–*27*). The most widely accepted 4 primary lineages have multiple and complex names: the BJ/94 lineage is also known as Y280-like, G9-like, or h9.4.2 (represented by A/chicken/Beijing/1/94, A/chicken/Hong Kong/G9/97, or A/duck/Hong Kong/Y280/97); the G1 lineage is also known as h9.4.1 (represented by A/quail/Hong Kong/G1/97); the Y439 lineage is also known as Korean or h9.3 (represented by

Author affiliations: Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy (A. Fusaro, L. Tassoni, I. Monne); China Agricultural University, Beijing, China (J. Pu, Y. Zhou, J. Liu); University of Edinburgh, Edinburgh, Scotland, UK (L. Lu); Chinese Center for Disease Control and Prevention, Beijing (Y. Lan); The University of Hong Kong, Hong Kong, China (T.T.-Y. Lam, Z. Song); University of Georgia, Athens, Georgia, USA (J. Bahl, J. Chen); Chinese Academy of Sciences, Beijing (G.F. Gao)

DOI: https://doi.org/10.3201/eid3008.231176

[1]These first authors contributed equally to this article.

[2]Members of The International H9 Evolution Consortium are listed at the end of this article.

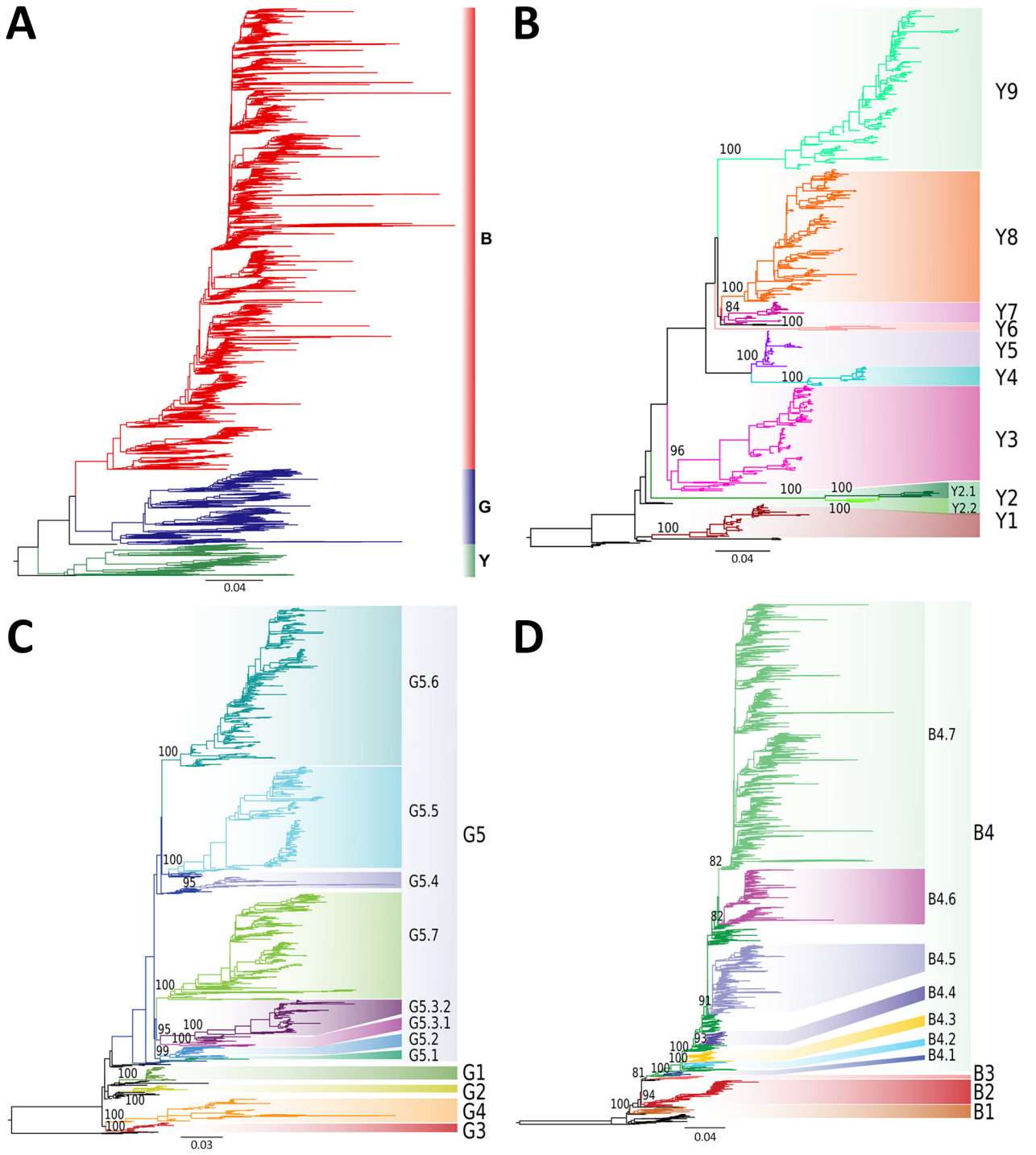


Figure 1. Phylogenetic trees and assigned clades as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. A) Maximum-likelihood (ML) phylogenetic tree obtained using the complete dataset. The branches are colored according to the 3 identified lineages. A/quail/Hong Kong/A28945/1988 and A/Quail/Hong Kong/AF157/1992 were used as an outgroup (black). B) ML phylogenetic tree of the Y lineage dataset. Six sequences of the G lineage were used as the outgroup. Numbers next to the clade-defining nodes represent ultra-fast bootstrap supports. Clades are labeled and marked in colors. C) ML phylogenetic tree of the G lineage. Six sequences of the Y lineage were selected as the outgroup. Numbers next to the clade-defining nodes represent ultra-fast bootstrap supports. Clades are labeled and marked in colors. D) ML phylogenetic tree of the B lineage. Six sequences of the Y lineage were used as the outgroup. Numbers next to the clade-defining nodes represent ultra-fast bootstrap supports. Clades are labeled and marked in colors. Scale bars indicate substitutions per site.

A/chicken/Korea/38349-p96323/96 or A/duck/Hong Kong/Y439/97); and the American lineage is also known as h9.1-h9.2 (represented by A/turkey/Wisconsin/1/1966 or A/turkey/Minnesota/511/1978). Additional sublineages and clades continue to emerge because of the rapid evolution of the A/H9 virus (*8*,*22*,*23*,*26*–*29*). The difference in clade classification and naming hinders comparison between studies and challenges critical discussion on the epidemiology, evolution, and spread of this subtype.

To address this issue, an international consortium of scientists from 22 laboratories in Europe, Asia, Africa, and the Americas was established to create a uniform system for classifying A/H9 viruses. This group was endorsed by the WOAH/FAO (World Organisation for Animal Health/Food and Agriculture Organization of the United Nations) Network of Expertise on Animal Influenza. By reconstructing a comprehensive phylogenetic history of all globally collected and publicly available HA sequences of A/H9 viruses, of any neuraminidase subtype, the consortium worked to create a harmonized classification and nomenclature system for A/H9 to be used in future epidemiologic and evolutionary studies.

Methods

Sequence Data and Metadata Preparation

To provide a comprehensive picture of A/H9 genetic diversity, we generated a dataset of the HA gene that included every H9-HA sequence available on the GISAID (https://www.gisaid.org) and GenBank public databases, which we accessed on July 18, 2022. We aligned the sequence dataset by using MAFFT v7.0 (*30*) and curated as described (Appendix 1, https://wwwnc.cdc.gov/EID/article/30/8/23-1176-App1.pdf). A final dataset containing 10,638 HA sequences and the related information, including accession numbers, resulted after our quality check process (Appendix 2 Table 1, https://wwwnc.cdc.gov/EID/article/30/8/23-1176-App2.xlsx).

Phylogenetic Characterization

We generated a maximum-likelihood (ML) phylogenetic tree from the final complete dataset (n = 10,638) of the A/H9 virus HA genes by using IQ-TREE v1.6 (*31*) and applying the best-fitted nucleotide substitution model selected by ModelFinder (*32*). We performed clustering of the ML phylogenies of the complete dataset by using the pathogen-agnostic clustering tool PhyCLIP (Appendix 1 Figure 1) (*33*).

We conducted validation of lineages and clades identified with PhyCLIP by using multiple datasets and software to perform phylogenetic analyses. We assessed the robustness of our inference on a down-sampled dataset of 1,000 sequences, and then we assessed the nodal supports on a down-sampled dataset of 2,000 sequences. We selected sequences on the basis of phylogenetic diversity and calculated them as the sum of branch lengths of the minimal subtree spanned by a set of taxa included in the phylogenetic tree. We conducted this selection by using the phylogenetic diversity analyzer tool (*34*), starting from the ML tree obtained from the complete dataset. We generated ML trees from the 2 down-sampled datasets in IQ-TREE v2.1.3 by performing ultrafast-bootstrap resampling analysis (1,000 replications) (*31*,*35*). We used the best-fitted nucleotide substitution model selected by Model Finder, implemented in IQ-TREE (*32*). We analyzed lineages separately to characterize the different clades within each lineage. We generated 3 datasets: Y_dataset, G_dataset, and B_dataset.

To assess the consistency of the clades, we conducted phylogenetic analyses for each dataset by using IQ-TREE v 2.1.3 (*31*) and PhyML v3.0 (*36*). We used ModelFinder, implemented in IQ-TREE (*32*), to select the best-fit nucleotide substitution model for each dataset. We assessed nodal supports in the IQ-TREE and PhyML analyses by using ultrafast-bootstrap and Shimodaira-Hasegawa (SH)–like branch supports (*35*,*36*).

To visualize the phylogenetic trees, we used FigTree version 1.4.2 (Figtree, http://tree.bio.ed.ac.uk/software/figtree). We calculated within and between clades average pairwise nucleotide distances (APD) by using MEGA X (p-distance) (*37*). We used TreeTime (https://treetime.readthedocs.io/en/latest) to reconstruct the ancestral sequence, which enabled us to infer the nonsynonymous nucleotide mutations leading to amino acid variations at each internal node (*38*).

Pilot Dataset

After the clade identification process, we created subsets for each lineage. The subsets are pilot_Y (86 sequences); pilot_G (105 sequences), and pilot_B (101 sequences) and pilot_complete, included all 3 lineages (292 sequences) (Appendix 3, http://wwwnc.cdc.gov/EID/article/30/8/23-1176-App3. xlsx). We created the 4 subsets by using the phylogenetic diversity analyzer tool (*34*) and then manually selected sequences of underrepresented lineages and clades (Appendix 3). We implemented ML phylogenetic analyses with 1,000 ultrafastbootstrap and 1,000 nonparametric bootstrap replicates to ensure that the

Figure 2. Pilot maximum-likelihood phylogenetic tree of the A/H9 influenza virus gene sequences obtained by using the representative datasets (Appendix 3, https://wwwnc.cdc.gov/EID/article/30/8/23-1176-App3.xlsx) for the Y lineage provided as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. Each clade is represented by ≥3 sequences, each labeled and colored according to the clade of belonging. Ultrafast-bootstrap supports >80% are indicated next to nodes. Scale bar indicates substitutions per site.

topology inferred from the larger datasets would be maintained with fewer sequences.

Online Tool Development

We created our online tool to automate the lineage classification of A/H9 influenza viruses. For the tool creation, we used Python (Python, https://www.python.org) and added an interactive display through the Django framework (Django, https://www.djangoproject.com). The tool asseses user-provided sequences to determine if they belong to the A/H9 subtype. If confirmed, the tool further classifies the sequences into their specific lineage and clade. The query sequences will be included in a pilot dataset, and a ML phylogenetic tree will be constructed by using IQ-TREE. The output from the tool includes the clade annotation and ML phylogenetic tree.

Results

Principles of the Nomenclature System

The genomic characterization of A/H9 is needed to support the surveillance, prevention, and risk assessment of this globally prevalent subtype. A unified criterion for A/H9 genotyping is necessary to compare molecular epidemiology data obtained from different laboratories. We propose a workable and practical lineage classification and nomenclature for A/H9 AIVs on the basis of a comprehensive phylogenetic analysis of all available HA gene sequences. Our proposed nomenclature of A/H9 AIVs meets multiple requirements by capturing local and global patterns of virus genetic diversity in a timely and coherent manner, being robust and flexible enough to accommodate emerging virus diversity, providing support to track emerging clades when they exhibit amino acid or biologic and epidemiologic changes, and being informative and easily traceable while enabling the incorporation of the major and minor clades over time.

Clade Classification of A/H9 Viruses

We used a curated dataset of 10,638 global A/H9 HA sequences (Appendix 2 Table 1) to assess the phylogenetic relationships among globally circulating A/H9 viruses and establish a unified classification system for the virus. We applied 5 steps to assign and identify clades (Appendix 1 Figure 2) by integrating phylogenetic topology, branch support, genetic distances, epidemiologic information, and amino acid variants shared within each group. We defined clades on the basis of the following criteria: clades are assigned to monophyletic groups including ≥3 sequences obtained from viruses sampled over a period ≥3 years, indicating sustained transmission of the clade in ≥2 influenza seasons; clades have an ultrafast-bootstrap value ≥80% for the clade-defining node; HA proteins within the clade must share ≥1 amino acid mutation; an APD of ≥6% between each clade is recommended. However, to meet all other criteria, the APD may be slightly lower for certain clades.

Using those criteria, we defined 3 lineages, 18 first-order clades, 16 second-order clades, and 2 third-order clades (Figure 1). Nodal supports for each clade obtained from different analysis methods of each complete or down-sampled datasets are available (Appendix 1 Table 1). Each clade contains ≥1 amino acid change at the internal nodes (Appendix 1 Table 2). We found many mutations associated with host tropism, virulence, or antigenicity of the influenza virus (*8*), indicating that A/H9 AIVs of these clades may have undergone changes in biologic characteristics.

Nomenclature Criteria of A/H9 Lineages and Clades

To maintain consistency with traditional and commonly used lineage names for A/H9 (Y439, American, G1, and BJ/94 lineages) that convey information on spatial and biologic characteristics, we opted to label the 3 lineages as Y (to correspond with Y439 and American lineages), G (to correspond with G1 lineages), and B (to correspond with BJ/94 lineages). The numeric values, such as Y1 or G1 for the first-order clade, Y1.1 or G1.1 for the second-order clade, Y1.1.1 or G1.1.1 for the third-order clade, are used to identify clades that descended from the Y, G, and B lineages. In addition, viruses are classified as Y-like, G-like, or B-like if they belong to 1 of the identified lineages but do not fall within any clade. We provided a comparison between our proposed nomenclature system and some of the previously used A/H9 nomenclatures that enables rapid mapping of our lineages and clades to the name of the previously used lineages and clades (Appendix 1 Table 3).

Pilot Datasets of Lineages and Clades

We generated 3 small datasets of ≈100 sequences containing representative viruses for each clade of the Y, G, and B lineages. To ensure that topology is maintained with a lower number of sequences, we performed ML phylogenetic reconstructions. All lineage and clade-defining branches are strongly supported (Appendix 1 Table 1) and the topologies of the generated pilot trees (Figures 2–4; Appendix 1 Figure 4) are consistent with the trees generated from larger datasets. All pilot datasets that contain sequences labeled

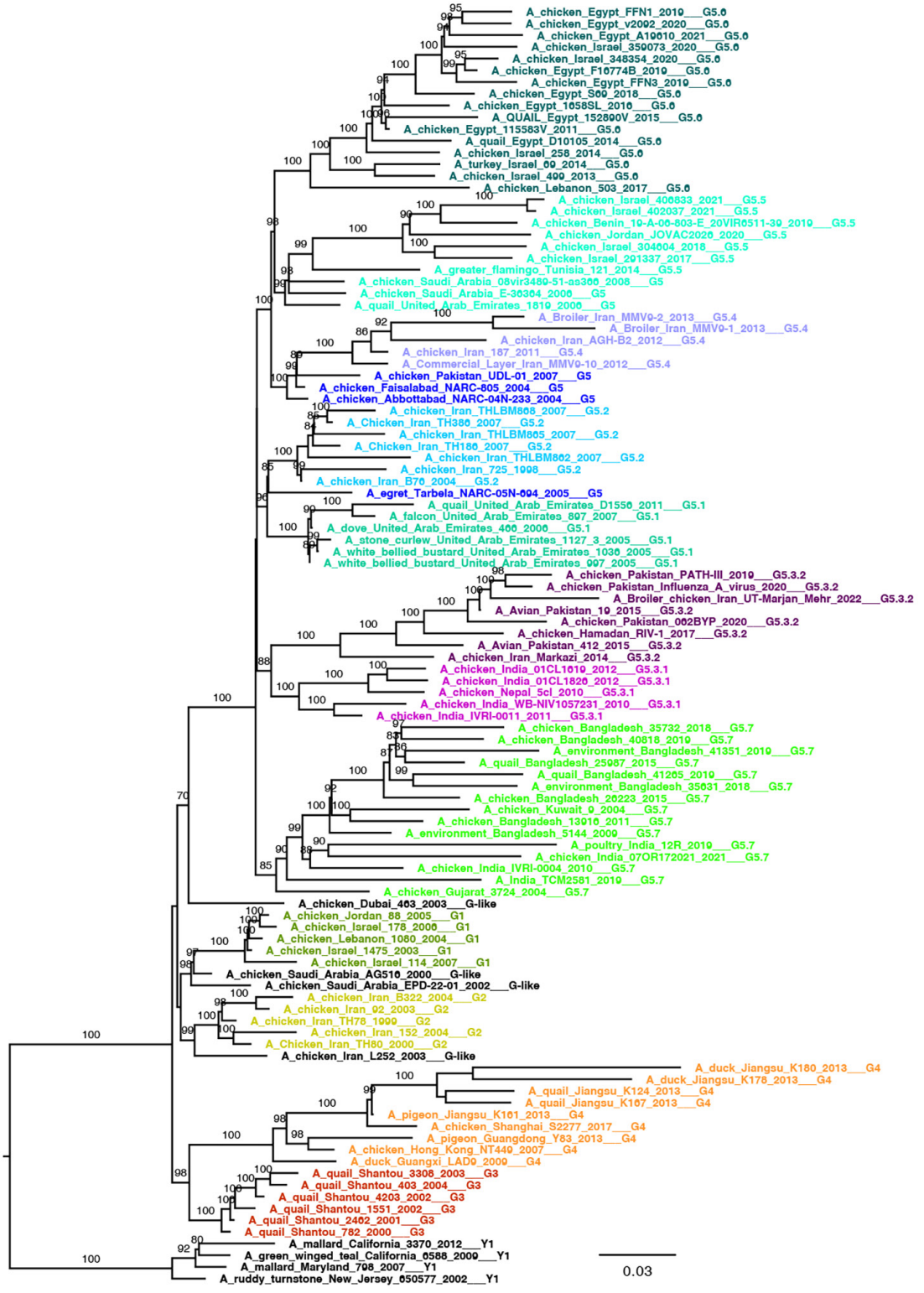


Figure 3. Pilot maximum-likelihood phylogenetic tree of the A/H9 influenza virus gene sequences obtained by using the representative dataset (Appendix 3, https://wwwnc.cdc.gov/EID/article/30/8/23-1176-App3.xlsx) for the G lineage provided as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. Each clade is represented by ≥3 sequences, each labeled and colored according to the clade of belonging. Ultrafast-bootstrap supports >80% are indicated next to nodes. Scale bar indicates substitutions per site.

according to lineage and clade are available (Appendix 3). Those pilot datasets can be used for rapid classification of newly sequenced A/H9 viruses according to our proposed classification scheme. If the lineage is unknown, the nucleotide A/H9 HA gene sequence or sequences under investigation can be analyzed by using the complete pilot dataset that includes all 3 lineages (Appendix 3). The complete pilot dataset will identify both the lineage and the specific clade according to clustering. If the lineage is known, the phylogenetic analysis can be performed by using the small and easily manageable pilot dataset specific for the lineage of interest, either pilot G, pilot Y, or pilot B (Appendix 3). The lineage and clade of an H9 sequence can also be identified by using the online tool.

Genetic and Epidemiologic Characteristics of Individual A/H9 Lineages and Clades

To determine whether the nomenclature system can reflect the real epidemiologic characteristics of A/H9 influenza viruses, we compiled a detailed description for those lineages and clades (Appendix 1 Tables 4–6; Appendix 2 Table 2) that includes the temporal, spatial, and host distribution. However, considering the limited number of sequences available for certain countries and the limited sampling data from wild birds, sampling bias cannot be excluded.

Y Lineage

The Y lineage (n = 622), including the previously named Y439 (or Korean, h9.3) and American (or h9.1–h9.2) lineages (*27*,*39*), was first identified in 1966 and is the oldest and most widespread lineage of A/H9 influenza viruses (*40*). In our dataset, the group contains sequences identified after 1976 because we removed earlier strains such as A/turkey/Wisconsin/1/1966 (Figure 5; Appendix 1 Table 4). The Y lineage has spread to every continent (Figure 6; Appendix 1 Table 4) and has evolved into 9 first-order (Y1–Y9) and 2 second-order (Y2.1 and Y2.2) well-supported clades (ultrafast-bootstrap >84, SH-like >0.85) (Figure 1, panel B). Clades of the Y lineage exhibited a high between-clade APD ranging from 7.10% to 18.30% (Appendix 1 Figure 3; Appendix 2 Table 2). Clades Y2.1 (South America, 2007–2017) and Y2.2 (United States, 1993–1996) showed the highest genetic distance from all other clades (15.37%–18.30%). Within-clade APD was <6% for all Y lineage clades, ranging from clade Y5 at 1.19% (United States, 2005–2007) to clade Y6 at 5.71% (Cambodia, Vietnam, 2009–2018).

Clades Y3, Y4, Y7, and Y8 were detected in ≥3 countries on ≥2 different continents, and clade Y8 has the widest geographic distribution: Australia, Asia, Europe, Africa, and North America. The remaining clades were identified in fewer countries. Clades Y1, Y2, and Y5 are exclusively from the Americas, and clade Y6 is specific to Southeast Asia (Cambodia and Vietnam). Detections of Y lineage viruses are equally distributed between wild and domestic birds, and 1 detection occurred in a mammal host (swine). The proportion of wild and domestic hosts varies across the identified clades. Domestic poultry accounts for 83% of clade Y6 detections and 96% of clade Y9 detections, whereas the other clades have been identified from mostly wild birds (64%–100%) (Figure 7; Appendix 1 Table 4).

G Lineage

The G lineage (n = 1643), previously known as G1 or h9.4.1 (*27*,*39*), has been circulating since 1997 (Figure 5; Appendix 1 Table 5). Domestic birds account for 85% of G lineage sequences (Figure 7), and the lineage is found in countries in Asia (61%) and Africa (38%) (Figure 6; Appendix 1 Table 5). According to the proposed clade classification criteria, we distinguished 5 first-order clades (G1–G5). Between-clade APD of the G lineage ranges from 4.81% to 13.59% (Appendix 1 Figure 3; Appendix 2 Table 2). Clade G5 shows the highest genetic distance from the other clades (10.94%–13.59%). Clade G5 was detected in 29 countries and was further divided into second-order (G5.1–G5.7) and third-order (G5.3.1–G5.3.2) subclades (Figure 1, panel C). The genetic distance between G1 and G2 is 5.81% and between G5.1 and G5.2 is 4.81%; those distances are less than the established 6% cutoff but are considered acceptable because they do not share a stable common root node. Within-clade APD was <6% for all clades, ranging from clade G1 (Israel, Jordan, Lebanon, 2003–2007) at 1.30% to clade G5.7 at 5.84% (Bangladesh, India, Kuwait, Pakistan, 2003–2022) (Appendix 1 Figure 3; Appendix 2 Table 2). All the identified clades are well supported (ultrafast-bootstrap >95; SH-like >0.73).

Most of the G clade viruses originate from domestic poultry, except for G5.1, which was found in 70% of the viruses sampled from wild avian species (Figure 7; Appendix 1 Table 5). Since 1999, human infections of viruses belonging to clades G4 (n = 4, 1999–2009), G5.3.2 (n = 1, 2015), G5.5 (n = 2, 2019), and G5.7 (n = 2, 2011–2019) have been detected.

B Lineage

The B lineage (n = 8,373), previously known as BJ/94, Y280, G9, or h9.4.2 (*27*,*39*), has been circulating since 1994 in domestic birds (92% of the sequences) from Asia, and ≥93% of the sequenced viruses originated

Figure 4. Pilot maximum-likelihood phylogenetic tree of the A/H9 influenza virus gene sequences obtained by using the representative dataset (Appendix 3, https://wwwnc.cdc.gov/EID/article/30/8/23-1176-App3.xlsx) for the B lineage provided as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. Each clade is represented by ≥3 sequences, each labeled and colored according to the clade of belonging. Ultrafast-bootstrap supports >80% are indicated next to nodes. Scale bar indicates substitutions per site.

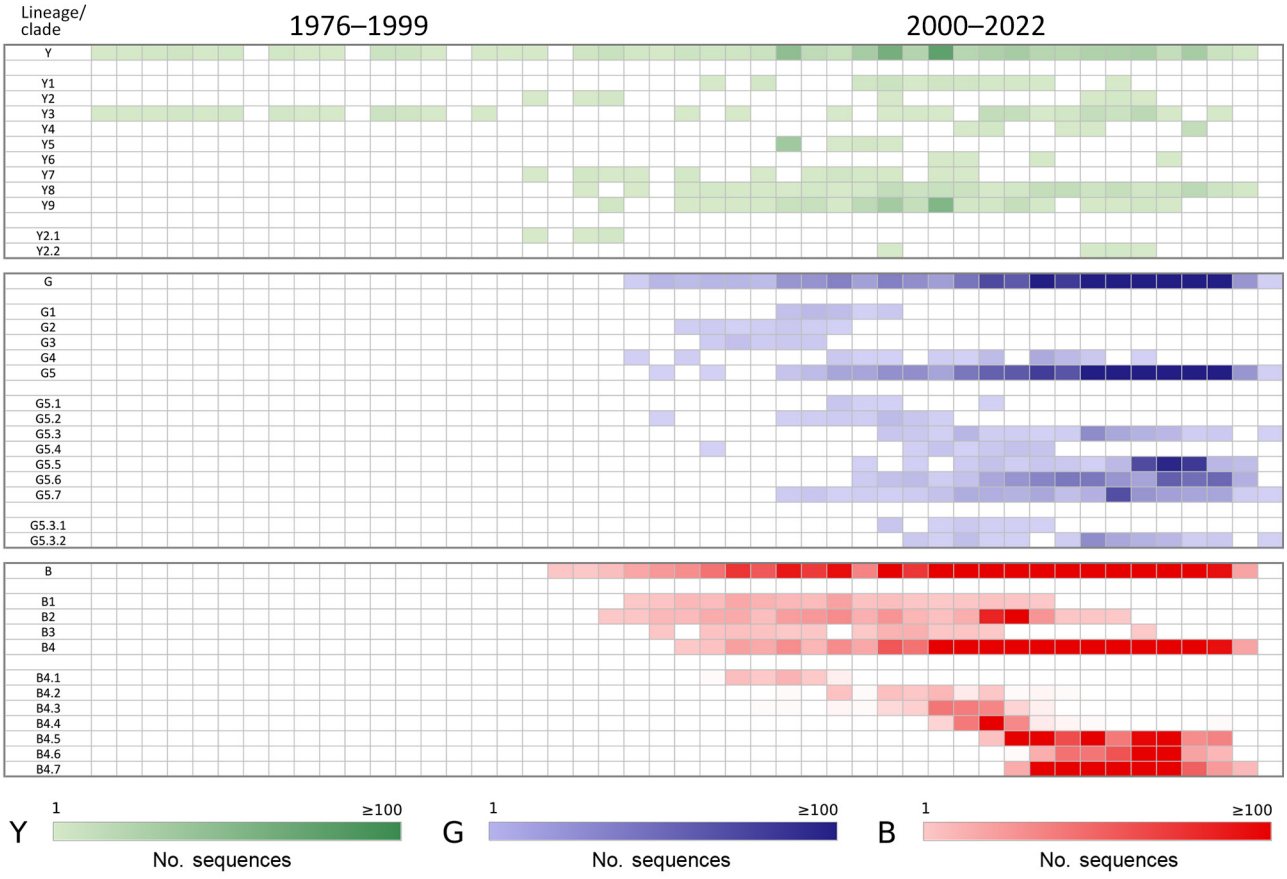


Figure 5. Temporal distribution of each lineage and clade for A/H9 influenza viruses as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. The heat map displays the number of sequences for each lineage and clade per year.

from China (Figures 5–7; Appendix 1 Table 6). We divided lineage B into 11 well supported (ultrafast-bootstrap >81; SH-like >0.9) clades, consisting of 4 first-order clades (B1–B4) and 7 second-order clades (B4.1–B4.7) (Figure 1, panel D). Compared with the other lineages, clades of the B lineage exhibited lower between-clade APD, ranging from 4.48% to 12.06% (Appendix 1 Figure 3; Appendix 2 Table 2). Clade B4 shows the highest genetic distance from all the other clades (10.54%–12.06%). Clade B4 has had a considerable geographic expansion in the past 10 years. This clade affects 11 distinct countries in Asia and has further diversified into 7 second-order clades (B4.1–B4.7). As with the G lineage, the B lineage displayed a genetic distance between some clades of <6%. Within-clade APD was <6% for all clades, ranging from clade B4.4 at 2.18% (China, 2009–2020) to clade B4.7 at 5.04% (Cambodia, China, Japan, Laos, Myanmar, Russian Federation, Tajikistan, Vietnam, 2012–2021) (Figure 6; Appendix 1 Table 6; Appendix 2 Table 2).

Within the B lineage, all viruses sequenced in the past 5 years fall into clades B4.4–B4.7. Most of the viruses in each B clade were identified in poultry (Figure 7; Appendix 1 Table 6), and almost all clades contain viruses collected from mammal species. Most viruses recovered from human infections (82.9%, 34/41) fall within clade B4.7 (2014–2021).

Online Tool for Automatic Sequence Classification

To enhance the accessibility of our A/H9 influenza virus clade classification and nomenclature system, we developed an intuitive online tool (https://nmdc.cn/influvar/tools/H9aiv) with the aim to provide a user-friendly interface, making it easier for researchers and stakeholders to use and navigate the classification system. Users will be able to access the website and submit their A/H9 sequence file in the appropriate format. The tool will assign the corresponding clade and will provide a phylogenetic tree.

Discussion

Establishing a unified nomenclature system for A/H9 influenza viruses is essential for the provision of communication to address the increasing global threat

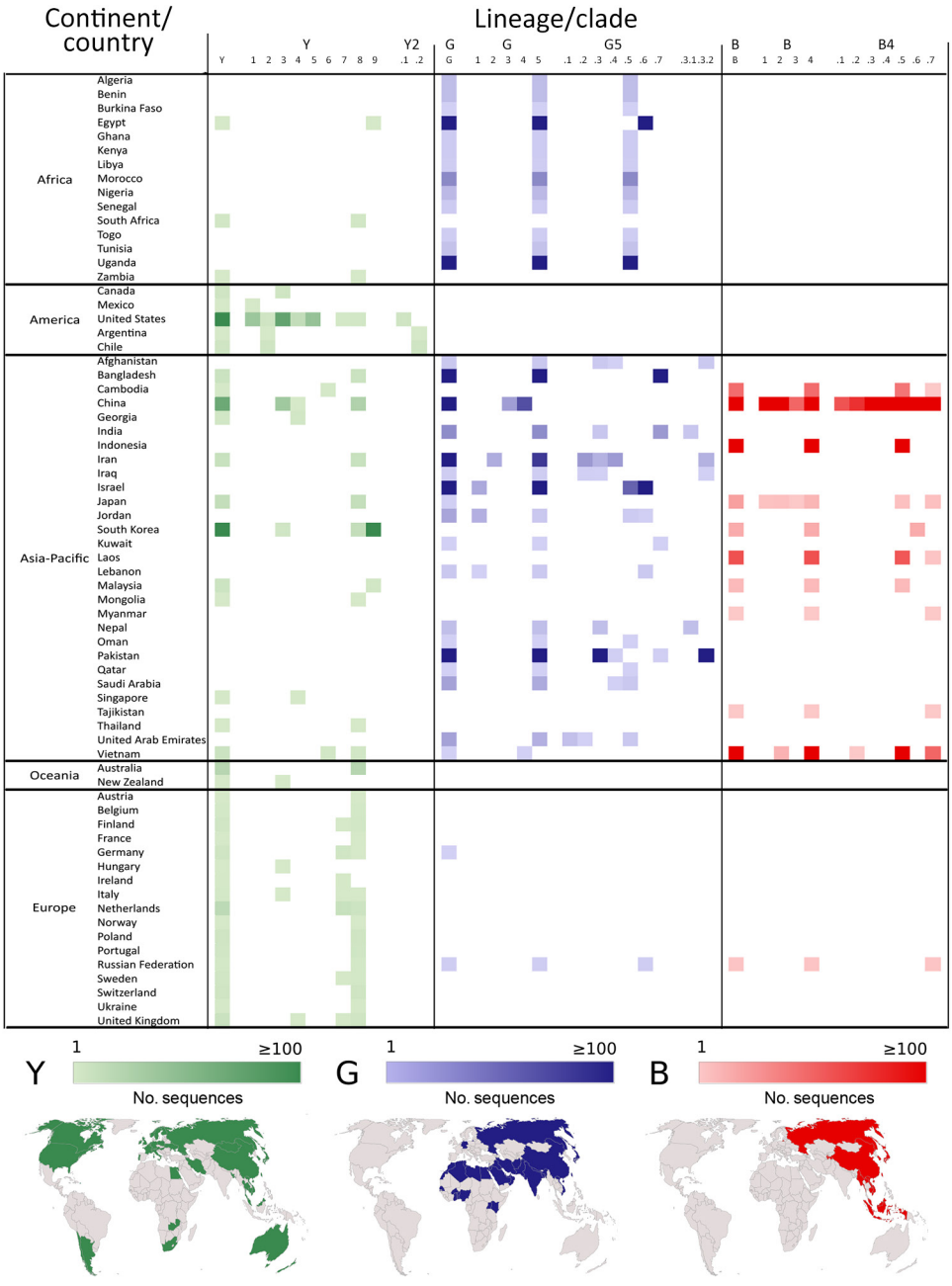


Figure 6. Geographic distribution of each lineage and clade for A/H9 influenza viruses as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. The heat map displays the number of sequences for each lineage and clade per country. Countries were ordered by their macro-region (upper panels). Each country displaying >1 sequence was colored on the map in green (Y lineage), blue (G lineage), or red (B lineage) (bottom panels).

of the viruses. This system is characterized by 4 key hallmarks: comprehensiveness, robustness, inclusiveness, and practicality.

For comprehensiveness, we have created a system based on the HA gene sequence data of all the A/H9 viruses (including different neuraminidase subtypes) collected worldwide during 1976–2022. Those data enable understanding of the complete evolution of A/H9 AIVs on a regional and global scale, as well as the pattern of geographic and host distribution of lineages and clades. For example, the G and Y lineages are found in multiple continents, indicating a wide geographic diffusion, whereas the B lineage is isolated primarily in Asia and has occasional spillover to Europe through Russia. In addition, we present a record of all A/H9 AIVs clades, past and presently circulating, to provide a comprehensive understanding of the virus evolution trajectory in different host and geographic contexts.

Our system demonstrates robustness by being built from phylogenetic analysis of high-quality, nearly complete (>75%) HA gene sequences from

multiple datasets (complete and down-sampled) and software packages. We have included all unique publicly available sequences, from a wide variety of origins and collection dates, to ensure the comprehensiveness and representativeness of our dataset. We only accepted high-supported clades consistently recognized across all trees, regardless of the datasets and approaches used, and characterized by a specific set of shared amino acid mutations. For example, there are 4 widely and internationally recognized lineages of A/H9 viruses, American, Y439, G1, and BJ/94 (Y280). However, because our classification approach revealed that the HA sequences of American-like and Y439-like lineages are phylogenetically closely related and share an ancestral node, they were combined into a single lineage and renamed as the Y lineage.

The inclusiveness of our system stems from the establishment of multiple criteria to identify and classify novel clades that emerge, incorporating phylogenetic evidence, molecular evidence, and epidemiologic features. Accommodating all the criteria can be challenging, because natural virus evolution is unlikely to always produce discrete boundaries. Among the criteria, the phylogenetic relationship and epidemiologic characteristics are the most useful for classification, whereas a certain degree of flexibility can be acceptable for bootstrap or APD value, provided there is strong evidence that the monophyletic clade exists and is epidemiologically relevant. This flexibility should enable the monitoring of ongoing epidemics in that clade, especially the emergence and evolution of antigenically distinct strains. The balance between the fixed criteria and topology or epidemiologic characters is also applied in the classification of H5 influenza viruses and SARS-CoV-2 (*41*,*42*).

Practicality is the most relevant hallmark of our system. We have fully considered the needs of users and have made major efforts to achieve full applicability of our system. Users can follow simple and clear criteria to identify novel clades, understand the biologic and epidemiologic significance of clades, and communicate results of their investigations through an informative nomenclature system. In assigning lineage names, we considered the traditional nomenclature in use to create consistency with historic classifications and to enable comparisons with previous studies.

In summary, our proposed nomenclature system and online tool for A/H9 AIVs is valuable for understanding the ongoing evolution and spread of this influenza subtype. Our nomenclature system will improve communication among international and local organizations and laboratories working on A/H9 viruses. The use of our system will enhance the global prevention and control capacity of A/H9 AIVs outbreaks.

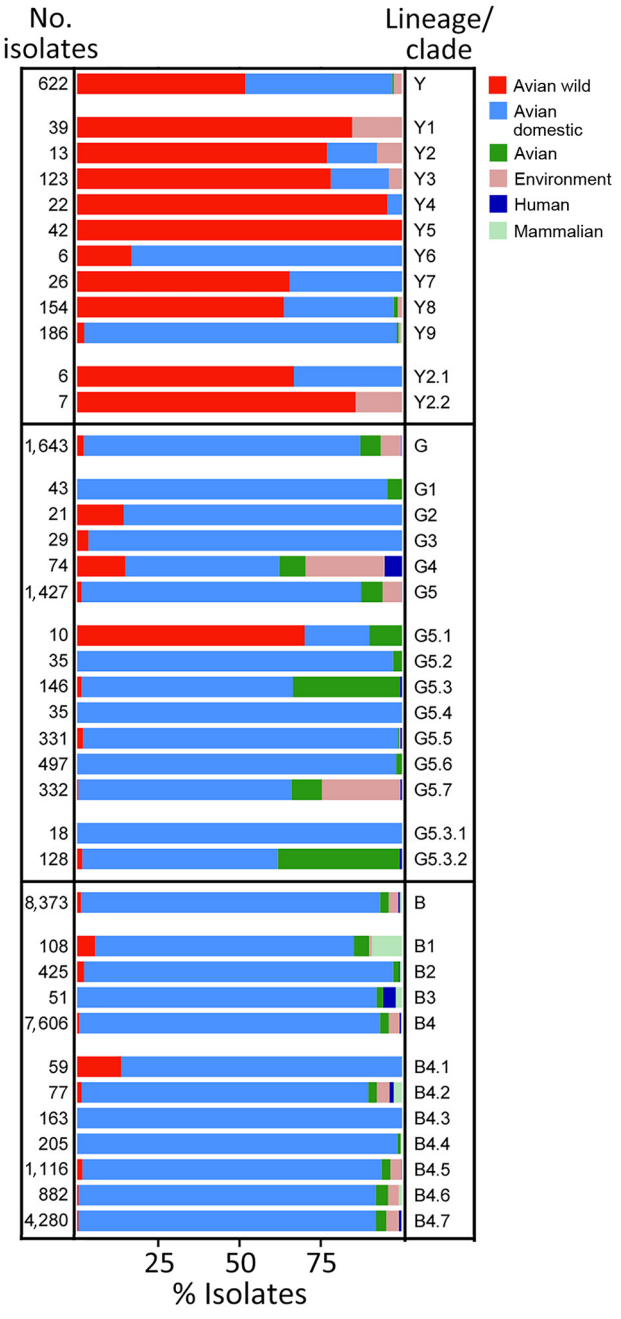


Figure 7. Host distribution of each lineage and clade for A/H9 influenza viruses as part of a proposed global classification and nomenclature system for A/H9 influenza viruses. The bar chart illustrates the percentage of host composition for each lineage and clade. Hosts are grouped into the following categories, represented as colors on each bar: avian wild, avian domestic, avian (birds that are not identified as wild or domestic), environment, human, and mammalian (other than human).

Members of the International H9 Evolution Consortium: Celia Abolnik (South Africa), Ian Brown (United Kingdom), Charles Todd Davis (United States), Xiangjun Du (China), Shangang Jia (China), Erik A. Karlsson (Cambodia), Dong-Hun Lee (South Korea), Nicola S. Lewis (United Kingdom), Hualei Liu (China), Yudong Li (China), Yoshihiro Sakoda (Japan), Jianzhong Shi (China), David E Swayne (United States), Frank Y.K.Wong (Australia), and Jinfeng Zeng (China).

Acknowledgments

We thank the WOAH/FAO Network of Expertise on Animal Influenza for their technical expertise and leadership on this project. We thank Jie Cui, Xiang Li, and Francesca Ellero for their support and assistance.

Funding has been provided from the National Key Research and Development Program of China (grant nos. 2021YFD1800202 and 2022YFF0802403) and the National Natural Science Foundation of China (grant no. 32192450). Partial funding was provided by the EU through the NextGeneration EU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (project no. PE00000007, INF-ACT), the European Union's Horizon 2020 Research and Innovation program (grant no. 874735), the Biological Sciences Research Council ecology and evolution of infectious diseases project (grant no. BB/V011286/1), Biological Sciences Research Council Institute Strategic Program Grant for the control of infectious diseases (grant no. BBS/E/D/20002173), and the Innovation and Technology Commission of the Hong Kong Special Administrative Region, China.

About the Author

Dr. Fusaro is a scientist at the European Union Reference Laboratory for Avian Influenza and Newcastle Disease in Italy. Her primary research interests include molecular epidemiology, intra- and inter-host evolution, gene flow, and cross-species transmission of RNA viruses.

References

1. Zhang J, Huang L, Liao M, Qi W. H9N2 avian influenza viruses: challenges and the way forward. Lancet Microbe. 2023;4:e70–1. https://doi.org/10.1016/S2666-5247(22)00305-6
2. Li X, Shi J, Guo J, Deng G, Zhang Q, Wang J, et al. Genetics, receptor binding property, and transmissibility in mammals of naturally isolated H9N2 Avian Influenza viruses. PLoS Pathog. 2014;10:e1004508. https://doi.org/10.1371/journal.ppat.1004508
3. Wang X, Liu K, Guo Y, Pei Y, Chen X, Lu X, et al. Emergence of a new designated clade 16 with significant antigenic drift in hemagglutinin gene of H9N2 subtype avian influenza virus in eastern China. Emerg Microbes Infect. 2023; 12:2249558. https://doi.org/10.1080/22221751.2023.2249558
4. Bonfante F, Mazzetto E, Zanardello C, Fortin A, Gobbo F, Maniero S, et al. A G1-lineage H9N2 virus with oviduct tropism causes chronic pathological changes in the infundibulum and a long-lasting drop in egg production. Vet Res (Faisalabad). 2018;49:83. https://doi.org/10.1186/s13567-018-0575-1
5. Pu J, Wang S, Yin Y, Zhang G, Carter RA, Wang J, et al. Evolution of the H9N2 influenza genotype that facilitated the genesis of the novel H7N9 virus. Proc Natl Acad Sci U S A. 2015;112:548–53. https://doi.org/10.1073/pnas.1422456112
6. Li C, Yu K, Tian G, Yu D, Liu L, Jing B, et al. Evolution of H9N2 influenza viruses from domestic poultry in mainland China. Virology. 2005;340:70–83. https://doi.org/10.1016/j.virol.2005.06.025
7. Dong J, Zhou Y, Pu J, Liu L. Status and challenges for vaccination against avian H9N2 influenza virus in China. Life (Basel). 2022;12:12. https://doi.org/10.3390/life12091326
8. Carnaccini S, Perez DR. H9 Influenza viruses: an emerging challenge. Cold Spring Harb Perspect Med. 2020;10:10. https://doi.org/10.1101/cshperspect.a038588
9. Zhang N, Quan K, Chen Z, Hu Q, Nie M, Xu N, et al. The emergence of new antigen branches of H9N2 avian influenza virus in China due to antigenic drift on hemagglutinin through antibody escape at immunodominant sites. Emerg Microbes Infect. 2023;12:2246582. https://doi.org/10.1080/22221751.2023.2246582
10. Yan W, Cui H, Engelsma M, Beerens N, van Oers MM, de Jong MCM, et al. Molecular and antigenic characterization of avian H9N2 viruses in southern China. Microbiol Spectr. 2022;10:e0082221. https://doi.org/10.1128/spectrum.00822-21
11. Pu J, Yin Y, Liu J, Wang X, Zhou Y, Wang Z, et al. Reassortment with dominant chicken H9N2 influenza virus contributed to the fifth H7N9 virus human epidemic. J Virol. 2021;95:95. https://doi.org/10.1128/JVI.01578-20
12. Peacock THP, James J, Sealy JE, Iqbal M. A global perspective on H9N2 avian influenza virus. Viruses. 2019;11:11. https://doi.org/10.3390/v11070620
13. Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Mirinavičiūtė G, Niqueux É, et al.; European Food Safety Authority; European Centre for Disease Prevention and Control; European Union Reference Laboratory for Avian Influenza. Avian influenza overview September–December 2023. EFSA J. 2023;21:e8539.
14. Chan RWY, Chan LLY, Mok CKP, Lai J, Tao KP, Obadan A, et al. Replication of H9 influenza viruses in the human ex vivo respiratory tract, and the influence of neuraminidase on virus release. Sci Rep. 2017;7:6208. https://doi.org/10.1038/s41598-017-05853-5
15. Yang R, Sun H, Gao F, Luo K, Huang Z, Tong Q, et al. Human infection of avian influenza A H3N8 virus and the viral origins: a descriptive study. Lancet Microbe. 2022; 3:e824–34. https://doi.org/10.1016/S2666-5247(22)00192-6
16. Wang Y, Niu S, Zhang B, Yang C, Zhou Z. The whole genome analysis for the first human infection with H10N3 influenza virus in China. J Infect. 2021:S0163–4453(21)00318-2.
17. Bi Y, Chen Q, Wang Q, Chen J, Jin T, Wong G, et al. Genesis, evolution and prevalence of H5N6 avian influenza viruses in China. Cell Host Microbe. 2016;20:810–21. https://doi.org/10.1016/j.chom.2016.10.022
18. Qi W, Zhou X, Shi W, Huang L, Xia W, Liu D, et al. Genesis of the novel human-infecting influenza A(H10N8) virus and potential genetic diversity of the virus in poultry, China. Euro Surveill. 2014;19:20841. https://doi.org/10.2807/1560-7917.ES2014.19.25.20841

19. Liu D, Shi W, Gao GF. Poultry carrying H9N2 act as incubators for novel human avian influenza viruses. Lancet. 2014;383:869. https://doi.org/10.1016/S0140-6736(14)60386-X
20. Liu D, Shi W, Shi Y, Wang D, Xiao H, Li W, et al. Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. Lancet. 2013;381:1926–32. https://doi.org/10.1016/S0140-6736(13)60938-1
21. Liu WJ, Wu Y, Bi Y, Shi W, Wang D, Shi Y, et al. Emerging HxNy influenza A viruses. Cold Spring Harb Perspect Med. 2022;12:12. https://doi.org/10.1101/cshperspect.a038406
22. Zhuang Q, Wang S, Liu S, Hou G, Li J, Jiang W, et al. Diversity and distribution of type A influenza viruses: an updated panorama analysis based on protein sequences. Virol J. 2019;16:85. https://doi.org/10.1186/s12985-019-1188-7
23. Li C, Wang S, Bing G, Carter RA, Wang Z, Wang J, et al. Genetic evolution of influenza H9N2 viruses isolated from various hosts in China from 1994 to 2013. Emerg Microbes Infect. 2017;6:e106. https://doi.org/10.1038/emi.2017.94
24. Shepard SS, Davis CT, Bahl J, Rivailler P, York IA, Donis RO. LABEL: fast and accurate lineage assignment with assessment of H5N1 and H9N2 influenza A hemagglutinins. PLoS One. 2014;9:e86921. https://doi.org/10.1371/journal.pone.0086921
25. Jiang W, Liu S, Hou G, Li J, Zhuang Q, Wang S, et al. Chinese and global distribution of H9 subtype avian influenza viruses. PLoS One. 2012;7:e52671. https://doi.org/10.1371/journal.pone.0052671
26. Fusaro A, Monne I, Salviato A, Valastro V, Schivo A, Amarin NM, et al. Phylogeography and evolutionary history of reassortant H9N2 viruses with potential human health implications. J Virol. 2011;85:8413–21. https://doi.org/10.1128/JVI.00219-11
27. Liu S, Ji K, Chen J, Tai D, Jiang W, Hou G, et al. Panorama phylogenetic diversity and distribution of type A influenza virus. PLoS One. 2009;4:e5022. https://doi.org/10.1371/journal.pone.0005022
28. Dalby AR, Iqbal M. A global phylogenetic analysis in order to determine the host species and geography dependent features present in the evolution of avian H9N2 influenza hemagglutinin. PeerJ. 2014;2:e655. https://doi.org/10.7717/peerj.655
29. Guan Y, Shortridge KF, Krauss S, Webster RG. Molecular characterization of H9N2 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? Proc Natl Acad Sci U S A. 1999;96:9363–7. https://doi.org/10.1073/pnas.96.16.9363
30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010
31. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300
32. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285
33. Han AX, Parker E, Scholer F, Maurer-Stroh S, Russell CA. Phylogenetic clustering by linear integer programming (PhyCLIP). Mol Biol Evol. 2019;36:1580–95. https://doi.org/10.1093/molbev/msz053
34. Chernomor O, Minh BQ, Forest F, Klaere S, Ingram T, Henzinger M, et al. Split diversity in constrained conservation prioritization using integer linear programming. Methods Ecol Evol. 2015;6:83–91. https://doi.org/10.1111/2041-210X.12299
35. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95. https://doi.org/10.1093/molbev/mst024
36. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21. https://doi.org/10.1093/sysbio/syq010
37. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9. https://doi.org/10.1093/molbev/msy096
38. Sagulenko P, Puller V, Neher RA. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. https://doi.org/10.1093/ve/vex042
39. Guo YJ, Krauss S, Senne DA, Mo IP, Lo KS, Xiong XP, et al. Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia. Virology. 2000;267:279–88. https://doi.org/10.1006/viro.1999.0115
40. Homme PJ, Easterday BC. Avian influenza virus infections. I. Characteristics of influenza A-turkey-Wisconsin-1966 virus. Avian Dis. 1970;14:66–74. https://doi.org/10.2307/1588557
41. World Health Organization/World Organization for Animal Health/Food and Agriculture Organization H5N1 Evolution Working Group. Toward a unified nomenclature system for highly pathogenic avian influenza virus (H5N1). Emerg Infect Dis. 2008;14:e1. https://doi.org/10.3201/eid1407.071681
42. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5:1403–7. https://doi.org/10.1038/s41564-020-0770-5

Address for correspondence: Jinhua Liu, China Agricultural University, No. 2 Yuanmingyuan West Rd, Beijing 100193, China; email: ljh@cau.edu.cn

Article DOI: https://doi.org/10.3201/eid3008.231176

EID cannot ensure accessibility for supplementary materials supplied by authors. Readers who have difficulty accessing supplementary content should contact the authors for assistance.

Proposal for a Global Classification and Nomenclature System for A/H9 Influenza Viruses

Appendix 1

Methods

Sequence Data and Metadata Preparation

To provide a comprehensive picture of the A/H9 genetic diversity, we generated a dataset of the hemagglutinin gene that included all the H9-HA sequences available on the GISAID (www.epicov.org) and GenBank (www.ncbi.nlm.nih.gov/nucleotide/) public databases (accessed on July 18, 2022). Multiple sequence entries (i.e., sequences obtained from the same sample, deposited multiple times or available in both databases) or sequences obtained from laboratory-derived viruses were discarded. Sequences were further filtered for length and quality. Specifically, all the sequences with more than 5 ambiguous bases and with a length <1275 bp (75% of the coding region) were removed. If no sequences matching these criteria were available for a specific country and collection year, sequences with a length >900pb were accepted to have the most exhaustive dataset as possible in terms of sequence representativeness and quality.

Sequences alignment, obtained using MAFFT v7.0 (*1*), was manually curated and nucleotides outside the coding region of the mature HA gene were trimmed. After removing sequences containing out-of-frame indels, a preliminary Maximum Likelihood (ML) phylogenetic tree using IQ-TREE v1.6 (*2*) was generated from this dataset to test 'clocklikeness'

of the dated-tip phylogeny using TempEst (*3*). A good correlation between the collection dates (year) of the virus and the divergence from the root was observed (r = 0.82). This analysis helped us to identify outlier sequences, which may be due to mislabeling of the virus (incorrect year of collection), sample contamination by an older virus or sequencing errors. All the outlier sequences were removed from the dataset. In the process, early strains such as A/turkey/Wisconsin/1/1966 were removed.

Furthermore, only the oldest sequence was kept among sequences with 100% identity that were collected in the same country. Finally, since mosaic influenza genome segments have previously been described as resulting from laboratory contamination or artifacts, or from a natural homologous recombination (*4*), the dataset was screened for mosaic structures using the RDP, Geneconv, Maxchi, BootScan, 3Seq and Chimaera methods available in the RDP package v.4 (*5*), applying default settings. The Simplot program v.3.5 was also used to define the locations of recombination breakpoints (*6*). The potential mosaic sequences identified by at least two methods with $p < 1\times10^{-10}$ were considered unreliable and were removed from the dataset. A final dataset (Complete dataset) containing 10,638 HA sequences and the related information, including accession numbers, were produced after the quality check process (Appendix 1 Table 1).

Testing and Selection of PhyCLIP Parameters

PhyCLIP utilizes an integer linear programming (ILP) approach that optimally delineates a tree into statistically principled clusters (*7*), to optimize our clustering results, we tested a range of values for three key parameters: 1) minimum number of sequences (S) that can be quantified as a cluster (S = 3, 5, 10), 2) false discovery rate (FDR) used to infer that the diversity observed for every combinatorial pair of output clusters is significantly distinct from one another (FDR range from 0.1 to 0.2 in increments of 0.05), and 3) multiple of deviations (γ) from the grand median of the mean pairwise sequence patristic distance that defines the within-cluster

divergence limit (WCL) (γ range from 1 to 3 in increments of 1). We used the clustering resulting from the optimal parameter (S = 5, FDR = 0.2 and γ = 3) combinations as a reference for the assignment of clades.

References:

1. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. PubMed https://doi.org/10.1093/molbev/mst010

2. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. PubMed https://doi.org/10.1093/molbev/msu300

3. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:vew007. PubMed https://doi.org/10.1093/ve/vew007

4. Lam TT, Chong YL, Shi M, Hon CC, Li J, Martin DP, et al. Systematic phylogenetic analysis of influenza A virus reveals many novel mosaic genome segments. Infect Genet Evol. 2013;18:367–78. **PMID: 23548803**

5. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26:2462–3. PubMed https://doi.org/10.1093/bioinformatics/btq467

6. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73:152–60. PubMed https://doi.org/10.1128/JVI.73.1.152-160.1999

7. Han AX, Parker E, Scholer F, Maurer-Stroh S, Russell CA. Phylogenetic clustering by linear integer programming (PhyCLIP). Mol Biol Evol. 2019;36:1580–95. PubMed https://doi.org/10.1093/molbev/msz053

8. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. PubMed https://doi.org/10.1093/ve/vex042

9. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34:4121–3. PubMed https://doi.org/10.1093/bioinformatics/bty407

10. Carnaccini S, Perez DR. H9 influenza viruses: an emerging challenge. Cold Spring Harb Perspect Med. 2020;10:a038588. PubMed https://doi.org/10.1101/cshperspect.a038588

11. Guan Y, Shortridge KF, Krauss S, Webster RG. Molecular characterization of H9N2 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? Proc Natl Acad Sci U S A. 1999;96:9363–7. PubMed https://doi.org/10.1073/pnas.96.16.9363

12. Liu S, Ji K, Chen J, Tai D, Jiang W, Hou G, et al. Panorama phylogenetic diversity and distribution of Type A influenza virus. PLoS One. 2009;4:e5022. PubMed https://doi.org/10.1371/journal.pone.0005022

13. Fusaro A, Monne I, Salviato A, Valastro V, Schivo A, Amarin NM, et al. Phylogeography and evolutionary history of reassortant H9N2 viruses with potential human health implications. J Virol. 2011;85:8413–21. PubMed https://doi.org/10.1128/JVI.00219-11

14. Dalby AR, Iqbal M. A global phylogenetic analysis in order to determine the host species and geography dependent features present in the evolution of avian H9N2 influenza hemagglutinin. PeerJ. 2014;2:e655. PubMed https://doi.org/10.7717/peerj.655

15. Li C, Wang S, Bing G, Carter RA, Wang Z, Wang J, et al. Genetic evolution of influenza H9N2 viruses isolated from various hosts in China from 1994 to 2013. Emerg Microbes Infect. 2017;6:e106. PubMed https://doi.org/10.1038/emi.2017.94

16. Zhuang Q, Wang S, Liu S, Hou G, Li J, Jiang W, et al. Diversity and distribution of type A influenza viruses: an updated panorama analysis based on protein sequences. Virol J. 2019;16:85. PubMed https://doi.org/10.1186/s12985-019-1188-7

Appendix 1 Table 1. Ultra-fast bootstrap (UFB) values, standard bootstraps (SB) and SH-like supports (aLRT SH-like) obtained for each clade nodes from the analyses of different datasets (complete and pilot datasets) using different software (IQ-TREE and PhyML).

Dataset		**Complete datasets**		**Pilot datasets**	
Software		**IQ-TREE**	**PhyML**	**IQ-TREE**	**IQ-TREE**
Lineage	**Clades**	**UFB**	**aLRT SH-like**	**UFB**	**SB**
G	G1	100	1	100	100
	G2	100	0.997	100	99
	G3	100	0.999	100	100
	G4	100	0.996	100	100
	G5.1	99	1	100	100
	G5.2	99	0.733	100	96
	G5.3.1	100	0.999	100	100
	G5.3.2	100	1	100	93
	G5.4	95	1	100	100
	G5.5	100	0.992	99	100
	G5.6	100	1	100	100
	G5.7	100	0.995	85	66
Y	Y1	100	1	100	100
	Y2.1	100	0.967	99	100
	Y2.2	100	1	100	100
	Y3	96	0.85	100	100
	Y4	100	1	100	100
	Y5	100	0.852	99	92
	Y6	100	1	100	100
	Y7	84	0.868	97	77
	Y8	100	1	100	100
	Y9	100	1	100	100

Dataset		Complete datasets		Pilot datasets	
Software		IQ-TREE	PhyML	IQ-TREE	IQ-TREE
Lineage	Clades	UFB	aLRT SH-like	UFB	SB
B	B1	100	0.959	99	86
	B2	94	0.91	99	96
	B3	81	0.927	99	91
	B4.1	100	0.998	100	98
	B4.2	100	0.912	97	60
	B4.3	100	0.925	100	99
	B4.4	93	0.988	100	100
	B4.5	91	0.908	99	90
	B4.6	82	0.92	97	73
	B4.7	82	0.952	100	100

Appendix 1 Table 2. Representative amino acid residues of each clade.

Lineage	Clade	Amino acid mutations based on ancestral reconstruction using TreeTime *(8, 9)*
Y	Y1	V288I, V317A, I451V, R479K
	Y2	N45D, F104L, N109R, Q112K, V153F, N161T, T182N, I249V, K276R, V288I, D319N, R358K, V365I, K487R
	Y3	V153F, N267D, V352T
	Y4	T54K, E72T, H146Q, V153I, N264K, R320K, V365I, E501D, K505R
	Y5	S103A, N398S
	Y6	K131A, H146Q, V153F, D155N, E162N, N165S, A317V, D319G, E363V, N398S, I451M
	Y7	V269I
	Y8	I451R
	Y9	I20V, N94R, N109R, Q112K, I114L, Q115L, I116L, T120R, I121T, V153I, N161D, E162W, T182I, V194I, D319G, N455K, Q480L, Q483K, G502E, L527M
	Y2.1	L69I, I166V, V206M, V302A
	Y2.2	K164E, E459D, F523L
G	G1	S83P, V95I, G135D, T195A, I202V, V213A, V249I, N264K, V300I, N374S, V393I
	G2	G135D, D178E, S370T
	G3	N264T, V411I

Lineage	Clade	Amino acid mutations based on ancestral reconstruction using TreeTime *(8, 9)*
	G4	H34Q, A132S, S165N, E180D, N183T, D198E, L216Q, R301K
	G5	A108S, A132S, S140N, S148N, I186T, D198N, N200D, M206L
	G5.1	N148S, L150F, T182R, T186I, L216Q, I217T, V226I, I288V, I365V
	G5.2	S486A
	G5.3	S150L, N165D
	G5.4	R317K, N374T
	G5.5	K483T
	G5.6	L150V
	G5.7	D262N, T295N, V496I
	G5.3.1	Q112R, R162Q, Q467H, L488I
	G5.3.2	V24I, H28Q, H48R, S108A, T120A, T127D, D135N, V153I, V169I, R317K, D359G, I365V, V376I, K381R
B	B1	N395A
	B2	I153V
	B3	N148S
	B4	N264K, V269A
	B4.1	T395N
	B4.2	K481R
	B4.3	K492R
	B4.4	S125T
	B4.5	D221N, R236K
	B4.6	D221N
	B4.7	D135G, E163G

Note: all HA positions follow the H9 numbering. The red color indicates that these sites are associated with host tropism, virulence or identified antigenic sites (*10*).

Appendix 1 Table 3. Comparison of previous and current nomenclature systems.

Previous studies			Current study
Publication	**Clade classification and nomenclature**		**Clade classification and nomenclature**
Guan Y et al. (1999) (*11*)	G1		G1-G5
	BJ94(Y280)		B1-B4
	Y439		Y4-Y9
	TY66		Y1-Y3
Liu S et al. (2009) (*12*)	h9.1, h9.2		Y1-Y3
	h9.3		Y4-Y9
	h9.4	h9.4.1	G1-G5
		h9.4.2	B1-B4
Fusaro A et al. (2011) (*13*)	G1-A		G1
	G1-B		G5
	G1-C		G-like
	G1-D		G2
Dalby AR et al. (2014) (*14*)	Clade A		Y1-Y9
	Clade B		G1-G5
	Clade C, Main Chinese Clade		B1-B4
Li C et al. (2017) (*15*)	0–15	0, 5–7, 9–11, 13	B-like
		1	Y-like
		2	Y3, Y8
		3	G3
		4	G4
		8	B1
		12	B2
		14	B3
		15	B4
Zhuang Q et al. (2019) (*16*)	H9.1		Y1-Y9
	H9.2a		G1-G5
	H9.2b		B1-B4
Carnaccini S et al. (2020) (10)	h9.1 (h9.1.2)		Y1-Y3
	Y439-h9.2 (Korea h9.2.2)		Y4-Y9
	BJ94-h9.3		B1-B4
	G1-h9.4.1 Eastern		G1-G4
	G1-h9.4.2 Western		G5

Appendix 1 Table 4. Temporal, spatial and host distribution characteristics of lineage Y.

Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
Y	1976–2021	Argentina, Australia, Austria, Bangladesh, Belgium, Cambodia, Canada, Chile, China, Egypt, Finland, France, Georgia, Germany, Hungary, Iran, Ireland, Italy, Japan, Malaysia, Mexico, Mongolia, Netherlands, New Zealand, Norway, Poland, Portugal, Russian Federation, Singapore, South Africa, South Korea, Sweden, Switzerland, Thailand, Ukraine, United Kingdom, United States, Vietnam, Zambia	Avian, Avian domestic, Avian wild, Environment, Mammalian	622	A_Duck_Hong_Kong_86_1976
Y1	2000–2016	Mexico, United States	Avian wild, Environment	39	A_shorebird_Delaware_Bay_277_2000
Y2	1993–2017	Argentina, Chile, United States	Avian domestic, Avian wild, Environment	13	A_Quail_Arkansas_29209–1_1993
Y3	1976–2020	Canada, China, Hungary, Italy, New Zealand, South Korea, United States	Avian domestic, Avian wild, Environment	123	A_Duck_Hong_Kong_86_1976
Y4	2010–2019	China, Georgia, Singapore, United Kingdom, United States	Avian domestic, Avian wild	22	A_chicken_England_1415–51184_2010
Y5	2003–2007	United States	Avian wild	42	A_ruddy_turnstone_Delaware_1016406_2003
Y6	2009–2018	Cambodia, Vietnam	Avian domestic, Avian wild	6	A_duck_Vietnam_OIE-2313_2009
Y7	1993–2010	Finland, Germany, Ireland, Italy, Netherlands, Sweden, United Kingdom, United States	Avian domestic, Avian wild	26	A_mallard_Ireland_PV46B_1993

Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
Y8	1995–2021	Australia, Austria, Bangladesh, Belgium, China, Finland, France, Germany, Iran, Italy, Japan, Mongolia, Netherlands, Norway, Poland, Portugal, Russian Federation, South Africa, South Korea, Sweden, Switzerland, Thailand, Ukraine, United Kingdom, United States, Vietnam, Zambia	Avian, Avian domestic, Avian wild, Environment	154	A_ostrich_South_Africa_9508103_1995
Y9	1996–2018	Egypt, Malaysia, South Korea	Avian, Avian domestic, Avian wild, Environment, Mammalian	186	A_chicken_Korea_25232–96006_1996
Y-like	1978–2001	China, Japan, Malaysia	Avian domestic	11	A_Duck_Hong_Kong_366_1978
Y2.1	1993–1996	United States	Avian domestic Avian wild	6	A_Quail_Arkansas_29209–1_1993
Y2.2	2007–2017	Argentina, Chile	Avian wild Environment	7	A_rosy-billed_pochard_Argentina_CIP051–559_2007

Appendix 1 Table 5. Temporal, spatial and host distribution characteristics of lineage G.

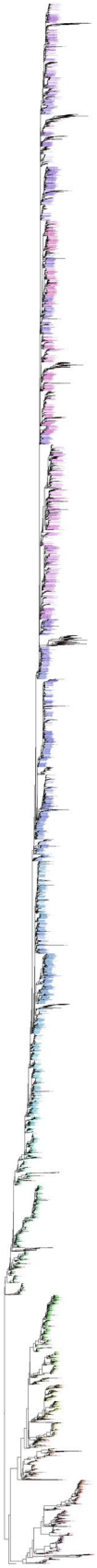
Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
G	1997–2022	Afghanistan, Algeria, Bangladesh, Benin, Burkina Faso, China, Egypt, Germany, Ghana, India, Iran, Iraq, Israel, Japan, Jordan, Kenya, Kuwait, Lebanon, Libya, Morocco, Nepal, Nigeria, Oman, Pakistan, Qatar, Russian Federation, Saudi Arabia, Senegal, Togo, Tunisia, Uganda, United Arab Emirates, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human	1643	A_quail_Hong_Kong_G1_1997
G1	2003–2007	Israel, Jordan, Lebanon	Avian, Avian domestic	43	A_chicken_Jordan_12_2003
G2	1999–2005	Iran	Avian domestic, Avian wild	21	A_chicken_Iran_705_1999
G3	2000–2004	China	Avian domestic, Avian wild	29	A_quail_Shantou_782_2000
G4	1997–2017	China, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human	74	A_quail_Hong_Kong_G1_1997
G5	1998–2022	Afghanistan, Algeria, Bangladesh, Benin, Burkina Faso, Egypt, Ghana, India, Iran, Iraq, Israel, Jordan, Kenya, Kuwait, Lebanon, Libya, Morocco, Nepal, Nigeria, Oman, Pakistan, Qatar, Russian Federation, Saudi Arabia, Senegal, Togo, Tunisia, Uganda, United Arab Emirates	Avian, Avian domestic, Avian wild, Environment, Human	1427	A_chicken_Iran_725_1998
G-like	1997–2007	Germany, Iran, Israel, Japan, Pakistan, Saudi Arabia, United Arab Emirates	Avian, Avian domestic	49	A_parakeet_Chiba_1_1997
G5.1	2005–2011	United Arab Emirates	Avian, Avian domestic, Avian wild	10	A_white_bellied_bustard_United_Arab_Emirates_1127_1_2005

Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
G5.2	1998–2009	Iran, Iraq, United Arab Emirates	Avian, Avian domestic	35	A_chicken_Iran_725_1998
G5.3	2007–2022	Afghanistan, India, Iran, Iraq, Nepal, Pakistan	Avian, Avian domestic, Avian wild, Human	146	A_Chicken_Nepal_JHAPA_28_2007
G5.4	2000–2013	Afghanistan, Iran, Pakistan, Saudi Arabia	Avian domestic	35	A_chicken_Saudi_Arabia_2525_2000
G5.5	2006–2021	Algeria, Benin, Burkina Faso, Ghana, Israel, Jordan, Kenya, Libya, Morocco, Nigeria, Oman, Qatar, Saudi Arabia, Senegal, Togo, Tunisia, Uganda, United Arab Emirates	Avian, Avian domestic, Avian wild, Environment, Human	331	A_avian_Libya_RV35D_2006
G5.6	2006–2021	Egypt, Israel, Jordan, Lebanon, Russian Federation	Avian, Avian domestic	497	A_chicken_Israel_1638_2006
G5.7	2003–2022	Bangladesh, India, Kuwait, Pakistan	Avian, Avian domestic, Avian wild, Environment, Human	332	A_chicken_Chandigarh_2048_2003
G5.3.1	2007–2013	India, Nepal	Avian domestic	18	A_Chicken_Nepal_JHAPA_28_2007
G5.3.2	2008–2022	Afghanistan, Iran, Iraq, Pakistan	Avian, Avian domestic, Avian wild, Human	128	A_chicken_Afghanistan_329–6vir09-AFG-Khost9_2008

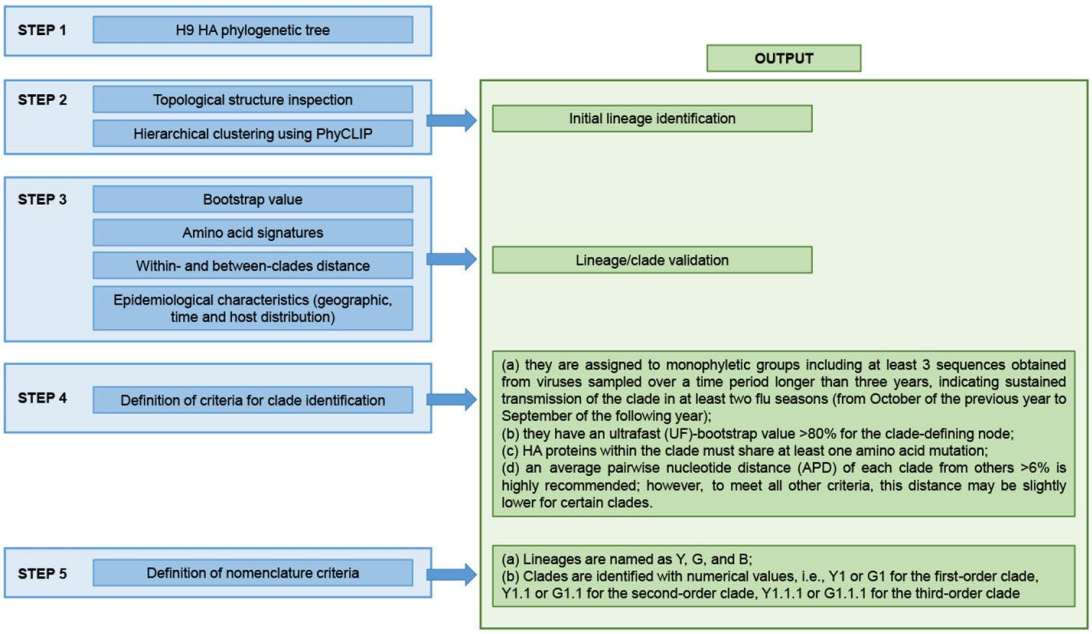
Appendix 1 Table 6. Temporal, spatial and host distribution characteristics of lineage B.

Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
B	1994–2021	Cambodia, China, Indonesia, Japan, Laos, Malaysia, Myanmar, Russian Federation, South Korea, Tajikistan, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human, Mammalian	8373	A_chicken_Beijing_1_1994
B1	1997–2013	China, Japan	Avian, Avian domestic, Avian wild, Environment, Mammalian	108	A_Chicken_Sichuan_5_1997
B2	1996–2016	China, Japan, Vietnam	Avian, Avian domestic, Avian wild, Human, Mammalian	425	A_Quail_Shanghai_8_1996
B3	1998–2017	China, Japan	Avian, Avian domestic, Human, Mammalian	51	A_Shaoguan_408_1998
B4	1999–2021	Cambodia, China, Indonesia, Japan, Laos, Malaysia, Myanmar, Russian Federation, South Korea, Tajikistan, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human, Mammalian	7606	A_chicken_Shandong_JN_1999
B-like	1994–2017	China, Japan	Avian, Avian domestic, Avian wild, Human, Mammalian	183	A_chicken_Beijing_1_1994
B4.1	2000–2005	China	Avian domestic, Avian wild	59	A_partridge_Shantou_5692_2000
B4.2	2003–2014	China, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human, Mammalian	77	A_chicken_Guangdong_B7_2003
B4.3	2002–2013	China	Avian domestic	163	A_chicken_Guangdong_A7_2002
B4.4	2009–2020	China	Avian, Avian domestic, Mammalian	205	A_Duck_Fujian_1753_2009
B4.5	2011–2020	Cambodia, China, Indonesia, Japan, Laos, Malaysia, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human, Mammalian	1116	A_chicken_Anhui_A12_2011

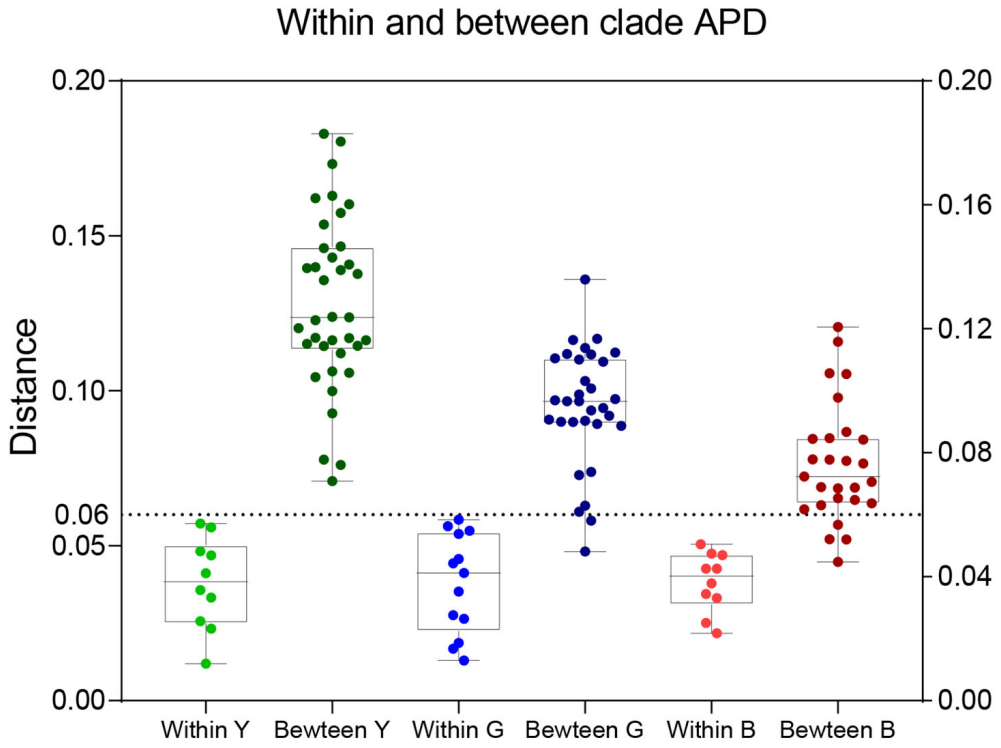
Lineage/Clade	Time range	Countries of origin	Type of host	Number of taxa	The earliest strain
B4.6	2012–2020	China, South Korea	Avian, Avian domestic, Avian wild, Environment, Mammalian	882	A_chicken_Shandong_QD6_2012
B4.7	2012–2021	Cambodia, China, Japan, Laos, Myanmar, Russian Federation, Tajikistan, Vietnam	Avian, Avian domestic, Avian wild, Environment, Human, Mammalian	4280	A_chicken_Anhui_A225_2012



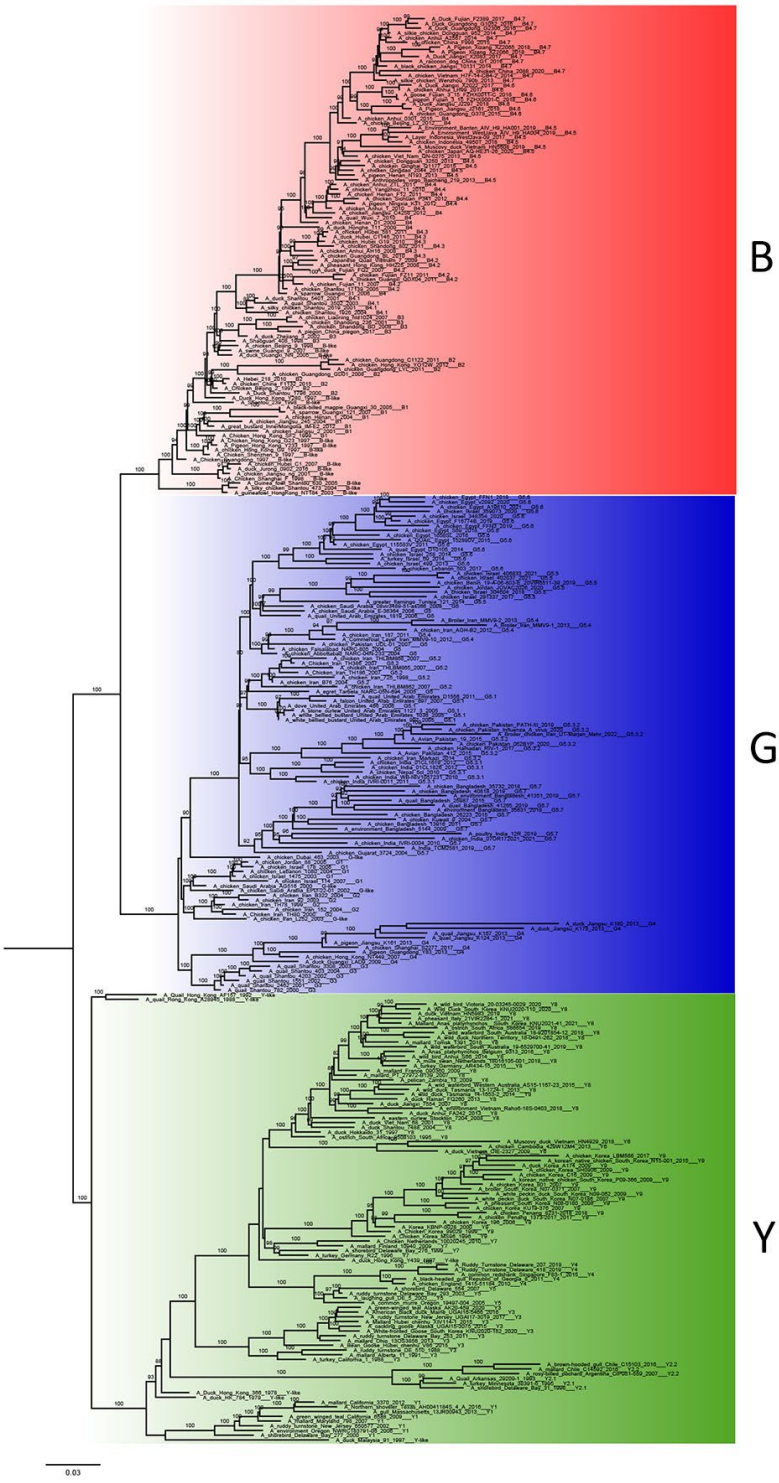
Appendix 1 Figure 1. Clustering based on optimal parameters of PhyCLIP. The numbers on the tree indicate the clade classification of PhyCLIP.



Appendix 1 Figure 2. Scheme of the strategy adopted to classify A/H9 hemagglutinin sequences into lineages and clades.



Appendix 1 Figure 3. Average pairwise distances within- and between-clades of A/H9 influenza viruses. The boxplot displays the average pairwise distances (APD) calculated within and between each clade of the three lineages.



Appendix 1 Figure 4. Pilot Maximum likelihood phylogenetic trees of the H9-HA gene sequences obtained by using the complete small representative dataset available in Appendix 3 for all 3 lineages (Appendix 3 “Pilot Complete Genomes”).