Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatiotemporal genomic epidemiology of Eurasian avian-like H1N1 swine influenza viruses

Hanlin Liu
Sun Yat-sen University
Yao Zhang
Sun Yat-sen University
Jiangnan Yi
South China Agricultural University
Zhijian Zhou
Sun Yat-sen University
Chunyuan Li
Southern China Center for Disease Control and Prevention
Fengxiang Xu
South China Agricultural University
Ruiwen Ren
Southern China Center for Disease Control and Prevention
Ming Liao
South China Agricultural University
Hailiang Sun
hsun@scau.edu.cn

South China Agricultural University

Research Article

Keywords: Swine influenza A virus, Eurasian avian-like H1N1, reassortment, Bayesian phylodynamics, hemagglutinin evolution, positive selection

Posted Date: October 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11044325/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Eurasian avian-like (EA) H1N1 swine influenza viruses pose a zoonotic threat, yet their genomic evolution and dissemination across Eurasia remain incompletely understood. This study investigated genotype diversity, temporal persistence, and geographic spread, and assessed the association between live-swine trade and viral migration to inform surveillance.

Methods

We analysed 1,703 publicly available EA H1N1 genomes sampled from 19 countries between 2009 and 2025. Lineage assignments across all eight gene segments were used to identify genotypes and describe their temporal and geographic distributions. Time-scaled Bayesian phylogenetics and discrete phylogeography were used to reconstruct evolutionary relationships and migration patterns. Generalized linear modelling assessed migration predictors, including directed live-swine trade and sampling intensity. The mixed effects model of evolution was used to detect episodic diversifying selection in hemagglutinin, followed by structural mapping of candidate residues.

Results

We identified 19 genotypes. The wholly EA-derived genotype G1 accounted for 57.0% of genomes and was detected across nearly the entire study period. Many reassortant genotypes carrying A(H1N1)pdm09- or triple-reassortant-derived segments were detected over short intervals, although several showed prolonged persistence. Genotype composition differed geographically, with reassortant genotypes enriched in Asia and predominantly EA-derived constellations more common in Europe. Phylogeographic reconstruction indicated asymmetric dissemination within and between Europe and Asia, with Western Europe occupying a central position in the inferred migration network. Directed live-swine trade received the strongest support among migration predictors after accounting for sampling intensity. Sixteen candidate hemagglutinin codons showed evidence of episodic diversifying selection. Most candidate HA1 residues mapped to exposed regions of the globular head near antigenic regions.

Conclusions

EA H1N1 evolution across Eurasia was characterized by a persistent EA genomic background, regional genotype diversification through reassortment, and spread associated with live-swine trade. These findings support integrating genomic surveillance with animal movement data to guide surveillance along trade networks and improve the detection and assessment of emerging variants of zoonotic concern.

Background

Influenza A viruses (IAVs) are segmented, negative-sense RNA viruses that infect diverse avian and mammalian hosts, including humans [1, 2]. Reassortment during coinfection can lead to the generation of viruses that have altered antigenic and biological properties and contribute to the emergence of pandemics [3–5]. Pigs are susceptible to avian, swine, and human IAVs, which can lead to opportunities for coinfection and reassortment, as well as facilitate the emergence and transmission of viruses that have zoonotic or pandemic potential [6, 7]. The emergence of a virus from swine caused the 2009 H1N1 pandemic, which has demonstrated the consequences of this risk and reinforces the need for continuous surveillance of swine influenza viruses (SIVs) [8, 9].

Long-term persistence, geographic expansion, and ongoing reassortment make the Eurasian avian-like H1N1 lineage a priority for surveillance [10–12]. EA H1N1 viruses were first recognized in European pigs in 1979 following an transmission event from avians to swine [13]. Subsequent reassortment among viruses originating from avians, humans, and swine have contributed to the genetic diversification of European swine influenza viruses [14, 15]. EA H1N1 viruses have spread to Asia and are now predominant in Chinese pigs [10, 16, 17]. After 2009, the acquisition of A(H1N1)pdm09-derived and triple-reassortant-derived internal genes generated diverse EA H1N1 reassortants in China, including genotype 4 viruses with mammalian-adaptation and zoonotic-risk features [18, 19].

In Europe, EA H1 viruses continue to be prevalent and diversify through reassortment and adaptive evolution [12, 20, 21]. Human EA H1N1 infections have been reported in the Netherlands, and reassortant H1 viruses have recently emerged in Danish pigs, which are drawing more attention to the zoonotic relevance of this lineage [22–24]. Nevertheless, the global evolutionary dynamics and dissemination pathways of EA H1N1 viruses have not been completely resolved.

Studies have mainly focused on national or continental populations or have examined global swine IAV spread without specifically resolving the EA lineage [25–28]. Therefore, is not yet clear how reassortant genotypes emerge and persist across regions and whether viral exchange is continuing to occur between Europe and Asia. Questions also remain in regard to which ecological and anthropogenic factors shape interregional dissemination and how the *HA* gene is evolving within swine reservoirs. Addressing these questions requires geographically representative genomic datasets and phylodynamic approaches that integrate temporal, spatial, and ecological information [29, 30].

We have analysed 1,703 publicly available EA H1N1 genomes that were collected from 19 European and Asian countries between 2009 and 2025. We characterized their segment-level lineage composition, genotype diversity, temporal turnover, and geographic structure. Using a geographically and temporally balanced subsample, we reconstructed population dynamics and viral dissemination with time-calibrated Bayesian phylogenetic and discrete phylogeographic methods. A generalized linear model diffusion analysis was performed to evaluate the associations of viral migration with live-swine trade, pig population size, geographic distance, climatic and genotype similarity, ground connectivity, and sampling intensity. Finally, we identified positively selected HA sites and mapped them onto the HA structure. These analyses provide a transcontinental view of the evolution of EA H1N1 and its dissemination and

could provide useful information for genomic surveillance, risk assessment, and pandemic preparedness.

Methods

Data acquisition and preparation

Publicly available influenza A(H1N1) virus genome sequences sampled between 1 January 2009 and 1 August 2026 were retrieved from the NCBI Influenza Virus Resource and the GISAID EpiFlu database. Records were excluded if they (i) lacked sampling-date or geographic metadata, (ii) were from laboratory-derived strains, or (iii) were duplicates. Putative recombinant SIV genomes were identified and removed using RDP4, which incorporates the RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ algorithms [31]. Only genomes containing complete sequences for all eight genomic segments were included in downstream analyses. In total, the downstream analyses included 1,703 EA H1N1 genomes that were sampled between 2009 and 2025 from 19 countries.

Maximum likelihood phylogenies and lineage assignment

Using MAFFT v7.149, multiple-sequence alignments were generated independently for each of the eight IAV genome segments and were manually inspected when necessary [32]. The best-fitting nucleotide-substitution model was selected with ModelFinder according to the Akaike information criterion [33]. GTR + F+G4 was selected for all eight segments. Maximum-likelihood phylogenies were inferred using IQ-TREE 2 (Nguyen et al. 2015), and branch support was assessed using the ultrafast bootstrap approximation with 10,000 replicates [34, 35].

By comparison with curated reference datasets, each genomic segment was assigned to one of five lineages: EA H1N1, A(H1N1)pdm09, triple-reassortant (TR), human seasonal H1N1, or avian [17, 36]. Genotypes G1-G19 were defined as unique combinations of the eight segment lineages. Phylogenetic trees were annotated and visualized using the ggtree package in R [37].

Genotype persistence and geographic structure

For each genotype, the observed persistence span was calculated as the last detection year minus the first detection year plus one, and the number of distinct detection years was recorded. Genotype abundance was defined as the total number of eligible genomes assigned to each genotype. Genotypes represented by fewer than five genomes were pooled into an “Other genotypes” category before country-level genotype proportions and Bray–Curtis dissimilarities were calculated; no genome records were removed. The resulting country-level profiles were ordinated using classical principal coordinates analysis.

Countries were assigned to Europe or Asia according to a predefined regional classification (Russia was assigned to Europe). Differences in genotype composition between the two regions were assessed using one-factor PERMANOVA, and homogeneity of multivariate dispersion was evaluated using PERMDISP, each with 9,999 label permutations [38]. Regional enrichment was tested for genotypes that were represented by at least five genomes using two-sided Fisher's exact tests based on 2 × 2 contingency tables, while all genome records were retained in the regional background. P values were adjusted using the Benjamini–Hochberg procedure. Log2 odds ratios and approximate 95% confidence intervals were calculated using the Haldane–Anscombe correction.

Geographically and temporally balanced subsampling

To address preferential and heterogeneous spatial and temporal sampling, eligible HA sequences were subsampled to improve geographic and temporal representation [39–41]. Eligible sequences were at least 1,600 nucleotides long, contained no more than 1% ambiguous bases, had valid collection dates, and included complete country and genotype metadata. Countries were assigned to six predefined macroregions: Eastern Asia, South-eastern Asia, Northern Europe, Western Europe, Southern Europe, and Eastern Europe. A weighted greedy set-cover procedure was used to represent each observed macroregion-year combination, country, and genotype. The coverage weights were 5.0, 3.0, and 1.5 for macroregion-year combinations, countries, and genotypes, respectively.

Macroregions received approximately equal quotas, and annual quotas were allocated according to the square root of annual sequence availability. Genotype targets were proportional to eligible genotype counts raised to the power of 0.75. In each macroregion-year stratum, up to 32 candidates were ranked according to genotype-target pressure, genotype novelty, and genetic distinctness. Genetic distinctness was estimated from Jaccard distances between custom alignment-free 9-mer MinHash sketches, which contained up to 192 hash values [42].

Up to 500 within-stratum swaps were performed to reduce deviation from genotype targets while preserving coverage and quotas. Exact local duplicates were collapsed, and eligible nonduplicate sequences were added when necessary. After downsampling, the temporal signal was assessed by regressing root-to-tip genetic distances from the maximum-likelihood tree against exact sampling dates using TempEst v1.5.3 [43]. Temporal outliers were excluded, and the resulting dataset of 311 sequences was used for downstream molecular-clock and phylogeographic analyses.

Bayesian evolutionary analysis and phylodynamics

Time-scaled phylogenies were inferred using BEAST X v10.5.0 [44] under the GTR + F+G4 substitution model, an uncorrelated relaxed molecular clock with a lognormal distribution [45], and a Gaussian Markov random field (GMRF) Bayesian Skyride coalescent prior [46]. Markov chain Monte Carlo (MCMC)

chains were run for 200,000,000 iterations and sampled every 10,000 iterations. Independent runs were combined using LogCombiner after removal of burn-in.

Convergence was assessed in Tracer v1.7, and effective sample sizes greater than 200 were required for all relevant parameters [47]. Posterior means, standard deviations, and 95% highest posterior density (HPD) intervals were summarized. MCC trees were generated from the posterior tree distribution using TreeAnnotator v1.10.4. Changes in relative genetic diversity through time were reconstructed using a GMRF Bayesian Skyride model and visualized in Tracer.

Phylogeographic inference

Geographic diffusion was reconstructed using the same sequence dataset (n = 311) and an asymmetric discrete-trait phylogeographic model with BSSVS [29]. The MCMC settings, run combination, and convergence assessment were the same as in the procedures described above. The dispersal through time was visualized using SpreaD3 v0.9.6, which was also used to calculate Bayes factor (BF) support for individual transitions between geographic states [48]. Transition support was classified as strong (BF = 3–100), very strong (BF = 100–1,000), or decisive (BF > 1,000). The expected number of transitions between location states was estimated using Markov-jump counting [49]. The total inbound and outbound jump counts were summarized for each of the six macrogeographic regions to quantify asymmetry in viral movement among regions.

Testing predictors of viral migration

An asymmetric discrete-trait phylogeographic diffusion model was parameterized with a generalized linear model (GLM) in BEAST to evaluate factors associated with EA H1N1 virus migration [30]. The transition rates among the six geographic regions were modelled as functions of climatic similarity, geographic distance, pig population size in the origin and destination regions, sampling intensity in the origin and destination regions, and directed live-swine trade. Geographic distance was calculated as the mean great-circle distance between all country pairs belonging to each pair of regions. Directed live-swine trade data for the item “Swine and pigs” were obtained from the FAOSTAT Detailed Trade Matrix for 2009–2021 (https://www.fao.org/faostat/en/#data/TM).

Annual pig-stock data for 2009–2024 were obtained from FAOSTAT (https://www.fao.org/faostat/en/#data/QCL), and sampling intensity was represented by regional viral sequence counts. Climatic similarity was calculated from ERA5 reanalysis data on 2-m air temperature and total precipitation for 2009–2025. ERA5 data were obtained from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). The GLM was used to evaluate period-level associations rather than year-specific effects. Each predictor was represented by a single fixed 6 by 6 matrix, with rows denoting the origin region and columns denoting the destination region. Diagonal entries were set to zero. Positive-valued predictors were log-transformed as log(x + 1) and standardized. Bayesian stochastic

search variable selection was used to estimate posterior inclusion probabilities and regression coefficients conditional on predictor inclusion.

Detection of episodic diversifying selection and structural mapping

Codon-based selection analyses of the HA gene were performed using a custom analysis pipeline. The 14-amino-acid signal peptide was identified and removed before numbering. An amino-acid alignment of the mature HA protein containing 552 aligned sites was used to generate a corresponding codon alignment by back-translation. The HA1–HA2 cleavage boundary was defined immediately upstream of the conserved fusion-peptide motif GLFGAIAGFIEGG. According to the H3 numbering scheme used in this study, HA1 comprised sites 1–330 and HA2 comprised sites 331–552.

A maximum-likelihood phylogeny was inferred from the resulting nucleotide alignment using IQ-TREE 2 under the GTR + F+G4 model. The inferred tree and codon alignment were then used as input for the mixed-effects model of evolution (MEME) in HyPhy to test for site-specific episodic diversifying selection [50, 51]. MEME allows selective pressures to vary among phylogenetic branches at individual codon sites. Sites with a nominal likelihood-ratio test P value < 0.05 were considered to show evidence of episodic positive selection. Site-specific MEME likelihood-ratio test statistics were plotted according to H3 numbering [52]. Positively selected residues were mapped to the HA crystal structure (PDB ID: 8GV5) using the corresponding H3-numbered positions and visualized with UCSF ChimeraX [53, 54].

Results

Expansion and lineage cmposition of EA H1N1 vruses

The curated dataset comprised 1,703 EA H1N1 genomes that were sampled between 2009 and 2025 from 19 countries in Europe and Asia. The maximum-likelihood phylogenies of the eight genome segments showed that EA-derived surface genes were predominant (Fig. 1A and Table S1). All 1,703 *HA* sequences belonged to the EA lineage, whereas 1,700 NA sequences (representing 99.8% of the dataset) were of EA origin. The remaining three genomes carried an A(H1N1)pdm09-derived NA segment (Fig. 1B).

Nine HA clades were identified. Clades 1C.2.2 and 1C.2.1 were predominant and accounted for 1,041 genomes (61.1%) and 545 genomes (32.0%), respectively. Together, they covered 93.1% of all genomes. The NA segment was distributed among four EA-derived lineages. EA-AVIAN-4 was the most frequent lineage and comprised 776 genomes (45.6%), followed by EA-AVIAN-1 with 454 genomes (26.7%), EA-AVIAN-2 with 445 genomes (26.1%), and EA-AVIAN-3 with 25 genomes (1.5%). HA and NA backgrounds showed uneven and incomplete associations. The 1C.2.1 HA clade was mainly associated with the EA-AVIAN-1 and EA-AVIAN-2 NA lineages, and the 1C.2.2 was predominantly associated with EA-AVIAN-4. These patterns indicate that the major HA and NA backgrounds were frequently linked but were not invariant throughout the dataset (Fig. 1B and C).

Genotype diversity, temporal turnover, and persistence

Lineage assignments across all eight segments resolved the 1,703 genomes into 19 genotype constellations, designated G1–G19 (Fig. 2A). All genotypes retained an EA-derived HA, and all except G17 retained an EA-derived NA, indicating that reassortment predominantly involved internal gene segments. G1 was the only entirely EA-derived genotype, whereas the remaining genotypes contained one to seven non-EA segments. These segments were predominantly derived from the A(H1N1)pdm09 and triple-reassortant lineages, with occasional contributions from human seasonal H1N1 and avian influenza viruses.

Genotype frequencies were highly uneven. G1 accounted for 970 genomes (57.0%), followed by G6 with 324, G10 with 191, G4 with 115, and G3 with 51. The remaining 14 genotypes accounted for 52 genomes, and 12 were represented by three or fewer genomes. The four most frequently sampled reassortant genotypes had distinct internal gene constellations (Fig. 2A). G4 differed from G1 only by an A(H1N1)pdm09-derived MP segment. G10 retained EA-derived HA, NA, and MP segments, together with A(H1N1)pdm09-derived PB2, PB1, PA, and NP segments and a triple-reassortant-derived NS segment. G6 shared this constellation except that its MP segment was also derived from A(H1N1)pdm09. In contrast, G3 carried A(H1N1)pdm09-derived segments at all six internal gene positions.

The relative abundance of genotypes changed substantially over time (Fig. 2B). Although G1 predominated across the dataset as a whole, its proportion decreased markedly during 2014–2016, coinciding with increased representation of G10 and G6. G10 was prominent during this interval, whereas G6 continued to account for a substantial proportion of sampled genomes in subsequent years. After 2022, the proportion of G1 declined again as G4 increased, with G4 accounting for the largest share of sampled genomes in 2024. Observed persistence varied substantially among genotypes (Fig. 2C). The median persistence span was one year, whereas G1 was detected in every study year from 2009 to 2025. G3, G4, G6, and G10 each had observed persistence spans of at least 10 years. Genotypes represented by more genomes generally occurred in more sampling years and had longer observed spans. However, the number of non-EA segments showed no consistent relationship with observed persistence: long spans occurred in both genotypes with few non-EA segments, such as G4, and those with multiple non-EA segments, such as G3 and G6.

The reassortment schematic placed these genotype constellations in the context of their detection chronology (Fig. 3). G2, G3, and G4 were detected early in the study period, followed by additional constellations during 2013–2021. The depicted constellations illustrated contributions from A(H1N1)pdm09 and triple-reassortant lineages throughout the timeline. Less frequently observed lineage contributions included a human seasonal H1N1-derived NP segment in G16 and an avian-derived MP segment in G19. These dates represent detection in the sampled dataset rather than the timing of the underlying reassortment events.

Human infection records included both the entirely EA-derived G1 genotype and the reassortant genotypes G10, G6, and G4 (Fig. 3; Table S2). G1 was identified in human cases in Germany, China, the Netherlands, and Italy during 2010–2020. Human infections involving G10 were recorded in China in 2015, while G6 was identified in China in 2016 and 2021. G4 was subsequently identified in a human infection in the Netherlands in 2023.

Geographic structuring of genotype diversity

Genotype composition varied geographically among the sampled countries. G1 was predominant in most sampled European countries, while a more heterogeneous composition occurred in China, with substantial contributions from G1, G6, G10, and other genotypes (Fig. 4A). Principal coordinates analysis was performed based on Bray–Curtis dissimilarities of country-level genotype compositions. The results largely distinguished European and Asian countries along PCoA1. PCoA1 and PCoA2 explained 37.4% and 21.6% of the total variation, respectively. PERMANOVA supported a significant regional effect on genotype composition (pseudo-F value = 5.14, R^2 = 0.23, and P = 0.002). PERMDISP did not indicate evidence of unequal within-region dispersion (F = 2.12 and P = 0.266), indicating that the observed regional contrast was consistent with differences in genotype composition rather than dispersion alone (Fig. 4B).

Regional enrichment analysis identified contrasting genotype distributions between Europe and Asia. After Benjamini–Hochberg correction, G1, G4, and G13 were enriched in Europe, while G2, G6, and G10 were enriched in Asia. All Asia-enriched genotypes contained internal segments that were derived from the A(H1N1)pdm09 or triple-reassortant lineages. The Europe-enriched group was more heterogeneous, including the completely EA-derived G1, the largely EA G4, and the more extensively reassorted G13. Therefore, reassortant constellations were more consistently represented among the Asia-enriched genotypes, whereas the Europe-enriched pattern was driven mainly by G1 and G4 (Fig. 4C).

Temporal signal and population dynamics

The subsampled panel of 311 HA sequences showed a strong temporal signal. Root-to-tip regression yielded a correlation coefficient of 0.90, an R2 value of 0.81, and a slope of 3.50 × 10 − 3 substitutions per site per year, which support the use of time-calibrated Bayesian phylogenetic models (Fig. 5A). Relative genetic diversity showed several temporal phases (Fig. 5B). The posterior median declined from 2009 to an initial trough around 2013, increased to a peak between 2016 and 2017, and then decreased to a second trough in 2020. The diversity remained relatively stable during 2021 and 2022 before declining again from 2023 onward. The lowest estimated diversity occurred in 2025, although the 95% HPD intervals widened substantially in the most recent years.

Reduced sampling intensity during the late study period may have contributed to the greater uncertainty of these recent estimates. The maximum clade credibility tree showed that reassortant segment constellations were distributed throughout the phylogeny rather than confined to separate clades (Fig. 5C). This distribution was consistent with recurrent reassortment occurring on a predominantly EA-derived HA background.

Spatial dissemination and predictors of viral migration

Discrete phylogeographic reconstruction identified a strongly asymmetric migration network across the six macroregions (Fig. 6A and B). Western Europe was connected to Southern, Northern, and Eastern Europe through transitions with decisive support, with Bayes factors greater than 1,000 and posterior probabilities of 1.0. The route from Western Europe to Southern Europe also had the highest estimated transition rate of 3.54. Additional support was observed for Eastern Asia to South-eastern Asia, Northern Europe to Southern Europe, Northern Europe to South-eastern Asia, and Western Europe to Eastern Asia, with Bayes factors of 522, 181, 79, and 4.5, respectively. These routes included supported connections between Europe and Asia, which is consistent with intercontinental viral exchange during the study period.

Markov jump counts further showed that Western Europe was the principal net exporter, with approximately 47 outbound and one inbound jump, whereas Southern Europe mainly acted as a sink, with approximately 27 inbound and one outbound jump. Among the evaluated predictors, live-swine trade showed decisive support for an association with viral transition rates (Fig. 6C). Its posterior inclusion probability was 1.0, and its conditional coefficient was positive, with an estimate of 1.86 and a 95% HPD interval of 1.29 to 2.55. Origin sampling intensity and origin pig population received limited support, whereas geographic distance had a negative but weakly supported coefficient. Climatic similarity, destination pig population, and destination sampling intensity received minimal support, with E(δ) values no greater than 0.02. The strong association with live-swine trade persisted after sampling intensity was included for both origin and destination regions.

HA evolution, episodic diversifying selection, and structural mapping

Temporal annotation of the HA maximum-likelihood phylogeny showed recurrent amino acid changes at a restricted set of positions between 2009 and 2025 (Fig. 7A). Based on H3 numbering, substitutions at positions 51, 72, 86, 144, 158, 225, 510, 531, and 536 occurred on multiple branches. Several positions showed multiple alternative substitutions, including I51S and I51T, S86P and P86S, A144V and A144T, G158E and G158A, and E225G, E225D, and E225N. MEME analysis identified 16 codons that met the nominal threshold for episodic diversifying selection defined as $P < 0.05$ (Fig. 7B). There were 12 selected sites located in HA1 (positions 6, 41, 51, 72, 86, 144, 158, 190, 225, 255, 322, and 323). Four additional sites were located in HA2 (positions 180, 188, 201, and 206).

The strongest signal occurred at HA1 site 6 with an LRT statistic of 32.80 and $P = 3.1 \times 10 - 8$. This was followed by HA2 site 201 (LRT statistic of 16.53, HA2 site 206, with an LRT statistic of 14.53) and HA1 site 322 (LRT statistic of 13.69). Selected HA1 residues at positions 144, 158, 190, and 225 mapped to or near annotated antigenic regions surrounding the receptor-binding pocket. Structural mapping onto the HA trimer structure represented by PDB 8GV5 showed that selected HA1 residues clustered on the exposed surface of the globular head in both front and rear views. The four selected HA2 sites localized

to the stalk region. Overall, the selected sites were concentrated in the exposed HA1 head, and additional sites were distributed in the HA2 stalk (Fig. 7C and D).

Discussion

Understanding how swine influenza viruses persist and disseminate between connected swine-production systems is important for effective control and risk assessment. EA H1N1 viruses are established in pig populations across Europe and Asia, undergo recurrent reassortment with other swine influenza lineages, and have been associated with human infections reported in multiple countries [10–12, 16, 24, 55, 56]. Recent genomic epidemiological work has also provided evidence of human exposure and zoonotic potential in China [17]. International live-swine trade connects production systems and may facilitate viral movement across regions [28]. These features place EA H1N1 at the intersection of viral evolution, animal movement, and zoonotic risk.

Genotype G1, comprising exclusively EA-derived segments, was the most abundant genotype and was detected throughout the study period. Reassortant genotypes G3, G4, G6, and G10 also had prolonged detection spans of approximately 10–16 years. In contrast, the median persistence span across all genotypes was one year, and many low-frequency genotypes were detected only once or within short intervals. These findings suggest that some reassortant constellations persisted while other genotypes continued to turn over. Human infections involving viruses assigned here to G1, including cases in Hebei, China, in 2012 and the Netherlands in 2019, show that entirely EA-derived viruses also pose a zoonotic risk [22, 55].

Experimental characterization of viruses previously designated G4 EA H1N1, corresponding to G6 in our classification based on segment lineage assignments, demonstrated human-type receptor binding, replication in human airway epithelial cells, and aerosol transmission in ferrets. Serological studies also provided evidence of exposure among swine workers [19]. Human infections have been reported with viruses classified in our study as G10 in Hunan in 2015 and G6 in Fujian in 2016 and Shandong in 2021 [57–59]. These reports show that several reassortant genotypes detected over extended periods were also associated with sporadic human infections, underscoring the need for continued surveillance in pigs and humans. In our dataset, genotypes enriched in Asia were mainly reassortants, whereas those enriched in Europe included both entirely EA-derived and reassortant viruses. This regional variation may reflect differences in the influenza lineages circulating in pigs and their opportunities for reassortment, shaped by pig production systems and animal movement [11, 20, 25].

Phylogeographic reconstruction identified Western Europe as the predominant net exporter in the sampled network, suggesting that it played a central role in inferred EA H1N1 dissemination during the study period. Well-supported transitions between Europe and Asia indicate sustained intercontinental connectivity rather than isolated regional circulation. The directionality of these transitions is consistent with evidence that animal movement contributes to the global circulation of SIVs [25]. Therefore, Western Europe may represent an important regional hub for monitoring introductions, onward

dissemination, and the circulation of reassortant genotypes. The inferred transitions describe relative patterns of regional connectivity, and the biological and commercial networks underlying these movements may include shared breeding systems, production links, and trade routes.

Several HA1 sites identified in our analysis as candidates for episodic diversifying selection overlap residues previously implicated in antigenic evolution, receptor binding, or glycan-mediated immune escape in human and swine H1 viruses, highlighting regions that warrant targeted functional investigation. Residue 190 lies in the 190-helix and the Sb antigenic region, and the N190D substitution in Eurasian avian-like H1N1 swine virus has been shown to affect antibody recognition and binding to human-type receptors [60]. Residue 225 is located near the receptor-binding interface; D225G alters recognition of α2,3- and α2,6-linked sialic acids in pandemic H1 HAs, whereas D225E has been associated with adaptation of EA H1N1 virus toward human-type receptor binding [61–64]. Residues 144 and 158 are also situated in antigenically important regions of the globular head. Glycosylation involving residue 144 can influence antibody recognition, while substitutions around residue 158, including G158E, G158K, and N159K, have been associated with antibody escape or antigenic variation in human and European swine H1 viruses [65–69]. Positions 51, 72, and 86 fall within or near Cb-like head epitopes in refined H1 epitope maps, while positions 6, 41, 255, 322, and 323 represent additional candidates for investigation [70]. Collectively, the HA1 pattern is consistent with recurrent immune- and receptor-related pressures shaping EA H1N1 diversification and identifies both previously characterized functional regions and additional sites for lineage-specific analysis.

Four additional candidate codons were identified in HA2 (180, 188, 201, and 206), indicating that episodic selection may also involve the stalk or membrane-proximal region of HA. The HA2 stalk is generally more evolutionarily constrained than the globular head, yet low-pH activation requires coordinated structural rearrangements throughout the HA2 subunit; variation in this region could therefore influence fusion activation, trimer stability, or host adaptation [71–73]. At the nominal MEME threshold of $P < 0.05$, these sites define a focused set of experimentally testable hypotheses. Reverse-genetics or pseudotyped-virus experiments combining serological, receptor-binding, fusion-pH, HA-stability, cleavage, and surface-expression assays will help determine the contributions of these substitutions to HA antigenicity, receptor usage, protein stability, membrane fusion, and host adaptation.

Conclusions

EA H1N1 viruses maintained an EA-derived genomic background while diversifying through reassortment, with several reassortant genotypes persisting over extended periods. Genotype composition differed between Europe and Asia, and Western Europe occupied a central position in the inferred migration network. Live-swine trade was the strongest supported predictor of viral migration. Candidate sites of episodic diversifying selection were concentrated in the HA1 globular head, suggesting a role for surface variation in viral diversification. These findings support integrating genomic surveillance with live-swine movement data to track emerging reassortants and inform surveillance at the swine–human interface.

Abbreviations

A(H1N1)pdm09
Influenza A(H1N1) pandemic 2009 virus
BEAST
Bayesian evolutionary analysis by sampling trees
BF
Bayes factor
BSSVS
Bayesian stochastic search variable selection
EA
Eurasian avian-like
GLM
Generalized linear model
GMRF
Gaussian Markov random field
GTR
General time-reversible
HA
Hemagglutinin
HPD
Highest posterior density
IAVs
Influenza A viruses
LRT
Likelihood-ratio test
MCC
Maximum clade credibility
MCMC
Markov chain Monte Carlo
MEME
Mixed effects model of evolution
ML
Maximum likelihood
MP
Matrix gene segment
NA
Neuraminidase
NCBI
National Center for Biotechnology Information

NP
Nucleoprotein
NS
Nonstructural gene segment
PA
Polymerase acidic protein
PB1
Polymerase basic protein 1
PB2
Polymerase basic protein 2
PCoA
Principal coordinates analysis
PDB
Protein Data Bank
PERMANOVA
Permutational multivariate analysis of variance
PERMDISP
Permutational analysis of multivariate dispersions
SIVs
Swine influenza viruses
TR
Triple-reassortant

Declarations

Acknowledgments

We gratefully acknowledge all the authors, all the originating laboratories, and all the submitting laboratories where SIV sequence data were generated and submitted, through GISAID's EpiFlu Database (https://gisaid.org/) and NCBI Influenza Virus Database (https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/).

Funding

This study was supported by the 14th Five-Year National Key R & D Program Subproject (2023YFD1800600).

Contributions

Conceptualization, H.-L.L., Y.Z., J.-N.Y., and H.-L.S.; Methodology, H.-L.L., Y.Z., C.-Y.L., and Z.-J.Z.; Investigation, H.-L.L., F.-X.X., and J.-N.Y.; Writing – Original Draft, H.-L.L., and H.-L.S.; Writing – Review

and Editing, All authors; Funding Acquisition, M.L., and H.-L.S.; Resources (Computational), H.-L.L.; Supervision, R.-W.R., M.L., and H.-L.S.

Availability of data and materials

The sequence data analysed in this study were obtained from the NCBI Influenza Virus Resource and the GISAID EpiFlu database. Accession numbers, strain names, collection dates, geographic origins, genotype assignments, and genomic segment annotations are provided in Table S1. The analysis files, including BEAST input files for the BSSVS and GLM analyses, predictor matrices, maximum-likelihood phylogenetic trees, tree annotation files, and the R plotting script, are openly available in Figshare at https://doi.org/10.6084/m9.figshare.33440737.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

1. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza Nat reviews Disease primers. 2018;4:3.
2. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992;56:152–79.
3. Wille M, Holmes EC. The Ecology and Evolution of Influenza Viruses. Cold Spring Harbor perspectives in medicine. 2020;10.
4. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459:1122–5.
5. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Volume 325. New York, NY: Science; 2009. pp. 197–201.
6. Ito T, Couceiro JN, Kelm S, Baum LG, Krauss S, Castrucci MR, et al. Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. J Virol. 1998;72:7367–73.
7. Ma W, Lager KM, Vincent AL, Janke BH, Gramer MR, Richt JA. The role of swine in the generation of novel influenza viruses. Zoonoses Public Health. 2009;56:326–37.
8. Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, Garten RJ, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med. 2009;360:2605–15.
9. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature. 2009;459:931–9.
10. Vijaykrishna D, Smith GJ, Pybus OG, Zhu H, Bhatt S, Poon LL, et al. Long-term evolution and transmission dynamics of swine influenza A virus. Nature. 2011;473:519–22.

11. Henritzi D, Petric PP, Lewis NS, Graaf A, Pessia A, Starick E, et al. Surveillance of European Domestic Pig Populations Identifies an Emerging Reservoir of Potentially Zoonotic Swine Influenza A Viruses. Cell Host Microbe. 2020;28:614–e6276.
12. Lycett SJ, Baillie G, Coulter E, Bhatt S, Kellam P, McCauley JW, et al. Estimating reassortment rates in co-circulating Eurasian swine influenza viruses. J Gen Virol. 2012;93:2326–36.
13. Pensaert M, Ottis K, Vandeputte J, Kaplan MM, Bachmann PA. Evidence for the natural transmission of influenza A virus from wild ducts to swine and its potential importance for man. Bull World Health Organ. 1981;59:75–8.
14. Brown IH, Harris PA, McCauley JW, Alexander DJ. Multiple genetic reassortment of avian and human influenza A viruses in European pigs, resulting in the emergence of an H1N2 virus of novel genotype. J Gen Virol. 1998;79(Pt 12):2947–55.
15. Marozin S, Gregory V, Cameron K, Bennett M, Valette M, Aymard M, et al. Antigenic and genetic diversity among swine influenza A H1N1 and H1N2 viruses in Europe. J Gen Virol. 2002;83:735–45.
16. Yang H, Chen Y, Qiao C, He X, Zhou H, Sun Y, et al. Prevalence, genetics, and transmissibility in ferrets of Eurasian avian-like H1N1 swine influenza viruses. Proc Natl Acad Sci USA. 2016;113:392–7.
17. Sun H, Liu H, Zhang J, Qu X, Pang Z, Xu F, et al. Genome-scale evolution and phylodynamics of swine influenza A viruses in China: a genomic epidemiology study. Lancet Microbe. 2025;6:101020.
18. Liang H, Lam TT, Fan X, Chen X, Zeng Y, Zhou J, et al. Expansion of genotypic diversity and establishment of 2009 H1N1 pandemic-origin internal genes in pigs in China. J Virol. 2014;88:10864–74.
19. Sun H, Xiao Y, Liu J, Wang D, Li F, Wang C, et al. Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. Proc Natl Acad Sci USA. 2020;117:17204–10.
20. Joseph U, Vijaykrishna D, Smith GJD, Su YCF. Adaptive evolution during the establishment of European avian-like H1N1 influenza A virus in swine. Evol Appl. 2018;11:534–46.
21. Chepkwony S, Parys A, Vandoorn E, Stadejek W, Xie J, King J, et al. Genetic and antigenic evolution of H1 swine influenza A viruses isolated in Belgium and the Netherlands from 2014 through 2019. Sci Rep. 2021;11:11276.
22. Parys A, Vandoorn E, King J, Graaf A, Pohlmann A, Beer M, et al. Human Infection with Eurasian Avian-Like Swine Influenza A(H1N1) Virus, the Netherlands, September 2019. Emerg Infect Dis. 2021;27:939–43.
23. Ryt-Hansen P, George S, Hjulsager CK, Trebbien R, Krog JS, Ciucani MM, et al. Rapid surge of reassortant A(H1N1) influenza viruses in Danish swine and their zoonotic potential. Emerg microbes infections. 2025;14:2466686.
24. Eggink D, Kroneman A, Dingemans J, Goderski G, van den Brink S, Bagheri M et al. Human infections with Eurasian avian-like swine influenza virus detected by coincidence via routine respiratory

surveillance systems, the Netherlands, 2020 to 2023. Euro surveillance: bulletin Europeen sur les maladies transmissibles = European. Commun disease Bull. 2025;30.

25. Nelson MI, Viboud C, Vincent AL, Culhane MR, Detmer SE, Wentworth DE, et al. Global migration of influenza A viruses in swine. Nat Commun. 2015;6:6696.
26. Watson SJ, Langat P, Reid SM, Lam TT, Cotten M, Kelly M, et al. Molecular Epidemiology and Evolution of Influenza Viruses Circulating within European Swine between 2009 and 2013. J Virol. 2015;89:9920–31.
27. Lam TT, Zhu H, Wang J, Smith DK, Holmes EC, Webster RG, et al. Reassortment events among swine influenza A viruses in China: implications for the origin of the 2009 influenza pandemic. J Virol. 2011;85:10279–85.
28. Nelson MI, Lemey P, Tan Y, Vincent A, Lam TT, Detmer S, et al. Spatial dynamics of human-origin H1 influenza A virus in North American swine. PLoS Pathog. 2011;7:e1002077.
29. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5:e1000520.
30. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog. 2014;10:e1003932.
31. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1:vev003.
32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
33. 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.
34. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4.
35. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35:518–22.
36. Chang J, Anderson TK, Zeller MA, Gauger PC, Vincent AL. octoFLU: Automated Classification for the Evolutionary Origin of Influenza A Virus Gene Sequences Detected in U.S. Swine. Microbiol resource announcements. 2019;8.
37. Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, et al. Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. iMeta. 2022;1:e56.
38. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics. 2006;62:245–53.
39. Karcher MD, Palacios JA, Bedford T, Suchard MA, Minin VN. Quantifying and Mitigating the Effect of Preferential Sampling on Phylodynamic Inference. PLoS Comput Biol. 2016;12:e1004789.

40. Dellicour S, Lemey P, Artois J, Lam TT, Fusaro A, Monne I, et al. Incorporating heterogeneous sampling probabilities in continuous phylogeographic inference - Application to H5N1 spread in the Mekong region. Bioinf (Oxford England). 2020;36:2098–104.
41. Layan M, Müller NF, Dellicour S, De Maio N, Bourhy H, Cauchemez S, et al. Impact and mitigation of sampling bias to determine viral spread: Evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations. Virus Evol. 2023;9:vead010.
42. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132.
43. 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.
44. Baele G, Ji X, Hassler GW, McCrone JT, Shao Y, Zhang Z, et al. BEAST X for Bayesian phylogenetic, phylogeographic and phylodynamic inference. Nat Methods. 2025;22:1653–6.
45. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4:e88.
46. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008;25:1459–71.
47. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67:901–4.
48. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol. 2016;33:2167–9.
49. Minin VN, Suchard MA. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol. 2008;56:391–412.
50. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8:e1002764.
51. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol. 2020;37:295–9.
52. Burke DF, Smith DJ. A recommended numbering scheme for influenza A HA subtypes. PLoS ONE. 2014;9:e112302.
53. Chen Y, Wang F, Yin L, Jiang H, Lu X, Bi Y, et al. Structural basis for a human broadly neutralizing influenza A hemagglutinin stem-specific antibody including H17/18 subtypes. Nat Commun. 2022;13:7603.
54. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein science: publication Protein Soc. 2021;30:70–82.
55. Wang DY, Qi SX, Li XY, Guo JF, Tan MJ, Han GY, et al. Human infection with Eurasian avian-like influenza A(H1N1) virus, China. Emerg Infect Dis. 2013;19:1709–11.

56. Dürrwald R, Wedde M, Biere B, Oh DY, Heßler-Klee M, Geidel C et al. Zoonotic infection with swine A/H1(av)N1 influenza virus in a child, Germany, June 2020. Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2020;25.
57. Zhu W, Zhang H, Xiang X, Zhong L, Yang L, Guo J et al. Reassortant Eurasian Avian-Like Influenza A(H1N1) Virus from a Severely Ill Child, Hunan Province, China, 2015. Emerging infectious diseases. 2016;22:1930–1936.
58. Xie JF, Zhang YH, Zhao L, Xiu WQ, Chen HB, Lin Q, et al. Emergence of Eurasian Avian-Like Swine Influenza A (H1N1) Virus from an Adult Case in Fujian Province, China. Virol Sin. 2018;33:282–6.
59. He Y, Song S, Wu J, Wu J, Zhang L, Sun L, et al. Emergence of Eurasian Avian-Like Swine Influenza A (H1N1) virus in a child in Shandong Province, China. BMC Infect Dis. 2024;24:550.
60. Xu C, Zhang N, Yang Y, Liang W, Zhang Y, Wang J, et al. Immune Escape Adaptive Mutations in Hemagglutinin Are Responsible for the Antigenic Drift of Eurasian Avian-Like H1N1 Swine Influenza Viruses. J Virol. 2022;96:e0097122.
61. Zhang W, Shi Y, Qi J, Gao F, Li Q, Fan Z, et al. Molecular basis of the receptor binding specificity switch of the hemagglutinins from both the 1918 and 2009 pandemic influenza A viruses by a D225G substitution. J Virol. 2013;87:5949–58.
62. Chen H, Wen X, To KK, Wang P, Tse H, Chan JF, et al. Quasispecies of the D225G substitution in the hemagglutinin of pandemic influenza A(H1N1) 2009 virus from patients with severe disease in Hong Kong, China. J Infect Dis. 2010;201:1517–21.
63. Zheng B, Chan KH, Zhang AJ, Zhou J, Chan CC, Poon VK et al. D225G mutation in hemagglutinin of pandemic influenza H1N1 (2009) virus enhances virulence in mice. Experimental biology and medicine (Maywood, NJ). 2010;235:981-8.
64. Zou J, Jiang M, Xiao R, Sun H, Liu H, Peacock T, et al. GGCX promotes Eurasian avian-like H1N1 swine influenza virus adaption to interspecies receptor binding. Nat Commun. 2025;16:670.
65. Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbò P, et al. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc Natl Acad Sci USA. 2011;108:E1417–22.
66. Kim JI, Lee I, Park S, Hwang MW, Bae JY, Lee S, et al. Genetic requirement for hemagglutinin glycosylation and its implications for influenza A H1N1 virus evolution. J Virol. 2013;87:7539–49.
67. O'Donnell CD, Vogel L, Wright A, Das SR, Wrammert J, Li GM et al. Antibody pressure by a human monoclonal antibody targeting the 2009 pandemic H1N1 virus hemagglutinin drives the emergence of a virus with increased virulence in mice. mBio. 2012;3.
68. Chastagner A, Hervé S, Quéguiner S, Hirchaud E, Lucas P, Gorin S et al. Genetic and Antigenic Evolution of European Swine Influenza A Viruses of HA-1C (Avian-Like) and HA-1B (Human-Like) Lineages in France from 2000 to 2018. Viruses. 2020;12.
69. Dinis JM, Florek KR, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, et al. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. J Virol. 2016;90:3355–65.

70. Deem MW, Pan K. The epitope regions of H1-subtype influenza A, with application to vaccine efficacy. Protein engineering, design & selection: PEDS. 2009;22:543-6.
71. Kirkpatrick E, Qiu X, Wilson PC, Bahl J, Krammer F. The influenza virus hemagglutinin head evolves faster than the stalk domain. Sci Rep. 2018;8:10432.
72. Benton DJ, Gamblin SJ, Rosenthal PB, Skehel JJ. Structural transitions in influenza haemagglutinin at membrane fusion pH. Nature. 2020;583:150–3.
73. Trovão NS, Khan SM, Lemey P, Nelson MI, Cherry JL. Comparative evolution of influenza A virus H1 and H3 head and stalk domains across host species. mBio. 2024;15:e0264923.

Figures

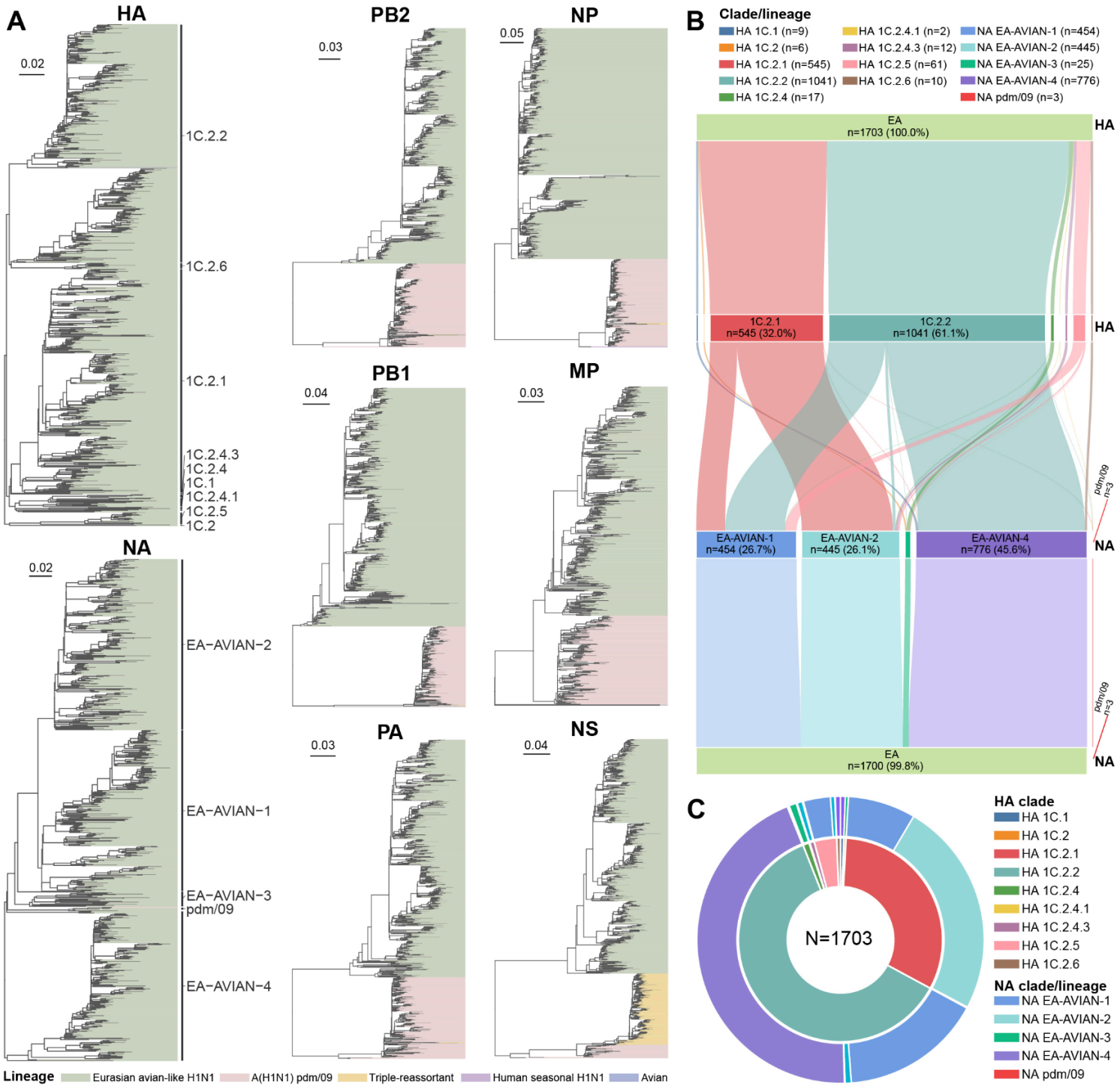


Figure 1

Phylogenetic classification and clade composition of EA H1N1 viruses sampled across Europe and Asia.

(A) Maximum-likelihood phylogenies of the eight influenza A virus genome segments, including PB2, PB1, PA, HA, NP, NA, MP, and NS. Branches are colored according to segment lineage, including Eurasian avian-like H1N1, A(H1N1)pdm09, triple-reassortant, human seasonal H1N1, and avian lineages. HA clades and NA clades or lineages are indicated beside the corresponding phylogenies. Scale bars represent nucleotide substitutions per site.

(B) Alluvial diagram showing the associations between HA clades and NA clades or lineages among 1,703 genomes. Flow widths are proportional to the numbers of genomes, and labels indicate the numbers and percentages of genomes assigned to the major clades or lineages.
(C) Concentric donut charts showing the relative proportions of HA clades and NA clades or lineages. The inner ring represents HA clades and the outer ring represents NA clades or lineages.

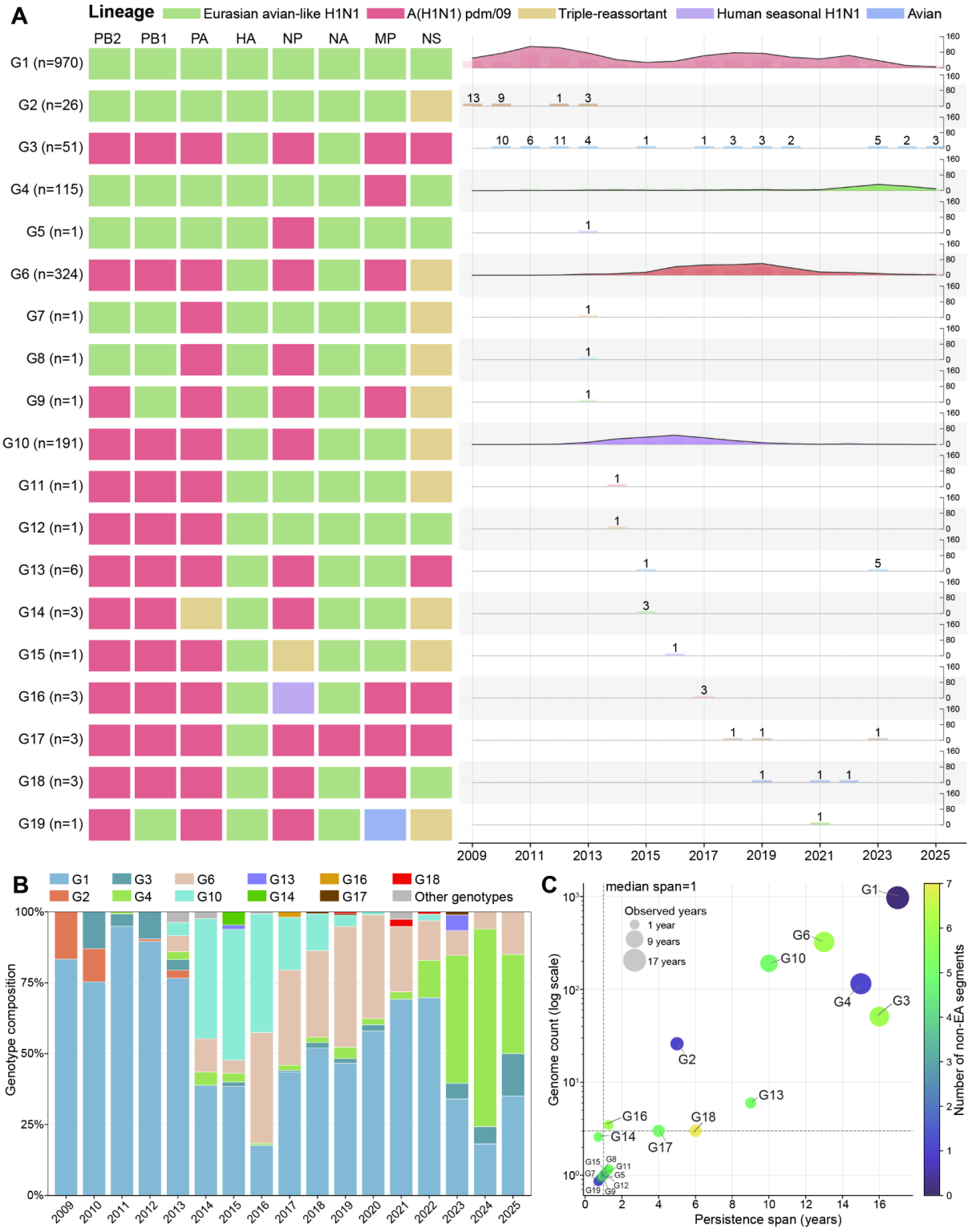


Figure 2

Genotype composition and temporal persistence of EA H1N1 viruses.
(A) Genomic segment constellations of the 19 identified genotypes, designated G1 to G19. Colored tiles indicate the inferred lineage of each segment, and the numbers in parentheses indicate the numbers of genomes assigned to each genotype. Annual detection profiles from 2009 to 2025 are shown beside the segment constellations.
(B) Annual changes in genotype composition from 2009 to 2025. Each stacked bar represents the proportion of genomes assigned to each genotype in a given year.
(C) Relationship between genotype persistence and genome abundance. The x-axis indicates the observed persistence span in years, and the y-axis indicates genome count on a logarithmic scale. Bubble size represents the number of years in which a genotype was detected, and color indicates the number of genomic segments not assigned to the EA lineage. The vertical dashed line represents the median persistence span.

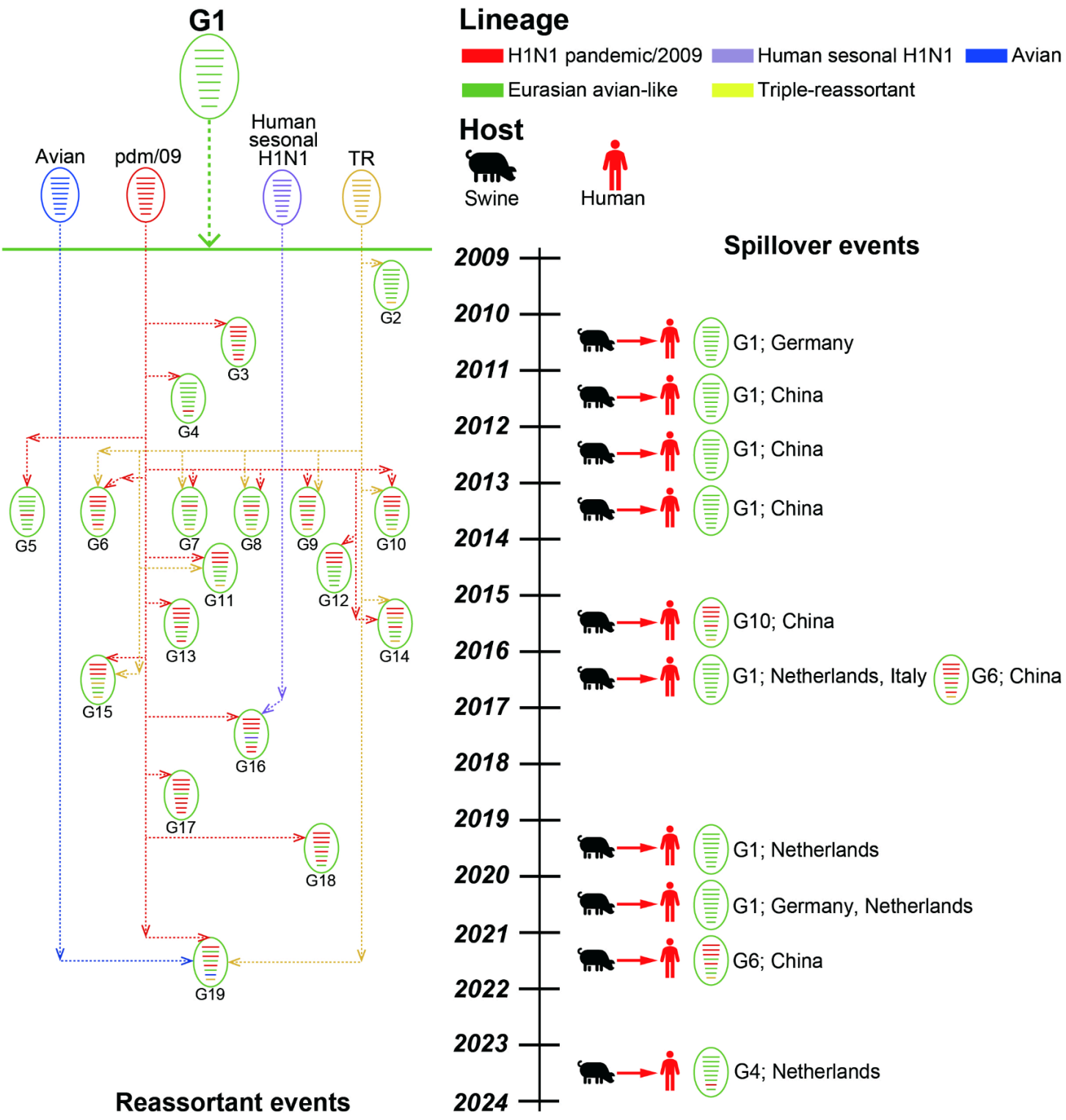


Figure 3

Reassortment history and human spillover events of Eurasian avian-like swine influenza A(H1N1) viruses.

The left panel illustrates the formation of different viral genotypes through reassortment between Eurasian avian-like (EA) swine influenza viruses and viruses of other lineages. Viral particles are

represented by ovals containing eight horizontal bars corresponding to the viral gene segments, from top to bottom: PB2, PB1, PA, HA, NP, NA, MP, and NS. Segments are color-coded according to their lineage of origin: EA (green), A(H1N1)pdm09 (red), triple-reassortant (yellow), human seasonal H1N1 (purple), and avian (blue). Dotted arrows indicate lineage contributions to the resulting reassortant genotypes. The central timeline indicates the detection years of the depicted reassortant genotypes and human spillover events. The right panel shows reported swine-to-human spillover events, together with the corresponding viral genome constellations, genotype designations, and countries of detection. Genotypes G1–G19 are defined in this study.

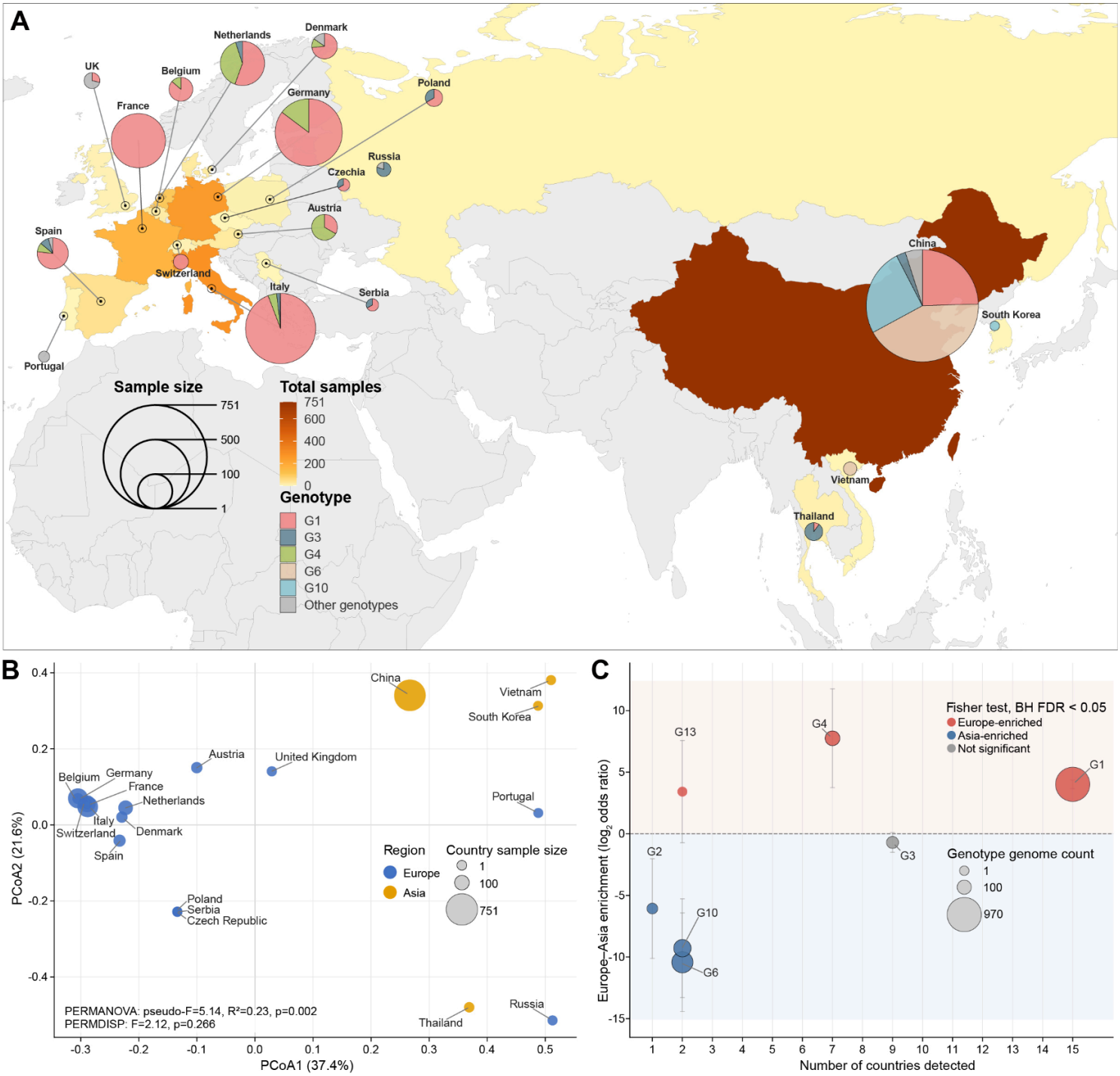


Figure 4

Geographic distribution and regional structuring of genotype diversity.
(A) Geographic distribution of sampled EA H1N1 viruses. Country shading indicates the total number of sampled genomes, circle size indicates country-specific sample size, and pie charts show genotype composition by country.
(B) Principal coordinates analysis of country-level genotype-composition dissimilarities based on Bray–Curtis distances. Points represent countries, colors indicate geographic regions, and point sizes are proportional to country-specific sample sizes. The first and second coordinates explain 37.4% and 21.6% of the total variation, respectively. Results of PERMANOVA and PERMDISP are shown.
(C) Regional enrichment of genotypes in Europe and Asia. The y-axis shows the Europe-to-Asia enrichment score expressed as a log2 odds ratio, and the x-axis indicates the number of countries in which each genotype was detected. Point size represents genome count. Red and blue points indicate genotypes significantly enriched in Europe and Asia, respectively, based on Fisher's exact tests with Benjamini–Hochberg correction.

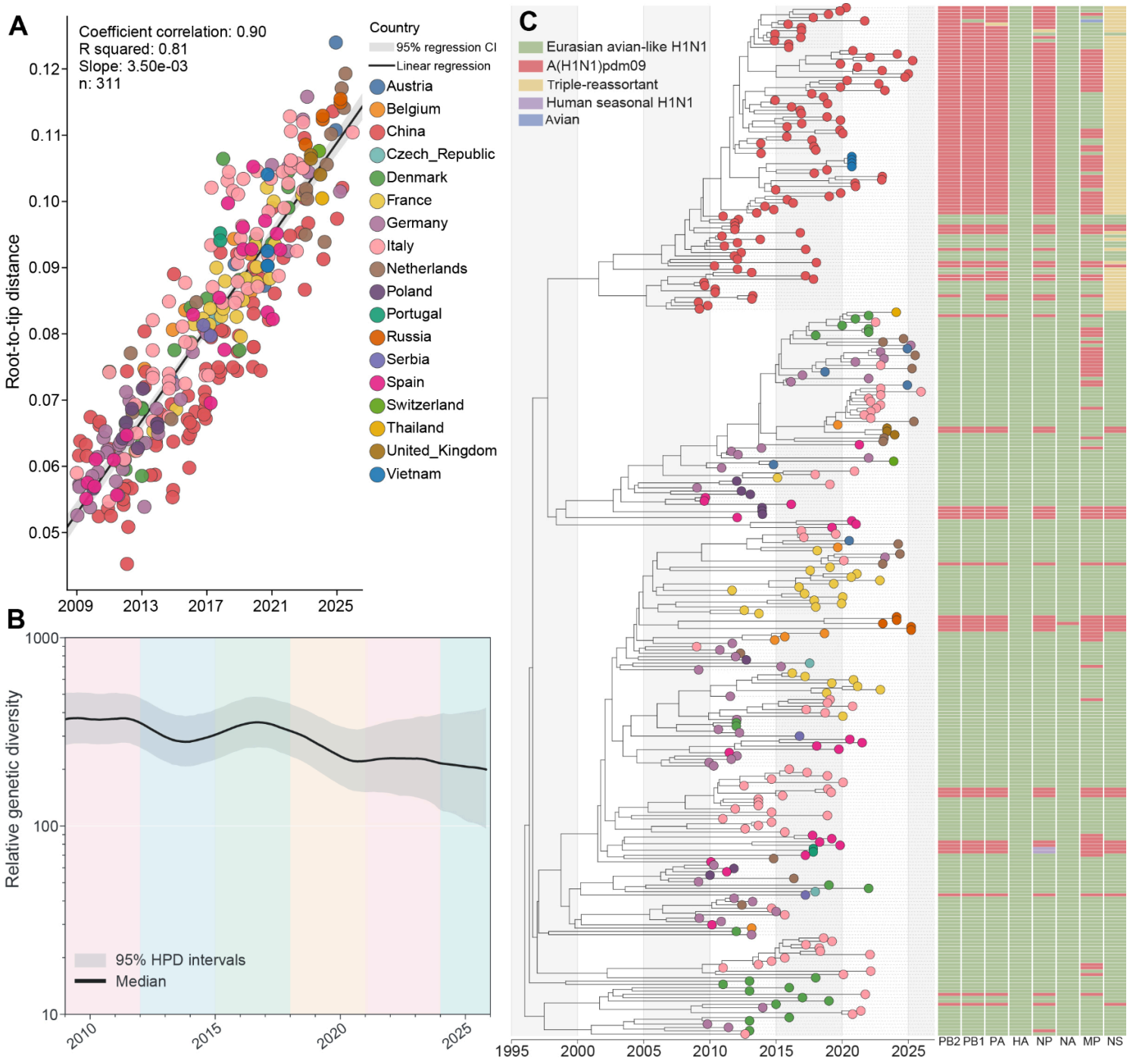


Figure 5

Temporal signal and population dynamics of EA H1N1 viruses.

(A) Root-to-tip regression analysis based on the 311-sequence time-calibrated HA dataset. Root-to-tip genetic distance is plotted against sampling year. Points are colored by country, and the regression line and its 95% confidence interval are shown. The correlation coefficient, R^2 value, regression slope, and sample size are indicated.

(B) Bayesian reconstruction of relative genetic diversity through time. The black line represents the posterior median, and the shaded area represents the 95% highest posterior density interval.

(C) Maximum clade credibility tree inferred using BEAST. Tip colors indicate sampling countries, and

adjacent colored bands indicate the inferred lineages of the PB2, PB1, PA, HA, NP, NA, MP, and NS segments. The horizontal axis represents calendar time.

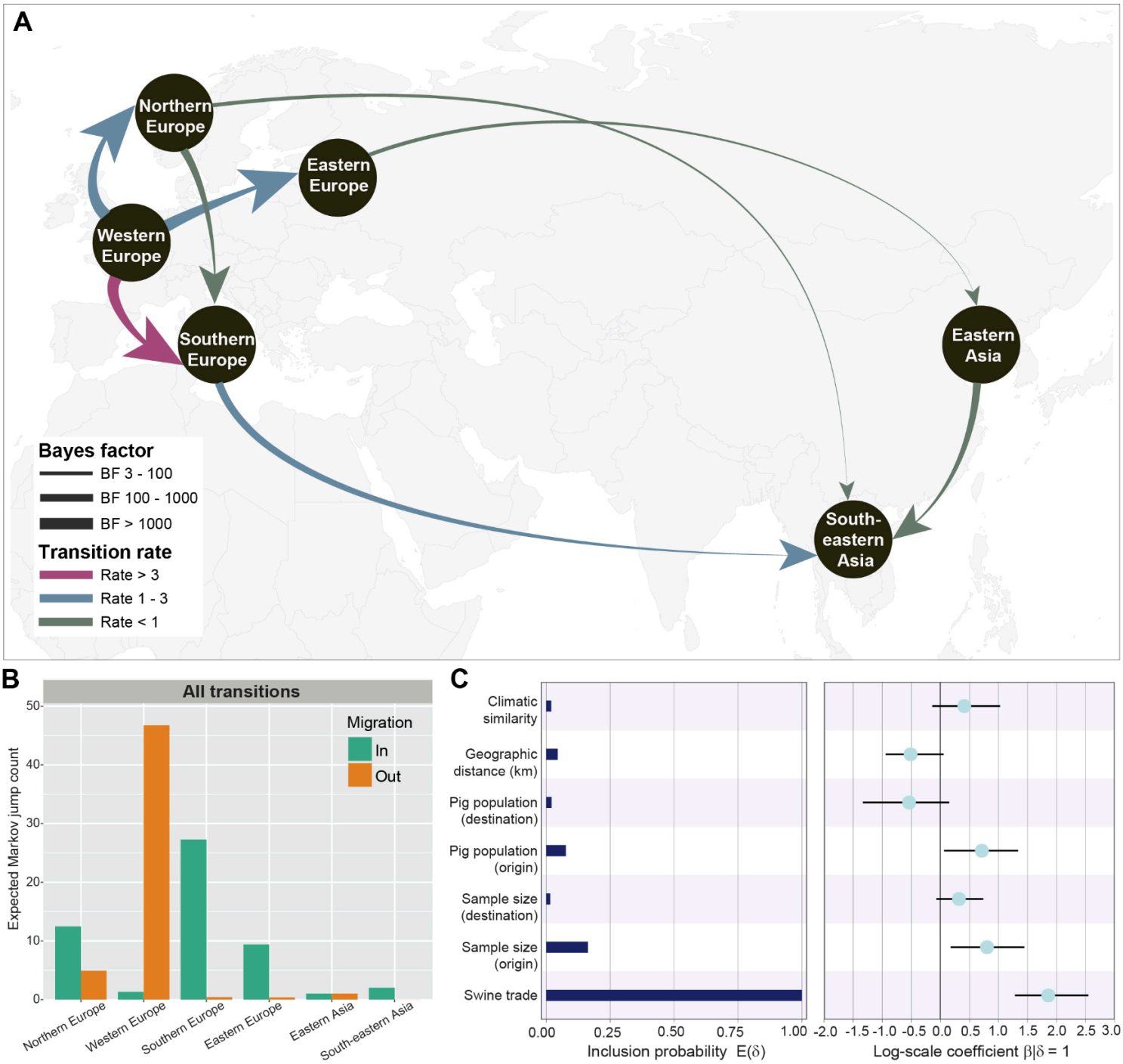


Figure 6

Spatial dissemination and predictors of viral migration.

(A) Bayesian discrete phylogeographic reconstruction with Bayesian stochastic search variable selection among six geographic regions. Arrows indicate supported migration pathways. Line width represents Bayes factor support, and line color represents the estimated transition rate.

(B) Expected numbers of Markov jumps associated with viral movement into and out of each geographic

region. Green and orange bars represent inbound and outbound transitions, respectively.
(C) Generalized linear model analysis of predictors associated with viral spatial diffusion. The left panel shows posterior inclusion probabilities, denoted by E(δ), for climatic similarity, geographic distance, pig population size, sampling intensity, and live-swine trade. Pig population size and sampling intensity are shown separately for the origin and destination regions. The right panel shows regression coefficients conditional on inclusion on the log scale. Points represent posterior estimates, and horizontal bars represent 95% credible intervals.

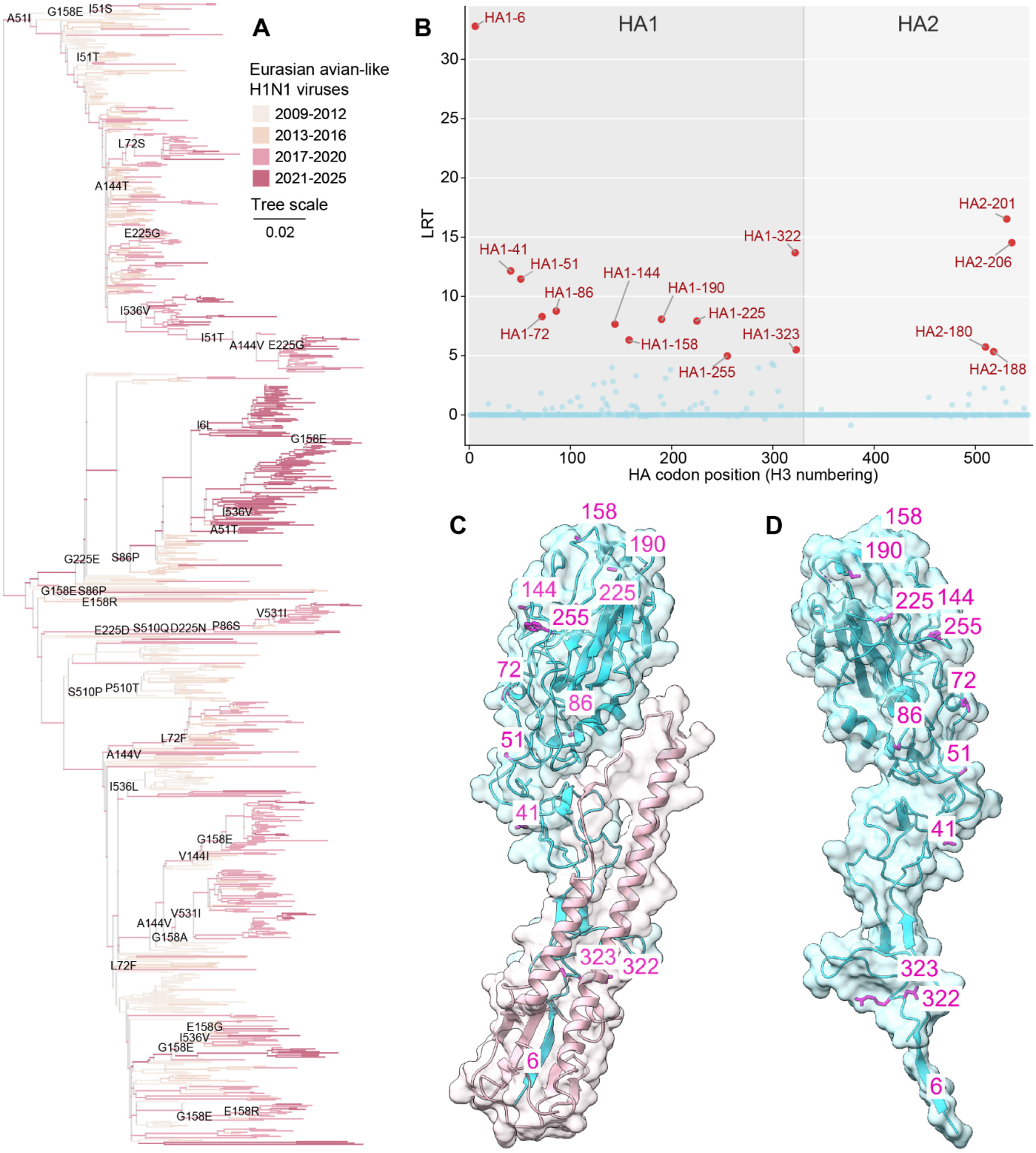


Figure 7

HA evolution, episodic diversifying selection, and structural distribution of selected residues.
(A) Maximum-likelihood phylogeny of HA sequences from EA H1N1 viruses collected between 2009 and 2025. Branches are colored according to sampling period, and recurrent amino acid substitutions are annotated at their inferred phylogenetic positions.
(B) Site-wise analysis of episodic diversifying selection across the HA gene using MEME. Likelihood-ratio statistics are plotted by codon position according to H3 numbering. Selected sites are highlighted and annotated, and the HA1 and HA2 domains are indicated.
(C) Structural mapping of selected HA residues onto the HA trimer structure, based on PDB 8GV5. The HA1 and HA2 domains are shown in distinct colors, and selected residues are highlighted in magenta.
(D) Rear view of the HA trimer showing the spatial distribution of selected residues within the HA1 domain.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- floatimage8.jpeg