Emergence of D1.1 reassortant H5N1 avian influenza viruses in North America

Alvin Crespo-Bellido[1§], Nídia S. Trovão[2,3], Alexander Maksiaev[1], Guy Baele[4], Simon Dellicour[4,5,6], Martha I. Nelson[1§]

[1] Division of Intramural Research, National Library of Medicine, Bethesda, MD USA
[2] Department of Pathobiology, College of Veterinary Medicine, University of Illinois Urbana-Champaign, Urbana, IL, USA
[3] Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA
[4] Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium
[5] Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Brussels, Belgium
[6] Interuniversity Institute of Bioinformatics in Brussels, Université Libre de Bruxelles, Vrije Universiteit Brussel, Brussels, Belgium

[§] **Corresponding authors:** Alvin Crespo-Bellido (alvin.crespobellido@nih.gov) and Martha I. Nelson (nelsonma@mail.nih.gov)

Keywords: influenza A virus, highly pathogenic avian influenza, H5N1, evolution, Bayesian phylodynamics, genomic reassortment

Summary (max 150 words, current 147 words)

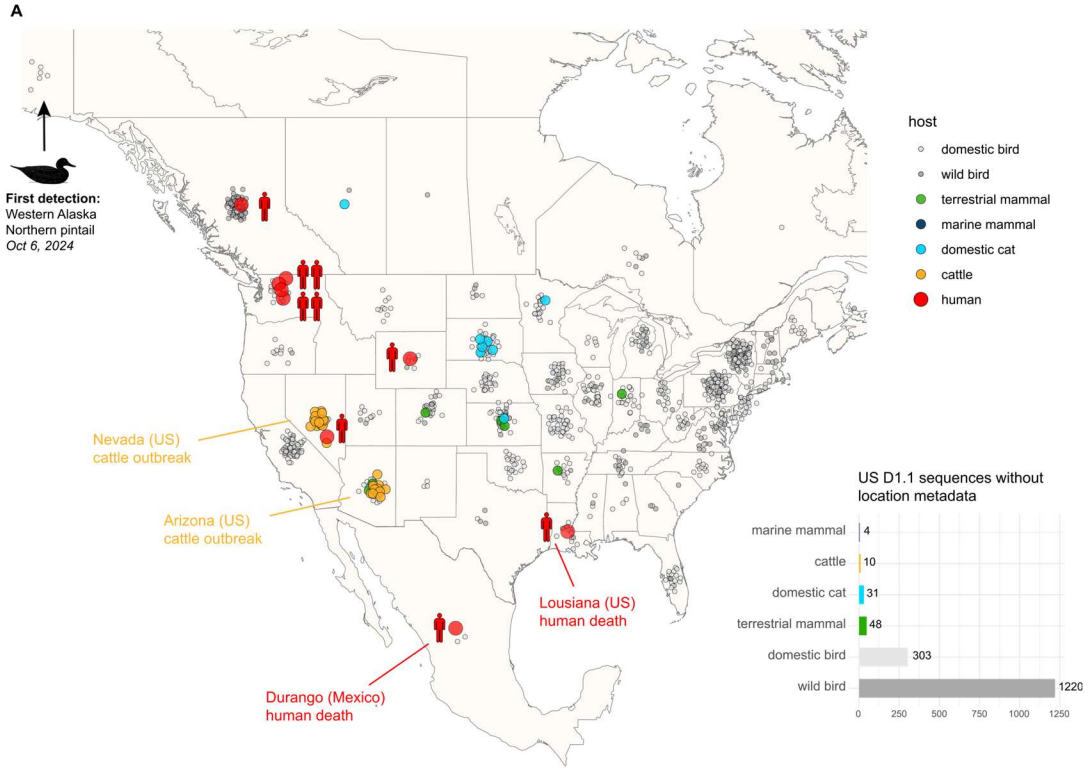
Since 2021, highly pathogenic avian influenza viruses (HPAIV) belonging to H5N1 clade 2.3.4.4b have caused high mortality in North American wild birds and poultry. In 2025, a new D1.1 genotype caused two human deaths and host-switched to dairy cattle. However, the evolutionary origins and dynamics of D1.1 have not been fully characterized. Here, our phylogenetic analysis of 17,516 H5N1 genome sequences uncovers how D1.1 introduced a major shift in the antigenic diversity and ecology of the H5N1 epizootic in North America. D1.1 is the first major H5N1 genotype to (a) emerge in the Pacific flyway and spread west-to-east faster than any prior genotype; (b) antigenically shift via reassortment with the North American N1 segment, displacing the previously fixed Eurasian N1; and (c) transmit to a broader range of host species than any H5N1 genotype to date, introducing mammalian adaptations.

Introduction

Highly pathogenic H5N1 avian influenza viruses from clade 2.3.4.4b have circulated in Eurasia since 2016[1], causing severe mortality events in wild birds[2], poultry[3,4], and farmed mink and foxes[5,6]. In November 2021, H5N1 clade 2.3.4.4b viruses introduced by wild birds migrating from Europe were detected in Newfoundland, Canada, representing the first detection of clade 2.3.4.4b in the Western hemisphere[7]. Eurasian-origin H5N1 viruses spread across North America's four migratory flyways during 2021-2022 and underwent genomic reassortment with North American low pathogenicity avian influenza (LPAI) viruses. Entire genome segments were swapped between H5N1 and LPAI viruses to produce new genotypes with novel combinations of genome segments[8]. To categorize the vast array of new H5N1 genotypes containing North American LPAI genes, researchers at the US Department of Agriculture developed a software tool to rapidly categorize new strains referred to as GenoFlu[8]. GenoFlu developed a nomenclature in which original non-reassorted Eurasian H5N1 viruses were referred to as "A" genotypes. Reassortants with H5N1 and LPAI segments were referred to as "B" or "C" genotypes. The GenoFlu nomenclature became an essential tool to describe new reassortant H5N1 genotypes emerging in the Americas, including the B3.2 viruses that caused mass mortality in South American coastal wildlife[9–12], the B3.6 genotype that infected goats in Minnesota in March 2024[13], and the B3.13 viruses that caused a multistate outbreak in US dairy cattle during 2024-2025[14–16] (still ongoing as of December 3, 2025).

A new genotype termed "D1.1" was first detected in a wild duck in western Alaska on October 6, 2024 (A/northern pintail/USA/IZ24-0555/2024(H5N1), accession PV602213; **Figure 1A**). In November 2024, a D1.1 virus hospitalized a teenager with no known animal exposure in British Columbia, Canada[17] (**Figure 1B**). In December 2024, the D1.1 genotype caused the first fatality from avian influenza in the Americas, which occurred in an adult with underlying medical conditions who was exposed to backyard poultry in Louisiana, USA[18]. In early 2025, D1.1 viruses were identified in dairy cattle in Nevada and Arizona, leading to sustained outbreaks involving multiple herds in both states[19,20]. In Nevada, a dairy worker infected with D1.1 following exposure to raw milk experienced mild conjunctivitis symptoms similar to the dairy workers infected with B3.13

in California, Colorado, and Michigan in 2024[21]. In March 2025, a D1.1 infection in Durango, Mexico – and the first reported H5N1 human infection in Mexico overall – led to the death of a young child with no underlying medical conditions, no travel history, nor exposure to infected animals[22].



B

D1.1 human samples

Virus name	Collection date	Location	Source	Disease severity	Age	Acquired mammalian-adaptive PB2 mutations	Accession (GISAID)
A/Washington/240/2024	10/18/2024	USA-WA	commercial poultry	mild (unspecified)	adult (>18)	--	EPI_ISL_19531299
A/Washington/UW32495/2024	10/18/2024	USA-WA	commercial poultry	mild (unspecified)	adult (>18)	--	EPI_ISL_19541397
A/Washington/254/2024	10/23/2024	USA-WA	commercial poultry	mild (unspecified)	adult (>18)	--	EPI_ISL_19666173
A/Washington/255/2024	10/29/2024	USA-WA	commercial poultry	mild (unspecified)	adult (>18)	--	EPI_ISL_19552697
A/British Columbia/PHL-2032/2024	11/9/2024	Canada-BC	unknown	hospitalization (with conjunctivitis)	teen (13-18)	E627K	EPI_ISL_19548836
A/Louisiana/12/2024	12/11/2024	USA-LA	backyard poultry	death	adult (>18)	--	EPI_ISL_19634828
A/Nevada/10/2025	2/4/2025	USA-NV	dairy cattle	mid (conjunctivitis)	adult (>18)	D701N	EPI_ISL_19726293
A/Wyoming/01/2025	2/11/2025	USA-WY	backyard poultry	hospitalization	adult (>18)	E627K	EPI_ISL_19749443
A/Mexico/Dur_Indre_2292/2025	3/15/2025	Mexico-Durango	unknown	death	child (<13)	--	EPI_ISL_20227034

Figure 1. H5N1 genotype D1.1 samples from North America, 2024-2025. (A) Map of North America showing the geographic distribution of publicly available D1.1 samples collected in 2024-2025 across the US, Canada and Mexico (*n* = 2734). The coordinates for each data point are not exact and are jittered around the centroid for each US state, Canadian province and the country of Mexico. Circles are colored according to the host type. The inset bar graph shows the number of US D1.1 sequences in the data set without available state metadata by host type. (B) Characteristics of nine D1.1 infections in humans that have been sequenced as of December 16, 2025.

There is an urgent need to track the changing ecology and genetics of H5N1 in North America. To compare D1.1 evolution to prior H5N1 genotypes, we performed a phylogenetic analysis using 17,516 H5N1 2.3.4.4b genome sequences that are available from all host species in the Americas, as of June 13, 2025. Here, we use Bayesian phylodynamic approaches to provide a comprehensive analysis of the central evolutionary processes that drive H5N1 diversity in the Americas, including (a) intercontinental migration from Asia and Europe, (b) spatial dissemination across the Americas, (c) genomic reassortment with LPAI, and (d) interspecies transmission followed by host adaptation. In particular, we examine how the evolutionary dynamics of D1.1 differ from prior H5N1 genotypes in all of these areas.

Results

Introductions of H5N1 clade 2.3.4.4b from Europe and Asia into North America. Of the nine total introductions of H5N1 clade 2.3.4.4b viruses into North America from Eurasia since 2021, most did not sustain long-term transmission or reassort with LPAI (**Figures 2-3**). On the east coast in the Atlantic flyway, the A2 genotype introduced from Europe in 2021 was the only introduction to transmit extensively in wild birds and reassort with LPAI to create new North American genotypes. GenoFlu[8] defines A1 and A2 as separate genotypes, distinguished by mutations, but a close inspection of the segment phylogenies reveals that A1 and A2 likely descend from a single introduction from Europe and that A1 descends from A2 (**Figure S1**). The B genotypes descend from A1, while the C genotypes descend from A2 (**Figure 2**). On the west coast in the Pacific flyway, most introductions of the A3 genotype from Asia do not sustain long-term transmission in the Americas (there have been at least six, **Figure 3B**). Only the third A3 introduction was transmitted extensively in North American wild birds and reassorted with North American LPAI, generating new D reassortant genotypes that include D1.1, D1.2, D1.3, and D2.1 (**Figures 2-3**). Of the four D genotypes, only D1.1 transmitted extensively in wild birds, causing a major outbreak in early 2025 (**Figure 4A**).

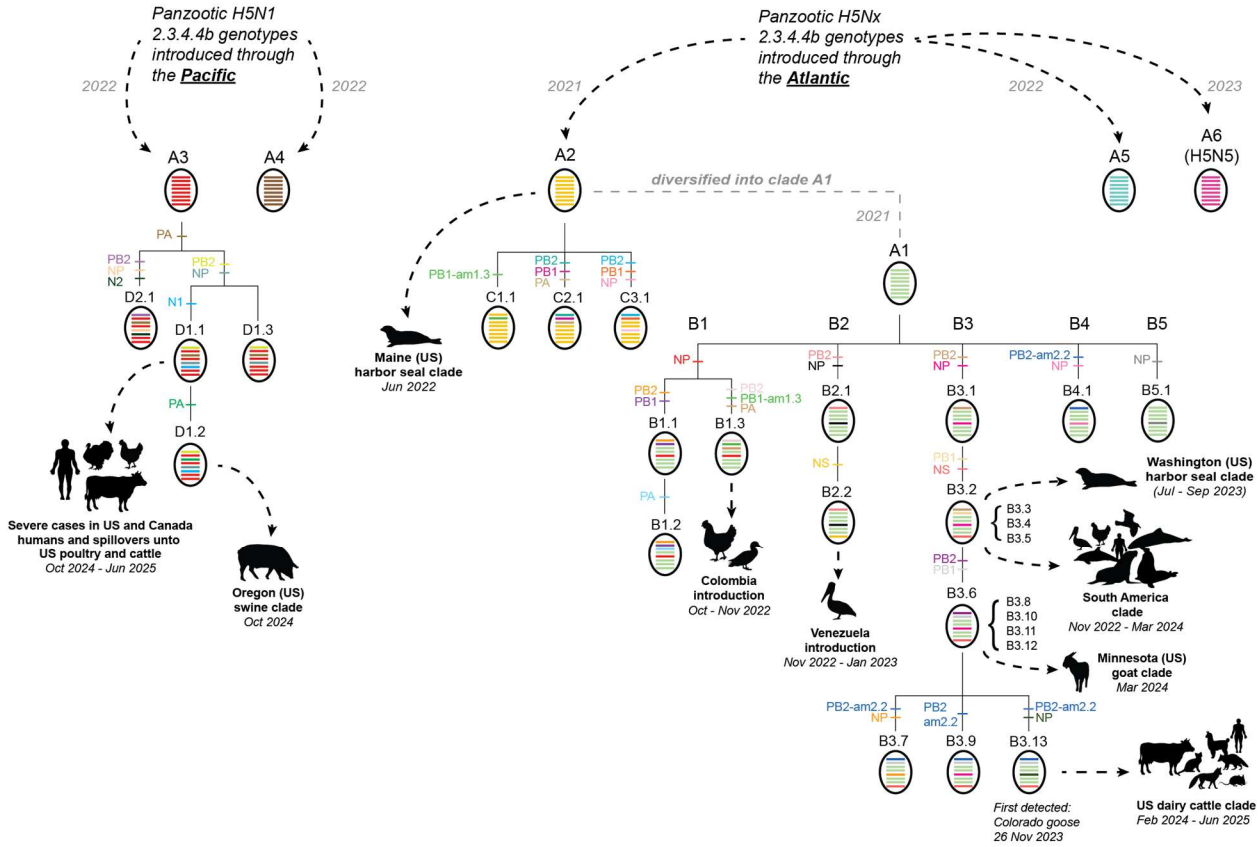


Figure 2. Reassortment history of H5N1 reassortant genotypes in the Americas. Each oval represents a genotype (A-D). Each bar inside the oval represents one of the eight gene segments of the IAV genome, ordered top-to-bottom from longest to shortest: PB2, PB1, PA, HA, NP, NA, MP, NS. Bars are shaded by lineage, as defined by GenoFlu (see "Genotype key" in **Supplemental File 1**) and confirmed by phylogenetic analysis (**Figure S2**). For simplicity, all bars from "A" genotypes are shaded the same color. Colored horizontal bars preceding each genotype indicate a reassorted segment of American LPAI origin, again colored by lineage. Asterisk indicates independent acquisition of the same lineage present in multiple genotypes. Notable outbreaks are identified with animal cartoons.

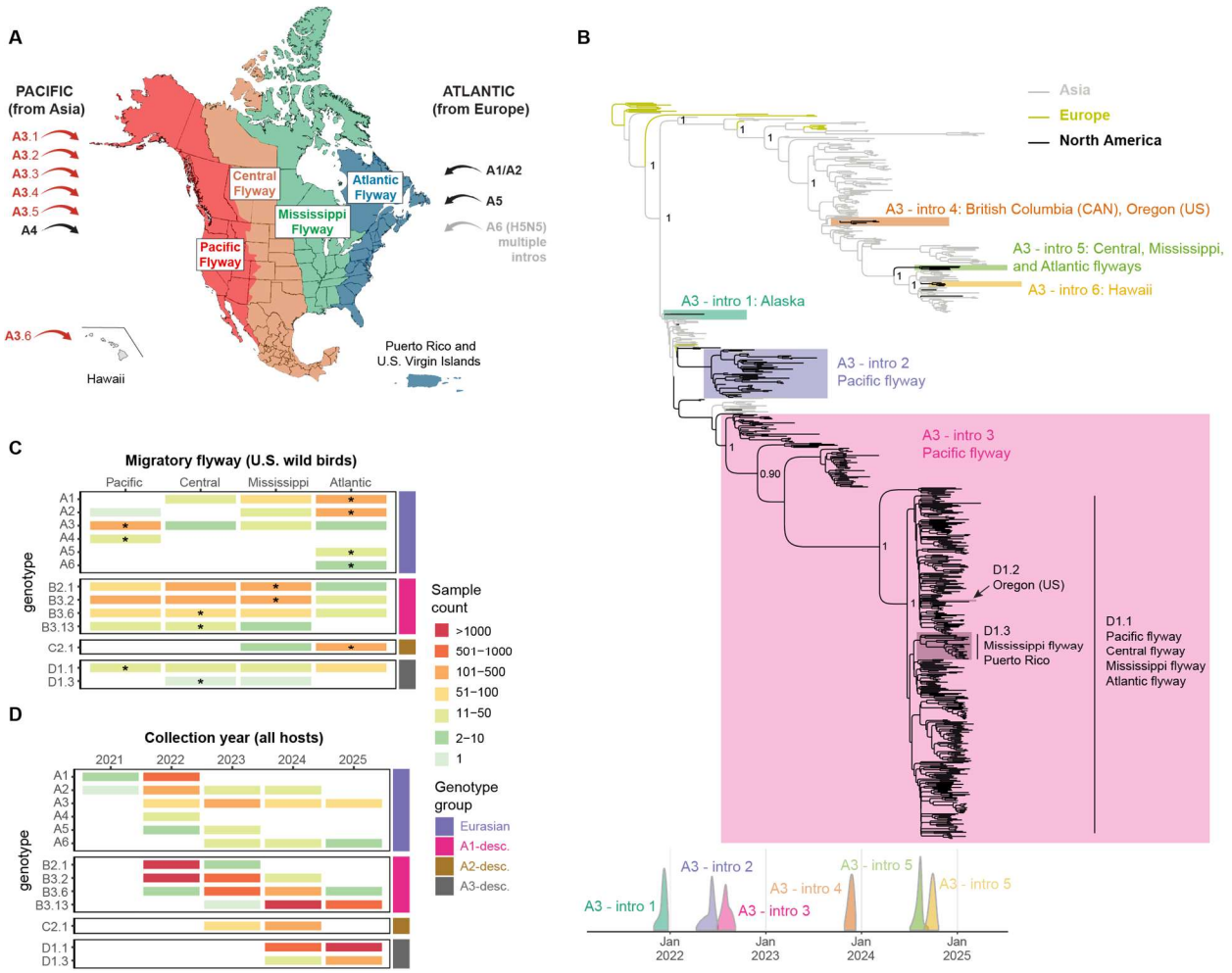


Figure 3. **Introductions of A genotypes into the Americas, 2021-2025.** (A) Map of North America's four administrative migratory flyways, defined by the US Fish and Wildlife Service. Arrows indicate independent introductions of Eurasian H5Nx clade 2.3.4.4b genotypes into North American migratory flyways, where black arrows indicate single H5N1 introductions, red arrows indicate repeated introductions of genotype A3, and the grey arrow indicates multiple introductions of H5N5 genotype A6. (B) Time-scaled maximum clade credibility (MCC) tree inferred for four concatenated genome segments (PB1, HA, MP and NS) that belong to the Eurasian lineage in the A3, D1.1, D1.2, and D1.3 genotypes. Branches are shaded by geographic location. Posterior probabilities are provided for key nodes. The 95% highest posterior density interval for the timing of each of the identified North American introductions is provided below the phylogeny. (C) Heatmap shaded by the number of sequences in each North American migratory flyway for U.S. wild birds (n = 4,501) and (D) heatmap shaded by collection year across all hosts (n = 14,443). Only select genotypes are shown. Asterisks denote the flyway where the initial detection occurred.

Genomic reassortment between H5N1 and North American LPAI. From the arrival of clade 2.3.4.4b in 2021 until the emergence of D1.1 in 2024, most reassortment events between H5N1 and North American LPAI involved the viral ribonucleoprotein (vRNP) complex, which includes the polymerase (PB2, PB1, PA) and NP segments (**Figures 4B-4C**). The NA and MP segments were rarely acquired during reassortment with LPAI. In

144 fact, 100% of H5N1 viruses circulating in the Americas prior to the emergence of D1.1 in
145 October 2024 still retained the Eurasian N1 (**Figure S3, Supplemental File S1**). Then,
146 the D1.1 genotype emerged with four segments from North American LPAI (PB2, PA, NP,
147 and NA (N1)) (**Figure 4B**) and four segments from the A3 virus introduced into the Pacific
148 flyway from Asia (introduction 3, **Figure 3B**). The LPAI PB2, PA, and NP fit the prior
149 pattern of frequent H5N1 reassortment with the LPAI vRNP, but no prior major
150 reassortants had acquired a new NA. The D1.3 genotype, which caused a large outbreak
151 in poultry in Ohio and Indiana, has the same genome constellation as D1.1 for seven
152 segments, but did not acquire the North American N1 (**Figure 4B**). D1.2 was sampled in
153 two outdoor pigs and some poultry in Oregon, U.S., and is descended from D1.1, but
154 subsequently acquired a different LPAI PA segment (**Figure 2**). Like D1.1, D2.1 is also
155 descended from A3 and shares the same LPAI PA segment, but acquired a PB2, NP and
156 NA (N2) from LPAI (**Figure 2**) and has not spread widely, having been detected once in
157 a chicken in British Columbia.

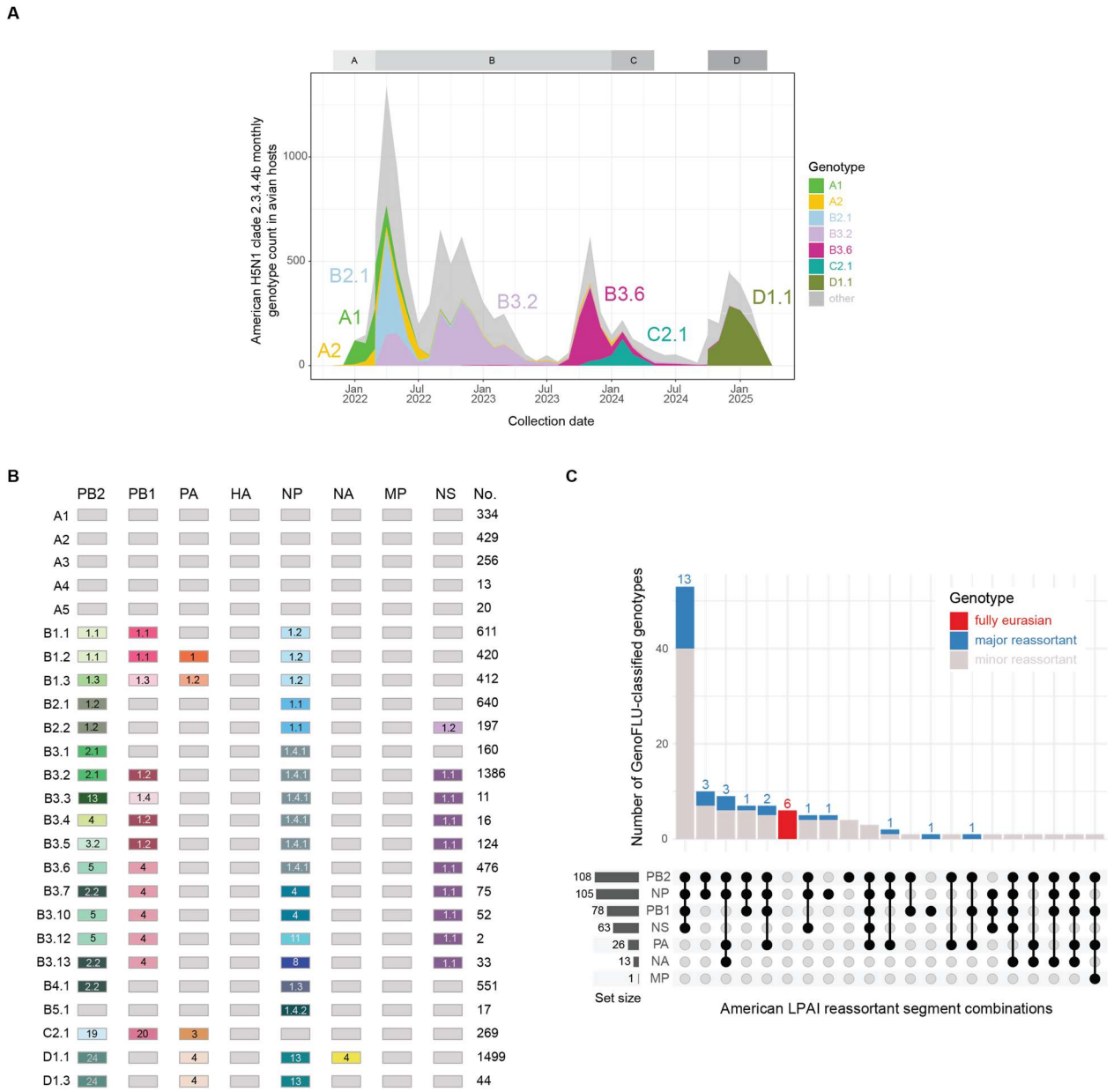


Figure 4. Reassortant H5N1-LPAI genotypes in the Americas. (A) Number of H5N1 clade 2.3.4.4b sequences detected for each genotype, by month, in avian hosts in the Americas during November 2021 - February 2025, based on 6,872 H5Nx 2.3.4.4b genome sequences with complete collection dates. (B) Genotype constellation for clade 2.3.4.4b H5N1 major genotypes with >10 sequences available. Grey squares indicate Eurasian lineage; colored squares represent reassorted North American LPAI lineages. The number of genome sequences available from wild birds in the Americas is provided for each genotype in the last column. (C) The y-axis represents the number of H5Nx genotypes (defined by GenoFlu; n = 121) with each combination of LPAI segments. Major reassortant genotypes (as defined by GenoFlu, see Methods) are shaded blue; minor reassortant genotypes are shaded grey; non-reassorted fully Eurasian genotypes are shaded red. The number of genotypes with LPAI derived from each segment is indicated on the left.

N1 amino acid substitutions in D1.1. The acquisition of an entire N1 genome segment from North American LPAI represents a major genetic change in a key antigenic protein

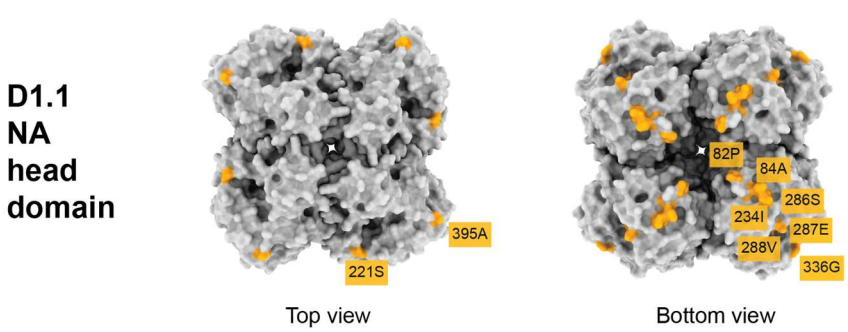
173 (**Figure 5A**). The LPAI N1 acquired by D1.1 viruses differed at 23 amino acid positions
174 compared to the Eurasian-origin N1 segment found in other H5N1 genotypes in the
175 Americas and globally. The 23 residues are distributed across all major functional
176 domains of the N1 protein (**Table S3**), including 12 in the transmembrane and stalk
177 domains (positions 1-84 in N1 numbering) that anchor the protein in the viral envelope
178 and influence its stability and 11 on the globular head domain (positions 85+ in N1
179 numbering) that is a primary target for the host immune system. Most D1.1-specific
180 globular head domain residues are arranged in a diagonal pattern on the underside
181 surface of the head, which typically is highly conserved (**Figure 5B**).

A

Segment origin: Eurasian H5N1 / American LPAI

Genotype	PB2	PB1	PA	HA	NP	NA	MP		NS	
							M1	M2	NS1	NEP
B3.2	340K, 464M	215K, 429R	--	--	--	--	--	--	--	--
B3.6	251K, 353R, 663R, 667I	--	--	--	230L	--	--	--	--	--
B3.13	109I, 362G, 441N, 495I, 631L, 649I, 676A	--	219I, 497R	--	482N	--	--	--	7L	7L
C2.1	129A, 470N	386K, 618G	104R, 609R	210T, 488Y	353I, 450S	--	--	--	70G, 111M, 114P, 161T, 176D	--
D1.1	--	16D, 154S, 207R, 614D	--	36A, 104L, 210V, 325R, 476D, 511I	--	**8I, 20I, 23V, 44N, 45H, 48T, 53V, 74L, 75I, 81D, 82P, 84A, 221S, 234I, 241I, 257R, 286S, 287E, 288V, 329S, 336G, 338V, 395A**	200A	61R	75G, 147I, 171N, 193Q	36K, 63E

B



182

183 **Figure 5. Mutational features of D1.1.** (A) Genotype-defining residues among dominant reassortant
184 genotypes. Only residues found in >90% of all sequences within the genotype but <20% of other major
185 genotypes in the table and <10% of all other clade 2.3.4.4b sequences overall are listed. All residues follow
186 sequential numbering except HA, which follows an H5 numbering scheme. (B) Tetrameric NA head domain
187 protein structure corresponding to H1N1 strain A/California/04/2009 (PDB: 3NSS) annotated with D1.1-
188 specific residues. The surface of the head protein structure is shown at the top while the underside is shown
189 at the bottom.

Rapid spatial diffusion of D1.1 in North America. Bayesian phylogeographic analyses estimate that the D1.1 genotype emerged in the Pacific flyway in mid-2024 (95% HPDI = 2024-07-19 to 2024-08-09; **Figure 6**), following the emergence of B3.2 and B3.6 in 2022, and C2.1 in 2023 (estimates of the posterior density distribution for the time of the most recent common ancestor (TMRCA) are shown in **Figures S4-S7**). D1.1 viruses then rapidly spread in wild birds, displacing the B and C genotypes and emerging as the dominant North American subtype in 2025 (**Figure 4A**). The D1.1 genotype dispersed across all four flyways during 2024-2025, with the highest weighted diffusion coefficient of North America's four major dominant genotypes (B3.2, B3.6, and C2.1) at ~3800 km^2/day (**Figure 6**, **Table S2**). During the initial invasion phase, the wavefront velocity of D1.1 was estimated at ~13,000 km/year, similar to the B3.2 genotype (~11,500 km/year), but higher than for the C2.1 and B3.6 wavefront velocities of ~4,000 km/year and ~2,500 km/year, respectively.

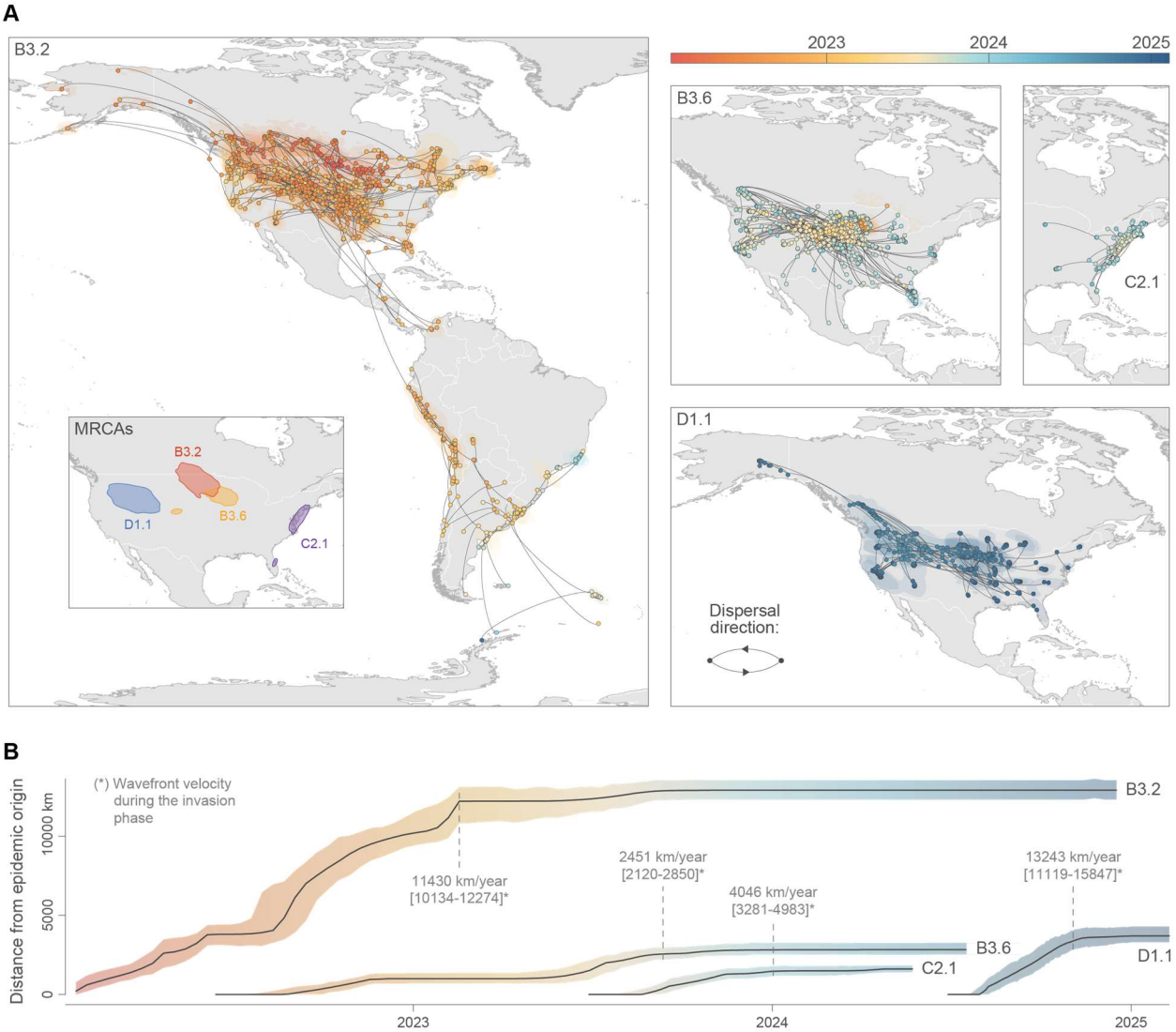


Figure 6. Dispersal history of H5N1 reassortant genotypes in the Americas. (A) Maps display the phylogeographic reconstruction inferred for each genotype. Curved line segments represent the branches of the maximum clade credibility (MCC) tree. Nodes are colored according to their time of occurrence, progressing from red (2022) to blue (2025). Each MCC tree is superimposed on 80% highest posterior density (HPD) polygons reflecting phylogeographic uncertainty, which are colored according to the same time scale. The inset map provides the estimated geographic origin for each genotype. These locations represent the 95% HPD region for the most recent common ancestor (MRCA) from which each genotype descended. (B) The plot quantifies the expansion dynamics by displaying the distance of the epidemic wavefront from its geographic origin (the MRCA) over time. Solid lines represent the median distance, while shaded areas represent the 95% HPD intervals. The annotated values are the median wavefront velocities in kilometers per year during the initial invasion phase for each genotype.

D1.1 spillover and adaptation to mammals. D1.1 viruses spilled over from wild birds into a wide range of host species in North America (**Figure 7A**) and were the only genotype to infect all seven host types: wild bird, terrestrial mammal, marine mammal, domestic bird, cattle, domestic mammal (other than cattle), and human (**Figure 7B**). D1.1

220 viruses also had a broad host range in wild birds, and were detected in 12 taxonomic
221 orders, equaling B3.2 which also spread widely across South America (**Figure 7B**).
222 Whereas reassortment generated most of the amino acid changes observed in N1,
223 additional amino acid changes were observed among D1.1 samples following spillover
224 into other hosts (**Figure 7C, Supplemental File 2**). Most spillovers from D1.1 into
225 humans, domestic cats, and terrestrial and marine carnivores involve direct exposure to
226 wild birds, and key PB2 mutations that improve replication efficiency in mammals (e.g.,
227 E627K and D701N) are frequently selected for. Of the nine D1.1 infections in humans,
228 three viruses acquired PB2 mammalian adaptations: D701N in A/Nevada/10/2025 and
229 E627K in A/Wyoming/01/2025 and A/British Columbia/PHL-2032/2024 (**Figure 1B**). In
230 contrast, B3.13 is rarely detected in wild birds and direct B3.13 spillovers from wild birds
231 into humans, wild carnivores, or domestic cats are not observed. Instead, B3.13 spillover
232 occurs through a bovine intermediary. Since the B3.13 bovine lineage already acquired
233 the PB2 M631L substitution that improves replication in mammals, no additional E627K
234 or D701N mutations were observed in B3.13 spillovers from cattle into other host species
235 (**Figure 7C**). The A/Texas/37/2024 strain containing the E627K mutation was collected
236 early in the B3.13 outbreak in the Texas panhandle, prior to the M631L mutation that later
237 became fixed in cattle.
238

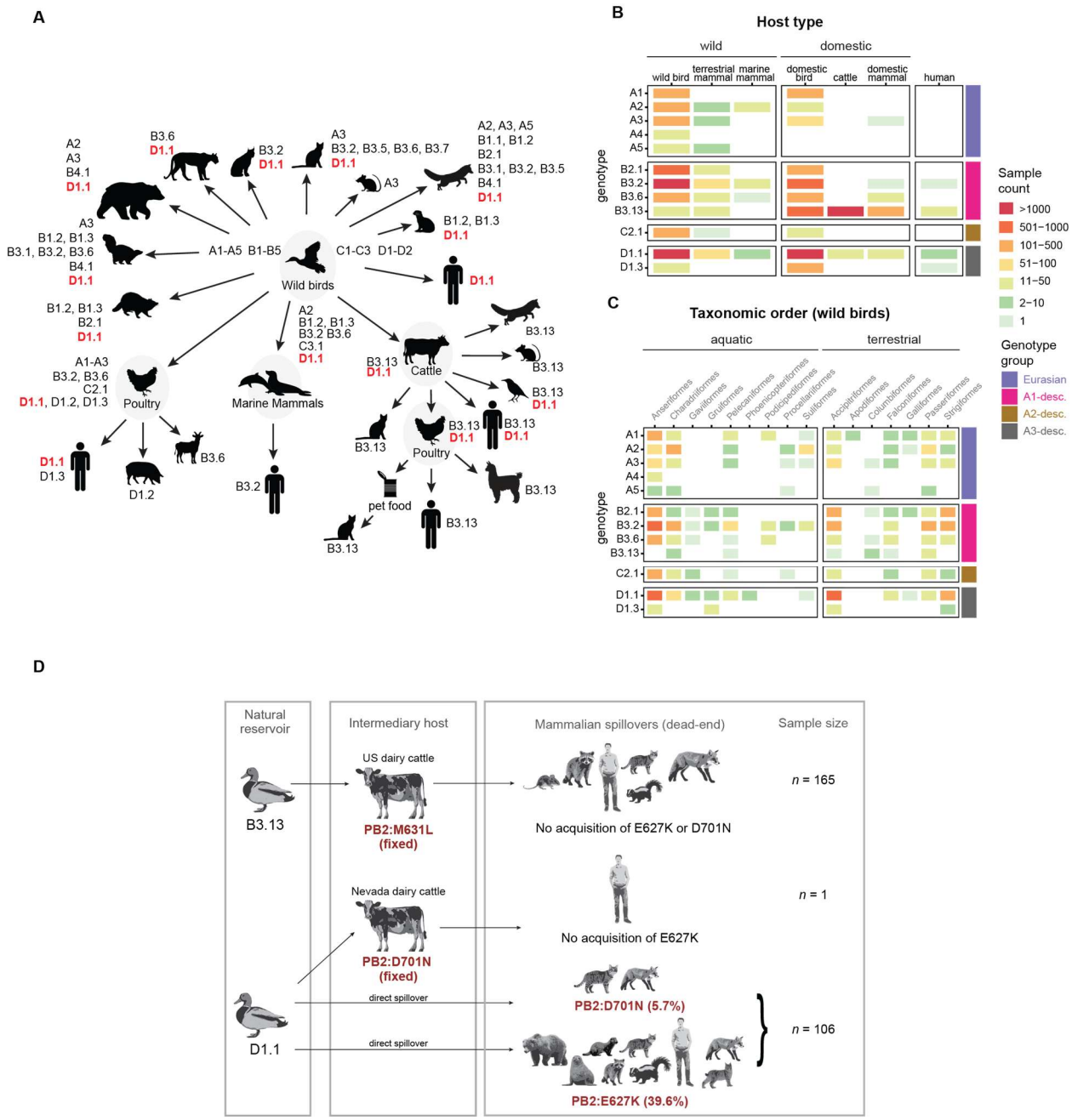


Figure 7. Host range of clade H5N1 clade 2.3.4.4b genotypes. (A) Spillover of H5N1 clade 2.3.4.4b genotypes between different host species. (B) Heatmap shaded by the number of sequences in each wild bird taxonomic order (n = 4,044) and (C) host type (n = 13,488). Only select genotypes are shown. (D) Acquisition of mammalian-adaptive amino acid substitutions PB2:E627K and PB2:D701N among B3.13 and D1.1 mammalian samples.

Discussion

The arrival of H5N1 clade 2.3.4.4b in the Americas in 2021 opened new pathways for virus evolution, including genomic reassortment with North American LPAI segments that are present only in the Western hemisphere. In contrast to the prior 2014-2015 H5Nx outbreak in North America that predominantly affected poultry, the 2021-2025 epizootic has sustained widespread transmission in wild birds[23], including the *Anseriformes* order of ducks and geese that constitute the natural reservoir for LPAI and provide ample opportunity for reassortment. The D1.1 genotype's acquisition of a new N1 neuraminidase from the North American LPAI lineage by reassortment in 2024 represents a turning point in the evolution of H5N1 in the Americas. Prior to the emergence of D1.1 in mid-2024, mostly LPAI internal gene segments had been acquired by H5N1 in the Americas. The rapid expansion of D1.1 viruses fundamentally altered the genetic composition of the H5N1 epizootic in North America, and 92.8% of clade 2.3.4.4b viruses sampled January 1 – June 13, 2025 contained the LPAI N1, mainly due to the emergence of D1.1 (**Figure S3C-S3D**).

The North American LPAI N1 segment acquired by D1.1 differs from the Eurasian N1 at numerous amino acid sites, and the effect of these changes on virus phenotype and pandemic risk requires further study. Human exposure to H1N1 seasonal influenza viruses containing an N1 segment from the 2009 H1N1 pandemic, originally sourced from Eurasian swine, is thought to provide a degree of cross-protection against most H5N1 clade 2.3.4.4b viruses circulating in birds globally. Whether the North American LPAI N1 elicits similar cross-protection needs to be examined. In contrast to previous H5 clades, clade 2.3.4.4b H5 preferentially pairs with long stalk N1[24], which includes the North American LPAI N1[25]. The amino acid changes observed in the North American LPAI N1 segment are concentrated primarily on the underside of the globular head domain – or the neuraminidase "dark side" (NDS) – which is positioned away from the enzymatic active site. This region has recently been identified as containing highly conserved epitopes that serve as effective targets for cross-reactive monoclonal antibodies against HxN1 infections[26].

The future trajectory of D1.1 in the Americas is uncertain, including whether D1.1 spreads to South America this winter with migrating wild birds. One reason B3.2 spread to South America could be the overlap in timing of its peak activity in North American wild birds during September-December, the same time that North American birds migrate to South America to overwinter in warmer climates. In that scenario, a genotype needs not only to be highly fit in North American wild birds, but to reach peak prevalence at the same time wild birds are migrating. D1.1's North American peak occurred during the winter of 2025, prior to southward migrations, and it remains unclear if D1.1 will peak again or be replaced by a new genotype. A recent human fatality involving an H5N5 virus in Washington State (reported on November 15, 2025) was surprising, given that H5N5 was introduced into North America in 2023[27] and was excluded from further examination in this study because it had been sparsely sampled up to this point. The human case indicates that H5N5 has reached the Pacific flyway in wild birds, even though the virus had never been detected outside the Atlantic flyway. If H5N5 is an emerging genotype that presents a new risk to humans, our surveillance is inadequate. Relatively small investments in data pipelines for avian influenza could provide a large return in tracking evolution and zoonotic risk. We are currently lagging even in D1.1 surveillance. Of the 1,722 D1.1 sequences available from North American wild birds that were deposited on NCBI's Sequence Read Archive (SRA) as of Sep 10, 2025 (**Figure S8**), only 8.9% (*n* = 154) have a corresponding GenBank entry that includes a collection date (month, day, and year) and more precise sampling location (US state). If complete metadata were available for even half of those samples, scientists would be able to track D1.1 with higher precision.

Additional background LPAI data are also needed to trace H5N1 reassortment histories. North American avian influenza surveillance has understandably prioritized sequencing H5N1 over LPAI in recent years, but a lack of LPAI sequences in public databases makes it difficult to resolve phylogenies and distinguish new reassortment events involving LPAI. Additional avian influenza sequence data from Europe are also needed as background context for the North American epizootic. Our phylogenies support a scenario in which the A1 and A2 genotypes represent a single introduction from Europe to North America, but additional background sequences from Europe are needed to

310 provide further confirmation. There is also a need for a universal H5N1 nomenclature.
311 The A1-A6 genotypes from Eurasia defined by GenoFlu have different names in the
312 nomenclature systems devised by European influenza researchers. As H5N1 expands its
313 geographical range globally, greater international coordination is needed.

314 As the epicenter for H5N1 evolution shifts to the Americas, there is an urgent need
315 to expand the region's infrastructure for real-time surveillance and monitoring. The
316 Americas present a different ecological landscape compared to Eurasia, including South
317 American marine mammals that are highly susceptible to H5N1 disease and farming
318 systems with long-range animal movements that sustain virus persistence in dairy cattle[16].
319 Our capacity to track the spread and evolution of H5N1 in the Americas is currently made
320 possible by the collective efforts of many government and academic partners, including
321 the U.S. Department of Agriculture[8], Centers for Disease Control and Prevention, US
322 Geological Survey[28], the Canadian government[29], South American researchers[10,12], and
323 NIH-funded Centers of Excellence for Influenza Research and Response (CEIRR)
324 network[30]. Still, further investment in animal health is needed to monitor sudden changes
325 in virus evolution and risk to humans, livestock, and wildlife in real-time. Thus far, Eurasia
326 has been a major source of H5N1 viruses for North America, and there is little evidence
327 of D1.1, B3.2, or other dominant North American genotypes migrating back to Europe or
328 Asia. However, H5N1 is routinely changing patterns. Understanding the risk that major
329 evolutionary changes in H5N1 in the Americas presents to the rest of the world is a central
330 question going forward.
331
332
333 **Methods**
334
335 *Dataset.* An H5Nx clade 2.3.4.4b dataset for North America, South America and
336 Antarctica was compiled by querying the Global Initiative on Sharing All Influenza Data
337 (GISAID; https://gisaid.org/)[31–33] and NCBI Virus[34] databases for all H5Nx clade 2.3.4.4b
338 samples collected between November 1, 2021 – June 13, 2025 with available genomic
339 sequences for all eight gene segments (accessed on June 13, 2025). Additionally,
340 consensus genomes from raw reads deposited to NCBI's Sequence Read Archive and

associated with BioProjects PRJNA1102327, PRJNA1122849, PRJNA1134696, PRJNA1219588, PRJNA1207547, PRJNA980729 were downloaded from the Andersen Lab's avian-influenza github repository (https://github.com/andersen-lab/avian-influenza). The samples in each dataset were assigned a genotype using the GenoFLU genotyping tool[8]. GenoFLU categorizes reassortant H5N1 genotypes as "major" (*n* = 25, as of August 27, 2025) if they meet the criteria of at least 20 wild bird detections, infection of two or more poultry premises, and/or detection in atypical mammalian host species. The vast majority of genotypes (88/113) are designated by GenoFlu as 'minor'. Samples with unassigned genotypes were removed from the dataset and duplicate samples across databases were filtered by isolate id. Of the 121 genotypes listed on GenoFLU's website, genetic sequence data was publicly available in SRA or GenBank for 83 of them (<70%). The final American clade 2.3.4.4b dataset consisted of 17,516 genome sequences. To place the clade 2.3.4.4b genotypes in phylogenetic context and identify reassortment involving American LPAI viruses, a background dataset was compiled by querying the GISAID database for LPAI samples with complete genomes collected in the Americas and Antarctica between January 1, 2015 – June 13, 2025 (accessed on June 13, 2025). Select samples belonging to distinct lineages were removed from the analysis which included H1N1 samples from humans, swine, mink and turkey, H3N2 samples from canines, swine, mink, cat and turkeys, and H1N2 swine samples. The search resulted in an LPAI dataset of 3,121 samples. All sequences from the American clade 2.3.4.4b and background LPAI were combined and grouped by segment resulting in eight segment datasets – PB2, PB1, PA, H5, NP, N1, MP and NS.

An additional background dataset was compiled to evaluate the number of North American clade 2.3.4.4b introductions of fully Eurasian genotypes A1-A6 by downloading all European and Asian H5Nx clade 2.3.4.4b and LPAI samples with complete genomes collected between January 1, 2015 and June 13, 2025 (accessed on June 13, 2025). To reduce the size of the datasets while keeping a representative Eurasian background sample set, the sequences were then grouped by segment and used to create local Basic Local Alignment Search Tool (BLAST) databases using BLAST v2.15.0[35]. The resulting Eurasian segment databases were BLASTed against using the American A1-A6 clade

372 2.3.4.4b datasets as queries, and the top 200 Eurasian BLAST hits with >85% sequence
373 identity across all American clade 2.3.4.4b sequences were included for subsequent
374 phylogenetic analysis. The final American A1-A6 + Eurasian sequence datasets
375 consisted of: H5 (*n* = *4,923*), MP (*n* = 5,035), N1 (*n* =3,935), NP (*n* = 5,035), NS (*n* =
376 5,035), PA (*n* = 5,035), PB1 (*n* = 5,035), PB2 (*n* = 5,035).
377
378 Each gene segment dataset was aligned using MAFFT[36,37] and trimmed to the coding
379 sequence region. Highly diverged sequences that were poorly aligned were removed from
380 each segment dataset. Maximum-likelihood (ML) phylogenies were inferred for each
381 dataset using IQ-TREE v2.47[38] with a general-time reversible (GTR) model of nucleotide
382 substitution with among-site rate heterogeneity modelled through a discretized gamma
383 distribution (+G) and 1,000 ultra-fast bootstrap (UFBOOT) replicates, using the high-
384 performance computational capabilities of the Biowulf Linux cluster at the National
385 Institutes of Health (http://biowulf.nih.gov).
386
387 To increase the readability of the American clade 2.3.4.4b and background LPAI
388 phylogenies consisting of >16,000 sequences, ML phylogenetic analyses were also
389 performed on representative, subsampled datasets. Each segment dataset was
390 subsampled using SAMPI (https://github.com/jlcherry/SAMPI)[39]. Subsampling was
391 stratified by collection date, location, host category (avian, cattle, environmental, feline,
392 human, and other mammals), and assigned genotype to maintain representative diversity
393 across these variables. The resulting subsampled datasets consisted of H5 (*n* = 1,160),
394 MP (*n* = 1,340), N1 (*n* =1,120), NP (*n* = 1,354), NS (*n* = 1,346), PA (*n* = 1,358), PB1 (*n* =
395 1,353), PB2 (*n* = 1,372). Phylogenies were visualized using the *ggtree* R package
396 v3.13.1[40].
397
398 *Bayesian phylodynamic analyses*. Bayesian phylodynamic reconstructions were
399 performed for temporally dominant genotypes B3.2, B3.6, C2.1 and D1.1 using
400 concatenated genome datasets (all eight segments were concatenated using
401 ConcatenatorGUI[41]). To assess the timing of North American A3 introductions and the
402 emergence of the D1 genotypes, an additional analysis was performed on a concatenated

403 shared segment dataset (i.e., PB1, HA, MP and NS) of all A3 sequences and the D1
404 genotypes. For computational efficiency, subsampled sequence datasets were first
405 obtained using SAMPI. ML phylogenies were inferred for each dataset and TempEst[42]
406 was used to exclude outlier sequences whose genetic divergence and sampling date
407 were incongruent. This resulted in the following final datasets – B3.2 (*n* = 535), B3.6 (*n* =
408 543), C2.1 (*n* = 211), D1.1 (*n* = 358), and A3-plus-D1 (*n* = 725). Bayesian phylogenetic
409 and phylogeographic analyses were performed using the software package BEAST X
410 v1.10.5[43], using the BEAGLE high-performance library for computational efficiency[44]. We
411 employed a GTR+Γ nucleotide substitution model, an uncorrelated relaxed molecular
412 clock with an underlying lognormal distribution to account for rate variation across
413 lineages[45], and a non-parametric coalescent tree prior for which we employ Hamiltonian
414 Monte Carlo sampling[46,47]. Empirical tree distributions were collected and subsequently
415 used in the phylogeographic reconstructions[48]. To reconstruct the finer-scale geographic
416 spread of the more recently emerged genotypes B3.2, B3.6, C2.1, and D1.1, we
417 conducted continuous phylogeographic analyses with the software package BEAST X[43].
418 This analysis employed a relaxed random walk (RRW) diffusion model, with a Cauchy
419 distribution to model dispersal velocity heterogeneity[49–51] which allows reconstructing the
420 dispersal history of viral lineages in two-dimensional space (latitude and longitude). For
421 the genomic samples that did not have associated precise sampling coordinates, we used
422 the polygon(s) of their administrative area of origin to define an homogeneous prior range
423 of sampling coordinates and estimate those sampling locations through Bayesian
424 phylogeographic inference[52,53]. We employed default prior settings in BEAUti X[43] for all
425 analyses. Bayesian inference using Markov chain Monte Carlo were run for at least 400
426 million iterations, sampled every 40,000 iterations. Parameter convergence and effective
427 sample size (ESS; required to be >200) were evaluated in Tracer v.1.7.3[54]. Using
428 LogCombiner X, at least 10% burn-in was removed from each run and independent runs
429 were combined to establish the maximum clade credibility (MCC) tree using
430 TreeAnnotator X[55]. We used the R package "seraphim"[56] to generate a visualization for
431 the different phylogeographic reconstructions and for estimating dispersal statistics based
432 on these reconstructions.
433

434 *Identification of genotype-specific amino acid residues and D1.1 phenotypic markers.*
435 American H5Nx clade 2.3.4.4b segment datasets were translated into protein alignments
436 with Nextclade CLI[57], using the following built-in reference datasets: H5N8
437 'A/Astrakhan/3212/2020(H5N8)' for HA, H1N1 'A/Wisconsin/588/2019' for NA, and H1N1
438 'A/California/07/2009' for all other segments. Sequences that failed to be aligned to the
439 reference were not considered for this analysis. A custom R script was used to identify
440 genotype-specific amino acid residues found among dominant genotypes B3.2, B3.6,
441 B3.13, C2.1 and D1.1. These genotype-specific residues are defined as residues found
442 in >90% of all sequences within the genotype but <20% of other dominant genotypes and
443 <10% of all other clade 2.3.4.4b sequences. Characterized phenotypic markers of
444 mammalian adaptation, pathogenicity, and drug resistance were identified within the
445 complete D1.1 (n = 2,734) and B3.13 (n = 4,830) nucleotide segment alignments using
446 FluMutGUI v3.2.0[58].
447
448 **Data availability**
449 All data (including supplemental files and GISAID acknowledgement table) and BEAST
450 X XML files with model parametrizations, phylogenetic trees and analysis output files
451 are available at https://github.com/nidiatrovao/H5N1Americas.
452
453 **Acknowledgements**
454 This research was supported in part by the Intramural Research Program of the National
455 Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of
456 the United States Government. The findings and conclusions presented in this paper are
457 those of the author(s) and do not necessarily reflect the views of the NIH or the U.S.
458 Department of Health and Human Services, or the United States government. We
459 gratefully acknowledge all data contributors, i.e., the Authors and their Originating
460 laboratories responsible for obtaining the specimens, and their Submitting laboratories
461 for generating the genetic sequence and metadata and sharing via the GISAID Initiative,
462 on which this research is based. GB and SD acknowledge support from the Research
463 Foundation — Flanders (*Fonds voor Wetenschappelijk Onderzoek — Vlaanderen*, FWO,
464 Belgium; grant n°G098321N), and from the European Union Horizon 2023 RIA project

LEAPS (grant agreement no. 101094685). GB also acknowledges support from the Research Foundation — Flanders (*Fonds voor Wetenschappelijk Onderzoek — Vlaanderen*, FWO, Belgium; grant n° G0E1420N), and from the DURABLE EU4Health project 02/2023-01/2027, which is co-funded by the European Union (call EU4H-2021-PJ4) under Grant Agreement No. 101102733. SD also acknowledges support from the *Fonds National de la Recherche Scientifique* (F.R.S.-FNRS, Belgium; including grant n°F.4515.22), and from the European Union Horizon 2020 projects MOOD (grant agreement n°874850).

References

1. Xie, R., Edwards, K.M., Wille, M., Wei, X., Wong, S.-S., Zanin, M., El-Shesheny, R., Ducatez, M., Poon, L.L.M., Kayali, G., et al. (2023). The episodic resurgence of highly pathogenic avian influenza H5 virus. Nature *622*, 810–817. https://doi.org/10.1038/s41586-023-06631-2.

2. Rijks, J.M., Leopold, M.F., Kühn, S., In ’t Veld, R., Schenk, F., Brenninkmeijer, A., Lilipaly, S.J., Ballmann, M.Z., Kelder, L., de Jong, J.W., et al. (2022). Mass Mortality Caused by Highly Pathogenic Influenza A(H5N1) Virus in Sandwich Terns, the Netherlands, 2022. Emerg. Infect. Dis. *28*, 2538–2542. https://doi.org/10.3201/eid2812.221292.

3. Fusaro, A., Zecchin, B., Giussani, E., Palumbo, E., Agüero-García, M., Bachofen, C., Bálint, Á., Banihashem, F., Banyard, A.C., Beerens, N., et al. (2024). High pathogenic avian influenza A(H5) viruses of clade 2.3.4.4b in Europe—Why trends of virus evolution are more difficult to predict. Virus Evol. *10*, veae027. https://doi.org/10.1093/ve/veae027.

4. European Food Safety Authority, European Centre for Disease Prevention and Control, European Union Reference Laboratory for Avian Influenza, Alexakis, L., Buczkowski, H., Ducatez, M., Fusaro, A., Gonzales, J.L., Kuiken, T., Ståhl, K., et al. (2025). Avian influenza overview December 2024-March 2025. EFSA J. Eur. Food Saf. Auth. *23*, e9352. https://doi.org/10.2903/j.efsa.2025.9352.

5. Agüero, M., Monne, I., Sánchez, A., Zecchin, B., Fusaro, A., Ruano, M.J., Del Valle Arrojo, M., Fernández-Antonio, R., Souto, A.M., Tordable, P., et al. (2023). Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. *28*, 2300001. https://doi.org/10.2807/1560-7917.ES.2023.28.3.2300001.

6. Kareinen, L., Tammiranta, N., Kauppinen, A., Zecchin, B., Pastori, A., Monne, I., Terregino, C., Giussani, E., Kaarto, R., Karkamo, V., et al. (2024). Highly pathogenic

avian influenza A(H5N1) virus infections on fur farms connected to mass mortalities of black-headed gulls, Finland, July to October 2023. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. *29*, 2400063. https://doi.org/10.2807/1560-7917.ES.2024.29.25.2400063.

7. Giacinti, J.A., Signore, A.V., Jones, M.E.B., Bourque, L., Lair, S., Jardine, C., Stevens, B., Bollinger, T., Goldsmith, D., British Columbia Wildlife AIV Surveillance Program (BC WASP), et al. (2024). Avian influenza viruses in wild birds in Canada following incursions of highly pathogenic H5N1 virus from Eurasia in 2021–2022. mBio *15*, e03203-23. https://doi.org/10.1128/mbio.03203-23.

8. Youk, S., Torchetti, M.K., Lantz, K., Lenoch, J.B., Killian, M.L., Leyson, C., Bevins, S.N., Dilione, K., Ip, H.S., Stallknecht, D.E., et al. (2023). H5N1 highly pathogenic avian influenza clade 2.3.4.4b in wild and domestic birds: Introductions into the United States and reassortments, December 2021-April 2022. Virology *587*, 109860. https://doi.org/10.1016/j.virol.2023.109860.

9. Leguia, M., Garcia-Glaessner, A., Muñoz-Saavedra, B., Juarez, D., Barrera, P., Calvo-Mac, C., Jara, J., Silva, W., Ploog, K., Amaro, Lady, et al. (2023). Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. Nat. Commun. *14*, 5489. https://doi.org/10.1038/s41467-023-41182-0.

10. Uhart, M.M., Vanstreels, R.E.T., Nelson, M.I., Olivera, V., Campagna, J., Zavattieri, V., Lemey, P., Campagna, C., Falabella, V., and Rimondi, A. (2024). Epidemiological data of an influenza A/H5N1 outbreak in elephant seals in Argentina indicates mammal-to-mammal transmission. Nat. Commun. *15*, 9516. https://doi.org/10.1038/s41467-024-53766-5.

11. Tomás, G., Marandino, A., Panzera, Y., Rodríguez, S., Wallau, G.L., Dezordi, F.Z., Pérez, R., Bassetti, L., Negro, R., Williman, J., et al. (2024). Highly pathogenic avian influenza H5N1 virus infections in pinnipeds and seabirds in Uruguay: Implications for bird-mammal transmission in South America. Virus Evol. *10*, veae031. https://doi.org/10.1093/ve/veae031.

12. Pardo-Roa, C., Nelson, M.I., Ariyama, N., Aguayo, C., Almonacid, L.I., Gonzalez-Reiche, A.S., Muñoz, G., Ulloa, M., Ávila, C., Navarro, C., et al. (2025). Cross-species and mammal-to-mammal transmission of clade 2.3.4.4b highly pathogenic avian influenza A/H5N1 with PB2 adaptations. Nat. Commun. *16*, 2232. https://doi.org/10.1038/s41467-025-57338-z.

13. Detwiler, L. H5N1 Highly Pathogenic Avian Influenza (HPAI) in Livestock - Information for small ruminant (sheep and goat) and camelid stakeholders (Animal and Plant Health Inspection Service).

14. Worobey, M., Gangavarapu, K., Pekar, J.E., Joy, J.B., Moncla, L., Kraemer, M.U.G., Dudas, G., Goldhill, D., Ruis, C., Serrano, L.M., et al. (2024). Preliminary report on genomic epidemiology of the 2024 H5N1 influenza A virus outbreak in U.S. cattle (Part 1 of 2) - Influenza virus H5N1-global. Preprint at Virological.

15. Burrough, E.R., Magstadt, D.R., Petersen, B., Timmermans, S.J., Gauger, P.C., Zhang, J., Siepker, C., Mainenti, M., Li, G., Thompson, A.C., et al. (2024). Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4b Virus Infection in Domestic Dairy Cattle and Cats, United States, 2024. Emerg. Infect. Dis. *30*, 1335–1343. https://doi.org/10.3201/eid3007.240508.

16. Nguyen, T.-Q., Hutter, C., Markin, A., Thomas, M., Lantz, K., Killian, M.L., Janzen, G.M., Vijendran, S., Wagle, S., Inderski, B., et al. (2024). Emergence and interstate spread of highly pathogenic avian influenza A(H5N1) in dairy cattle. Preprint at Microbiology, https://doi.org/10.1101/2024.05.01.591751 https://doi.org/10.1101/2024.05.01.591751.

17. Jassem, A.N., Roberts, A., Tyson, J., Zlosnik, J.E.A., Russell, S.L., Caleta, J.M., Eckbo, E.J., Gao, R., Chestley, T., Grant, J., et al. (2025). Critical Illness in an Adolescent with Influenza A(H5N1) Virus Infection. N. Engl. J. Med. *392*, 927–929. https://doi.org/10.1056/NEJMc2415890.

18. CDC (2025). Genetic Sequences of Highly Pathogenic Avian Influenza A(H5N1) Viruses Identified in a Person in Louisiana. Avian Influenza Bird Flu. https://www.cdc.gov/bird-flu/spotlights/h5n1-response-12232024.html.

19. Pekar, J.E., Peacock, T., Moncla, L., Gangavarapu, P., Venkatesh, D., Goldhill, D.H., Gangavarapu, K., Kraemer, M.U.G., Dudas, G., Joy, J.B.J., et al. (2025). Timing and molecular characterisation of the transmission to cattle of H5N1 influenza A virus genotype D1.1, clade 2.3.4.4b - Influenza virus. Virological. https://virological.org/t/timing-and-molecular-characterisation-of-the-transmission-to-cattle-of-h5n1-influenza-a-virus-genotype-d1-1-clade-2-3-4-4b/991.

20. US Department of Agriculture, A. and P.H.I.S. (2025). The Occurrence of Another Highly Pathogenic Avian Influenza (HPAI) Spillover from Wild Birds into Dairy Cattle.

21. Central Nevada Health District (2025). The Central Nevada Health District is Actively Monitoring for Spread of H5N1 in Northern Nevada. Cent. Nev. Health Dist. https://www.centralnevadahd.org/press-release/.

22. World Health Organization (2025). Avian Influenza A(H5N1) - Mexico. https://www.who.int/emergencies/disease-outbreak-news/item/2025-DON564.

571 23. Damodaran, L., Jaeger, A.S., and Moncla, L.H. (2025). Ecology and spread of the
572 North American H5N1 epizootic. Nature, 1–10. https://doi.org/10.1038/s41586-025-
573 09737-x.

574 24. Nataraj, R., Chandra, A., and Kesavardhana, S. (2024). Avian influenza virus
575 neuraminidase stalk length and haemagglutinin glycosylation patterns reveal
576 molecularly directed reassortment promoting the emergence of highly pathogenic
577 clade 2.3.4.4b A (H5N1) viruses. Preprint at bioRxiv,
578 https://doi.org/10.1101/2024.05.22.595329
579 https://doi.org/10.1101/2024.05.22.595329.

580 25. Hermann, E., and Krammer, F. (2025). Clade 2.3.4.4b H5N1 neuraminidase has a
581 long stalk, which is in contrast to most highly pathogenic H5N1 viruses circulating
582 between 2002 and 2020. mBio *16*, e03989-24. https://doi.org/10.1128/mbio.03989-24.

583 26. Hansen, L., McMahon, M., Turner, H.L., Zhu, X., Turner, J.S., Ozorowski, G.,
584 Stadlbauer, D., Vahokoski, J., Schmitz, A.J., Rizk, A.A., et al. (2023). Human anti-N1
585 monoclonal antibodies elicited by pandemic H1N1 virus infection broadly inhibit HxN1
586 viruses in vitro and in vivo. Immunity *56*, 1927-1938.e8.
587 https://doi.org/10.1016/j.immuni.2023.07.004.

588 27. Rahman, I., Erdelyan, C.N.G., Signore, A.V., Ayilara, I., Wight, J., Jones, M.E.B.,
589 Sullivan, D.S., Lung, O., Hisanaga, T., Wilhelm, S.I., et al. (2024). Recurring incursions
590 and dissemination of novel Eurasian-origin H5Nx avian influenza viruses in Atlantic
591 Canada. Virus Evol. *10*, veae111. https://doi.org/10.1093/ve/veae111.

592 28. Ramey, A.M., Scott, L.C., Ahlstrom, C.A., Buck, E.J., Williams, A.R., Kim Torchetti,
593 M., Stallknecht, D.E., and Poulson, R.L. (2024). Molecular detection and
594 characterization of highly pathogenic H5N1 clade 2.3.4.4b avian influenza viruses
595 among hunter-harvested wild birds provides evidence for three independent
596 introductions into Alaska. Virology *589*, 109938.
597 https://doi.org/10.1016/j.virol.2023.109938.

598 29. Signore, A.V., Giacinti, J., Jones, M.E.B., Erdelyan, C.N.G., McLaughlin, A., Alkie,
599 T.N., Cox, S., Lair, S., Jardine, C.M., Stevens, B., et al. (2025). Spatiotemporal
600 reconstruction of the North American A(H5N1) outbreak reveals successive lineage
601 replacements by descendant reassortants. Sci. Adv. *11*, eadu4909.
602 https://doi.org/10.1126/sciadv.adu4909.

603 30. Lowen, A.C., Baker, A.L., Bowman, A.S., García-Sastre, A., Hensley, S.E.,
604 Lakdawala, S.S., Moncla, L.H., Nelson, M.I., Pekosz, A., Poulson, R.L., et al. (2025).
605 Pandemic risk stemming from the bovine H5N1 outbreak: an account of the knowns
606 and unknowns. J. Virol. *99*, e0005225. https://doi.org/10.1128/jvi.00052-25.

31. Elbe, S., and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Chall. *1*, 33–46. https://doi.org/10.1002/gch2.1018.

32. Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance *22*, 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

33. Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., et al. (2021). GISAID's Role in Pandemic Response. China CDC Wkly. *3*, 1049–1051. https://doi.org/10.46234/ccdcw2021.255.

34. Hatcher, E.L., Zhdanov, S.A., Bao, Y., Blinkova, O., Nawrocki, E.P., Ostapchuck, Y., Schäffer, A.A., and Brister, J.R. (2017). Virus Variation Resource - improved response to emergent viral outbreaks. Nucleic Acids Res. *45*, D482–D490. https://doi.org/10.1093/nar/gkw1065.

35. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

36. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform *9*, 286–298. https://doi.org/10.1093/bib/bbn013.

37. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

38. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

39. Li, X., Trovão, N.S., Wertheim, J.O., Baele, G., and de Bernardi Schneider, A. (2023). Optimizing ancestral trait reconstruction of large HIV Subtype C datasets through multiple-trait subsampling. Virus Evol. *9*, vead069. https://doi.org/10.1093/ve/vead069.

40. Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T.T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. iMeta *1*, e56. https://doi.org/10.1002/imt2.56.

41. Vences, M., Patmanidis, S., Kharchev, V., and Renner, S.S. (2022). Concatenator, a user-friendly program to concatenate DNA sequences, implementing graphical user

642 interfaces for MAFFT and FastTree. Bioinforma. Adv. *2*, vbac050.
643 https://doi.org/10.1093/bioadv/vbac050.

644 42. Rambaut, A., Lam, T.T., Max Carvalho, L., and Pybus, O.G. (2016). Exploring the
645 temporal structure of heterochronous sequences using TempEst (formerly Path-O-
646 Gen). Virus Evol. *2*, vew007. https://doi.org/10.1093/ve/vew007.

647 43. Baele, G., Ji, X., Hassler, G.W., McCrone, J.T., Shao, Y., Zhang, Z., Holbrook, A.J.,
648 Lemey, P., Drummond, A.J., Rambaut, A., et al. (2025). BEAST X for Bayesian
649 phylogenetic, phylogeographic and phylodynamic inference. Nat. Methods *22*, 1653–
650 1656. https://doi.org/10.1038/s41592-025-02751-x.

651 44. Ayres, D.L., Cummings, M.P., Baele, G., Darling, A.E., Lewis, P.O., Swofford, D.L.,
652 Huelsenbeck, J.P., Lemey, P., Rambaut, A., and Suchard, M.A. (2019). BEAGLE 3:
653 Improved Performance, Scaling, and Usability for a High-Performance Computing
654 Library for Statistical Phylogenetics. Syst. Biol. *68*, 1052–1061.
655 https://doi.org/10.1093/sysbio/syz020.

656 45. Drummond, A.J., W, H.S.Y., J, P.M., and Andrew, R. (2006). Relaxed phylogenetics
657 and dating with confidence. PLoS Biol *4*, e88.
658 https://doi.org/10.1371/journal.pbio.0040088.

659 46. Baele, G., Gill, M.S., Lemey, P., and Suchard, M.A. (2020). Hamiltonian Monte Carlo
660 sampling to estimate past population dynamics using the skygrid coalescent model in
661 a Bayesian phylogenetics framework. Wellcome Open Res. *5*, 53.
662 https://doi.org/10.12688/wellcomeopenres.15770.1.

663 47. Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., and Suchard, M.A.
664 (2013). Improving Bayesian population dynamics inference: a coalescent-based model
665 for multiple loci. Mol. Biol. Evol. *30*, 713–724. https://doi.org/10.1093/molbev/mss265.

666 48. Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral
667 character states on phylogenies. Syst Biol *53*, 673–684.
668 https://doi.org/10.1080/10635150490522232.

669 49. Lemey, P., Andrew, R., J, W.J., and A, S.M. (2010). Phylogeography takes a relaxed
670 random walk in continuous space and time. Mol Biol Evol *27*, 1877–1885.
671 https://doi.org/10.1093/molbev/msq067.

672 50. Pybus, O.G., A, S.M., Philippe, L., J, B.F., Andrew, R., W, C.F., R, G.R., Nimalan, A., L,
673 S.S., P, B.M., et al. (2012). Unifying the spatial epidemiology and molecular evolution of
674 emerging epidemics. Proc Natl Acad Sci U A *109*, 15066–15071.
675 https://doi.org/10.1073/pnas.1206598109.

676 51. Dellicour, S., Durkin, K., Hong, S.L., Vanmechelen, B., Martí-Carreras, J., Gill, M.S.,
677 Meex, C., Bontems, S., André, E., Gilbert, M., et al. (2021). A Phylodynamic Workflow to

Rapidly Gain Insights into the Dispersal History and Dynamics of SARS-CoV-2 Lineages. Mol. Biol. Evol. *38*, 1608–1613. https://doi.org/10.1093/molbev/msaa284.

52. Dellicour, S., Lemey, P., Artois, J., Lam, T.T., Fusaro, A., Monne, I., Cattoli, G., Kuznetsov, D., Xenarios, I., Dauphin, G., et al. (2020). Incorporating heterogeneous sampling probabilities in continuous phylogeographic inference — Application to H5N1 spread in the Mekong region. Bioinformatics *36*, 2098–2104. https://doi.org/10.1093/bioinformatics/btz882.

53. Dellicour, S., Lemey, P., Suchard, M.A., Gilbert, M., and Baele, G. (2022). Accommodating sampling location uncertainty in continuous phylogeography. Virus Evol. *8*, veac041. https://doi.org/10.1093/ve/veac041.

54. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.

55. Baele, G., Carvalho, L.M., Brusselmans, M., Dudas, G., Ji, X., McCrone, J.T., Lemey, P., Suchard, M.A., and Rambaut, A. (2025). HIPSTR: highest independent posterior subtree reconstruction in TreeAnnotator X. Bioinformatics, btaf488. https://doi.org/10.1093/bioinformatics/btaf488.

56. Dellicour, S., Rose, R., Faria, N.R., Lemey, P., and Pybus, O.G. (2016). SERAPHIM: studying environmental rasters and phylogenetically informed movements. Bioinforma. Oxf. Engl. *32*, 3204–3206. https://doi.org/10.1093/bioinformatics/btw384.

57. Aksamentov, I., Roemer, C., Hodcroft, E., and Neher, R. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Softw. *6*, 3773. https://doi.org/10.21105/joss.03773.

58. Giussani, E., Sartori, A., Salomoni, A., Cavicchio, L., de Battisti, C., Pastori, A., Varotto, M., Zecchin, B., Hughes, J., Monne, I., et al. (2025). FluMut: a tool for mutation surveillance in highly pathogenic H5N1 genomes. Virus Evol. *11*, veaf011. https://doi.org/10.1093/ve/veaf011.

SUPPLEMENTARY MATERIAL

TABLES

Table S1: Timing of introductions of A3 introductions from Europe and Asia into North America. The 95% highest posterior density intervals (95 % HPDIs) are inferred from the time-scaled Bayesian phylogenetic analysis discussed in Figure 3B.

Points of entry	Introduction	Time to the Most Recent Common Ancestor 95% HPDI
Alaska	intro1	2021.85 - 2021.98
Pacific flyway	Intro2	2022.27 - 2022.50
Pacific flyway	Intro3	2022.48 - 2022.68
British Columbia and Oregon	Intro4	2023.85 - 2023.94
Central, Mississippi, and Atlantic flyways	Intro5	2024.51 - 2024.70
Hawaii	Intro6	2024.66 - 2024.81

742 **Table S2. Comparison of dispersal metrics estimated for recent H5N1 genotypes having spread in**
743 **North and South America.** We report both weighted diffusion coefficient (WDC) and isolation-by-distance
744 (IBD) signal estimates, the latter metric having been estimated by the Pearson correlation coefficient (r_P)
745 between the patristic and log-transformed great-circle geographic distances computed for each pair of virus
746 samples. For each dataset and metric, we report both the posterior median estimate and the 95% highest
747 posterior density (HPD) interval. For comparison, we also report the same dispersal metrics previously
748 estimated for the West Nile virus spread in North America.
749

Dataset	Weighted diffusion coefficient (km^2/day)	WDC during invasion (km^2/ day)	WDC after invasion (km^2/ day)	IBD signal (r_P)
HPAI H5N1 (B3.2)	3061 [2786, 3489]	3731 [3379, 4318]	1783 [1264, 2489]	0.48 [0.46, 0.50]
HPAI H5N1 (B3.6)	1942 [1761, 2148]	1975 [1417, 2566]	1171 [758, 1649]	0.23 [0.22, 0.25]
HPAI H5N1 (C2.1)	545 [442, 734]	1133 [798, 1717]	124 [69, 217]	0.21 [0.18, 0.24]
HPAI H5N1 (D1.1)	3820 [3165, 4784]	5868 [3930, 8287]	2866 [2014, 4089]	0.22 [0.19, 0.25]
West Nile virus (North America)	161 [151, 170]	312 [280, 373]	119 [107, 135]	-0.02 [-0.04, 0.00]

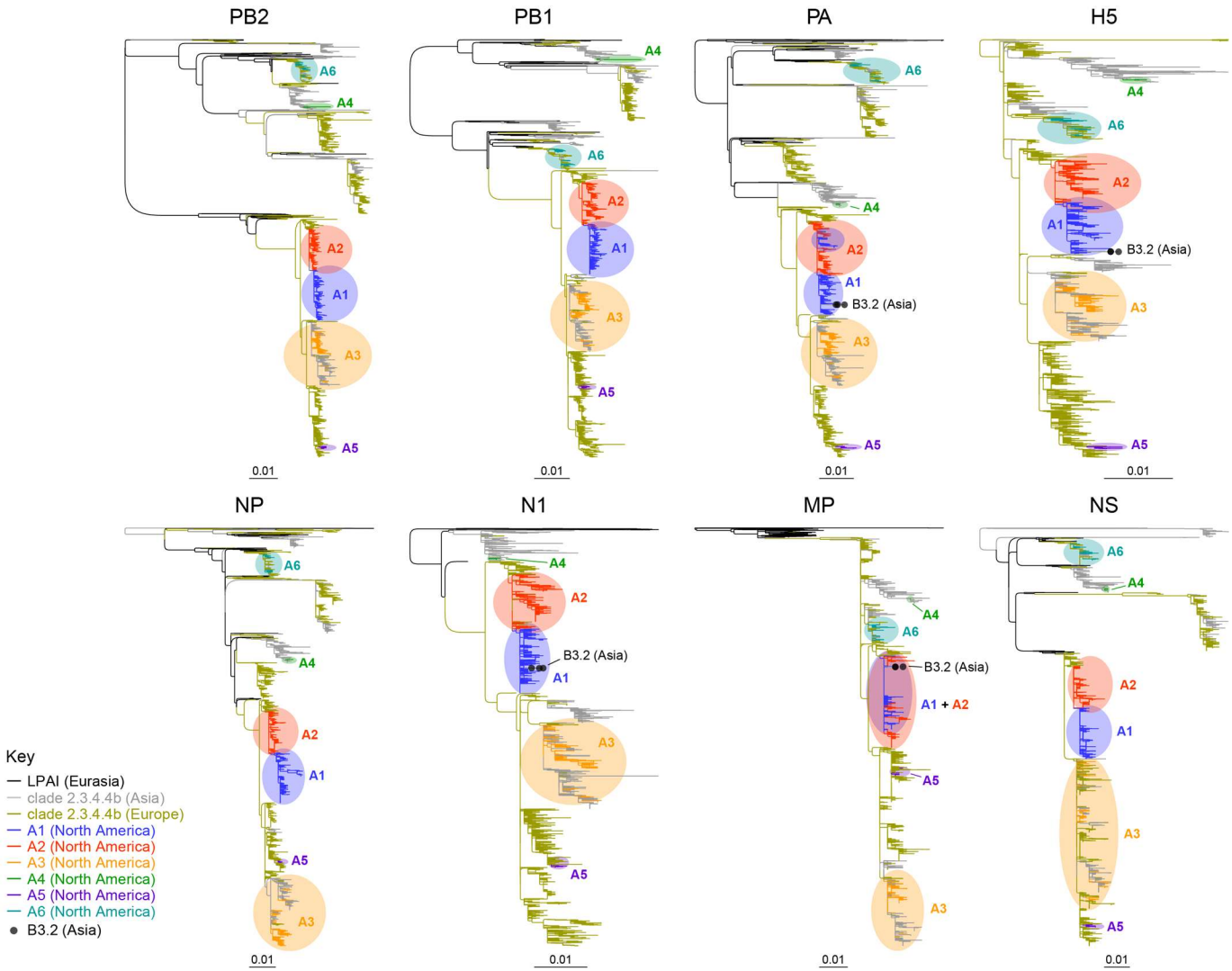
750

751 **Table S3. D1.1-specific NA segment residues by protein domain.**

752

NA domain (residue position range)	Sequential (N1) numbering	N2 numbering
Tail (1-6)	--	--
Transmembrane (7-29)	I8	I8
	I20	I20
	V23	V23
Stalk (30-84)	N44	N44
	H45	H45
	T48	insertion
	V53	V51
	L74	L72
	I75	L73
	D81	D79
	P82	P80
	A84	A82
Head (85+)	S221	S220
	I234	I233
	I241	I240
	R257	R256
	S286	S285
	E287	E286
	V288	V287
	S329	S329
	G336	G339
	V338	V341
	A395	A399

753 **FIGURES**



754

755 **Figure S1. Midpoint-rooted segment phylogenies for EA genotypes (A1-A6) sampled in the Americas, and Eurasian LPAI and clade**
756 **2.3.4.4b sequences (collected between January 1, 2015 and June 13, 2025). Tips for North American reassortant genotype B3.2 samples**
757 **collected in Asia are circled.** Full dataset tree files available at https://github.com/nidiatrovao/H5N1Americas.

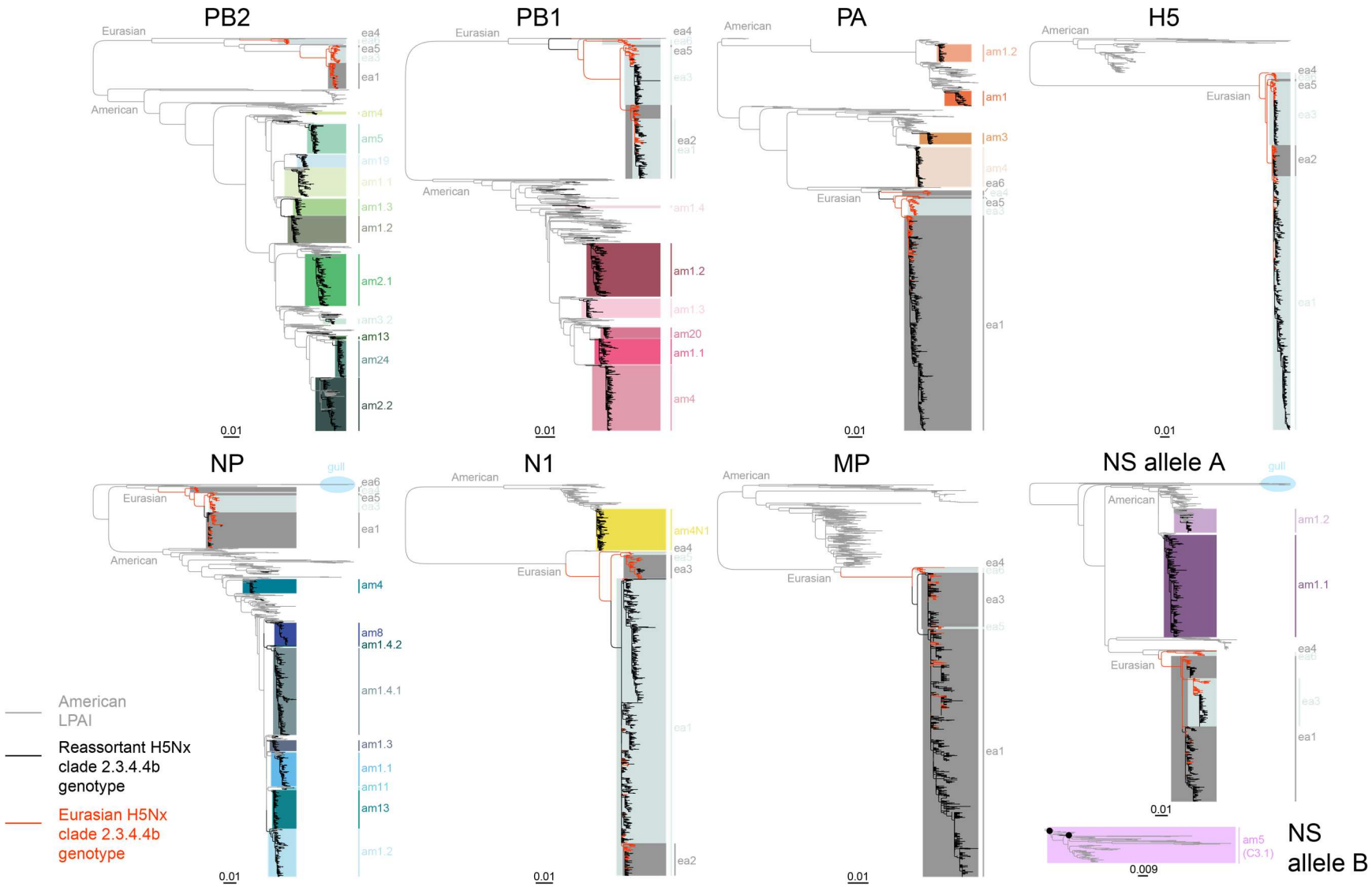


Figure S2. Subsampled, midpoint-rooted segment phylogenies for H5Nx clade 2.3.4.4b and LPAI viruses sampled in the Western hemisphere (collected between January 1, 2015 and June 13, 2025). Full and subsampled dataset tree files are available at.https://github.com/nidiatrovao/H5N1Americas

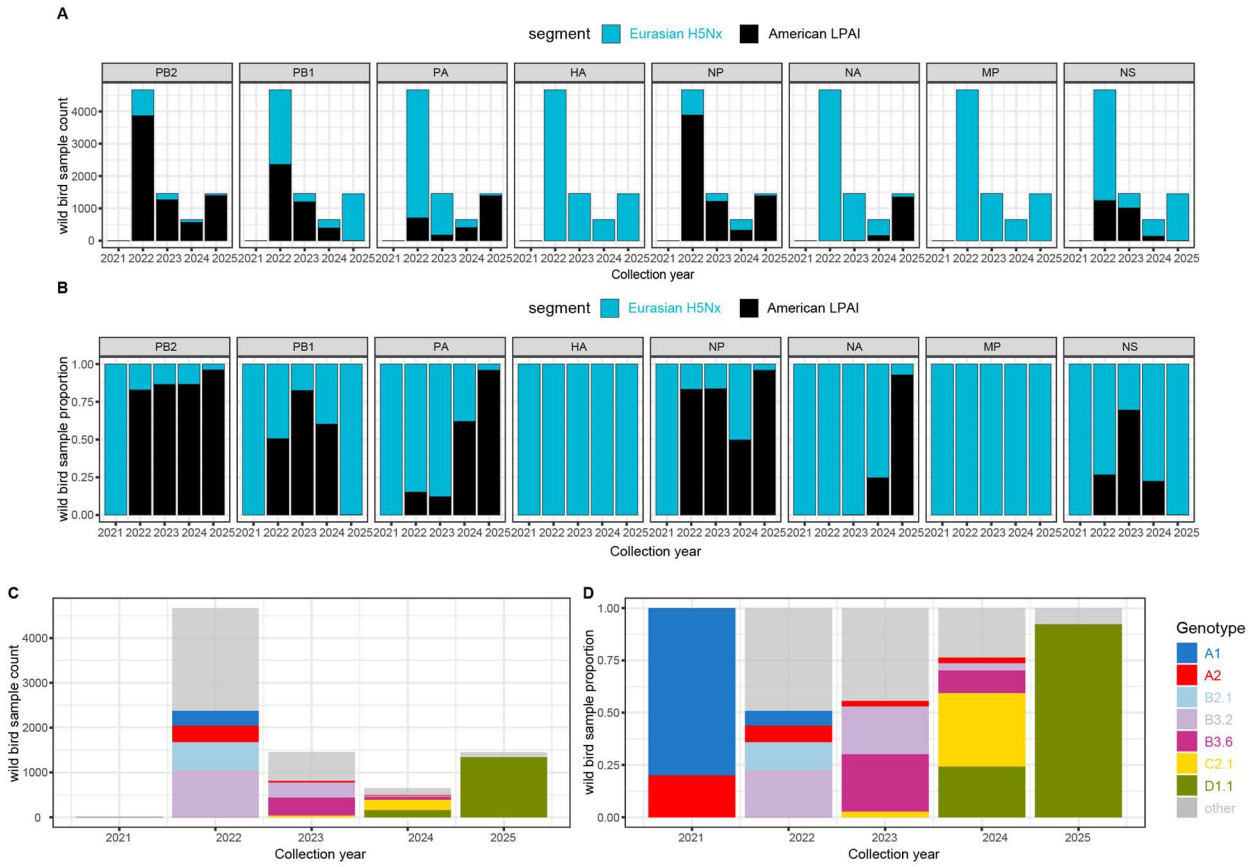


Figure S3. Yearly counts and proportions for H5Nx clade 2.3.4.4b segment sequences during 2021-2025. (A) Yearly counts and (B) proportions for H5Nx clade 2.3.4.4b wild bird segment sequences shaded by Eurasian H5Nx (blue) or American LPAI (black) origin. **(C)** Yearly counts and (D) proportions for H5Nx clade 2.3.4.4b wild bird samples shaded by genotype.

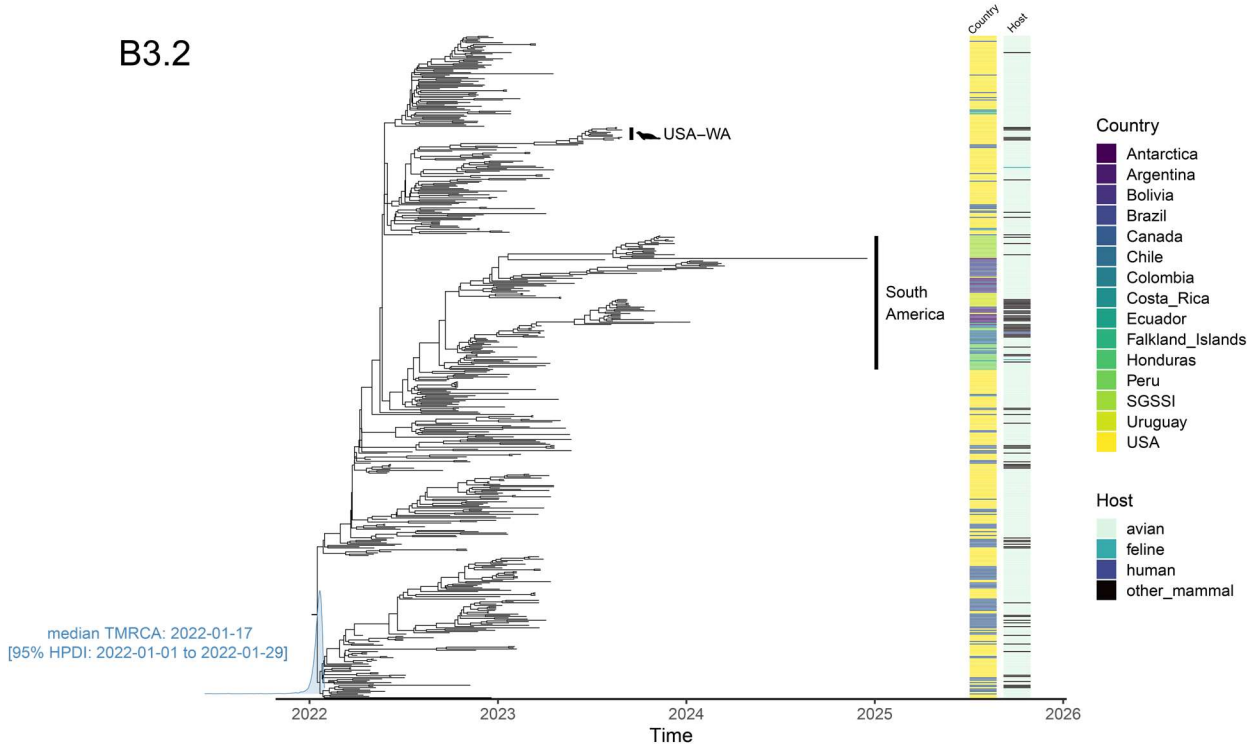


Figure S4. Maximum clade credibility (MCC) tree for H5N1 genotype B3.2. The consensus phylogenetic tree illustrates the evolutionary history of genotype B3.2, with the time-scaled X-axis in years. The blue curve at the base of the tree represents the posterior density distribution for the time of the most recent common ancestor (TMRCA). Annotations for the country and host species are shown in the colored bars next to the tips, with the corresponding keys provided in the legend. A large South American clade is highlighted, as is a sea lion outbreak from Washington, USA (USA-WA).

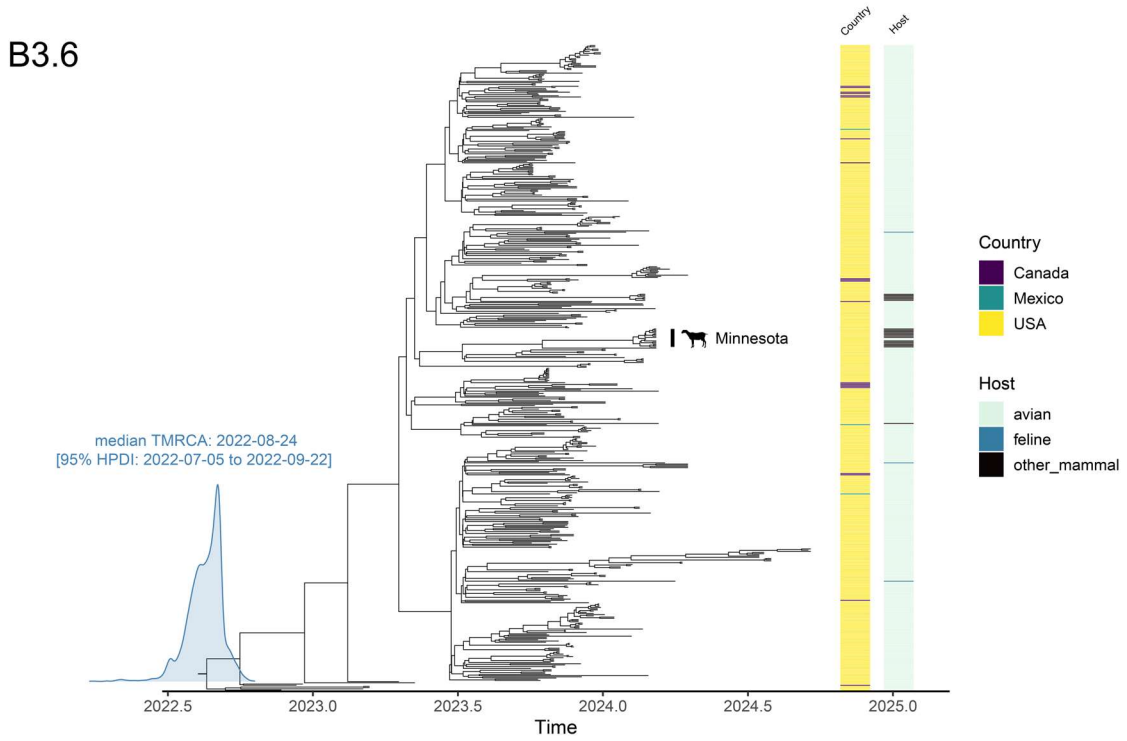


Figure S5. Maximum clade credibility (MCC) tree for H5N1 genotype B3.6. The consensus phylogenetic tree shows the evolutionary relationships of genotype B3.6 viruses, with the X-axis scaled in years. The blue curve at the base of the tree represents the posterior density distribution for the time of the TMRCA. The colored bars to the right of the tips indicate the country of origin and the host species for each sequence. The legend provides a key for these annotations. A clade corresponding to an outbreak in goats in Minnesota, USA is noted on the tree.

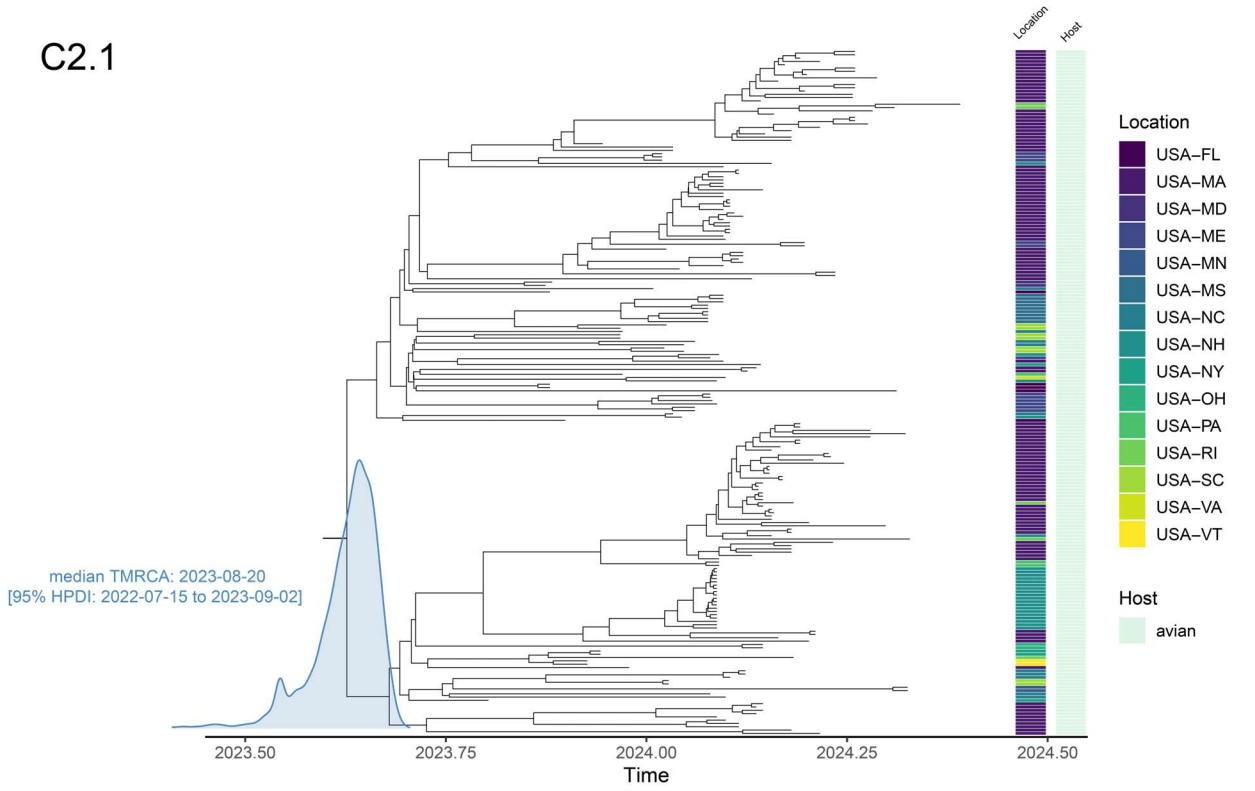


Figure S6. Maximum clade credibility (MCC) tree for H5N1 genotype C2.1. The evolutionary relationships of genotype C2.1 viruses are shown, with the X-axis representing time in years. The TMRCA estimate is represented by the blue posterior density curve at the root of the tree. Tips are annotated with colored bars indicating the U.S. state of collection and the host species, which is exclusively avian for this genotype.

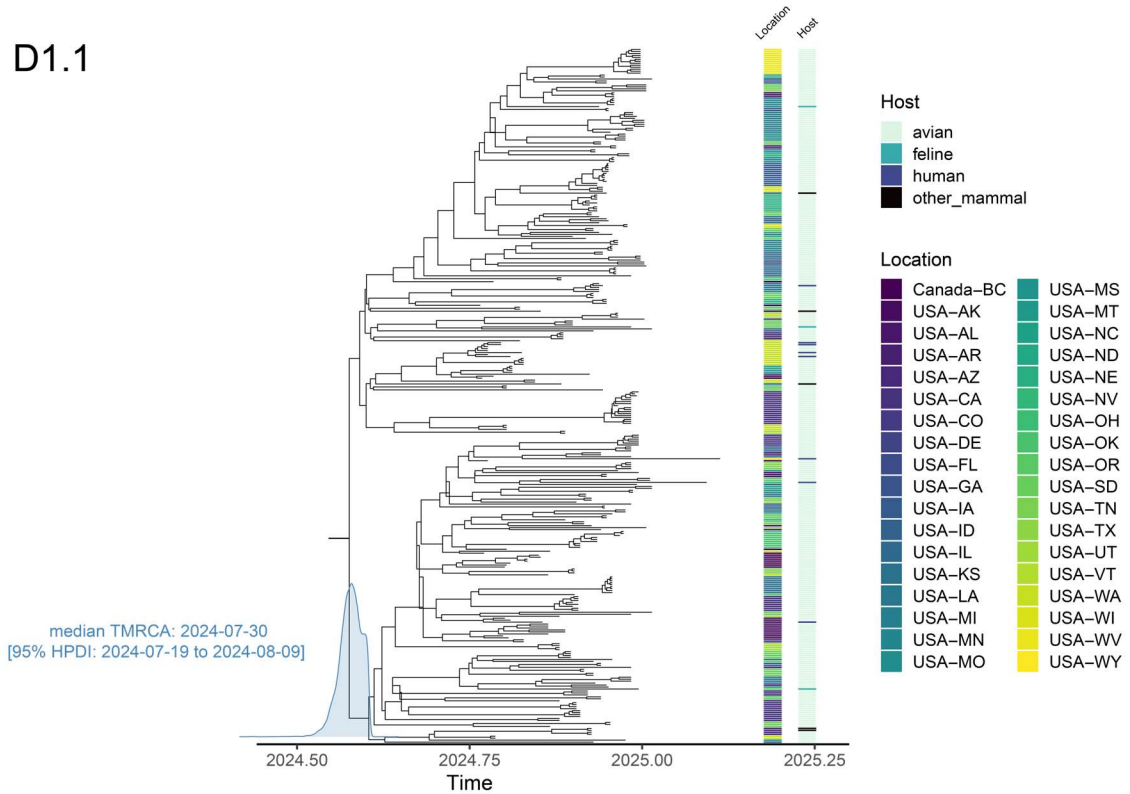


Figure S7. Maximum clade credibility (MCC) tree for H5N1 genotype D1.1. The MCC tree shows the evolutionary history of the D1.1 genotype, scaled by time in years on the X-axis. The blue distribution at the base of the tree shows the posterior probability for the TMRCA. The colored bars to the right of the tips indicate the location (U.S. state or Canadian province) and host species for each sequence, with keys provided in the legend. This genotype shows wide geographic dispersal across North America and spillovers into multiple mammalian hosts, including humans.

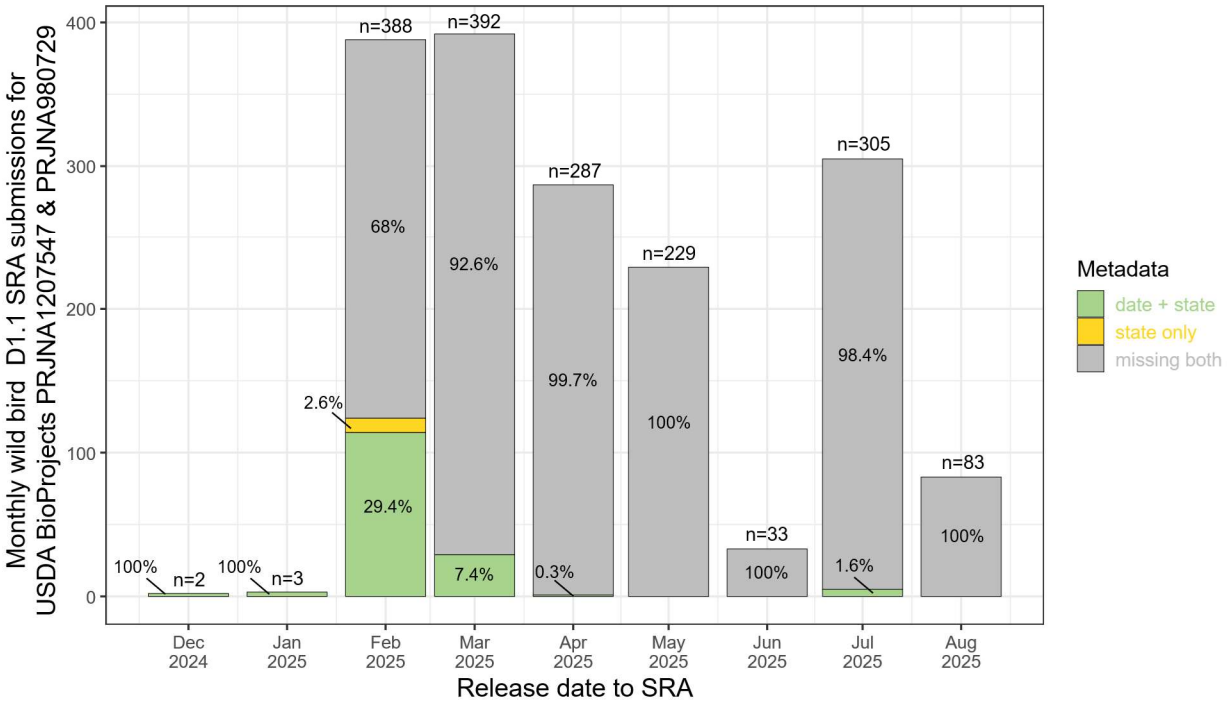


Figure S8. Monthly D1.1 SRA submissions for wild bird samples deposited to BioProjects PRJNA1207547 and PRJNA980729 (as of Sep 10, 2025). Stacked bars are colored according to metadata availability in GenBank, where "date + state" indicates that complete dates (i.e., including day, month and year) and U.S. state where the sample was collected are available, "state only" indicates that the U.S state is known yet the date of collection is missing, and "missing both" indicates that both pieces of metadata are missing.