

Article

Not peer-reviewed version

---

# Genotypic Clustering of H5N1 Avian Influenza Viruses Detected in North America

---

Patil Tawidian , Mia K. Torchetti , [Mary L. Killian](#) , Kristina Lantz , [Krista E. Dillione](#) , [Jourdan M. Ringenberg](#) , Sarah N. Bevins , Julianna B. Lenoach , [Hon S. Ip](#) \*

Posted Date: 31 October 2024

doi: 10.20944/preprints202410.2401.v1

Keywords: Avian influenza; High pathogenic influenza virus (HPAI); H5N1 virus; Virus reassortment; H5N1 genotypes; Virus nomenclature



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Genotypic Clustering of H5N1 Avian Influenza Viruses Detected in North America

Patil Tawidian <sup>1,2</sup>, Mia K. Torchetti <sup>3</sup>, Mary L. Killian <sup>3</sup>, Kristina Lantz <sup>3</sup>, Krista E. Dilione <sup>4</sup>, Jourdan M. Ringenberg <sup>4</sup>, Sarah N. Bevins <sup>4</sup>, Julianna B. Lenocho <sup>4</sup> and Hon S. Ip <sup>1,\*</sup>

<sup>1</sup> U.S. Geological Survey, National Wildlife Health Center Madison, WI 53711, USA

<sup>2</sup> Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>3</sup> National Veterinary Services Laboratories, U.S. Department of Agriculture, Ames, IA 50010, USA

<sup>4</sup> WS National Wildlife Disease Program, U.S. Department of Agriculture, Fort Collins, CO 80521, USA

\* Correspondence: hip@usgs.gov; Tel.: +1-608-270-2464

**Abstract:** The introduction of HPAI H5N1 clade 2.3.4.4b viruses to North America in late 2021 resulted in avian influenza outbreaks in poultry, mortality events in many wild bird species, and spillover into many mammalian species. Reassortment events with North American low pathogenic virus were identified as early as February 2022 and over 100 genotypes have been characterized. Such diversity increases the complexity and time required for monitoring virus evolution. Here, we performed ordination and clustering analyses on sequence data from H5N1 viruses identified in North America between January 2020 to December 2023 to visualize virus genotypic diversity in poultry and wildlife populations. Our results reveal that ordination and cluster-based approaches can complement traditional phylogenetic analyses specifically for the preliminary assignment of H5N1 viruses to genotypic groups or to identify novel genotypes. Our study expands current knowledge on genotype diversity of H5N1 viruses in North America and describes a rapid approach for early virus genotype assignment.

**Keywords:** avian influenza; high pathogenic influenza virus (HPAI); H5N1 virus; virus reassortment; H5N1 genotypes; virus nomenclature

## 1. Introduction

Wild migratory waterfowl are well known reservoir hosts for low pathogenic influenza (LPAI) viruses, either as asymptomatic carriers of the virus or exhibiting mild disease [1–3]. When H5 or H7 subtype LPAI viruses spillover from the reservoir host into non-host species such as gallinaceous birds (e.g., chickens, turkeys, pheasants) the virus has the potential to mutate to a highly pathogenic form (HPAI); however, spillback into the natural host is rare. The A/goose/Guangdong/1/1996 (Gs/GD) HPAI H5N1 lineage that emerged in poultry, spreading to poultry across other countries has generated two clades (most recently the H5 clade 2.3.4.4) that have spilled back into wild migratory birds resulting in intercontinental transmission across five continents and causing poultry outbreaks, wild bird mortality events, and spillovers into mammalian species [3,4]. This global expansion of Gs/GD lineage facilitated increased diversification of the HA gene [5].

During the 2020-2021 epidemic wave in Europe, the first report of HPAI H5N1 clade 2.3.4.4b viruses in migratory waterfowl was in the Netherlands during October 2020 [6,7]. Phylogenetic analyses revealed that the newly detected H5N1 virus was a result of reassortment among a circulating H5N8 2.3.4.4b virus with Eurasian avian lineage LPAI viruses [6]. Since its detection, H5N1 2.3.4.4b viruses have been detected across numerous countries within Africa, Asia, Europe, North, and South America [5,8]. H5N1 2.3.4.4b viruses, closely related to those in Europe, were detected in North America in December 2021 on the Atlantic coast of Canada [9–11]. Time-scaled phylogenetic analyses revealed that the spread of H5N1 to North America was likely due to migratory waterfowl movement from Iceland to Canada [11]. The H5N1 clade 2.3.4.4b virus has since

disseminated across all four North American flyways with detections in a myriad of wild and domestic bird populations and mammals [12–14].

Typically, the characterization of genotypic diversity in H5N1 viruses largely relies on phylogenetic analysis and tree construction by gene segment to assess existing and potential novel genotypes. These analyses and data visualization methods provide useful tools for the study of viral spatiotemporal evolution [9,13]. However, phylogenetic tree construction for large datasets can be time consuming and difficult to interpret [15,16]. Ordination analyses are powerful tools for the rapid visualization of evolutionary distances in low-dimensional space [16,17]. Ordination approaches are not a replacement for phylogenetic analysis but, when used in tandem, they provide useful means for efficient and targeted data analysis. Indeed, ordination and phylogenetic analyses have been used in parallel in numerous instances, including, for example, the study of evolution of HA and NA genes of influenza A viruses [18], human virome characterization [19], and fungal pathogenesis [20].

The overall goal of this study was to characterize the genotypic expansion of H5N1 2.3.4.4b virus in North America using rapid ordination approaches in concert with phylogenetics. To achieve this goal, we used ordination analyses on distance matrices generated from sequence alignments of over four thousand H5N1 clade 2.3.4.4b viruses detected in North America between its introduction in January 2020 and the end of 2023. Our ordination analyses paralleled the results observed with phylogenetic analyses, revealing the expansion of the A1 genotype into numerous reassortant genotypes upon its circulation in North America.

## 2. Materials and Methods

### 2.1. Viruses Used in This Study

A total of 4,752 H5N1 viruses within the 2.3.4.4b clade, collected between April 2020 and December 2023, were analyzed in this study (Table S1). Viral sequences were obtained from samples submitted to the U.S. Geological Survey (USGS) National Wildlife Health Center (NWHC) and U.S. Department of Agriculture (USDA) National Wildlife Disease Program, and sequenced at the USDA National Veterinary Services Laboratories (NVSL); and from two data repository banks, Global Initiative for Sharing All Influenza Database (GISAID) EpiFlu [21] and Genbank [22].

#### 2.1.1. NWHC/USDA-NVSL

From April 2020 to December 2023, 2,511 bird and four mammal samples were submitted to USDA/NWHC from bird mortality events, hunter-harvest surveillance, and live bird surveillance projects. The 2,515 samples used in this study were predominantly waterfowl (n=1,143), followed by non-host wild birds referred to as “other wild birds”, including raptors, corvids, and game birds (n=1,121), sea/shorebird (n=233), domestic waterfowl (n=12), mammal (n=4), and poultry (n=2). Tracheal/oropharyngeal swabs and cloacal swabs were collected from birds and mammals and tested for avian influenza virus, H5 and 2.3.4.4b H5 as described previously [13]. Samples positive by any of the three tests were sent to NVSL for repeat testing and whole genome sequencing.

#### 2.1.2. GISAID EpiFlu™ Database

To determine the diversity of H5N1 virus genotypes across North America (United States of America and Canada), we further included viruses submitted to GISAID that were identified in birds and mammals from January 2020 to December 2023 (n=2,182). These viruses were largely detected in poultry (n=831), followed by waterfowl (n=614), sea/shorebird (n=122), other wild bird (n=501), domestic waterfowl (n=61), and mammal (n=53). Virus sequences were obtained by downloading the eight gene segments for each viral submission (Tables S1 and S2). For comparison, ten H5N1 virus sequences identified in Europe from GISAID were downloaded and included in the downstream data analyses (Table S1).

### 2.1.3. Genbank Nucleotide Sequence Database

Additional H5N1 clade 2.3.4.4b virus submissions from January 2020 to December 2023 from the U.S., Canada, and Europe were downloaded from the Genbank nucleotide sequence database, cross-checked for duplicates in GISAID; only unique virus submissions to Genbank were retained (Table S1). The final Genbank dataset included gene segments of 55 H5N1 viruses identified in waterfowl (n=24), other wild bird (n=22), and sea/shorebird (n=9) (Table S1).

### 2.2. Sequence Pre-Processing

Prior to downstream phylogenetic analyses, sequences from each gene segment were assessed for primer trimming upstream and downstream of the gene open reading frames (ORFs). To detect the ORFs and trim the primer sequences, we used the findORFsFasta command in the package 'ORFik' [23] adapted on R Statistical Software (v4.2.2; R Core Team 2022). Each trimmed gene segment was sorted in descending order of size, as follows: PB2, PB1, PA, HA, NP, NA, MP, and NS. Thereafter, downstream phylogenetic and ordination analyses were performed on datasets consisting of either (1) sequences from each gene segment compared across all viruses or (2) concatenated virus genomes consisting of the ORFs in the same gene segment order.

### 2.3. Bioinformatic Data Analyses

#### 2.3.1. Phylogenetic Analysis

H5N1 viruses were first genotyped using GenoFlu v 1.02 [14]. This tool uses BLAST to identify North American H5NX genomes in the 2.3.4.4b clade from a curated database. Pre-defined genotypes are cross-referenced with the top segment identifications, and a genotype is assigned [24]. A cutoff of 2% difference from the closest curated sequence identifies new reassortment. New reassortants are reviewed using segment-based phylogenetic trees, and new segment sequences added to the curated database as new genotype assignments are identified. At the time of data analysis, H5N1 viruses (n=22) that were not assigned a genotype through the available GenoFLU version were categorized as 'unknown' in downstream analysis. As GenoFlu is continuously updated, these and emerging genotypes may be named in subsequent versions. Multiple sequence alignments on each virus gene segment and concatenated virus genome across all H5N1 viruses were performed using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) program using the R package 'msa' [25]. We then used the Generalized Time Reversible (GTR) substitution model and the discrete Gamma model to describe the rates of evolutionary change through fixed mutations among sequences. We visualized the evolutionary distance among sequences by constructing an ultrafast bootstrap maximum likelihood (ML) phylogenetic trees using the R package 'phangorn' [26]. Phylogenetic trees were constructed using the Interactive Tree of Life (iTOL) v6.0 [27]. To enhance readability of phylogenetic trees, we have presented the phylogenetic tree associated with gene segment PB2 in two formats: (1) phylogenetic tree with genotype annotated on the tree (Figure S2) and (2) phylogenetic tree without genotype annotation but with host type and flyway (Figure S3). Readers are encouraged to use these trees as a guideline to identify genotypes and associated color scheme for the remaining trees.

#### 2.3.2. Visualization of Evolutionary Distances Using Clustering Analysis

To allow for rapid and computationally feasible determination of existing and novel H5N1 virus genotypes, we adopted an ordination based approach for closely related sequences to accompany phylogenetic trees. We used the R package 'bios2mds' [17] to assign Euclidean distance-based difference scores to H5N1 virus sequences. Difference scores were visualized in a low dimensional space using multidimensional scaling (MDS). To identify genotypic clusters within the ordination analysis, K-means clustering was performed using the base R package "stats (version 3.6.2)". Ideal number of clusters for K-means clustering was determined via silhouette score analysis using the the R package 'bios2mds'. In brief, silhouette score values range from 0.0 to 1.0, where 0.0 is poor cluster

classification and 1.0 is optimal cluster classification [28]. We then used permutational multivariate analysis of variance (PERMANOVA) with 999 permutations on the distance matrices to determine whether significant differences were detected across host types and bird flyway. In addition, we assessed whether interactions were evident among the identified clusters and each of host type and bird flyway, separately. All statistical analyses and data visualization were performed on R Statistical Software (v4.2.2; R Core Team 2022).

### 2.3.3. Reference Dataset for H5N1 Virus Ordination and Clustering

To allow for future H5N1 virus ordination and clustering, we generated a reference dataset consisting of representative H5N1 virus sequences and associated genotypes identified through GenoFLU (Text S1 and Table S3). For genotypes with more than seventy virus assignments, we randomly selected seventy viruses as representatives. All viruses for genotypes with less than seventy virus assignments were included. We also provide an R script (Text S2) that will autonomously calculate multiple sequence alignment followed by Euclidean distance based difference score assignment of sequences, K-means clustering, ordination, and statistical analyses. The R script will require a FASTA file as an input and will output a csv file with the virus identifiers and the K-means cluster associated with each virus.

## 3. Results

### 3.1. H5N1 Virus Introductions to North America

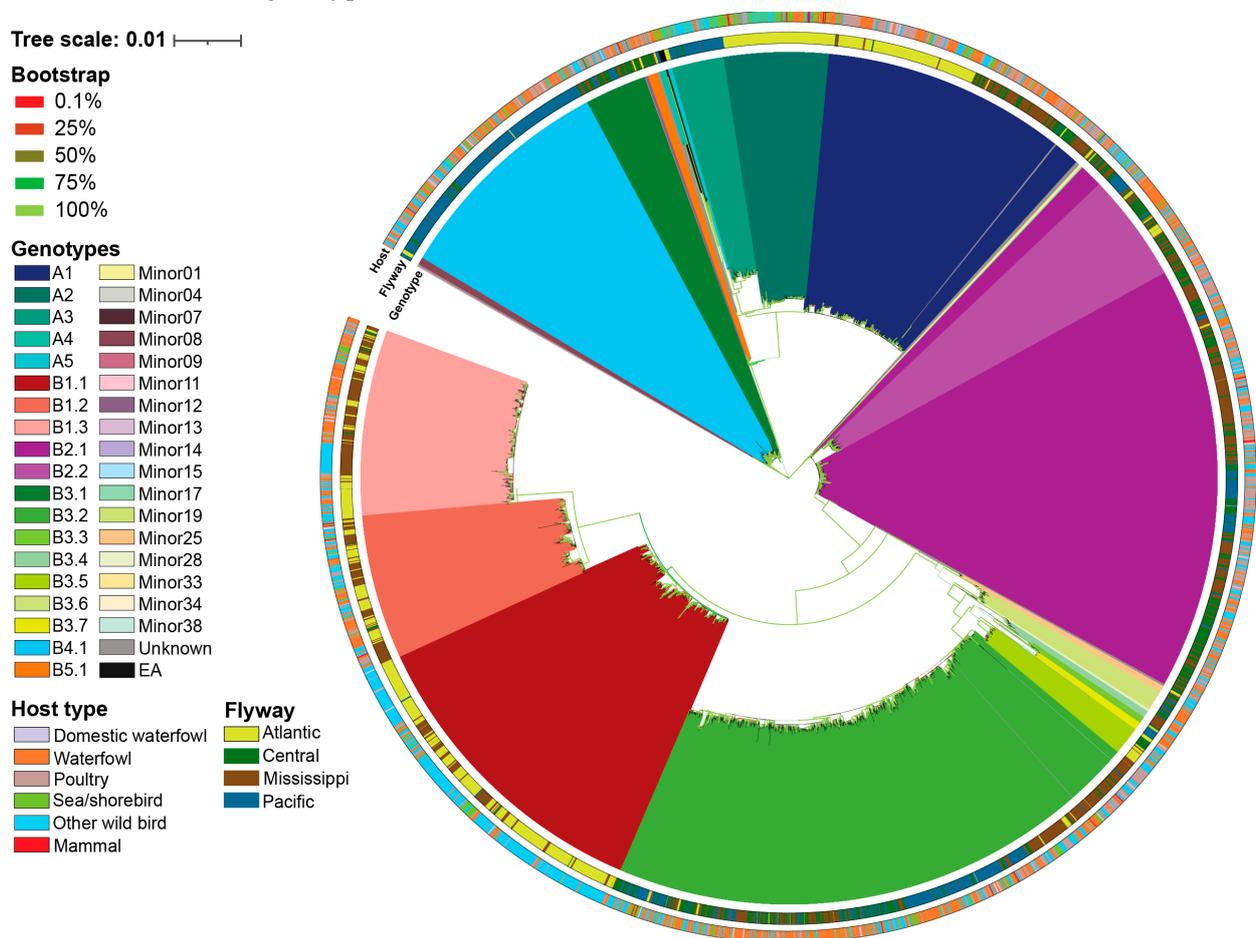
Our analysis of 4,752 H5N1 viruses within the 2.3.4.4b clade and their gene segments revealed five H5N1 virus introductions into North America. The earliest introduction (A1 genotype), through the Atlantic flyway, was detected in the United States of America (USA) on December 30, 2021 and last detected in our dataset on November 28, 2022. The A1 genotype accounted for 10.5% of total H5N1 viruses (n=500) detected in North America and was primarily associated with waterfowl (41.0%) followed by poultry (37.4%), other wild birds (14.0%), sea/shorebirds (6.0%), and domestic waterfowl (1.60%) (Table S1). On May 2022, a second introduction event was detected from the Atlantic flyway into USA known as the A2 genotype. The A2 genotype represents 3.9% of all H5N1 viruses (n=186) detected in our dataset between February 16, 2022 and September 30, 2023. In contrast to the A1 genotype, viruses within the A2 genotype were detected mostly in sea/shorebirds (45.2%) rather than waterfowl (22.0%) followed by other wild birds (21.5%), and poultry (3.8%). The A2 genotype was the only unassorted genotype among this dataset detected in mammals (7.5%) (Table S1). The third introduction was through the Pacific flyway (A3) and was first detected in Canada on February 3, 2022 in a deceased bald eagle (*Haliaeetus leucocephalus*). It was then detected in the USA on April 26, 2022 and last detected on September 19, 2023 of this dataset. This genotype accounted for 2.0% of total North American H5N1 viruses (n=96) in our dataset and was largely identified in other wild birds (59.4%) (Table S1). The fourth introduction (A4) was collected from waterfowl within Alaska, USA during October 2022. This genotype accounted for 0.1% (n=6) of total H5N1 viruses identified in this study (Table S1). The most recent introduction in our dataset (A5) accounted for 0.2% (n=8) of total viruses and was introduced into North America via the Atlantic flyway in sea/shorebirds.

### 3.2. Ordination Analysis Recapitulates H5N1 Genotypic Diversity Observed by Phylogenetic Analysis

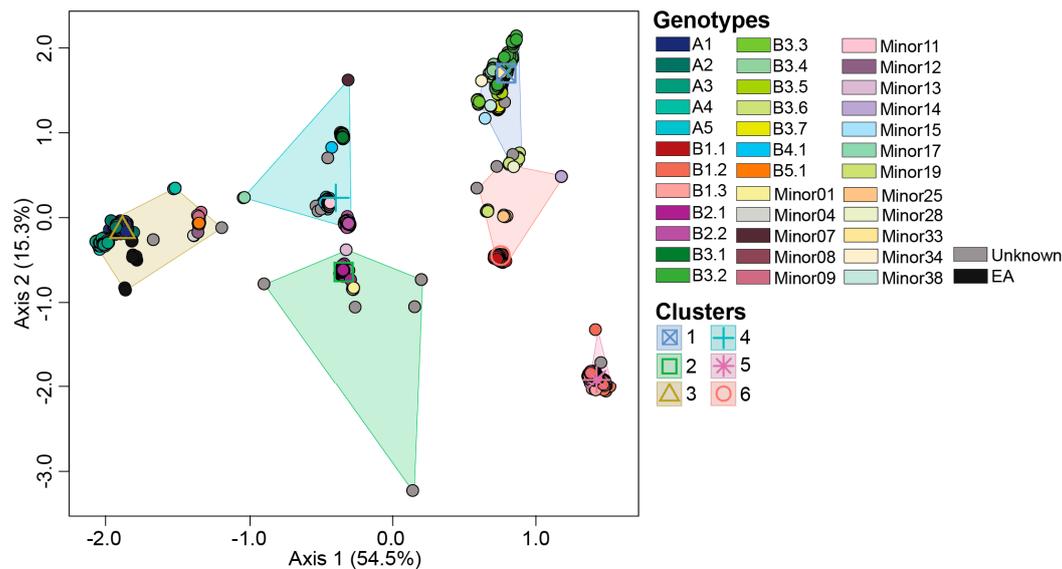
A total of 36 H5N1 genotypes were assigned using GenoFLU in our dataset. This was in agreement with a maximum likelihood phylogenetic analysis conducted on the multiple sequence alignment of the whole genome of H5N1 viruses (Figure 1). To assess whether the genotypic diversity observed using phylogenetic analyses was recapitulated with the ordination-based approach, we computed the Euclidean distances of full ORF genome H5N1 virus multiple sequence alignments. Our ordination analysis and the subsequent K-means clustering supported the genotypic diversity observed in the phylogenetic analysis, including five virus introductions to North America (genotypes A1, A2, A3, A4, and A5). Six distinct clusters were identified across the 36 genotypes in

our dataset (Figure 2, Table 1). The largest cluster was cluster 1 accounting for 23.5% of viruses identified by GenoFLU followed by clusters 2 (17.9%), 3 (17.7%), 4 (15.7%), 5 (12.8%), and 6 (12.5%). The A genotypes (unreassorted) clustered together in cluster 3 along with the ten European reference viruses. Cluster 3 also contained several minor genotypes as well as genotype B5.1 (Table 1). Other reassorted genotypes created distinct clusters. In addition, the ordination and cluster analyses assigned clusters to the viruses that were unassigned by GenoFLU (Figure 2), which was not unexpected but further supports the usefulness of an ordination-based tool for data visualization especially when identifying new genotypes. Unknown viruses were detected across all clusters: clusters 2 and 3 (n=6 each), cluster 4 (n=4), cluster 1 (n=3), cluster 6 (n=2), and cluster 5 (n=1) (Table S1). Some of these viruses (n=8) were detected in Canada while others were detected in the USA.

The observed clusters were significantly different from one another as supported by a PERMANOVA analysis (pseudo-F=70.3;  $P < 0.001$ ), indicating distinct clustering patterns. The ordination-based approach allows for rapid, albeit lower resolution, cluster identification of H5N1 viruses in less than two hours using a standard laptop, compared to a traditional full-scale phylogenetic approach with ultrafast bootstrapping, which can take up to five days using a High-Throughput Computing environment. Thus, the cluster-based approach was a robust and effective alternative for traditional phylogenetic analysis for the preliminary H5N1 virus classification and visualization into genotype clusters.



**Figure 1.** Maximum likelihood phylogenetic tree with ultrafast bootstraps of H5N1 2.3.4.4b viruses detected in North America since April 2020. Host and flyway for each virus are assigned as a color-coded ring on the outer edge of the phylogenetic tree. Color keys of genotypes, bootstraps, host category, and flyway are assigned at the left of the figure.



**Figure 2.** Multidimensional scaling to visualize the Euclidean distances computed on the multiple sequence alignment of H5N1 viruses detected in North America. Centroid of each K-means cluster is depicted in outline shapes. Genotypes and groups are color coded per the color key at the right portion of the figure.

**Table 1.** Summary of detection dates, flyways, and K-means cluster assignments of viruses that also have an assigned genotype by GenoFLU.

| Cluster | Genotype | First detected | Last detected | Flyways (N viruses) |         |             |         | Total viruses |
|---------|----------|----------------|---------------|---------------------|---------|-------------|---------|---------------|
|         |          |                |               | Atlantic            | Central | Mississippi | Pacific |               |
| 1       | B3.2     | 03/20/2022     | 12/22/2023    | 40                  | 261     | 348         | 343     | 1,111         |
|         | B3.3     | 10/08/2022     | 04/06/2023    | 3                   | 2       | 2           | x       |               |
|         | B3.4     | 11/23/2022     | 03/07/2023    | x                   | 2       | 15          | x       |               |
|         | B3.5     | 10/30/2022     | 05/18/2023    | 25                  | 17      | 4           | x       |               |
|         | B3.6     | 11/22/2022     | 10/12/2023    | x                   | 13      | 6           | 12      |               |
|         | B3.7     | 09/11/2023     | 10/11/2023    | x                   | 1       | x           | 12      |               |
|         | Minor15  | 04/29/2022     |               | x                   | x       | 1           | x       |               |
|         | Minor33  | 02/21/2023     |               | 1                   | x       | x           | x       |               |
|         | Minor34  | 09/17/2022     |               | x                   | x       | 1           | x       |               |
|         | Minor38  | 10/18/2022     |               | 2                   | x       | x           | x       |               |
| 2       | B2.1     | 03/01/2022     | 09/07/2023    | 5                   | 334     | 396         | 99      | 843           |
|         | Minor01  | 02/18/2022     | 04/27/2022    | 5                   | x       | 3           | x       |               |
|         | Minor13  | 03/15/2022     |               | x                   | 1       | x           | x       |               |
| 3       | A1       | 12/30/2021     | 11/28/2022    | 263*                | 74      | 162         | 1       | 824           |
|         | A2       | 2/16/2022      | 09/30/2023    | 186*                | x       | x           | x       |               |
|         | A3       | 02/03/2022     | 09/19/2023    | x                   | 6       | 1           | 89*     |               |
|         | A4       | 10/14/2022     |               | x                   | x       | x           | 6*      |               |
|         | A5       | 10/31/2022     | 02/03/2023    | 8*                  | x       | x           | x       |               |
|         | B5.1     | 03/01/2022     | 04/07/2022    | x                   | 13      | 6           | x       |               |
|         | Minor04  | 04/09/2022     |               | x                   | 2       | x           | x       |               |
|         | Minor09  | 02/01/2022     | 03/01/2022    | 4                   | x       | 1           | x       |               |
| 4       | Minor12  | 04/01/2022     |               | x                   | 2       | x           | x       | 742           |
|         | B2.2     | 03/28/2022     | 04/20/2023    | 22                  | 78      | 59          | 38      |               |
|         | B3.1     | 03/04/2022     | 11/01/2022    | 2                   | 65      | 24          | 23      |               |
|         | B4.1     | 03/05/2022     | 04/11/2023    | 1                   | 28      | 4           | 385     |               |
|         | Minor07  | 04/26/2022     |               | x                   | 1       | x           | x       |               |

|   |         |            |            |     |   |     |   |     |
|---|---------|------------|------------|-----|---|-----|---|-----|
|   | Minor08 | 02/01/2022 | 04/25/2022 | 8   | x | x   | x |     |
|   | Minor11 | 04/01/2022 | 04/04/2022 | x   | 2 | x   | x |     |
|   | Minor17 | 11/01/2022 |            | 2   | x | x   | x |     |
| 5 | B1.2    | 02/12/2022 | 03/06/2023 | 124 | 6 | 131 | x | 605 |
|   | B1.3    | 06/26/2022 | 04/28/2023 | 130 | 7 | 207 | x |     |
|   | B1.1    | 01/25/2022 | 04/06/2023 | 460 | 4 | 107 | 4 |     |
|   | Minor14 | 03/22/2022 | 03/25/2022 | 4   | x | x   | x |     |
| 6 | Minor19 | 11/18/2022 | 01/09/2023 | x   | 7 | 2   | x | 596 |
|   | Minor25 | 11/22/2022 | 12/13/2022 | x   | 5 | 1   | 1 |     |
|   | Minor28 | 11/21/2022 |            | x   | x | 1   | x |     |

\* Refers to flyway where this genotype was first detected in the USA; "x" indicates that no virus from our dataset was collected from the corresponding flyway.

### 3.3. Reassortant H5N1 Virus Genotypes

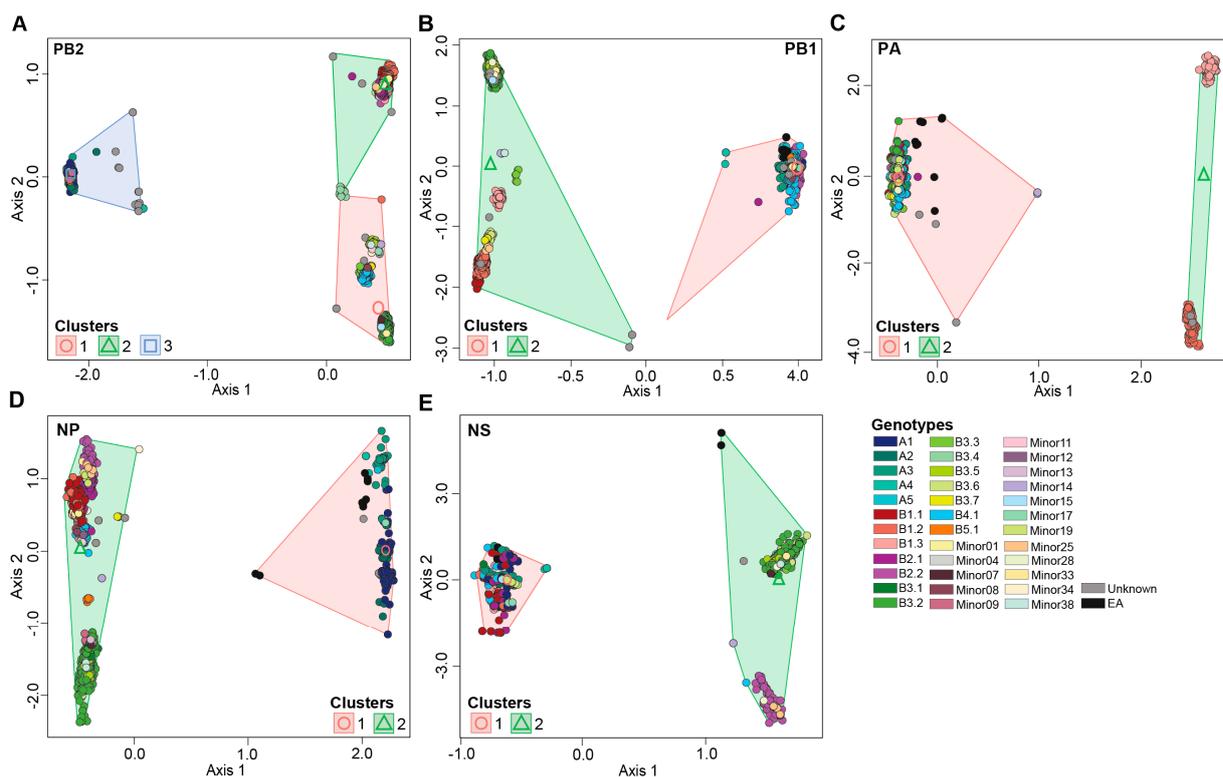
Our dataset contains 31 reassortant genotypes that were associated with the A1 genotype. Reassortment events associated with the A2, A3, A4, and A5 genotypes were not detected in this dataset. The majority of reassortant genotypes were collected during early 2022 and continued to circulate until December 2023. Fourteen reassortant genotypes were assigned to the group "B" while 17 were classified as "Minor" genotypes.

Within the reassortant genotype "B" viruses, genotypes B1, B2, B3, B4, and B5 were classified. The most common reassortant genotype was B3.2 and accounted for 25.3% of all H5N1 virus reassortant genotypes. Genotype B3.2 was first detected in a bald eagle collected from Mississippi flyway on March 20, 2022. Thereafter, it was detected largely in waterfowl (48.1%) and other wild birds (26.1%) (Table S1). The second most common reassortant genotype, B2.1 accounted for 21.3% of reassortant genotypes. This genotype was first detected on March 01, 2022 in bald eagle from the Mississippi flyway (Table S1). This genotype was primarily detected in other wild birds (32.9%) and waterfowl (31.9%). Genotype B1.1 was first detected in a bald eagle sample collected from the Atlantic flyway on January 25, 2022 (Table S1). This genotype represents the third most common reassortant (14.6%) and was largely associated with other wild birds (79.6%). Of the remaining genotypes, B4.1 was detected in a poultry sample in Canada during March 05, 2022 through the Pacific flyway. It was then detected in the USA starting April 04, 2022 (Table S1). This genotype accounted for 10.7% of reassortant genotypes. The least common genotype among this dataset was B5.1 (0.5%) and it was primarily detected in waterfowl starting March 01, 2022 (Table S1).

The genotypes assigned to the groups "Minor" in sum accounted for 1.5% of the reassortant genotypes detected in North America. The most common genotypes were Minor19 (n=9), Minor01 and Minor 08 (n=8 each), Minor07 (n=7), Minor09 (n=5), and Minor14 (n=4) (Table S1). While the remaining genotypes mostly detected in one or two hosts and as of this study remain sporadic.

### 3.4. H5N1 Virus Reassortants Were Limited to Five Viral Gene Segments

To identify gene segments involved in H5N1 virus reassortment, we performed ordination-based approaches of reassorted gene segments (Figures 3 and S1). In addition, we determined the patterns observed to those seen by maximum likelihood phylogenetic analysis on each individual gene segment (Figures S2–S7). H5N1 reassortment events among this dataset involved five gene segments: PB2, PB1, PA, NP, and NS.



**Figure 3.** Multidimensional scaling to visualize the dissimilarity distances computed on the multiple sequence alignment of H5N1 virus gene segments. (A) Axes 1 and 2 of gene segment PB2; (B) Axes 1 and 2 of gene segment PB1; (C) Axes 1 and 2 of gene segment PA; (D) Axes 1 and 2 of gene segment NP; (E) Axes 1 and 2 of gene segment NS. Genotypes and K-means clusters are color coded per the color key at the bottom right of the figure.

### 3.4.1. PB2 Gene Segment

PB2 gene segment had three K-mean clusters (1, 2, and 3) across the H5N1 viruses in this study (Figure 3A). PB2-clusters 1 and 2 were composed of reassortant viruses ( $n=3,912$ ; 82.3%) while PB2-cluster 3 was primarily composed of introduction genotypes. PB2-clusters 1 and 2 contain genotypes with a North American PB2 gene while PB2-cluster 3 contains genotypes with a Eurasian PB2 gene. PB2-cluster 1 accounted for most genotypes identified in this study ( $n=19$ ). Members within this cluster included viruses belonging to the genotypes B4.1, majority of B3 genotypes, with the exception of some B3.4 and all B3.6 viruses, and nine Minor genotypes (Minors 07, 08, 11, 14, 15, 17, 33, 34, and 38) (Figure 3A). The second largest cluster was PB2-cluster 2 ( $n=13$ ) encompassing all viruses within the genotypes B1, B2, B3.4, B3.6, and Minors 01, 13, 19, 25, and 28. PB2-cluster 3 was made up of A genotypes, B5.1, and three minor genotypes, Minors 04, 09, 12. PB2-cluster 3 was the most divergent cluster compared to 1 and 2.

To further investigate the genotypic diversity within PB2-clusters 1 and 2, we performed ordination analysis and K-mean clustering on a dataset without members of PB2-cluster 3 (Figure S1A). Our analysis reveals three novel reassortant clusters, which assign genotypes B3.5 and B4.1 to a distinct cluster separate from the B1/B2 and B3 genotypes. The patterns observed in our ordination assay were recapitulated in the phylogenetic tree associated with the PB2 gene segment with 100% bootstrap support (Figures S2 and S3).

### 3.4.2. PB1 Gene Segment

Phylogenetic analysis conducted on segment PB1, revealed four main clusters, one cluster containing the A viruses and three reassortant clusters ( $n=2,319$ ; 48.8%) (Figure S4). Similar patterns were observed in the ordination analyses (Figures 3B and S1B). Ordination analysis on the entire dataset revealed two significantly different clusters, PB1-cluster 1 contains genotypes with a Eurasian

PB1 gene (genotypes A, B2s, B3.1, B4.1, B5.1, and numerous Minor genotypes). Whereas PB1-cluster 2 viruses contain a North American PB1 gene segment, namely B1 and B3 genotypes (Figure 3B). In turn, ordination analysis performed on the genotypes within PB1-cluster 2, reveals further separation of reassortant genotypes into three sub-clusters (Figure S1B). PB1-cluster 2a was primarily associated with genotypes B1.3 and B3.3, while PB1-clusters 2b and 2c were associated with genotypes B1.1/B1.2 and B3s, respectively (Figure S1B).

#### 3.4.3. PA Gene Segment

Ordination and phylogenetic analyses revealed three clusters along the PA gene segment (Figures 3C and S5). PA-cluster 1 was the largest cluster and contained the Eurasian unassorted viruses and those retaining the Eurasian PA gene segment. While PA-cluster 2 was associated with the reassortant genotypes B1.2 and B1.3 representing a North American PA gene (Figure 3C). In addition, ordination analysis performed on genotypes B1.2 and B1.3, reveal distinct clustering patterns across the two genotypes (Figure S1C).

#### 3.4.4. NP Gene Segment

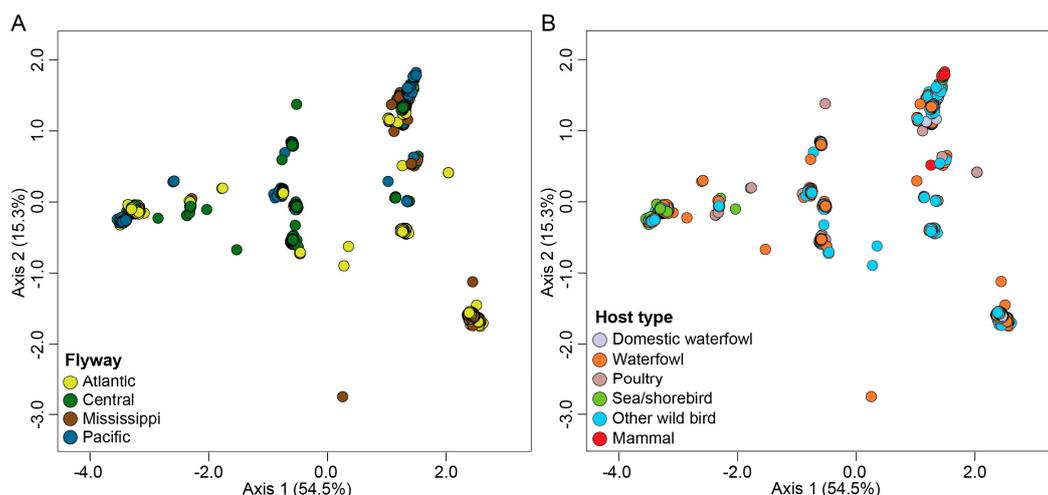
The gene segment NP had the highest number of reassortant viruses ( $n=3,937$ ; 82.8%) in addition to the original unassorted cluster (Figure S6). Ordination analysis revealed two main clusters separating the original and reassortant genotypes (Figure 3D). NP-cluster 1 was associated with the Eurasian unassorted viruses and Minor 17, while NP-cluster 2 consisted of all B genotypes and the majority of Minor genotypes. In turn, ordination analysis performed on the viruses within NP-cluster 2, identified four significantly different sub-clusters. These sub-clusters showed distinct clustering patterns of reassortant viruses that belonged to genotypes B1, B2, B3, B4, and B5 (Figure S1D).

#### 3.4.5. NS Gene Segment

Two reassortant clusters ( $n=1320$ , 27.8%) were identified in gene segment NS in addition to the unassorted cluster which accounted for the largest virus assignment (Figure 3E; Figure S7). NS-cluster 1 included the Eurasian unassorted genotypes while NS-cluster 2 included reassortant genotypes B2.2, B3.2, B3.3, B3.4, B3.5, B3.6, B3.7, and seven Minors (Figure 3E). In turn, genotype B2.2, Minors 14, 25, and 28 had significantly different clustering patterns that the remaining reassortant genotypes (Figure S1E).

### 3.5. Migratory Flyway and Host Type May Impact Viral Genotypic Diversity

Our results revealed a significant impact of bird flyway on H5N1 genotypic diversity in North America (pseudo- $F=32.2$ ;  $P<0.001$ ,  $R^2=0.100$ ) (Figure 4A). Our data show that the number of identified H5N1 viruses and their associated genotypes varies across subregions within individual bird flyways across the USA. For example, the Mississippi flyway accounted for the highest number of identified H5N1 viruses within North America ( $n=1,483$ ; 31.3%). However, more than 80% of these viruses ( $n=1,223$ ) were identified only in states within the midwestern region of this flyway compared to the northeastern, southeastern, and southcentral regions. Similar results were observed for the Central flyway ( $n=937$ ; 19.8%) where the north central regions accounted for the highest number of viruses ( $n=597$ ; 63.6%) identified within this flyway. Furthermore, several H5N1 virus clusters were specific to a subregion within a flyway as compared to other regions as supported by a significant interaction (pseudo- $F=6.42$ ;  $P<0.0001$ ;  $R^2=0.046$ ) between H5N1 genotypic clusters and bird flyway. In addition, a significant interaction was detected between host type and flyway (pseudo- $F=13.1$ ;  $P<0.001$ ;  $R^2=0.037$ ).

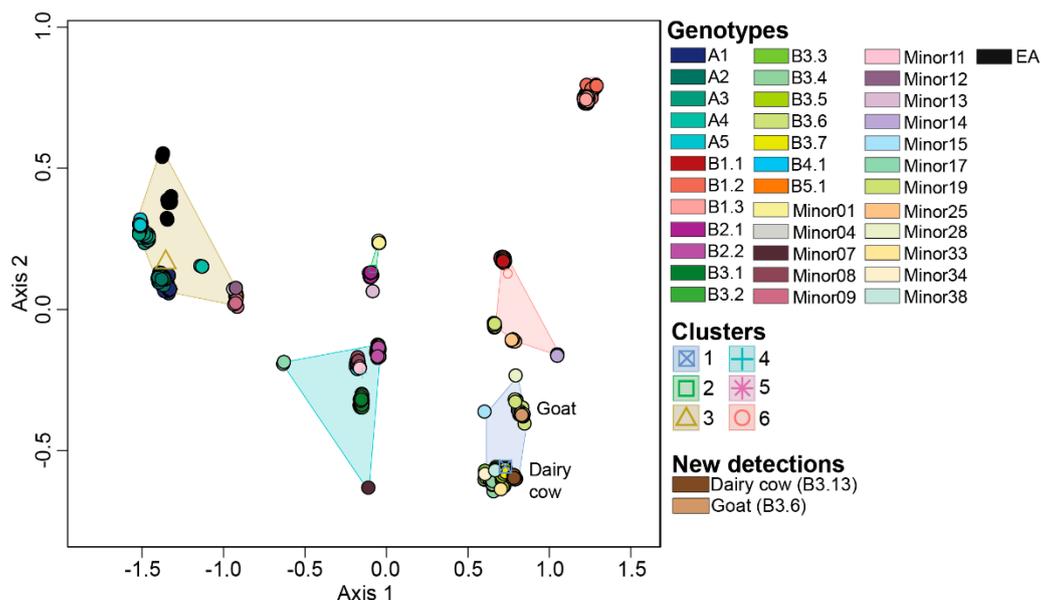


**Figure 4.** Multidimensional scaling to visualize genotypic diversity of H5N1 viruses in North America. (A) Flyway; (B) Host type. Each variable is color coded differently as represented by the color key at the bottom left of each panel.

Finally, we show differences in H5N1 virus distribution across host type (pseudo- $F=45.5$ ;  $P<0.001$ ;  $R^2=0.042$ ) (Figure 4B). As may be expected, the majority of H5N1 viruses analyzed in this study were identified in waterfowl (37.5%) and other wild birds (34.6%).

### 3.6. Reference Dataset Is Robust to Be Used for Ordinal Assignment of Newly Sequenced H5N1 Viruses

To allow for rapid genotypic assignment of newly detected H5N1 viruses in North America, we generated a reference dataset containing 983 viruses spanning all genotypes identified in this study. We ensured that the reference dataset recapitulates the genotypic clusters observed in the original dataset by performing ordination analysis on the dissimilarity distances followed by K-means clustering. We further validated this dataset by including recently detected H5N1 viruses from dairy cows ( $n=17$ ) collected from Ohio, South Dakota, and Texas and goats ( $n=6$ ) from Minnesota (Figure 5). As observed in the original dataset, six K-means clusters were detected in the reference dataset. Each cluster was associated with the same genotypes as the original dataset (Figure 5). The dairy cow and goat samples were identified within cluster 5 that was associated with the majority of B3 genotypes, and consistent with expectations based upon phylogeny. All dairy cow samples clustered closely with genotype B3.7, whereas the goat samples clustered with genotype B3.6 as expected as this virus was identical to what the chickens and ducks on the same premises had. The virus affecting dairy cows is genotype B3.13. These results demonstrate that use of the reference dataset generated by this study allows rapid grouping of new viruses based upon ordination.



**Figure 5.** Multidimensional scaling performed on the dissimilarity distances on the viruses within the reference dataset. Centroid of each K-means clusters is depicted in outline shapes. Genotypes and clusters are color coded per the color key at the right portion of the figure. Novel viruses downloaded from GISAID are categorized under “New detections” and annotated on the MDS plot.

#### 4. Discussion

Multiple genotyping schemes have been developed for A/goose/Guangdong/1/1996 (Gs/GD) H5N1 lineage viruses. Because of the complex circulation of these viruses within flyways the distinct genotyping schemes utilized in Asia [29], Europe [7], and North America [14] have utility for better understanding and monitoring the change and movement of these viruses within those geographic areas. The GenoFLU tool used in this study was developed to characterize introductions of the 2.3.4.4b clade in the USA as well as subsequent re-assortments of that virus with low pathogenic North American viruses [14]. This study aimed at implementing strategies for the efficient detection and visualization of H5N1 2.3.4.4b virus genotypic diversity across North America. To this end, we used multidimensional scaling (MDS) analysis as an alternative to phylogenetic analysis for the identification of H5N1 virus reassortants across North America between January 2020 and December 2023. Our results revealed that the MDS approach readily paralleled the results observed with phylogenetic analysis while reducing time and computational demand. To our knowledge, this is the first study that uses clustering approach in tandem with phylogenetic analyses for the genotypic assignment of H5N1 viruses in North America. Ordination-based approaches are used in addition to phylogenetic analyses in numerous host-virus systems, including the study of coronavirus evolution through the analysis of the spike protein [30]. Similarly, studies that assess the virome associated with human blood and plasma utilize cluster-based methods to show the evolutionary space of anelloviruses in relation to viruses within reference databases [19,31]. This approach is not limited to the study of viral evolution in mammalian systems but is also used in the identification of tick evolution across various geographical spans [32]. While ordination analysis effectively detects clustering patterns in a dataset, it has some drawbacks, including sensitivity to outliers, potential for overfitting, and loss of information. Thus, the utilization of MDS approaches in addition to phylogenetic analyses can provide a powerful compendium of tools to study evolutionary changes in targeted populations within and across host systems.

The dataset used in this study identified five distinct H5N1 2.3.4.4b virus introductions into North America [14]; three of which arrived through the Atlantic flyway (A1, A2, and A5) [9,10] and one through the Pacific flyway (A3 and A4) [33]. Interestingly, our analysis of viruses not assigned genotypes by GenoFLU revealed viruses that clustered with A1 genotype which were first detected in Canada in mid-December 2021 where the A1 genotype was then detected in the USA in late

December 2021. While some viruses collected from Canada were not assigned any genotype, the clustering pattern supports the circulation of the A1 genotype in Canada prior to the USA [34,35]. Similarly, the A3 genotype was first detected in a bald eagle sample in British Columbia, Canada collected in February 2022, followed by Alaska, USA in April 2022. The remaining A genotypes were only detected in the USA across 2022 and 2023 [14,36]. We acknowledge that virus sequences from Canada were under-represented in our study such as the H5N5 (provisionally A6 genotypes) that has been found in Atlantic Canada at the beginning of 2023 [37]. As such, we cannot exclude the possibility that genotypes A2, A4, A5 or A1 reassortants were circulating in Canada prior to their detection in the USA. Waterfowl migratory flyways span North America in a north-south orientation [38], making it likely that common H5N1 genotypes in the United States also occur in Canada. Indeed, a previous study reports the identification of a reassortant H5N1 genotype containing gene segments closely related to the North American A1 genotype in broiler chicken in Canada [39], revealing the potential circulation of reassortant H5N1 genotypes in Canadian regions. Thus, future efforts to obtain historical H5N1 virus sequences from Canada would be beneficial to better understand the introduction and emergence of reassortant H5N1 2.3.4.4b viruses in North America. In addition, while over a thousand H5N1 genomes from the year 2022 introduction into North America have been made available, further sequencing of yet untested samples may reveal additional introductions, such as the H5N5 [40], and help elucidate the evolutionary history of these viruses following their introduction.

Reassortment events within H5N1 2.3.4.4b viruses are not surprising and have been reported across numerous continents [8,29]. Indeed, studies provide evidence of numerous reassortment events involving one or more gene segments of the virus. For example, a recent study characterized 16 distinct reassortant virus genotypes identified in 233 H5N1 2.3.4.4b virus sequences across countries within Africa, North Americas, Asia, and Europe [29,41]. While one genotype classified as G1, based on the Asian classification scheme [29], was widely spread across countries from all four continents, numerous other genotypes were associated with specific geographic locales. Furthermore, a recent study of genotypic diversity of H5N1 virus in the USA, revealed 21 distinct clusters of which six reassortant genotypes accounted for 92% of all viruses identified in the USA [14]. Our results parallel those described by Youk et al., [14] for North America, supporting the emergence of reassortant genotypes from wild bird LPAI viruses.

Our study identified that reassortment events were associated with five gene segments of H5N1 viruses. These data were supported by studies that report reassortment events associated with the same gene segments across the USA and other countries [14,42,43]. Reassortment events associated with avian influenza viruses are shown to generate significant viral diversity and facilitate host specificity and virulence [44] and may have contributed to the rapid spread of influenza viruses and their emergence in novel hosts [5]. Our data curation from public repositories revealed numerous examples of H5N1 viruses detected in North America from mammalian hosts, such as bobcats (*Lynx rufus*), red foxes (*Vulpes vulpes*), bottlenose dolphins (*Tursiops truncatus*), skunks, and seals. Similar host expansion is increasingly reported across North America, with novel detections in marine mammals including dolphins [45], and the recent spillover into dairy cows [46–48] with an associated increase in human detections [46]. Interestingly, in most cases, viruses detected in mammals were associated with reassortant genotypes, especially genotypes with PB2 reassorted. In addition, reassortant H5N1 2.3.4.4b viruses previously detected in North America were also detected in marine mammals in Peru [42].

The circulation of LPAI H5N1 in poultry populations allowed for the initial emergence of the HPAI H5N1 goose/Guangdong lineage viruses and their spillover to wild bird populations [49,50]. Following introduction into the United States, wide distribution of HPAI viruses in wild waterfowl species that are natural hosts for North American LPAI viruses created the potential for reassortment events and the emergence of novel genotypes [51]. While some mammalian hosts were detected, our study found that most H5N1 HPAI viruses in North America during this period were associated with avian hosts, particularly waterfowl and other wild birds. In addition, we showed varied H5N1 virus clustering patterns as measured by host type. For example, genotype B4.1 was mostly detected in

waterfowl followed by other wild birds. In addition, several reassortant genotypes were identified in mammals but were in turn not identified within the A1 genotype. This dataset suggests an association of viral genotypic diversity with host species; however, genotype success or frequency as well as potential sampling biases in the analyzed dataset are acknowledged. For example, 1) while genotype B1.1 was associated with other wild birds, this genotype was largely associated with black vulture (*Coragyps atratus*) mortality events (n=350/456). These animals roost together and are known to feed on members of the roost, perpetuating the infection [52], 2) bald eagles are large birds with high visibility, potentially increasing the sampling of bald eagle samples within mortality events of raptors, and 3) limited surveillance in mammals.

We also show that avian migratory flyways and sub-regions within a flyway may further contribute to H5N1 genotypic diversity in North America. Four introductions of H5N1 to North America were limited to the Atlantic flyway, while two introductions were limited to the Pacific flyway. However, reassortant genotypes had a higher rate of detection within subregions of the Central and Mississippi flyways. Our results agree with previous reports of genotypic diversity associated with geographic ranges, including bird flyways [14,53]. This result may reveal a geographic signature of H5N1 virus detection; however, some degree of sampling bias and data heterogeneity may also affect the magnitude in effect size and the observed distribution of genotypes. While surveillance mechanisms exist for monitoring of avian influenza in wild migratory waterfowl, the absence of a targeted and consistent surveillance strategy in other wild bird species and mammals provides an incomplete understanding of the full transmission dynamics and distribution of different genotypes.

In conclusion, our study reveals that ordination and cluster-based approaches can complement traditional phylogenetic analyses specifically for the preliminary assignment of H5N1 viruses to genotypic groups or to identify novel genotypes. Rapid genotypic assignment of newly detected H5N1 viruses was achieved using a reference dataset that recapitulates the genotypic clusters followed by K-means clustering. Using this approach, we provide fundamental insight into the genotypic diversity and spread of H5N1 2.3.4.4b viruses across the USA which expands current knowledge on genotype diversity of H5N1 viruses in North America. This work also highlights the need for more widespread and standardized surveillance strategies for the accurate depiction of reassortant genotype circulation within non-target host populations, including mammalian hosts.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Multidimensional scaling to visualize genotypic diversity of reassortant H5N1 viruses associated with each gene segment. (A) Axes 1 and 2 of gene segment PB2; (B) Axes 1 and 2 of gene segment PB1; (C) Axes 1 and 2 of gene segment PA; (D) Axes 1 and 2 of gene segment NP; (E) Axes 1 and 2 of gene segment NS. Genotypes and K-means clusters are color coded per the color key at the bottom right of the figure. Figure S2: Phylogenetic tree with ultrafast bootstraps of PB2 gene segment. Genotypes, flyway, and host category are color coded differently and bootstrap support values are indicated categorically. Figure S3: Phylogenetic tree with ultrafast bootstraps of PB2 gene segment with genotypes annotation on graph. Genotypes are color coded differently and bootstrap support values are indicated categorically. Figure S4: Phylogenetic tree with ultrafast bootstraps of PB1 gene segment. Genotypes, flyway, and host category are color coded differently and bootstrap support values are indicated categorically. Figure S5: Phylogenetic tree with ultrafast bootstraps of PA gene segment. Genotypes, flyway, and host category are color coded differently and bootstrap support values are indicated categorically. Figure S6: Phylogenetic tree with ultrafast bootstraps of NP gene segment. Genotypes, flyway, and host category are color coded differently and bootstrap support values are indicated categorically. Figure S7: Phylogenetic tree with ultrafast bootstraps of NS gene segment. Genotypes, flyway, and host category are color coded differently and bootstrap support values are indicated categorically. Table S1: Genotypic assignment of each virus detected in North America with description of collection date, host species, and bird flyway. Table S2: Accession numbers for H5N1 sequences from different repositories used in this study. Text S1: R Script for Bioinformatics Analysis of Avian Influenza Virus Data

**Author Contributions:** Conceptualization, Patil Tawidian and Hon Ip; Data curation, Patil Tawidian, Krista Dilione and Jourdan Ringenberg; Formal analysis, Patil Tawidian and Kristina Lantz; Funding acquisition, Mia Torchetti and Hon Ip; Investigation, Patil Tawidian; Methodology, Patil Tawidian, Mia Torchetti, Mary Killian, Sarah Bevins and Julianna Leno; Project administration, Hon Ip; Software, Patil Tawidian; Supervision, Hon Ip; Validation, Patil Tawidian; Visualization, Patil Tawidian; Writing – original draft, Patil Tawidian; Writing – review & editing, Mia Torchetti, Mary Killian, Kristina Lantz, Sarah Bevins and Julianna Leno. All authors

will be informed about each step of manuscript processing including submission, revision, revision reminder, etc. via emails from our system or assigned Assistant Editor.

**Funding:** This research was funded by a U.S. Department of Agriculture Farm Bill grant and by the U.S. Geological Survey Ecosystems Mission Area.

**Data Availability Statement:** The data presented in this study are available in GISAID and NCBI Genbank, accession number [To be filled upon receipt]. These data were derived from the following resources available in the public domain: GISAID (<http://www.gisaid.org>) and NCBI (<https://www.ncbi.nlm.nih.gov>).

**Acknowledgments:** We thank Katy Griffin for her expert assistance and members of the National Veterinary Services Laboratories, including Cameron Norris, Jessica Hicks, and Tod Stuber for their excellent technical and bioinformatic assistance. The authors gratefully acknowledge all data contributors, including wildlife biologists, researchers, and their originating laboratories. In addition, we acknowledge laboratories for generating sequences and sharing them on public repositories, such as GISAID Initiative and Genbank. Finally, phylogenetic analyses were performed using the computer resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Disclaimer:** Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the U.S. Department of Agriculture.

## References

1. Blagodatski, A.; Trutneva, K.; Glazova, O.; Mityaeva, O.; Shevkova, L.; Kegeles, E.; Onyanov, N.; Fede, K.; Maznina, A.; Khavina, E.; et al. Avian Influenza in Wild Birds and Poultry: Dissemination Pathways, Monitoring Methods, and Virus Ecology. *Pathogens*. **2021**, *10*, 630. doi:10.3390/pathogens10050630.
2. Boulonier, T. Avian Influenza Spread and Seabird Movements between Colonies. *TREE*. **2023**, *38*, 391–395. doi:10.1016/j.tree.2023.02.002.
3. Sonnberg, S.; Webby, R.J.; Webster, R.G. Natural History of Highly Pathogenic Avian Influenza H5N1. *Virus Res*. **2013**, *178*, 63–77. doi:10.1016/j.virusres.2013.05.009.
4. Lee, D.-H.; Bertran, K.; Kwon, J.-H.; Swayne, D.E. Evolution, Global Spread, and Pathogenicity of Highly Pathogenic Avian Influenza H5Nx Clade 2.3.4.4. *J. Vet. Sci.* **2017**, *18*, 269–280. doi:10.4142/jvs.2017.18.S1.269.
5. Verhagen, J.H.; Fouchier, R.A.M.; Lewis, N. Highly Pathogenic Avian Influenza Viruses at the Wild–Domestic Bird Interface in Europe: Future Directions for Research and Surveillance. *Viruses*. **2021**, *13*, 212. doi:10.3390/v13020212.
6. Lewis, N.S.; Banyard, A.C.; Whittard, E.; Karibayev, T.; Al Kafagi, T.; Chvala, I.; Byrne, A.; Meruyert (Akberovna), S.; King, J.; Harder, T.; et al. Emergence and Spread of Novel H5N8, H5N5 and H5N1 Clade 2.3.4.4 Highly Pathogenic Avian Influenza in 2020. *Emerg. Microbes & Infect.* **2021**, *10*, 148–151. doi:10.1080/22221751.2021.1872355.
7. Fusaro, A.; Zecchin, B.; Giussani, E.; Palumbo, E.; Agüero-García, M.; Bachofen, C.; Bálint, Á.; Banihashem, F.; Banyard, A.C.; Beerens, N.; et al. 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.* **2024**, *10*, veae027. doi:10.1093/ve/veae027.
8. Gu, W.; Shi, J.; Cui, P.; Yan, C.; Zhang, Y.; Wang, C.; Zhang, Y.; Xing, X.; Zeng, X.; Liu, L.; et al. Novel H5N6 Reassortants Bearing the Clade 2.3.4.4b HA Gene of H5N8 Virus Have Been Detected in Poultry and Caused Multiple Human Infections in China. *Emerg. Microbes & Infect.* **2022**, *11*, 1174–1185. doi:10.1080/22221751.2022.2063076.
9. Caliendo, V.; Lewis, N.S.; Pohlmann, A.; Baillie, S.R.; Banyard, A.C.; Beer, M.; Brown, I.H.; Fouchier, R. a. M.; Hansen, R.D.E.; Lameris, T.K.; et al. Transatlantic Spread of Highly Pathogenic Avian Influenza H5N1 by Wild Birds from Europe to North America in 2021. *Sci. Rep.* **2022**, *12*, 11729. doi:10.1038/s41598-022-13447-z.
10. Bevins, S.N.; Shriner, S.A.; Cumbee, J.C.; Dilione, K.E.; Douglass, K.E.; Ellis, J.W.; Killian, M.L.; Torchetti, M.K.; Leno, J.B. Intercontinental Movement of Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4 Virus to the United States, 2021. *Emerg. Infect. Dis.* **2022**, *28*, 1006–1011. doi:10.3201/eid2805.220318.
11. Günther, A.; Krone, O.; Svansson, V.; Pohlmann, A.; King, J.; Hallgrímsson, G.T.; Skarphéðinsson, K.H.; Sigurðardóttir, H.; Jónsson, S.R.; Beer, M.; et al. Iceland as Stepping Stone for Spread of Highly Pathogenic Avian Influenza Virus between Europe and North America. *Emerg. Infect. Dis.* **2022**, *28*, 2383–2388. doi:10.3201/eid2812.221086.
12. Elsmo, E.; Wünschmann, A.; Beckmen, K.; Broughton-Neiswanger, L.; Buckles, E.; Ellis, J.; Fitzgerald, S.; Gerlach, R.; Hawkins, S.; Ip, H.; et al. Highly Pathogenic Avian Influenza A(H5N1) Virus Clade 2.3.4.4b

- Infections in Wild Terrestrial Mammals, United States, 2022. *Emerg. Infect. Dis.* **2023**, *29*, 2451-2460. doi: 10.3201/eid2912.230464.
13. Puryear, W.; Sawatzki, K.; Hill, N.; Foss, A.; Stone, J.J.; Doughty, L.; Walk, D.; Gilbert, K.; Murray, M.; Cox, E.; et al. Highly Pathogenic Avian Influenza A(H5N1) Virus Outbreak in New England Seals, United States. *Emerg. Infect. Dis.* **2023**, *29*, 786–791. doi:10.3201/eid2904.221538.
  14. Youk, S.; Torchetti, M.K.; Lantz, K.; Lenocho, J.B.; Killian, M.L.; Leyson, C.; Bevins, S.N.; Dilione, K.E.; Ip, H.S.; Stallknecht, D.E.; et al. 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.* **2023**, *587*, 109860. doi: 10.1016/j.virol.2023.109860.
  15. Blackshields, G.; Sievers, F.; Shi, W.; Wilm, A.; Higgins, D.G. Sequence Embedding for Fast Construction of Guide Trees for Multiple Sequence Alignment. *Algorithms Mol. Biol.* **2010**, *5*. doi:10.1186/1748-7188-5-21.
  16. Higgins, D.G. Sequence Ordinations: A Multivariate Analysis Approach to Analysing Large Sequence Data Sets. *Bioinformatics.* **1992**, *8*, 15–22. doi: 10.1093/bioinformatics/8.1.15.
  17. Pelé, J.; Bécu, J.-M.; Abdi, H.; Chabbert, M. Bios2mids: An R Package for Comparing Orthologous Protein Families by Metric Multidimensional Scaling. *BMC Bioinform.* **2012**, *13*, 133. doi:10.1186/1471-2105-13-133.
  18. Shi, W.; Lei, F.; Zhu, C.; Sievers, F.; Higgins, D.G. A Complete Analysis of HA and NA Genes of Influenza A Viruses. *PLoS One.* **2010**, *5*, e14454. doi:10.1371/journal.pone.0014454.
  19. Thijssen, M.; Khamisipour, G.; Maleki, M.; Devos, T.; Li, G.; Van Ranst, M.; Matthijssens, J.; Pourkarim, M.R. Characterization of the Human Blood Virome in Iranian Multiple Transfused Patients. *Viruses.* **2023**, *15*, 1425. doi:10.3390/v15071425.
  20. Shang, Y.; Xiao, G.; Zheng, P.; Cen, K.; Zhan, S.; Wang, C. Divergent and Convergent Evolution of Fungal Pathogenicity. *GBE.* **2016**, *8*, 1374–1387. doi:10.1093/gbe/evw082.
  21. GISAID. Available online: <http://www.gisaid.org> (accessed on [11 June 2024]).
  22. NCBI. Available online: <https://www.ncbi.nlm.nih.gov> (accessed on [11 June 2024]).
  23. Tjeldnes, H.; Labun, K.; Torres Cleuren, Y.; Chyżyńska, K.; Świrski, M.; Valen, E. ORFik: A Comprehensive R Toolkit for the Analysis of Translation. *BMC Bioinform.* **2021**, *22*, 336. doi:10.1186/s12859-021-04254-w.
  24. Hicks, J.; Stuber, T.; Lantz, K.; Torchetti, M.; Robbe-Austerman, S. vSNP: A SNP Pipeline for the Generation of Transparent SNP Matrices and Phylogenetic Trees from Whole Genome Sequencing Data Sets. *BMC Genomics.* **2024**, *25*, 545. doi:10.1186/s12864-024-10437-5.
  25. Bodenhofer, U.; Bonatesta, E.; Horejš-Kainrath, C.; Hochreiter, S. Msa: An R Package for Multiple Sequence Alignment. *Bioinformatics.* **2015**, *31*, 3997–3999. doi:10.1093/bioinformatics/btv494.
  26. Schliep, K.P. Phangorn: Phylogenetic Analysis in R. *Bioinformatics.* **2011**, *27*, 592–593. doi:10.1093/bioinformatics/btq706.
  27. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Res.* **2024**, *52*, gkae268. doi:10.1093/nar/gkae268.
  28. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. doi:10.1016/0377-0427(87)90125-7.
  29. Cui, P.; Shi, J.; Wang, C.; Zhang, Y.; Xing, X.; Kong, H.; Yan, C.; Zeng, X.; Liu, L.; Tian, G.; et al. Global Dissemination of H5N1 Influenza Viruses Bearing the Clade 2.3.4.4b HA Gene and Biologic Analysis of the Ones Detected in China. *Emerg. Microbes & Infect.* **2022**, *11*, 1693–1704. doi:10.1080/22221751.2022.2088407.
  30. Telele, N., Fikrie Predicting Interspecies Transmission and Pandemic Risks of Coronaviruses [Masters thesis], University of skovde, Skövde, **2020**.
  31. Thijssen, M.; Tacke, F.; Van Espen, L.; Cassiman, D.; Naser Aldine, M.; Nevens, F.; Van Ranst, M.; Matthijssens, J.; Pourkarim, M.R. Plasma Virome Dynamics in Chronic Hepatitis B Virus Infected Patients. *Front. Microbiol.* **2023**, *14*. doi: 10.3389/fmicb.2023.1172574.
  32. Mohamed, W.M.A.; Moustafa, M.A.M.; Thu, M.J.; Kakisaka, K.; Chatanga, E.; Ogata, S.; Hayashi, N.; Taya, Y.; Ohari, Y.; Naguib, D.; et al. Comparative Mitogenomics Elucidates the Population Genetic Structure of *Amblyomma Testudinarium* in Japan and a Closely Related *Amblyomma* Species in Myanmar. *Evol. Appl.* **2022**, *15*, 1062–1078. doi:10.1111/eva.13426.
  33. Isoda, N.; Onuma, M.; Hiono, T.; Sobolev, I.; Lim, H.; Nabeshima, K.; Honjyo, H.; Yokoyama, M.; Shestopalov, A.; Sakoda, Y. Detection of New H5N1 High Pathogenicity Avian Influenza Viruses in Winter 2021–2022 in the Far East, Which Are Genetically Close to Those in Europe. *Viruses.* **2022**, *14*, 2168. doi:10.3390/v14102168.
  34. Gass, J.D.; Hill, N.J.; Damodaran, L.; Naumova, E.N.; Nutter, F.B.; Runstadler, J.A. Ecogeographic Drivers of the Spatial Spread of Highly Pathogenic Avian Influenza Outbreaks in Europe and the United States, 2016–Early 2022. *Int. J. Environ. Res. Public Health.* **2023**, *20*, 6030. doi:10.3390/ijerph20116030.
  35. Harvey, J.A.; Mullinax, J.M.; Runge, M.C.; Prosser, D.J. The Changing Dynamics of Highly Pathogenic Avian Influenza H5N1: Next Steps for Management & Science in North America. *Biol. Conserv.* **2023**, *282*, 110041. doi:10.1016/j.biocon.2023.110041.
  36. Ramey, A.M.; Scott, L.C.; Ahlstrom, C.A.; Buck, E.J.; Williams, A.R.; Kim Torchetti, M.; Stallknecht, D.E.; Poulson, R.L. Molecular Detection and Characterization of Highly Pathogenic H5N1 Clade 2.3.4.4b Avian

- Influenza Viruses among Hunter-Harvested Wild Birds Provides Evidence for Three Independent Introductions into Alaska. *Virology*. **2024**, 589, 109938. doi:10.1016/j.virol.2023.109938.
37. Erdelyan, C.N.G.; Kandeil, A.; Signore, A.V.; Jones, M.E.B.; Vogel, P.; Andreev, K.; Bøe, C.A.; Gjerset, B.; Alkie, T.N.; Yason, C.; et al. Multiple Transatlantic Incursions of Highly Pathogenic Avian Influenza Clade 2.3.4.4b A(H5N5) Virus into North America and Spillover to Mammals. *Cell Rep*. **2024**, 43, 114479. doi:10.1016/j.celrep.2024.114479.
  38. Fourment, M.; Darling, A.E.; Holmes, E.C. The Impact of Migratory Flyways on the Spread of Avian Influenza Virus in North America. *BMC Evol. Biol*. **2017**, 17, 118. doi:10.1186/s12862-017-0965-4.
  39. Alkie, T.N.; Lopes, S.; Hisanaga, T.; Xu, W.; Suderman, M.; Koziuk, J.; Fisher, M.; Redford, T.; Lung, O.; Joseph, T.; et al. A Threat from Both Sides: Multiple Introductions of Genetically Distinct H5 HPAI Viruses into Canada via Both East Asia-Australasia/Pacific and Atlantic Flyways. *Virus Evol*. **2022**, 8, veac077. doi:10.1093/ve/veac077.
  40. World Organisation of Animal Health (WAHIS) Available online: <https://wahis.woah.org/#/in-review/5065> (accessed on 26 July 2023).
  41. Authority, E.F.S.; European Centre for Disease Prevention and Control; Influenza, E.U.R.L. for A.; Alexakis, L.; Fusaro, A.; Kuiken, T.; Mirinavičiūtė, G.; Ståhl, K.; Staubach, C.; Svartström, O.; et al. Avian Influenza Overview March–June 2024. *EFSA Journal* **2024**, 22, e8930. doi:10.2903/j.efsa.2024.8930.
  42. 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. Highly Pathogenic Avian Influenza A (H5N1) in Marine Mammals and Seabirds in Peru; *Nat. Commun*. **2023**, 14, 5489. doi:10.1038/s41467-023-41182-0.
  43. Kandeil, A.; Patton, C.; Jones, J.C.; Jeevan, T.; Harrington, W.N.; Trifkovic, S.; Seiler, J.P.; Fabrizio, T.; Woodard, K.; Turner, J.C.; et al. Rapid Evolution of A(H5N1) Influenza Viruses after Intercontinental Spread to North America. *Nat. Commun*. **2023**, 14, 3082. doi:10.1038/s41467-023-38415-7.
  44. Shao, W.; Li, X.; Goraya, M.; Wang, S.; Chen, J.-L. Evolution of Influenza A Virus by Mutation and Re-Assortment. *Int. J. Mol. Sci*. **2017**, 18, 1650. doi:10.3390/ijms18081650.
  45. Murawski, A.; Fabrizio, T.; Ossiboff, R.; Kackos, C.; Jeevan, T.; Jones, J.C.; Kandeil, A.; Walker, D.; Turner, J.C.M.; Patton, C.; et al. Highly Pathogenic Avian Influenza A(H5N1) Virus in a Common Bottlenose Dolphin (*Tursiops truncatus*) in Florida. *Commun. Biol*. **2024**, 7, 476. doi:10.1038/s42003-024-06173-x.
  46. Garg, S. Outbreak of Highly Pathogenic Avian Influenza A(H5N1) Viruses in U.S. Dairy Cattle and Detection of Two Human Cases – United States, 2024. *MMWR*. **2024**, 73. doi:10.15585/mmwr.mm7321e1.
  47. 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. Emergence and Interstate Spread of Highly Pathogenic Avian Influenza A(H5N1) in Dairy Cattle 2024. *bioRxiv*. **2024**. doi: 10.1101/2024.05.01.591751.
  48. Hu, X.; Saxena, A.; Magstadt, D.R.; Gauger, P.C.; Burrough, E.; Zhang, J.; Siepker, C.; Mainenti, M.; Gorden, P.J.; Plummer, P.; et al. Genomic Characterization of Highly Pathogenic Avian Influenza A H5N1 Virus Newly Emerged in Dairy Cattle. *Emerg. Microbes & Infect.* **2024**, 13, 2380421. doi: 10.1080/22221751.2024.2380421.
  49. Beerens, N.; Heutink, R.; Harders, F.; Bossers, A.; Koch, G.; Peeters, B. Emergence and Selection of a Highly Pathogenic Avian Influenza H7N3 Virus. *J. Virol*. **2020**, 94, e01818-19. doi:10.1128/JVI.01818-19.
  50. Monne, I.; Fusaro, A.; Nelson, M.I.; Bonfanti, L.; Mulatti, P.; Hughes, J.; Murcia, P.R.; Schivo, A.; Valastro, V.; Moreno, A.; et al. Emergence of a Highly Pathogenic Avian Influenza Virus from a Low-Pathogenic Progenitor. *J. Virol*. **2014**, 88, 4375–4388. doi:10.1128/JVI.03181-13.
  51. Engering, A.; Hogerwerf, L.; Slingenbergh, J. Pathogen–Host–Environment Interplay and Disease Emergence. *Emerg. Microbes & Infect.* **2013**, 2, 1–7. doi:10.1038/emi.2013.5.
  52. Laughlin, A.J.; Hall, R.J.; Taylor, C.M. Ecological Determinants of Pathogen Transmission in Communally Roosting Species. *Theor. Ecol*. **2019**, 12, 225-235. doi:10.1007/s12080-019-0423-6.
  53. Lam, T.T.-Y.; Ip, H.S.; Ghedin, E.; Wentworth, D.E.; Halpin, R.A.; Stockwell, T.B.; Spiro, D.J.; Dusek, R.J.; Bortner, J.B.; Hoskins, J.; et al. Migratory Flyway and Geographical Distance Are Barriers to the Gene Flow of Influenza Virus among North American Birds. *Ecol. Lett*. **2012**, 15, 24–33. doi:10.1111/j.1461-0248.2011.01703.x.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.