Ecogeographic drivers of the spatial spread of highly pathogenic avian influenza outbreaks in Europe and North America, 2016-2022

Jonathon D. Gass, Jr.[1], Nichola J. Hill[1], Elena N. Naumova[1], Felicia B. Nutter[1], and Jonathan A. Runstadler[1]

[1]Affiliation not available

October 27, 2022

Title: Ecogeographic drivers of the spatial spread of highly pathogenic avian influenza outbreaks in Europe and North America, 2016-2022

Authors: Jonathon D. Gass, Jr.[1,2], Nichola J. Hill[3], Elena N. Naumova[4], Felicia B. Nutter[1], Jonathan A. Runstadler[1]

[1] Department of Infectious Disease and Global Health, Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA, USA

[2] Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, USA

[3] Department of Biology, University of Massachusetts, Boston, Boston, MA, USA

[4] Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

Corresponding author email: Jonathon.gass@tufts.edu

Summary H5Nx highly pathogenic avian influenza (HPAI) viruses of clade 2.3.4.4. have caused outbreaks in Europe since 2016, and most recently these viruses were introduced to North America via avian migratory transport in December 2021. We sought to analyze the spatiotemporal extent of these viruses within the two continent outbreak system, and to characterize predictors of virus spread between geographic regions through a Bayesian phylodynamic generalized linear model (phylodynamic-GLM). Findings from this study reveal for the first time the geographic extent and directionality of the H5Nx HPAI virus outbreak system in Europe and North America during 2016-2022. Data demonstrate localized epidemics of H5Nx throughout Europe in the first several years of the epizootic, followed by a singular branching point where H5N1 viruses were introduced to North America via wild migratory birds, likely via stopover locations in the North Atlantic. Once in the US, H5Nx HPAI viruses spread at a greater rate between US-based regions along migratory flyways, and no evidence points to spread back to any European region. Our GLM demonstrated that geographic proximity is a predictor of virus spread between regions, which implies that inter-continental transport across the Atlantic is relatively rare and may require northward virus movement that coincides with spring migration of susceptible avian species to regions in the North Atlantic, possibly Iceland and Greenland. Finally, mean change in precipitation at destinations of viral transitions along phylogenetic branches was predicted to reduce H5Nx HPAI virus spread, which may reflect the effect of climate change on declines in host species abundance or changes in migratory patterns as a result of ecological alterations. Our data provide new knowledge about the spread and directionality of H5Nx HPAI virus dispersal in Europe and North America during an actively evolving outbreak, including predictors of virus movement between

regions, which will contribute to surveillance and mitigation strategies as the outbreak unfolds, and in future instances of uncontained avian spread of HPAI viruses.

Keywords: Influenza A virus, North Atlantic, wild birds, phylodynamic-GLM, outbreak, spatial spread

Introduction

Highly pathogenic avian influenza (HPAI) viruses of clade 2.3.4.4 (H5Nx) emerged in Southeast Asia in 2014 prior to spreading across much of Asia, Europe, North America, and Africa, causing frequent outbreaks and high rates of mortality in wild and domestic birds [1-8]. H5Nx viruses, which are descendants of A/Goose/Guangdong/1/1996(H5N1) (Gs/GD) first detected in 1996 in China, frequently reassort with other HPAI and locally endemic low pathogenic subtypes resulting in a constellation of novel reassortant virus lineages that have been isolated from a wide range of avian species [1]. Following their emergence in 2014, H5Nx viruses migrated with wild aquatic birds across the Pacific into North America later in the same year, causing outbreaks in wild and domestic birds until mid-2015 [4, 9]. Despite the substantial, though short-lived, spread of H5Nx (specifically H5N8, H5N1, H5N2) in North America from this original Pacific incursion event, H5Nx viruses have continuously circulated throughout Eurasia until present, with a gradual but steady northward trajectory of wild bird and virus movement in Europe observed between 2014-2021 [10, 11]. In December 2021, the first ever documented incursion of HPAI via the Atlantic route into North America was detected in St. John, Newfoundland and Labrador, Canada, and subsequently caused significant mortality among wild birds throughout the United States and Canada [12]. This novel introduction of H5Nx viruses into North America via the Atlantic route raises questions about factors that govern virus movement and spread within this multi-continent outbreak system, and how changes to these factors over time may have facilitated the first documented inter-hemispheric introduction of H5Nx viruses to North America in late 2021.

Environmental, ecological, and anthropogenic factors have been investigated as drivers of host and virus movement via Bayesian phylodynamics previously, including H9N2 in Asia, H5N1 in Egypt, multiple subtypes of Influenza A virus (IAV) among wildlife in North America, and Ebola virus during the 2013-2016 West African epidemic [13-18]. Northern temperate zones at the margins of the Atlantic Ocean are undergoing shifts in climate regimes (i.e. air temperature, precipitation, sea-surface temperature) due to global climate change (GCC), which has been linked to alterations in avian host ecology including migratory patterns, reproduction cycles, and trophic interactions [19-21]. Northward shifts in population distributions, for example, may increase the density of susceptible hosts for IAV infection year-round and the risk for more frequent inter-hemispheric virus movement via short-distance flights across the Arctic perimeter [20, 22-24]. Much remains unknown, however, about the impact of environmental factors on the dispersal of viral lineages, particularly during active epizootics of HPAI viruses at a multi-continental scale [21, 25-30].

This study combines Bayesian phylodynamic and generalized linear modeling (Phylodynamic-GLM) to uncover ecological and environmental predictors of H5Nx HPAI virus diffusion in Europe and North America between 2016 and 2022. We hypothesize that geographic proximity between regions and higher latitude are predictive of greater of H5Nx HPAI virus movement. Additionally, we project that changes in air temperature and precipitation will be predictive of greater virus movement due to ecological disturbance these changes may cause, resulting in avian hosts competing for suitable habitat and trophic resources. Uncovering environmental and ecological factors that predict HPAI virus dispersal will provide increased insights into how current and future ecosystem shifts may impact host and pathogen ecology and will demonstrate the importance of climate-aware surveillance and mitigation strategies.

Methods

H5Nx surveillance in Massachusetts In response to the introduction of H5N1 to Northeastern Canada in 2021 and subsequent outbreak among wild birds throughout North America from 2021 to present, our research group began conducting surveillance of wild birds for H5N1 in Massachusetts in collaboration with Tufts University Wildlife Clinic, multiple local wildlife rehabilitation centers, and through active surveillance in Nantucket and Monomoy, Massachusetts. From the March 1-18, 2022, viral swab specimens were obtained

from a) live birds using a single polyester-tipped swab of the cloaca or oral cavity and b) fecal material using a direct swabbing method in Massachusetts. All live birds were immediately released following completion of sampling. Each swab was immediately placed in individual cryovials containing 1.25 ml viral transport media [31]. Vials were stored in freezers between -20 ° and -80° C until analysis. Included in this analysis are 15 hemagglutinin (HA) sequences isolated from viral swabs from Great Horned Owl (*Bubo virginianus*) (n=1), Sanderling (*Calidris alba*) (n=8), Red-tailed Hawk (*Buteo jamaicensis*) (n=1), Herring Gull (*Larus smithsonianus*) (n=1), and American Crow (*Corvus brachyrhynchos*) (n=4) (Supplementary table 1).

Publicly available H5Nx sequence data and phylogenetic analysis

All publicly available avian-derived (domestic and wild) H5Nx HA segment sequences from Europe and North America between 2016 and 2022 were downloaded from the Influenza Research Database (IRD) [32] on May 12, 2022 (n=321). We added 170 publicly available H5Nx HA sequences from 2021-2022 downloaded from GISAID on May 15, 2022 as these were unavailable on IRD at the time of sequence acquisition, and 15 unpublished H5N1 HA sequences from avian surveillance in Massachusetts, USA by our research group in 2022 (described above), totaling 546 HA sequences. Metadata for each sequence was collected, including sampling date, season, host species, and geographic sampling location. Only IAV sequences from wild avian species or environmental matrices were included. Duplicate sequences, sequences with less than 75% unambiguous bases, all vaccine derivative and recombinant sequences, and sequences with unavailable isolation date, location, or host species were excluded, resulting in 506 sequences. Downsampling was performed to ensure relative evenness of geographic state groupings while preserving genetic diversity of the dataset, using geographic state and year for random stratification. To root and historically time-calibrate the tree, H5 subtype HA avian sequences from IRD were downloaded for the period 1979-2015 from Europe and North America and randomly downsampled by year, resulting in 33 historic sequences. These sequences were 'masked' to ensure their contribution to the tree structure but not to quantification of diffusion rates or the GLM [33]. The total downsampled dataset, including the outgroup (GISAID sequences from North America (n=170), unpublished Massachusetts sequences acquired by our group (n=15), publicly available H5Nx sequences from Europe 2016-2022 (n=162)), and historic sequences (n=33) resulted in a total of 380 sequences (Supplementary table 1). Multiple sequence alignments were performed using MUSCLE in Geneious Prime 2022.05.14 (https://www.geneious.com) and trimmed to the open reading frame.

Time-scaled Bayesian phylogenetic analyses

Bayesian molecular clock analyses were conducted using the Markov chain Monte Carlo (MCMC) method in BEAST v.1.10.4 [34] to construct time-scaled phylogenetic trees. Phylogenetic analyses implemented a Generalized Time-Reversible model (GTR) of nucleotide substitution [35] with a gamma plus invariant sites distribution of site heterogeneity (the Yang96 model [36]), a with a lognormal uncorrelated relaxed molecular clock [37], and a constant coalescent population model [38]. The BEAGLE library, which optimizes computational efficiency, was used [39]. Eight independent MCMC analyses were run for 200 million generations, sampled every 20,000 runs, and parameter convergence and effective sample size (ESS) (required to be >200) were evaluated in Tracer v.1.7.1 [40]. Using LogCombiner v.1.10.4, 10% or greater burn-in was removed from each run and independent runs were combined to establish the maximum clade credibility (MCC) tree, from which the last 500 trees from the posterior distribution were extracted and used as the empirical tree for all subsequent phylodynamic analyses [41]. Trees were visualized using Figtree v1.4.4 [42].

Discrete trait diffusion analyses between geographic regions

To infer significant discrete trait transition rates along phylogenetic tree branches of H5Nx subtype HA sequences between geographic regions, discrete Bayesian phylodynamic analyses were performed using BEAST v.1.10.4 [34]. We used an asymmetric substitution model with Bayesian stochastic search variable selection (BSSVS) and a strict clock model to estimate diffusion between discrete states [43]. Sampling locations were grouped into geographic state categories, based on modified National Oceanic and Atmospheric Administration (NOAA) historical climate regions (locations in the United States) and grouped countries by latitude (locations in Europe). Posterior inference of the complete Markov jump history through time was evaluated

by quantifying transitions between discrete states (Markov jumps, i.e. the frequency of transitions from one geographic state to another along phylogenetic branches) and the duration of time viruses spend in each discrete state (Markov rewards) [44].

Generalized linear models and empirical predictors

As an extension to the discrete trait diffusion models, GLMs were implemented to quantify and evaluate predictors of transitions between geographic states along phylogenetic branches. GLM models parameterize transition rates between discrete states as outcomes of a log-linear combination of matrixed covariate predictors [45]. Specifically, we modeled diffusion between geographic region states using a non-reversible continuous-time Markov chain (CTMC) process expressed as a K x K infinitesimal rate matrix of discrete state change (Λ) among K discrete states (geographic regions) [45]. The rate of transition from discrete state i to discrete trait j (Λij) is modeled through a linearized log function which incorporates all pairwise predictors (p1, . . . , pn) in the following equation:

???Λ?? = ?1?1 log(?1{??}) + ?2?2 log(?2{??}) + [?] + ???? log(??{??}),

where βi is the relative contribution of predictor pi to the whole GLM across the empirical phylogenetic tree space and δ is a binary indicator of a predictor's inclusion in the simulation [46]. A Bernoulli prior probability distribution was used to equally weight the probability that a given predictor would be included or excluded from the model [45]. The probability that a single parameter is included in the model is:

$qq = 1 - e\hat{}[ln(p)/n]$,

where p is the probability no coefficients are included (0.5 as default) and n is the number of total coefficients. Statistical support for diffusion among discrete states was determined by Bayes Factors (BF): 3 [?] BF < 20, 20 [?] BF <150, and BF [?] 150, denoting positive, strong, and very strong support, respectively [47]. BFs represent the odds of the posterior probability (pp) of a coefficient's inclusion in the model over its prior probability (qq):

$BF = [pp / (1 - pp)] / [qq / (1 - qq)]$,

where pp is calculated from BSSVS results and qq is calculated based on the prior assumption that there is a 50% probability that none of the coefficients are included in the model. In addition to BFs, 95% Highest Posterior Density (HPD) credible intervals were derived for each predictor, which provide information on the certainty of each parameter value. Posterior probabilities were calculated to demonstrate a predictor's inclusion in the model, and GLM coefficients provide conditional effect sizes for each predictor. All non-binary (i.e. not labeled as 0,1) continuous predictors were log-transformed and standardized prior to implementing the model in BEAST v.1.10.4 [34], therefore, a GLM coefficient of 1.0 can be interpreted as an increase of one transition per year for every one unit increase in the log-transformed predictor. Each GLM was performed with at least four independent MCMC runs, containing 200 million generations which were sampled and logged every 20,000 runs.

To inform the diffusion of H5Nx viruses between geographic regions, several environmental, ecological, and geographical predictors were selected. Predictors were selected following review of the literature regarding ecological and environmental factors that have been or are hypothesized to be associated with the movement and spread of IAV and other pathogens by wildlife [15, 16, 18, 48, 49]. We also included predictors not previously evaluated, including predictor value change through time (i.e. change in precipitation between years (mm)). These were selected based on hypotheses associated with the relationship between climate change and alterations in host-pathogen ecology of IAVs [20, 23, 50, 51]. Variables reflecting the relationship between regions (i.e. distance between centroids, shared borders) were included in the model once, whereas all other predictors were included twice, to measure the directionality of rates of virus transitions between geographic regions. For example, average precipitation at both the geographic region of origin and destination were included to determine whether precipitation at the region of origin or the region of destination was associated with the movement of H5Nx between geographic states. Continuous variables represent the average value for each geographic region and timeframe. All predictors are further described in Table 1.

Ethical approvals

This research, including surveillance screening, was conducted under the approval of the Tufts University IACUC protocol G2020-108, previously G2017-118. All swab samples are collected by outside investigators, wildlife professionals, or veterinarians who have approval through their institution or agency such as the National Oceanic and Atmospheric Administration (NOAA), US Department of Fish and Wildlife, or the United States Geological Survey (USGS). Samples obtained from wildlife clinics are collected as part of their internal diagnostics or outbreak response. As HPAI is a reportable disease, all samples that are screened positive are communicated to the USDA, State Veterinarian, Wildlife officials, the initial submitter, and the National Veterinary Services Laboratory (NVSL). Diagnostic samples are handled under enhanced Biosafety Level 2 conditions using additional PPE, disinfection, and containment protocols approved and permitted by the USDA. Samples that are identified as HPAI are transferred or destroyed within 21 days of confirmation as required by the USDA.

Results

Bayesian phylogeography of the H5Nx HPAI virus clade 2.3.4.4 outbreak in Europe and North America, 2016-2022

H5Nx viruses circulated in Europe for almost 5 years prior to a singular inter-continental introduction event of H5N1 from Central Europe to the United States (most likely via Canada, as detailed in the Discussion), specifically to South Carolina in the Midwest and Mid-Atlantic region (BF=76) (Figures 1 & 2). Prior to and following the inter-continental incursion of H5N1 to North America in late 2021, H5Nx viruses circulated within Europe, transitioning with the most intensity and highest statistical support from northern to central Europe (BF=47475). Central Europe served as a highly significant source region to both southern (BF=2152) and northern (BF= 14) regions. The southern European region acted as a sink of virus and not a source back to either other European region (Figures 1 & 2). H5Nx viruses circulated in Northern Europe for the greatest duration of time between 2016-2022 (32.8% of this time period), as measured by Markov rewards, which represent the mean proportion of time that viruses circulate in each geographic region during an outbreak, followed by Central Europe (30.0%), and Southern Europe (10.2%). H5Nx viruses continuously circulated in mainland Europe for 73.0% of the time between 2016 and February 2022 (Supplementary Table 2).

Following the inter-continental introduction to North America, H5N1 viruses spread rapidly between geographic regions in the United States. The Midwest and Mid-Atlantic region served as a significant source of virus to both the Upper Midwest (BF= 858) and Northern Rockies and Plains (BF=1526). Transitions along an east-west axis occurred, for example between the Upper Midwest and the Northern Rockies and Plains (eastward movement (BF= 9), westward movement (BF=12)), though to a lesser extent than overall transitions along a predominantly northward in direction south-north axis. Following approximately December 2021, H5N1 viruses circulated for the greatest duration of time in the Midwest & Mid-Atlantic USA (32.8%), 26.4% of time Northeastern USA, 21.9% of time in the Northern Rockies & Plains, and 18.9% of time in the Upper Midwest region. Of all geographic viral transitions between US regions (represented by Markov jumps, which are defined as the percent of regional transitions between regional states along phylogenetic tree branches), 37.1%, 33.1%, and 11.8% migrated from the Midwest & Mid-Atlantic to the Northern Rockies & Plains, Upper Midwest, and Northeast, respectively. All other transitions between US regions occurred at rates less than 7%. Between December 2021- May 2022, H5N1 viruses circulated and spread within the United States with no statistically supported transitions back to any European regions (Figures 1 & 2).

Generalized linear model of ecological and environmental predictors of H5Nx diffusion in Europe and North America, 2016-2022

Among continuous predictor variables used to inform the phylodynamic-GLM, several ranged widely between geographic regions, including distance between centroids (km), precipitation (mm), precipitation change (mm), and temperature (C), reflecting the variability in distances and latitudes between geographic regions.

Given this analysis was completed using all publicly available H5N1 HA sequences in the United States during an ongoing outbreak, downsampling the dataset to ensure relatively similar sequence counts by region was not possible. To account for this variability (sample size per region ranged from 33 to 80) sample size was included as a covariate predictor in the model to evaluate whether inclusion probabilities of other predictors were sensitive to the sample sizes in each geographic state (Table 2). The final model contained 27 predictor variables, six of which demonstrated to be statistically supportive of inclusion within the model of viral spread between geographic regions.

Two predictors related to geographic proximity, a) whether geographic regions share borders, and b) the distance between geographic centroids of each region, demonstrated very strong support for model inclusion, each with a BF score of 81657. Median coefficients for each predictor demonstrated that virus movement between regions was greater among regions with shared borders (median coefficient = 2.83) and less distance between geographic centroids (median coefficient = -2.42). Significantly greater virus movement (BF= 2626) was also measured when the destination of the viral diffusion transition was located in one of the four regions in the US. In fact, virus diffusion to any of the four US geographic regions resulted in an average increase of 2.77 transitions per year, revealing significantly greater viral spread in the United States following the December 2021 introduction of H5N1 viruses via wild bird migration to North America. Precipitation change at destination (BF= 15.88) demonstrated to be negatively correlated with H5Nx spread in Europe and North America, specifically that greater change in precipitation at the destination of a transition between geographic regions is associated with less virus migration. Two other environmental predictors, precipitation change at origin (BF=3.51), and NDVI value at origin during summer (BF= 3.76) were statistically supported for model inclusion, however 95% HPD intervals for each predictor crossed the null, demonstrating uncertainty in the direction of their effects. The 95% HPD interval's upper value for NDVI value at origin during summer only crosses the null by 0.13, indicating that this predictor is most likely negatively correlated with virus movement, however the lower and upper 95% HPD interval values for precipitation change at origin is too wide to make conclusions about the directionality of the effect. In addition, both predictors have low posterior probability values, suggesting these variables contribute far less than others to the outcome of the GLM. The additional 21 predictors included in the model failed to demonstrate statistical support with BFs > 3.0 (Figure 3, Supplementary table 4).

Discussion

Findings from this study provide novel data on the migration of H5Nx HPAI viruses among geographic regions within and between Europe and North America during 2016-2022 as well as ecological and environmental predictors of virus spread. Specifically, we found that greater spread was associated with virus migration to US regions and between geographically proximal regions, and virus migration was negatively correlated with precipitation change in the destination region of viral diffusion. Other variables, NDVI during summer and precipitation change in the source region of virus diffusion were included in the model, however the directionality of their effects was ambiguous. Given the unprecedented geographic scope of the outbreak among wild and domestic birds, as well as terrestrial and marine mammals, and the first detected introduction of H5N1 viruses of clade 2.3.4.4 to North America via the Atlantic route, these data provide important ecological context regarding factors predictive of virus spread within the outbreak system [1, 2, 52-54].

The first recorded introduction of H5N1 viruses from Europe to North America, most likely via long distance migratory birds, was detected following an unusual mortality event on an exhibition farm in Newfoundland and Labrador, Canada in December 2021. This was the first detection of H5Nx HPAI viruses in North America since the 2014-2015 outbreaks, spread by wild birds migrating across the Pacific route from east Asia [12]. Our Bayesian phylogeographic analysis of the HA segment of H5Nx HPAI viruses in Europe and North America establishes that a divergence event occurred sometime during 2020-2021, where North American viruses split from their most recent common ancestor with European lineage H5Nx viruses (Figures 1, Supplementary figure 1). Though H5N1 viruses were not detected in Newfoundland and Labrador until the end of 2021, circulation across the Atlantic may have been facilitated by a) the seasonal spring migration of Anseriformes (ex. Eurasian Wigeon, Barnacle Geese, or Greylag Geese) from mainland Europe to breeding

grounds in Iceland or Greenland, and b) the subsequent autumn migration of several gull species (ex. Greater Black-backed, Lesser Black-backed, and Black-headed Gulls), whose pelagic migratory patterns link these regions with Northeastern Canada [12, 55, 56]. Though a direct trans-Atlantic incursion is possible, it is far more probable that inter-species transmission events on breeding grounds in the North Atlantic, particularly to immune-naïve juvenile birds, enabled the necessary conditions for inter-continental spread [12, 24, 57]. Our phylogeographic analysis established that the first detected introduction to the United States occurred in South Carolina in the Midwest & Mid-Atlantic region in late December 2021. H5N1 was detected in an American Wigeon and American Blue-winged Teal (Supplementary table 1). Though the ancestral and epidemiological relationship between H5N1 isolates detected in Canada and South Carolina remain unknown, it is most likely that the virus was introduced to late-season birds migrating southward along the Atlantic flyway in late 2021, which then seeded populations of birds migrating northward in the subsequent spring, which is supported by our phylogeographic analysis depicted in Figure 2.

Geographic proximity has been previously shown to be predictive of virus movement between global regions [15, 16, 18]. Specifically, regions that share borders and less distance between geographic centroids both demonstrated to increase diffusion, with 2.8 and 2.4 times increases per year for every one unit increase in the log-transformed predictor value. Our findings support observations that short-distance transmission drives global spread of H5Nx HPAI and IAVs in general, and that long-distance trans-ocean or trans-continental virus movement is a less frequently detected ecologic phenomenon. Gravity models, in which infectious disease transmission is a function of population size and geographic distance, were first described for Influenza epidemics by Viboud, et al. in 2006, and later adopted by Dudas, et al. (2017) regarding the spread of Ebolavirus during the West African epidemic during 2013-2016 [18, 58]. Our finding that geographic proximity predicts viral spread of H5Nx HPAI follows the same logic, whereby the degree of HPAI viral migration between regions is governed by distance between regions and the population density and distribution of susceptible short- and long-distance migratory birds. That is to say, highly synchronized local epidemics occurred throughout mainland Europe during 2016-2021 with significant northward virus movement. This trend, along with timely interactions between infected and susceptible wild birds, may have facilitated the gradual migration of H5Nx HPAI viruses across the North Atlantic to North America.

Our findings also reveal an increase in virus dispersal when transitions transpired from any region to a geographic destination region in the US. Reasons for this could include the a) native species composition of North American birds, b) high rate of susceptibility of diverse avian species, or c) geographic size of the continent, which may encompass the entire annual cycle for many bird species [4]. There are roughly four times as many individual species of birds in North America as compared with Europe (2059 species versus 544 species) [59, 60]. While evidence is lacking with respect to the relationship between avian species diversity and IAV spread, recent data have demonstrated that transmission relies on ecologically divergent bird hosts, and taxonomic diversity is associated with differences in H5N1-associated wild bird mortality due between Europe and Asia [61, 62]. Increased virus diffusion when viruses transition to US regions may also relate to relative immune-naivety of avian species in North America versus Europe, given that H5Nx HPAI viruses did not circulate in the North American region for approximately 6 years between 2015 and 2021 [1]. Comparatively, the endemicity of H5Nx viruses in Europe could have created a cycle of largely immune species due to frequent exposure during 2016-2022, lessening opportunities for viral spread compared to the novel introduction event to North America and subsequent rapid spread. Phylogeographic data from the US also indicate far more south to north virus migration than east-west migrations in either direction. While data on interactions between migratory flyways in North America are often contradictory, one study demonstrated that IAV dispersal within flyways was up to 13 times greater within flyways than between flyways, which suggests that the predominant gradient of diffusion of IAVs transpires along the north-south axis of within-flyway migration in North America, which our findings support [15, 63].

Environmental factors, and their average change over time, have been shown to impact both host and pathogen ecology for a variety of wildlife species and infectious diseases [51, 62, 64, 65]. The role of precipitation changes in the destination of viral transitions, defined as the difference (in mm) in 2020 from the mean precipitation recorded during 1901-2000, was revealed by our GLM as a predictor of decreased geographic

viral spread of H5Nx HPAI viruses for the first time. It is important to note that our GLM was unable to distinguish between increased precipitation and decreased precipitation at the virus transition destination, therefore our discussion focuses on difference, rather than decrease or increase in mm of this predictor value between 2020 as compared with the mean value for the period of 1901-2000. Our findings, therefore, indicate that increased and/or decreased mean precipitation rates in destination regions are negatively correlated with H5Nx HPAI virus diffusion, however given predictions that global average annual precipitation will increase though the year 2100, our findings are most relevant in this context [66]. There are multiple explanations for this finding. First, increasing or decreasing precipitation may drive species declines altogether, due to habitat and dietary resource loss related to increased aridity or flooding [67]. Second, avian species may shift migratory strategy (timing, length of stay) to regions with more stable and less extreme environmental trends. Previous research has proposed that climate change, including increased precipitation and extreme weather events, will alter the distribution and migration behavior of IAV hosts, bring species in greater contact due to resource competition, which would in-turn increase transmission opportunities [20]. Third, it is possible that changes to mean precipitation in destination regions will decrease the persistence of IAVs in environmental matrices, like water, soil, and ice, which have been demonstrated to be seasonal sources of abiotic transmission in wetlands and breeding ranges [68-70]. It should be noted that environmental factors are unlikely to impact host and pathogen ecology in isolation, and interactions between precipitation and air temperature may be heterogenous between global regions, which adds complexity that our GLM is not designed to measure [71].

We plan multiple research avenues derived from the findings of this study. First, while our data present important findings regarding H5Nx HPAI diffusion in Europe and North America, we will next compare these findings to similar analytical output from the 2014-2015 outbreak in North America which was seeded by virus migration from Asia via the Pacific route. We expect that differences in species diversity and the rate of transmission between North American regions will elucidate important ecological differences between outbreaks seeded by Asian- versus European-origin migratory birds. This comparative analysis will provide novel insights into the ecology of HPAI outbreaks in North America, and will contribute to future predictive models and surveillance strategies. Second, we plan to develop an application that will integrate with public repositories of IAV sequence data to extract and publish ecological and environmental data associated with IAV sequences. For instance, in current public repositories, meta-data associated with sequences are most often limited to origin location of isolate, host species, sampling date, and subtype. We plan to develop an algorithm (similar to GeoBoost2, a natural language processing pipeline for location extraction of molecular sequences [72]) to extract from location (either location name or latitude-longitude coordinates), the following environmental metadata: elevation, mean air temperature in sampling season, mean precipitation in sampling season, NDVI value on date of sampling, mean wild avian population density, mean domestic avian farm density, among others. These data will be easily downloadable using GenBank accession IDs, and will facilitate ecological and environmental analyses associated with IAV movement, transmission, spread, persistence, among many other topics.

Limitations

This study does have limitations. First, the availability of published sequence data for H5Nx of clade 2.3.4.4 from Europe and North America is limited due to non-uniform practices related to publishing virus sequence data, among other reasons. As such, regional groupings were necessarily geographically vast, encompassing multiple countries (Europe) and States (US). For Europe, regional groupings were determined based on relative latitude, given the assumption that environmental factors will generally cluster by a gradient along a north-south axis. For the US, regional groupings were determined by modifying NOAA's US Climate Regions to ensure that all publicly accessible data from the US would be included in the model to maintain a reasonable sample size, while preserving climactic consistency. US regions of Northeast, Upper Midwest, and Northern Rockies and Plains were well-represented in terms of sample size and geographic coverage; however the Midwest & Mid-Atlantic zone includes States within three distinct climate zones. While these States generally lie within the same latitudinal range, the Midwest & Mid-Atlantic regional grouping may have limited the model's power regarding transitions involving this region or the GLM's predictors themselves.

Second, host immune defenses exert more selective pressures on HA than other non-external genes, therefore unmeasured non-ecological forces may have influenced our evolutionary phylogenetic reconstruction moreso than if we had chosen to model another gene. Third, Markov rewards, the duration of time viruses will circulate in a given region relative to all global regions, indicated that H5Nx HPAI viruses circulated in Europe for 73% of the time in the sampling timeframe of 2016-May 2022. Inversely, during 27% of the latter time period, viruses were circulating in the United States. While H5N1 was first detected in December 2021, these viruses could have been introduced prior to this date, however it is highly doubtful introduction occurred approximately 1.6 years prior (the equivalent of 27% of time during 2016-May 2022). We believe this underestimation of duration of virus circulation in Europe is due to our downsampling strategy, which attempted to preserve virus diversity while ensuring relative evenness of sequence counts per region. In reality, the vast majority of viruses circulating during this time period are doing so in Europe, therefore we could have downsampled the dataset relative to prevalence by region, however this was not possible given the dearth of available prevalence data by region in the North American outbreak. Fourth, given the ongoing nature of the outbreak, data included and findings from this study are limited to cross-section of time from January 2016 to May 2022. As the outbreak evolves beyond May 2022, new data will emerge and either support or alter study findings.

Conclusions

Findings from this study reveal for the first time the geographic extent and directionality of the H5Nx HPAI virus outbreak in Europe and North America during 2016-2022. Data demonstrate localized epidemics of H5Nx throughout Europe in the first several years of the epizootic, followed by a singular branching point where H5N1 viruses were introduced to North America via wild migratory birds. Once in the US, H5Nx HPAI viruses spread at a greater rate between US-based regions along migratory flyways, and no evidence points to spread back to any European region. Overall, our GLM demonstrated that geographic proximity is a predictor of virus diffusion between regions, which implies that inter-continental spread across the Atlantic is relatively rare and may require northward virus movement that coincides with spring migration of susceptible avian species to regions in the North Atlantic. Finally, mean change in precipitation at virus transition destinations was predicted to reduce H5Nx HPAI virus spread, which may reflect the effect of climate change on declines in host species abundance or changes in migratory patterns as a result of ecological alteration. Our data provide new knowledge about the spread and directionality of H5Nx HPAI virus dispersal in Europe and North America, including predictors of virus movement between regions, which will contribute to surveillance and mitigation strategies as the outbreak unfolds, and in future instances of uncontained avian spread of HPAI viruses.

Acknowledgements Thank you to Dr. Joseph Hicks, Lambodhar Damodaran, and Nidia Trovao for providing expert guidance on constructing the phylodynamic-GLM model in BEAST 1.4.10 and R. We appreciate all in the surveillance and research community who provided genomic sequence data included in this analysis.

Ethics Statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to and the appropriate ethical review committee approval has been received. This research, including surveillance screening, was conducted under the approval of the Tufts University IACUC protocol G2020-108, previously G2017-118, and in accordance with USDA policy.

Conflict of Interest Statement

All co-authors confirm there are no conflicts of interest to declare.

Data Sharing and Accessibility

All data supporting the findings of this study are openly accessible. Genetic sequences from IAV surveillance in North America and Europe are deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank) under accession numbers listed in Supplemental table 1. Raw data used for the analysis are deposited in the Dryad Digital Repository (https://doi.org/10.5061/dryad.m37pvmd2m).

References

1. Lee, D.H., et al., *Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4.* J Vet Sci, 2017. **18** (S1): p. 269-280.

2. Antigua, K.J.C., et al., *The Emergence and Decennary Distribution of Clade 2.3.4.4 HPAI H5Nx.* Microorganisms, 2019.**7** (6): p. 156.

3. Dhingra, M.S., et al., *Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3.4.4 viruses with spatial cross-validation.* eLife, 2016. **5** : p. e19571.

4. Hill, N., et al., *Reassortment of Influenza A Viruses in Wild Birds in Alaska before H5 Clade 2.3.4.4 Outbreaks.* Emerging infectious diseases, 2017. **23** (4): p. 654-657.

5. Lee, D.H., et al., *Reoccurrence of Avian Influenza A(H5N2) Virus Clade 2.3.4.4 in Wild Birds, Alaska, USA, 2016.* Emerg Infect Dis, 2017. **23** (2): p. 365-367.

6. Lee, D.-H., et al., *Novel Reassortant Clade 2.3.4.4 Avian Influenza A(H5N8) Virus in Wild Aquatic Birds, Russia, 2016.* Emerging Infectious Disease journal, 2017. **23** (2): p. 359.

7. Shin, J., et al., *Highly pathogenic H5N6 avian influenza virus subtype clade 2.3.4.4 indigenous in South Korea.* Scientific Reports, 2020. **10** (1): p. 7241.

8. Yamaji, R., et al., *Pandemic potential of highly pathogenic avian influenza clade 2.3.4.4 A(H5) viruses.* Reviews in Medical Virology, 2020. **30** (3): p. e2099.

9. Stallknecht, D.E., et al., *LIMITED DETECTION OF ANTIBODIES TO CLADE 2.3.4.4 A/GOOSE/GUANGDONG/1/1996 LINEAGE HIGHLY PATHOGENIC H5 AVIAN INFLUENZA VIRUS IN NORTH AMERICAN WATERFOWL.* Journal of Wildlife Diseases, 2020. **56** (1): p. 47-57, 11.

10. Pavon-Jordan D, *Waterbirds in a changing world: Effects of climate, habitat, and conservation policy on European waterbirds* , in*Doctoral dissertation* , U.o. Helsinki, Editor. 2017.

11. Verhagen, J.H., R.A.M. Fouchier, and N. Lewis, *Highly Pathogenic Avian Influenza Viruses at the Wild–Domestic Bird Interface in Europe: Future Directions for Research and Surveillance.* Viruses, 2021. **13** (2): p. 212.

12. Caliendo, V., et al., *Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021.* Scientific Reports, 2022. **12** (1): p. 11729.

13. Yang, J., et al., *Bayesian phylodynamics of avian influenza A virus H9N2 in Asia with time-dependent predictors of migration.* PLoS computational biology, 2019. **15** (8): p. e1007189-e1007189.

14. Yang, J., et al., *Inferring host roles in bayesian phylodynamics of global avian influenza A virus H9N2.* Virology, 2019.**538** : p. 86-96.

15. Hicks, J.T., et al., *Host diversity and behavior determine patterns of interspecies transmission and geographic diffusion of avian influenza A subtypes among North American wild reservoir species.* PLOS Pathogens, 2022. **18** (4): p. e1009973.

16. Magee, D., et al., *Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion.*Archives of virology, 2015. **160** (1): p. 215-224.

17. Hicks, J.T., et al., *Agricultural and geographic factors shaped the North American 2015 highly pathogenic avian influenza H5N2 outbreak.* PLOS Pathogens, 2020. **16** (1): p. e1007857.

18. Dudas, G., et al., *Virus genomes reveal factors that spread and sustained the Ebola epidemic.* Nature, 2017. **544** (7650): p. 309-315.

19. Morin, C.W., et al., *Avian influenza virus ecology and evolution through a climatic lens.* Environment International, 2018.**119** : p. 241-249.

20. Gilbert, M., J. Slingenbergh, and X. Xiao, *Climate change and avian influenza.* Revue scientifique et technique (International Office of Epizootics), 2008. **27** (2): p. 459-466.

21. Vandegrift, K.J., et al., *Ecology of avian influenza viruses in a changing world.* Annals of the New York Academy of Sciences, 2010.**1195** : p. 113-128.

22. Hinzman, L.D., et al., *Evidence and Implications of Recent Climate Change in Northern Alaska and Other Arctic Regions.* Climatic Change, 2005. **72** (3): p. 251-298.

23. Arctic Monitoring and Assessment Programme (AMAP), *AMAP Climate Change Update 2019: An Update to Key Findings of Snow, Water, Ice and Permafrost in the Arctic* . 2019, Arctic Monitoring and Assessment Programme (AMAP): Oslo, Norway.

24. Gass J.D., et al., *Global dissemination of Influenza A virus is driven by wild bird migration through arctic and subarctic zones*Authorea, 2022.

25. Herrick, K.A., F. Huettmann, and M.A. Lindgren, *A global model of avian influenza prediction in wild birds: the importance of northern regions.* Veterinary Research, 2013. **44** (1): p. 42.

26. Hall, J., et al., *Avian influenza virus prevalence in marine birds is dependent on ocean temperatures.* Ecological Applications, 2020. **n/a** (n/a).

27. Ferenczi, M., et al., *Avian influenza infection dynamics under variable climatic conditions, viral prevalence is rainfall driven in waterfowl from temperate, south-east Australia.* Veterinary Research, 2016. **47** (1): p. 23.

28. Bahl, J., et al., *Ecosystem Interactions Underlie the Spread of Avian Influenza A Viruses with Pandemic Potential.* PLOS Pathogens, 2016. **12** (5): p. e1005620.

29. Rogers, S.O., W.T. Starmer, and J.D. Castello, *Recycling of pathogenic microbes through survival in ice.* Medical Hypotheses, 2004.**63** (5): p. 773-777.

30. Kirby, J.S., et al., *Key conservation issues for migratory land- and waterbird species on the world's major flyways.* Bird Conservation International, 2008. **18** (S1): p. S49-S73.

31. Docherty, D.E. and P.G. Slota, *Use of muscovy duck embryo fibroblasts for the isolation of viruses from wild birds.* Journal of tissue culture methods, 1988. **11** (3): p. 165-170.

32. Bao, Y., et al., *The influenza virus resource at the National Center for Biotechnology Information.* J Virol, 2008. **82** (2): p. 596-601.

33. Hicks, J.T., et al., *Global phylodynamic analysis of avian paramyxovirus-1 provides evidence of inter-host transmission and intercontinental spatial diffusion.* BMC Evolutionary Biology, 2019.**19** (1): p. 108.

34. Suchard, M.A., et al., *Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.* Virus Evolution, 2018.**4** (1).

35. Tavaré, S. and R.M. Miura. *Lectures on mathematics in the life sciences* . in *Am. Math. Soc* . 1986.

36. Yang, Z., *Maximum-likelihood models for combined analyses of multiple sequence data.* Journal of Molecular Evolution, 1996.**42** (5): p. 587-596.

37. Drummond, A.J., et al., *Relaxed phylogenetics and dating with confidence.* PLoS biology, 2006. **4** (5): p. e88.

38. Drummond, A.J., et al., *Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data.* Genetics, 2002. **161** (3): p. 1307-1320.

39. Baele, G., et al., *High-Performance Computing in Bayesian Phylogenetics and Phylodynamics Using BEAGLE* , in *Evolutionary Genomics: Statistical and Computational Methods* , M. Anisimova, Editor. 2019, Springer New York: New York, NY. p. 691-722.

40. Rambaut, A., et al., *Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7.* Systematic Biology, 2018.**67** (5): p. 901-904.

41. Pagel, M., A. Meade, and D. Barker, *Bayesian Estimation of Ancestral Character States on Phylogenies.* Systematic Biology, 2004.**53** (5): p. 673-684.

42. Rambaut A., *FigTree v 1.4. 4.* 2018.

43. Lemey, P., et al., *Bayesian Phylogeography Finds Its Roots.*PLOS Computational Biology, 2009. **5** (9): p. e1000520.

44. Hong, S.L., et al., *Bayesian Phylogeographic Analysis Incorporating Predictors and Individual Travel Histories in BEAST.*Current Protocols, 2021. **1** (4): p. e98.

45. Lemey, P., et al., *Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2.* PLOS Pathogens, 2014. **10** (2): p. e1003932.

46. Kuo, L. and B. Mallick, *Variable Selection for Regression Models.* Sankhyā: The Indian Journal of Statistics, Series B (1960-2002), 1998. **60** (1): p. 65-81.

47. Kass, R.E. and A.E. Raftery, *Bayes Factors.* Journal of the American Statistical Association, 1995. **90** (430): p. 773-795.

48. Duchatel, F., B.M.d.C. Bronsvoort, and S. Lycett,*Phylogeographic Analysis and Identification of Factors Impacting the Diffusion of Foot-and-Mouth Disease Virus in Africa.* Frontiers in Ecology and Evolution, 2019. **7** (371).

49. Müller, N.F., G. Dudas, and T. Stadler, *Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations.* Virus Evolution, 2019.**5** (2).

50. Brown, V.L. and P. Rohani, *The consequences of climate change at an avian influenza hotspot.* Biology Letters, 2012. **8** (6): p. 1036-1039.

51. Harvell, D., et al., *Climate change and wildlife diseases: When does the host matter the most?* Ecology, 2009. **90** (4): p. 912-920.

52. Banyard, A.C., et al., *Detection of Highly Pathogenic Avian Influenza Virus H5N1 Clade 2.3.4.4b in Great Skuas: A Species of Conservation Concern in Great Britain.* Viruses, 2022. **14** (2): p. 212.

53. Shin, D.-L., et al., *Highly Pathogenic Avian Influenza A(H5N8) Virus in Gray Seals, Baltic Sea.* Emerging infectious diseases, 2019.**25** (12): p. 2295-2298.

54. Lewis, N.S., et al., *Emergence and spread of novel H5N8, H5N5 and H5N1 clade 2.3.4.4 highly pathogenic avian influenza in 2020.*Emerging Microbes & Infections, 2021. **10** (1): p. 148-151.

55. Wille, M., et al., *Reassortment of American and Eurasian genes in an influenza A virus isolated from a great black-backed gull (Larus marinus), a species demonstrated to move between these regions.* Arch Virol, 2011. **156** (1): p. 107-15.

56. Dusek, R.J., et al., *North Atlantic Migratory Bird Flyways Provide Routes for Intercontinental Movement of Avian Influenza Viruses.* PLOS ONE, 2014. **9** (3): p. e92075.

57. Bevins, S.N., et al., *Intercontinental Movement of Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4 Virus to the United States, 2021.* Emerging infectious diseases, 2022. **28** (5): p. 1006-1011.

58. Viboud, C., et al., *Synchrony, waves, and spatial hierarchies in the spread of influenza.* Science, 2006. **312** (5772): p. 447-51.

59. BirdLife International, *European Red List of Birds 2021* . 2021.

60. Cornell Lab of Ornithology. *Species of birds in North America* . 2021; Available from: https://www.birds.cornell.edu/home/.

61. Hill, N.J., et al., *Ecological divergence of wild birds drives avian influenza spillover and global spread.* PLOS Pathogens, 2022.**18** (5): p. e1010062.

62. Stallknecht, D.E., J.D. Brown, and D. Swayne, *Ecology of avian influenza in wild birds.* Avian influenza, 2008. **1** : p. 43-58.

63. Fourment, M., A.E. Darling, and E.C. Holmes, *The impact of migratory flyways on the spread of avian influenza virus in North America.* BMC evolutionary biology, 2017. **17** (1): p. 118-118.

64. van Dijk, J.G.B., et al., *Host and virus ecology as determinants of influenza A virus transmission in wild birds.* Current Opinion in Virology, 2018. **28** : p. 26-36.

65. Gallana, M., et al., *Climate change and infectious diseases of wildlife: Altered interactions between pathogens, vectors and hosts.*Current Zoology, 2013. **59** (3): p. 427-437.

66. Intergovernmental Panel on Climate, C., *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* . 2014, Cambridge: Cambridge University Press.

67. Rosamond, K.M., et al., *Not Singing in the Rain: Linking Migratory Songbird Declines With Increasing Precipitation and Brood Parasitism Vulnerability.* Frontiers in Ecology and Evolution, 2020.**8** .

68. Coombe, M., et al., *A SYSTEMATIC REVIEW AND NARRATIVE SYNTHESIS OF THE USE OF ENVIRONMENTAL SAMPLES FOR THE SURVEILLANCE OF AVIAN INFLUENZA VIRUSES IN WILD WATERBIRDS.* Journal of Wildlife Diseases, 2021. **57** (1): p. 1-18, 18.

69. Rohani, P., et al., *Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion.* Proceedings of the National Academy of Sciences, 2009.**106** (25): p. 10365-10369.

70. Hood, G., et al., *A literature review of the use of environmental sampling in the surveillance of avian influenza viruses.*Transboundary and Emerging Diseases, 2021. **68** (1): p. 110-126.

71. Tingley, M.W., et al., *The push and pull of climate change causes heterogeneous shifts in avian elevational ranges.* Global Change Biology, 2012. **18** (11): p. 3279-3290.

72. Magge, A., et al., *GeoBoost2: a natural languageprocessing pipeline for GenBank metadata enrichment for virus phylogeography.*Bioinformatics (Oxford, England), 2020. **36** (20): p. 5120-5121.

73. Yang, J., et al., *Bayesian phylodynamics of avian influenza A virus H9N2 in Asia with time-dependent predictors of migration.* PLOS Computational Biology, 2019. **15** (8): p. e1007189.

74. Dellicour, S., et al., *Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework.* Nature Communications, 2020. **11** (1): p. 5620.

Tables and Figures

Table 1. Environmental, ecological, and geographic predictors of virus diffusion in the Europe-North American system of H5Nx of clade 2.3.4.4

Predictor	**Justification**	**Value**	**Data source**
Distance between centroids	Decreased distance between states has been shown to relate to viral spread between geographic states in several phylogeographic-GLM models [48, 73]	Great circle distance in kilometers (km) between geographic state centroids	Google Earth
Latitude	Given many migratory avian species breed in northern latitudes, higher latitudes may increase global dispersal of IAVs due to transmission dynamics between adults and juveniles at breeding ranges, spreading virus globally [13, 15, 24]	Decimal latitude at exact centroid of geographic region	Google Earth
Shared borders	Shared borders have been implicated in the spread of viruses between both humans and animals due to geographic proximity and movement patterns [13]	Binary 0/1 (no/yes)	Google Earth
Northward movement	Northward movement has been associated with global spread of IAVs due to the condensed land masses around the circumpolar perimeter of the Arctic that connect hemispheres, particularly following breeding season [10, 11]	Binary 0/1 (no/yes) whether geographic state of origin is north of state of destination	Google Earth
Precipitation	Precipitation has been modeled in GLMs previously, with varying significance [13, 48]	Mean yearly precipitation in mm, 2020	NOAA. National Centers for Environmental Prediction North American Regional Reanalysis. https://www.esrl.noaa.gov/psd/

Predictor	Justification	Value	Data source
Change in precipitation	Change in precipitation has not been previously used as a predictor in standard GLMs, however it has been used in a circuitscape/Seraphim "skygrid-GLM" [74]	Difference between mean precipitation 1901-2000 compared with 2020	NOAA. National Centers for Environmental Prediction North American Regional Reanalysis. https://www.esrl.noaa.gov/psd/
Air temperature	Air temperature has been modeled in GLMs previously, with varying significance [13, 48]	Mean yearly air temperature in Celsius, 2020	NOAA. National Centers for Environmental Prediction North American Regional Reanalysis. https://www.esrl.noaa.gov/psd/
Change in air temperature	Change in air temperature has not been previously used as a predictor in standard GLMs, however only in a circuitscape/Seraphim "skygrid-GLM" [74]	Difference from mean temperature 1901-2000 compared with 2020	NOAA. National Centers for Environmental Prediction North American Regional Reanalysis. https://www.esrl.noaa.gov/psd/
Normalized difference vegetation index (NDVI) (Summer and Winter)	Origin summer normalized difference vegetation index (NDVI) has been positively associated with viral dispersal [15]	Mean NDVI value 2017, data values range from -0 to 1	Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MOD13A3) Version 6 [52], and https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_-browseByCountry.php.
Change in normalized difference vegetation index (NDVI)	Has not been previously modeled, but changes in land cover and vegetation may impact avian migratory ecology	Difference in mean NDVI value in 2017 from 1982	Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MOD13A3) Version 6 [52], and https://www.star.nesdis.noaa.gov/
Sample size	Commonly included in GLMs to account for different sample sizes by state, which can bias results. GLM assumes that the sample sizes across sub-populations are proportional to the subpopulation sizes [13]	Count of virus sequences from each geographic state	Downsampled dataset

Predictor	Justification	Value	Data source
Continental location of sequence: Europe North America	There may be geographic differences in virus diffusion due to land size, proximity to nearby regions, or composition of host diversity and abundance in given regions. This has not been modelled previously.	Binary (0/1)	Downsampled dataset meta-data

Figure 1. Markov Chain Monte Carlo (MCC) time-scaled phylogeographic tree of H5Nx Influenza A viruses (IAV) of clade 2.3.4.4. HA gene segments, color-coded by geographic source region. Tree with 95% Highest Posterior Density intervals featured in Supplementary figure 1.

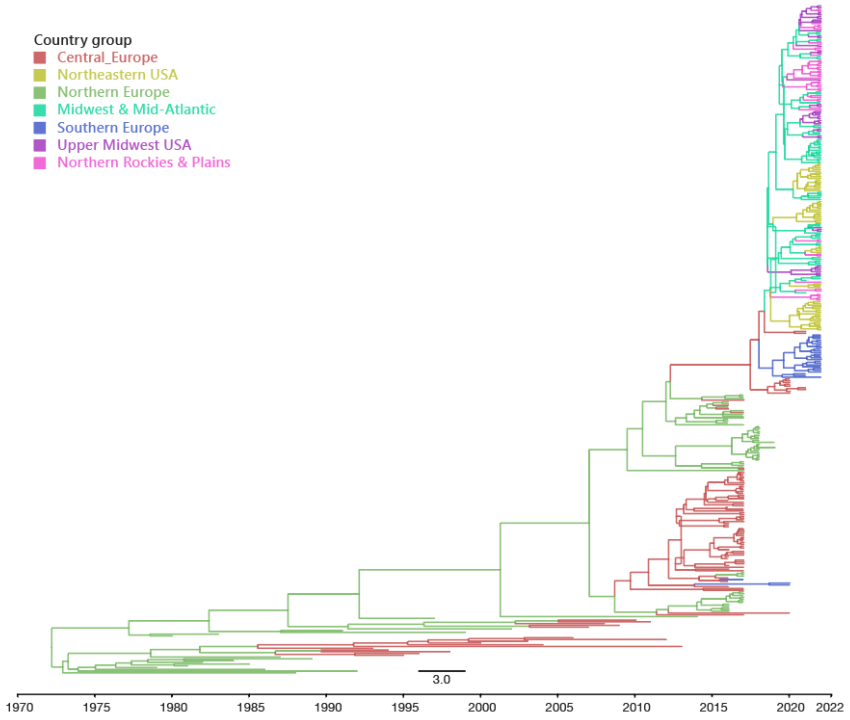


Figure 2. H5Nx circulation in Europe and North America during outbreaks between 2016 and 2022. Significant discrete phylogeographic transitions between regions are represented by arrows from Northern Europe to Central Europe (BF= 47575), Central Europe to Southern Europe (BF=2152), Midwest and Mid-Atlantic to Northern Rockies & Plains (BF=1526), Midwest and Mid-Atlantic to Upper Midwest (BF=858), Midwest & Mid-Atlantic to Northeast USA (BF= 692), Central Europe to Midwest &Mid-Atlantic (BF= 76), Central Europe to Northern Europe (BF=14), Upper Midwest USA to Northern Rockies & Plains (BF=12), and Northern Rockies & Plains to Upper Midwest USA (BF=9). Only BFs > 3 with corresponding posterior probability (PP) estimates > 0.5 are presented as statistically supported. Arrows signify directionality and greater arrow width corresponds to higher BF support of phylogeographic transitions between intracontinental geographic states. BFs and PPs for state transitions between all regions can be found in Supplementary Table 2. Map is not drawn to scale.

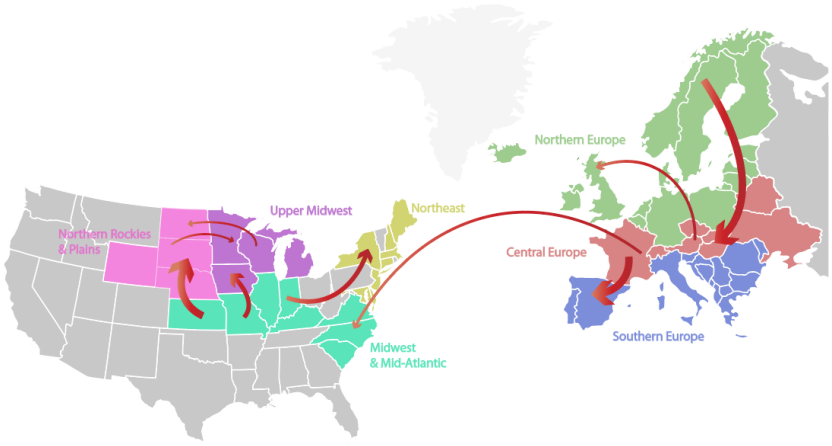
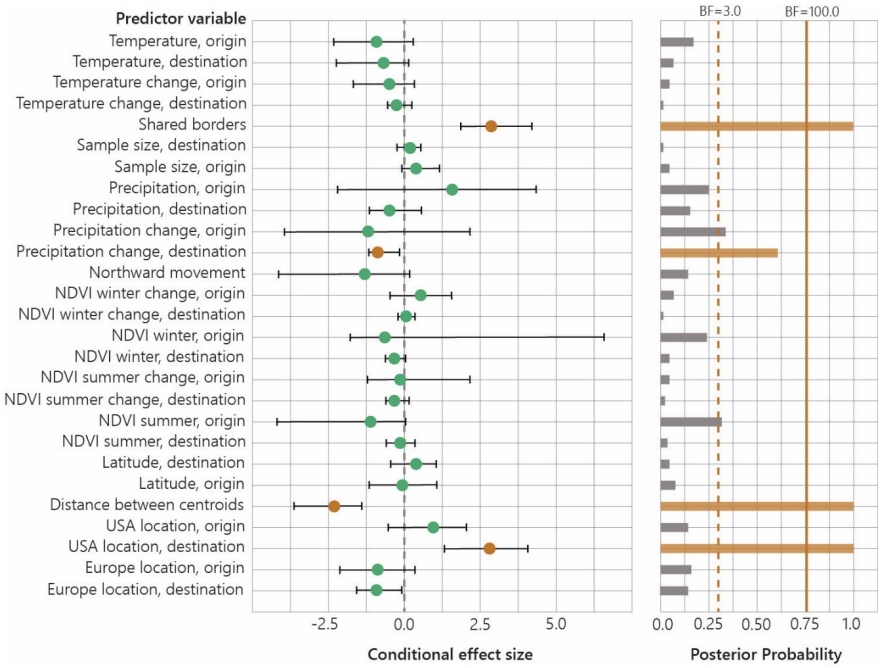


Table 2. Summary statistics of continuous predictor variables used to inform the Bayesian discrete diffusion generalized linear model describing H5Nx HPAI of clade 2.3.4.4 diffusion in Europe and North America, 2016-2022. Additional binary predictors not featured in the table include: shared borders, continent location USA (origin and destination), continent location Europe (origin and destination), and northward movement.

Variable name
Distance between centroids (km) Precipitation (mm) Precipitation change (mm) NDVI summer NDVI summer change NDV

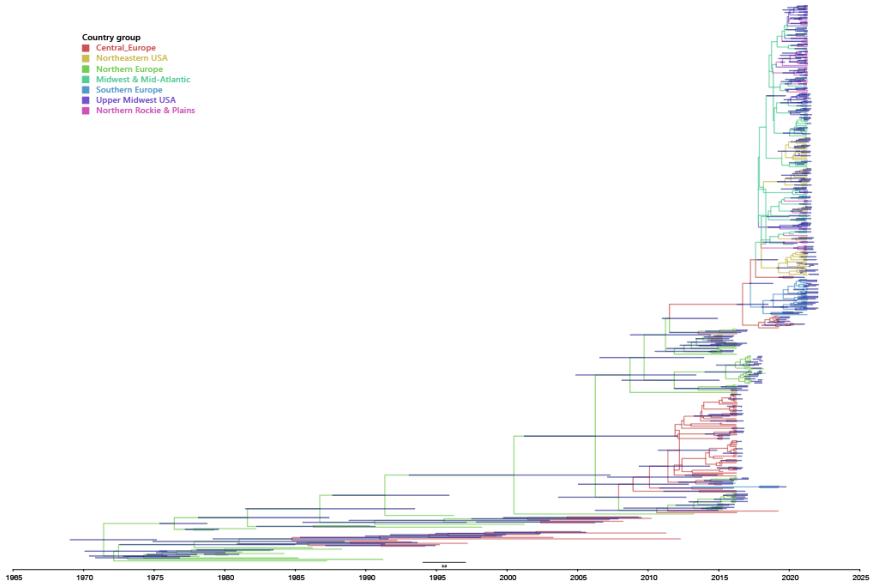
Figure 3. Predictors of migration of H5Nx HPAI viruses within and between Europe and North American regions during outbreaks, 2016-2022. In the left panel, conditional effect size (green circles) measures a predictor's effect on the rate of migration, and 95% Highest Posterior Density credible intervals (black bars through green circles) demonstrate certainty of the conditional effect. In the right panel, Posterior probability represents whether a predictor is included in the model, as determined by Bayes factors (BF) > 3.0 (orange dashed vertical line) and BF > 100 (solid orange vertical line). Conditional effect size circles and posterior probability bars associated with significant predictors of virus migration appear in orange.



Supplemental Materials

Supplementary table 1: Downsampled dataset of 380 H5Nx HA sequences, including sequences from H5Nx outbreak in North America, December 2021- May 2022 (n=170), unpublished Massachusetts sequences from outbreak in North America acquired by our group in May 2022 (n=15), publicly available H5Nx sequences from Europe 2016-2022 (n=162), and historic sequences (n=33) resulting in a total of 380 H5 HA sequences included in the analysis.

Supplementary figure 1: 95% Highest Posterior Density Tree of phylogeographic transitions between geographic regions.



Supplementary table 2: Statistically supported transitions between geographic regions. Only BFs > 3 with corresponding posterior probability (PP) estimates > 0.5 are presented as statistically supported (noted in bold italic).

ORIGIN	Northern Europe	Central Europe	Midwest & Mid-Atlantic	Midwest & Mid-Atlantic	Midwest & M

Supplementary Table 3: Markov rewards, by region. The mean proportion of time viruses spend within each global region between 2016-2022.

Region	Mean Markov Reward	% of total
Central Europe	225.57	30.0%
Northeast USA	53.08	7.1%
Northern Europe	246.46	32.8%
Midwest & Mid-Atlantic	66.71	8.9%
Southern Europe	76.63	10.2%
Upper Midwest USA	38.25	5.1%
Northern Rockies & Plains	44.82	6.0%
Total	751.52	100

Supplementary table 4. Predictors of H5Nx clade 2.3.4.4 virus diffusion within Europe and North America, 2016-2022. Variables with Bayes factors (BF) and posterior probabilities demonstrating statistical support are in bold. NDVI = Normalized difference vegetation index. HPD = Highest Posterior Density.

Variable name

Median coefficient

Lower 95% HPD

Upper 95% HPD

Posterior

probability

Bayes factor

Shared borders Distance between centroids Continent location USA destination Precipitation change destination NDVI summer origin Precipitation change origin Precipitation origin NDVI winter origin Precipitation destination Continent location USA origin Continent location Europe origin Northward movement Continent location Europe destination Temperature origin Latitude Sample size origin Temperature destination NDVI summer change origin NDVI winter destination Temperature change origin NDVI winter change origin Latitude destination NDVI summer destination Sample size destination NDVI summer change destination Temperature change destination NDVI winter change destination

2.836994468 -2.429007743 2.775413603 -0.70094621 -1.192317855 -1.032574784 1.600422557 -0.637124093 -0.516516735 0.83445501 -0.828740825 -1.348031447 -0.851190809 -1.079921952 -0.081684538 0.357703151 -0.626116589 0.089125815 -0.340697815 -0.56612304 0.496625062 0.344575298 -0.194016838 0.149744707 -0.290124094 -0.144364285 0.081708633

1.708556428 -3.640128621 1.365174183 -1.129137985 -4.088595203 -3.819073703 -1.955917441 -1.78780816 -1.028104295 -0.476028119 -1.954698269 -4.074666203 -1.748689222 -2.465237895 -1.080607284 -0.169982981

-1.266024604 -1.187185877 -0.808838284 -1.866089516 -0.40387291 -0.41823344 -0.56821217 -0.265220739 -0.72574173 -0.422896113 -0.232594075

4.10198173 -1.419502129 4.1987317 -0.24317806 0.137182122 2.304277683 4.470960891 6.602792845 0.591209359 1.900294218 0.410909015 0.240655393 -0.079186217 0.35274607 0.899454747 1.115345739 0.149792367 2.273894852 0.124650097 0.489636947 1.64596763 0.931233893 0.382439019 0.582553419 0.148756127 0.479178947 0.445216538

1 1 0.99690031 0.660533947 0.315768423 0.300869913 0.251274873 0.237476252 0.181481852 0.158884112 0.157984202 0.154284572 0.154084592 0.129187081 0.066093391 0.051894811 0.053594641 0.03029697 0.02859714 0.03439656 0.02729727 0.02609739 0.01439856 0.01219878 0.01209879 0.00489951 0.00369963

81657.95149 81657.95149 2626.225085 15.88902585 3.768461359 3.514141534 2.740470514 2.543110868 1.81052018 1.542492688 1.532116891 1.489692825 1.487410204 1.211414322 0.577900492 0.446957149 0.380836011 0.255128473 0.24039294 0.239557652 0.229159342 0.218816482 0.119293345 0.1008429 0.100006196 0.033111553 0.030322603