VIRUS EVOLUTION

Virus Evolution, 2022, **8(2)**, 1–10

DOI: https://doi.org/10.1093/ve/veac073
Advance access publication date: 18 August 2022
Research Article

Disentangling the role of poultry farms and wild birds in the spread of highly pathogenic avian influenza virus in Europe

Claire Guinat,[1,2,*,†] Cecilia Valenzuela Agüí,[1,2,‡] Timothy G. Vaughan,[1,2,§] Jérémie Scire,[1,2] Anne Pohlmann,[3,**] Christoph Staubach,[3] Jacqueline King,[3,††] Edyta Świętoń,[4,‡‡] Ádám Dán,[5] Lenka Černíková,[6] Mariette F. Ducatez,[7,§§] and Tanja Stadler[1,2,***]

[1]Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse, Basel 4058, Switzerland, [2]Swiss Institute of Bioinformatics, Quartier Sorge, Lausanne 1015, Switzerland, [3]Friedrich-Loeffler-Institut, Suedufer 10, Greifswald – Insel Riems 17489, Germany, [4]Department of Poultry Diseases, National Veterinary Research Institute, Al. Partyzantow 57, Pulawy 24-100, Poland, [5]DaNAm Vet Molbiol, Herman Ottó utca 5, Kőszeg 9730, Hungary, [6]State Veterinary Institute Prague, Sidlistni 136/24, Prague 165 03, Czech Republic and [7]IHAP, Université de Toulouse, INRAE, ENVT, 23 chemin des capelles, Toulouse 31076, France

[†]https://orcid.org/0000-0002-8245-9290
[‡]https://orcid.org/0000-0002-2774-4822
[§]https://orcid.org/0000-0001-6220-2239
[**]https://orcid.org/0000-0002-5318-665X
[††]https://orcid.org/0000-0003-1857-7660
[‡‡]https://orcid.org/0000-0002-1571-3279
[§§]https://orcid.org/0000-0001-9632-5499
[***]https://orcid.org/0000-0001-6431-535X
*Corresponding author: E-mail: claire.guinat@bsse.ethz.ch

Abstract

In winter 2016–7, Europe was severely hit by an unprecedented epidemic of highly pathogenic avian influenza viruses (HPAIVs), causing a significant impact on animal health, wildlife conservation, and livestock economic sustainability. By applying phylodynamic tools to virus sequences collected during the epidemic, we investigated when the first infections occurred, how many infections were unreported, which factors influenced virus spread, and how many spillover events occurred. HPAIV was likely introduced into poultry farms during the autumn, in line with the timing of wild birds' migration. In Germany, Hungary, and Poland, the epidemic was dominated by farm-to-farm transmission, showing that understanding of how farms are connected would greatly help control efforts. In the Czech Republic, the epidemic was dominated by wild bird-to-farm transmission, implying that more sustainable prevention strategies should be developed to reduce HPAIV exposure from wild birds. Inferred transmission parameters will be useful to parameterize predictive models of HPAIV spread. None of the predictors related to live poultry trade, poultry census, and geographic proximity were identified as supportive predictors of HPAIV spread between farms across borders. These results are crucial to better understand HPAIV transmission dynamics at the domestic–wildlife interface with the view to reduce the impact of future epidemics.

Key words: phylodynamics; highly pathogenic avian influenza; virus transmission and evolution; domestic–wildlife interface.

Introduction

Since the beginning of the 21st century, the highly pathogenic avian influenza (HPAI) H5N8 virus (clade 2.3.4.4b) represents one of the most serious threats to animal health, wildlife conservation, and livestock economic sustainability. In June 2016, the virus was detected in wild birds in regions of Central Asia (at the Ubsu-Nur and Qinghai lakes, known as migration stopovers) and subsequently spread to other Asian countries and Europe (Napp et al. 2018). By the end of 2017, the virus had caused one of the most severe epidemics in Europe in terms of the number of poultry outbreaks, wild bird cases, and affected countries (Napp et al. 2018). Most of the poultry outbreaks occurred in France ($n = 420$, 37.8 per cent), followed by Hungary ($n = 239$, 21.5 per cent), Germany ($n = 94$, 8.5 per cent), Poland ($n = 69$, 5.8 per cent), and the Czech Republic ($n = 43$, 3.9 per cent) (Napp et al. 2018). While no human cases were observed, the control strategies that were implemented in the affected countries resulted in the culling of several million poultry, causing devastating socioeconomic impacts for the poultry industry.

The emergence of H5N8 virus in Europe was likely attributable to infected migratory wild birds from Northern Eurasia, leading to occasional or multiple viral incursions into poultry farms (Lycett et al. 2016; Beerens et al. 2017; Fusaro et al. 2017; Globig et al. 2018; Mulatti et al. 2018; Świętoń and Śmietanka 2018). After emergence, farm-to-farm transmission likely occurred during the epidemic, with contact with infected poultry and contaminated fomites,

such as vehicles or equipment, being a major risk factor for farm infection (Globig et al. 2018; Mulatti et al. 2018; Napp et al. 2018; Świętoń and Śmietanka 2018; Andronico et al. 2019; Guinat et al. 2020b). In a number of cases, high poultry density and substantial gaps in farm biosecurity were also identified as potential risk factors for farm infection (Globig et al. 2018; Mulatti et al. 2018; Napp et al. 2018; Guinat et al. 2020a). The possibility of airborne transmission between poultry farms was also suggested, without being conclusively demonstrated (Guinat et al. 2018; Scoizec et al. 2018). Moreover, there is evidence that wild birds were another source of farm infection (Beerens et al. 2017; Napp et al. 2018; Świętoń and Śmietanka 2018).

While such epidemiological and phylogenetic studies have generated important clues about the H5N8 virus transmission patterns in Europe, they remained opaque to the specific role of poultry farms and wild birds in disease spread. In particular, understanding the viral transmission dynamics among these two subpopulations is crucial to determining which of these two has the greatest potential to drive the viral transmission during epidemics, which, in turn, represents critical information to better target control strategies. When appropriate pathogens' genetic and epidemiological data are collected, phylodynamic methods can fill this critical gap (Volz, Koelle, and Bedford 2013; du Plessis and Stadler 2015; Guinat et al. 2021). By fitting population dynamic models to genetic sequences collected during epidemics, these tools aim at quantifying disease transmission dynamics and have been particularly used to study the spread of infectious diseases in structured populations, be they stratified by time, species, or geography (Dudas et al. 2018; Faria et al. 2018; Nadeau et al. 2021). Importantly, birth–death model-based approaches (Kühnert et al. 2016) explicitly allow for the direct estimation of key epidemiological parameters, such as the effective reproduction number R_e (which captures the number of secondary infections generated at any time during an epidemic in a partially immune population) (Anderson and May 1979), while taking into account the sampling effort.

Using a phylodynamic framework, this study aimed at disentangling the role of poultry farms and wild birds in the spread of H5N8 in Europe during the 2016–7 epidemic. We fitted a phylodynamic model with geographical and host structure to H5N8 genome sequences of the Hemagglutinin (HA) segment collected from both host types (190 from poultry farms and 130 from wild birds) in four severely affected European countries (Czech Republic, Germany, Hungary, and Poland) to (i) estimate the early patterns of virus spread; (ii) infer the number of unreported infections; (iii) provide R_e estimates; (iv) discriminate the number of new infections arising from local transmission versus importation events; and (v) identify factors driving virus spread between farms across borders.

Results

We fitted a multi-type birth–death (MTBD) model to the aligned sequences (Kühnert et al. 2016; Scire et al. 2020) to co-infer epidemiological parameters and phylogenetic trees (Vaughan et al. 2014). The MTBD model was structured into five discrete subpopulations according to the host type and geographical location, referred to here as demes: 'poultry farms in the Czech Republic', 'poultry farms in Germany', 'poultry farms in Hungary', 'poultry farms in Poland', and 'wild birds in the four countries'. From a subsampled set of phylogenetic trees and model parameters inferred with the MTBD analysis (Vaughan, 2022), we simulated epidemic trajectories (i.e. the number of newly infected hosts per deme over time, due to within-deme and between-deme transmissions (Vaughan et al. 2019) (see the section 'Methods').

Early patterns of H5N8 virus spread in Europe

The maximum clade credibility (MCC) tree reconstructed using the MTBD model is shown in Fig. 1. Sequences from poultry farms of the same country for Germany, Hungary, and Poland were generally clustered together in the tree, while sequences from poultry farms for the Czech Republic were more scattered in the tree, as were wild birds' sequences. To inform the timeline of the first virus introduction, we inferred the date of the first virus introduction per trajectory and per deme for comparison with the date of the first officially reported infection per deme (FAO 2021) (Fig. 2). Overall, the inferred dates of the first virus introduction were before the date of the first officially reported infection in each deme, with a higher delay in the Czech Republic deme (median: 118 days, 95 per cent High Posterior Density (HPD): 63–156, i.e. approximately 16 weeks) compared to the other demes (from median: 20 days, 95 per cent HPD: 3–48 to median: 49.5 days, 95 per cent HPD: 11–83, i.e. approximately 3–7 weeks) (SI Appendix Table S1).

Number of unreported H5N8 infections

To assess the reporting rate during the epidemic, we inferred the number of poultry farms and wild birds that became noninfectious per trajectory and per deme (following death, culling, or recovery of the poultry flock/wild bird) and compared this quantity to the number of officially reported infections per deme (FAO 2021) (Fig. 3). The cumulative number of officially reported poultry farm outbreaks (94 for Germany, 240 for Hungary, and 65 for Poland) were within the inferred 95 per cent HPD (median: 84, 95 per cent HPD: 28–984 for Germany; median: 406.5, 95 per cent HPD: 197–987 for Hungary; and median: 98, 95 per cent HPD: 25–350 for Poland). More discrepancies were observed for poultry farms in the Czech Republic and wild birds in the four countries, with the cumulative number of officially reported infections (43 for the Czech Republic and 372 for wild birds) being outside the inferred 95 per cent HPD (median: 139, 95 per cent HPD: 45–289 for the Czech Republic and median: 4222, 95 per cent HPD: 1,112–13,713 for wild birds).

Key epidemiological parameters of H5N8 virus spread

Figure 4A and B shows the posterior distributions for the within-deme and between-deme R_e values, respectively, together with the prior (SI Appendix Table S2) for comparison. The within-deme R_e was estimated across four time intervals, corresponding to the four phases of the epidemic (SI Appendix Fig. S1). Note that only one sequence was available per poultry farm and per wild bird, implying that the within-deme R_e represents the farm-to-farm/wild bird-to-wild bird virus transmission and the between-deme R_e represents the cross-species and cross-country virus transmission (i.e. farm-to-wild bird/wild bird-to-farm/farm-to-farm across countries). For most demes, the median within-deme R_e posteriors were greater than or close to 1 during the first time period but decreased throughout the subsequent time periods (Fig. 4A, SI Appendix Table S3). However, these R_e estimates slightly increased again during the fourth time period (Feb–May 2017) in Germany, Hungary, and Poland. The highest median R_e estimates were observed between poultry farms in Hungary and between wild birds in the four countries during the first time period (Oct–Nov 2016). Overall, the between-deme R_e estimates were much lower than the within-deme R_e estimates, with extremely low values (median within the range of 10^{-3}–10^{-2}) for those representing farm-to-farm across countries

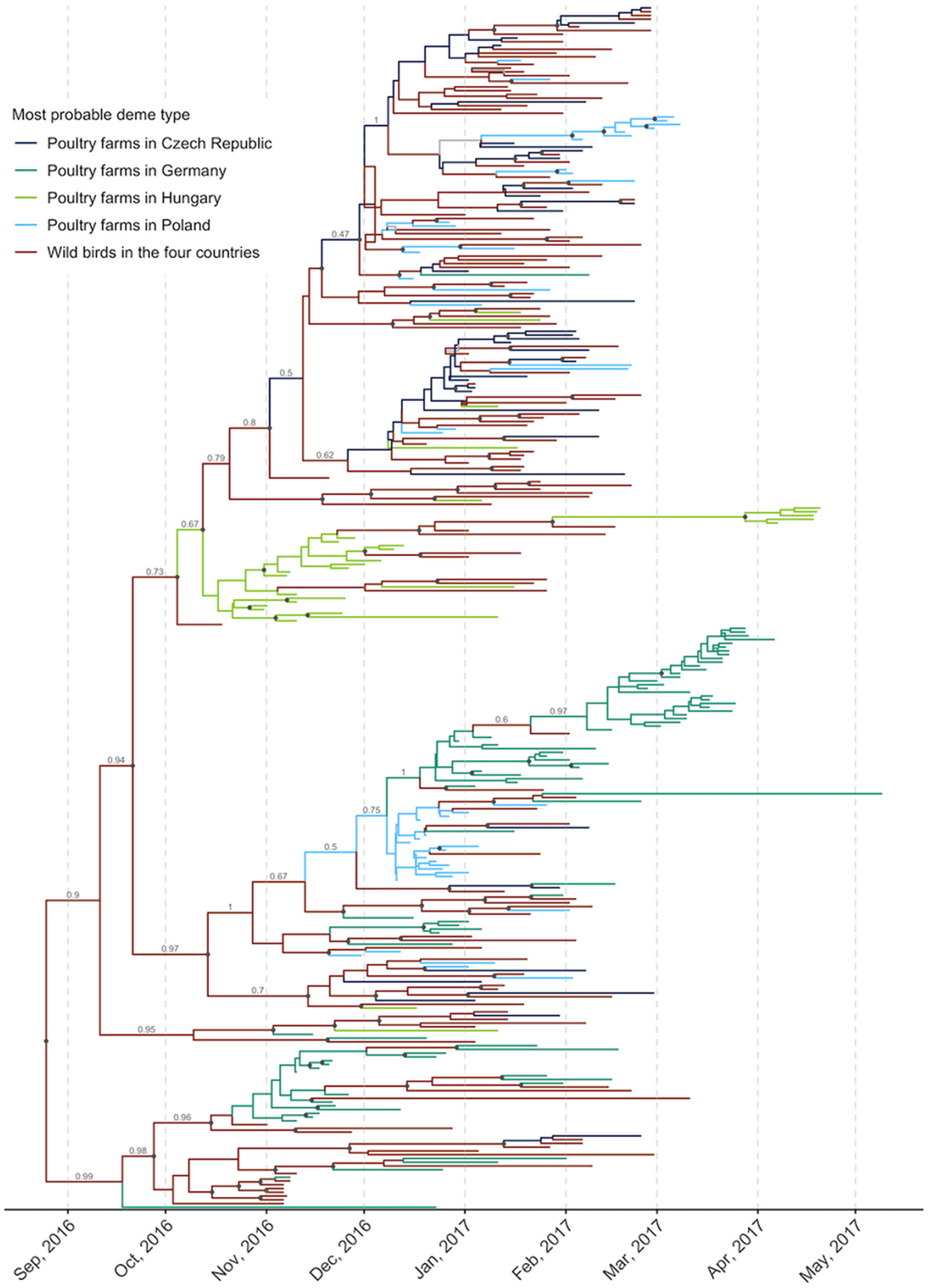

Figure 1. Time-scaled MCC phylogenetic tree of the HA segment of HPAI H5N8 virus sequenced from poultry farms and wild birds during the 2016–7 epidemic in the Czech Republic, Germany, Hungary, and Poland. The color of the tree branches shows the deme type with the highest probability (see legend). Circles at internal nodes indicate clade posterior probabilities above 0.50. For selected nodes, numbers show the posterior probabilities of the most probable deme type. There is evidence for virus spread among neighboring poultry farms illustrated by the presence of the clusters of H5N8 sequences from poultry farms of the same country (mainly Germany, Hungary, and Poland) in the tree, with the possibility of wild birds' movements facilitating virus spread between poultry farms across countries, illustrated by the dispersal distribution of H5N8 sequences from wild birds.

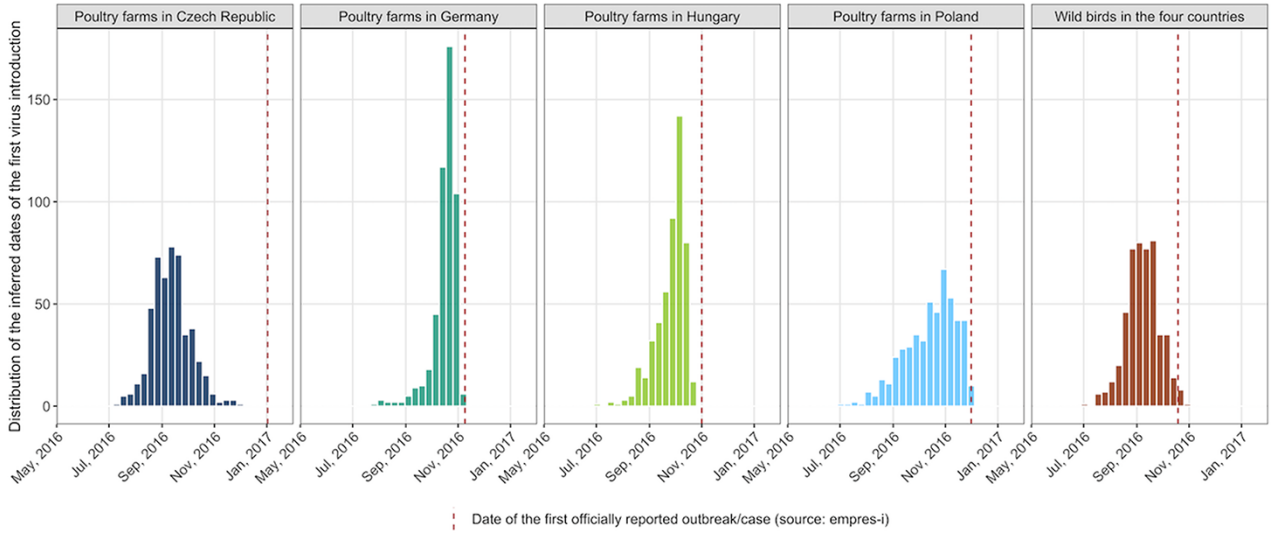

Figure 2. Temporal distribution of the inferred dates of the first virus introduction per deme. The dashed line represents the date of the first officially reported infection per deme for comparison. In this graph, for each trajectory and each deme, we extracted the date of the first virus introduction event and summarized them over time.

Figure 3. Temporal distribution of the inferred cumulative number of infections per deme (log scale). The solid line represents the median inferred number, and the colored areas represent the 95 per cent HPD. The dashed line represents the cumulative number of officially reported infections (log scale). In this graph, for each trajectory and each deme, we extracted the cumulative number of infections and summarized them over time in log scale.

and wild bird-to-farm transmission (Fig. 4B, SI Appendix Table S3). Slightly higher values were found for those representing farm-to-wild bird transmission (median of 0.1–0.4). One exception was found for the between-deme R_e estimate representing farm-to-wild bird transmission in the Czech Republic, with a median of 4.6 (95 per cent HPD: 0.9–9.2). The infectious period was also inferred for each deme, with the highest median found for poultry farms in the Czech Republic (median: 14 days, 95 per cent HPD: 7–23) and wild birds in the four countries (median: 14 days, 95 per cent HPD: 11–19) (SI Appendix Fig. S2 and Table S3). The infectious period was slightly lower for poultry farms in Germany (median: 10, 95 per cent HPD: 6–16), Hungary (median: 8 days, 95 per cent HPD: 4–14), and Poland (median: 7 days, 95 per cent HPD: 4–11). Using different priors on the R_e did not qualitatively change the results (SI Appendix Figs S3–S5).

Number of local H5N8 transmission versus importation events

To better target control strategies, we inferred the number of infections arising from transmission within the same deme (local transmission) as opposed to transmission from another deme (imported transmission) per deme. Figure 5 illustrates the temporal distribution of the inferred median number of local transmission and imported transmission events per deme. In Germany, Hungary, and Poland, the epidemic was dominated by farm-to-farm transmission events (median: 113, 95 per cent HPD: 52–4,284; median: 316, 95 per cent HPD: 137–790; and median: 79, 95 per cent HPD: 31–255, respectively) with an increase around March 2017, November 2016, and December 2016, respectively (SI Appendix Table S4). In the Czech Republic, the epidemic was dominated by wild bird-to-farm transmission events (median: 116,

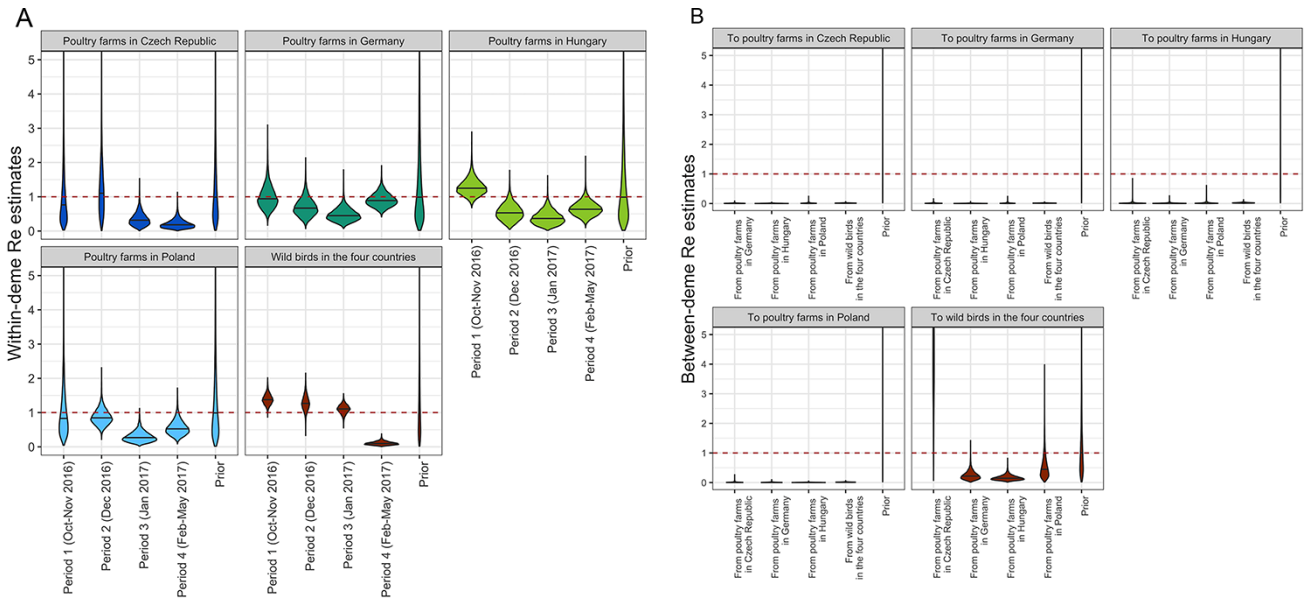

Figure 4. (A) Posterior distributions for the within-deme R_e values across four time intervals. Solid horizontal lines represent median values, and the dashed line represents the threshold between epidemic growth and fade out. (B) Posterior distributions for the between-deme R_e values. Solid horizontal lines represent median values, and the dashed line represents the threshold between epidemic growth and fade out.

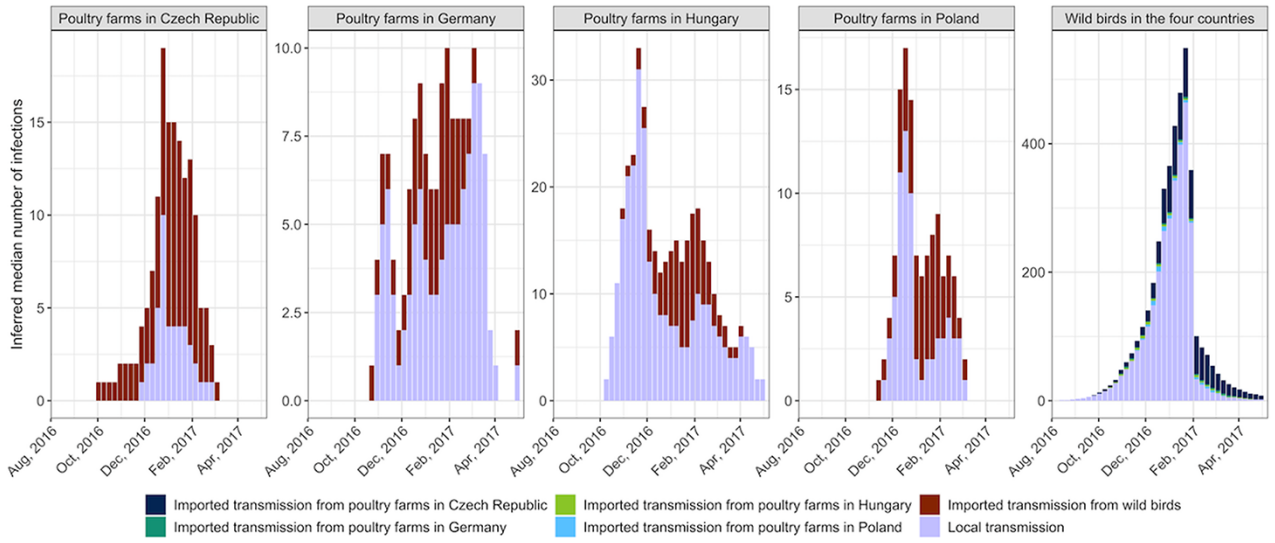

Figure 5. Temporal distribution of the inferred median number of infections arising from transmission within the same deme (local transmission) as opposed to transmission from another deme (imported transmission) per deme. In this graph, for each trajectory and each deme, we computed the median number of within-deme and between-deme transmission events over time.

95 per cent HPD: 55–218). For all countries, an increase in the number of wild bird-to-farm transmission events was observed around January–February 2017. There were a limited number of transmission events between poultry farms across countries. The epidemic in wild birds was also dominated by wild bird-to-wild bird transmission events (median: 3,237, 95 per cent HPD: 869–9,034), and the highest number of imported transmission events came from poultry farms in the Czech Republic (median: 935, 95 per cent HPD: 66–4,495).

Predictors of H5N8 virus spread between poultry farms across borders

Alongside inferring the transmission dynamics of H5N8, potential drivers of virus spread between poultry farms across the four countries were investigated by quantifying the corresponding R_e parameter in the MTBD model with a generalized linear model (GLM) (Lemey et al. 2014; Müller, Dudas, and Stadler 2019) (see the 'Methods' section). There were eight predictors included in the model: the 2016 live poultry trade (FAOSTAT, 2016), the 2016 poultry density in the source and destination deme (FAOSTAT, 2016), the 2014 poultry farm density in the source and destination deme (EFSA, European Centre For Disease Prevention And Control, European Union Reference Laboratory For Avian Influenza et al. 2017), the 2017 farm outbreak density in the source and destination deme (FAO 2021), and the distance between countries' centroids (SI Appendix Table S5). Figure 6A shows, for each predictor, the inclusion probability, which represents the proportion of the posterior samples in which the given predictor was included in the model, and the Bayes Factor (BF), which quantifies which of the

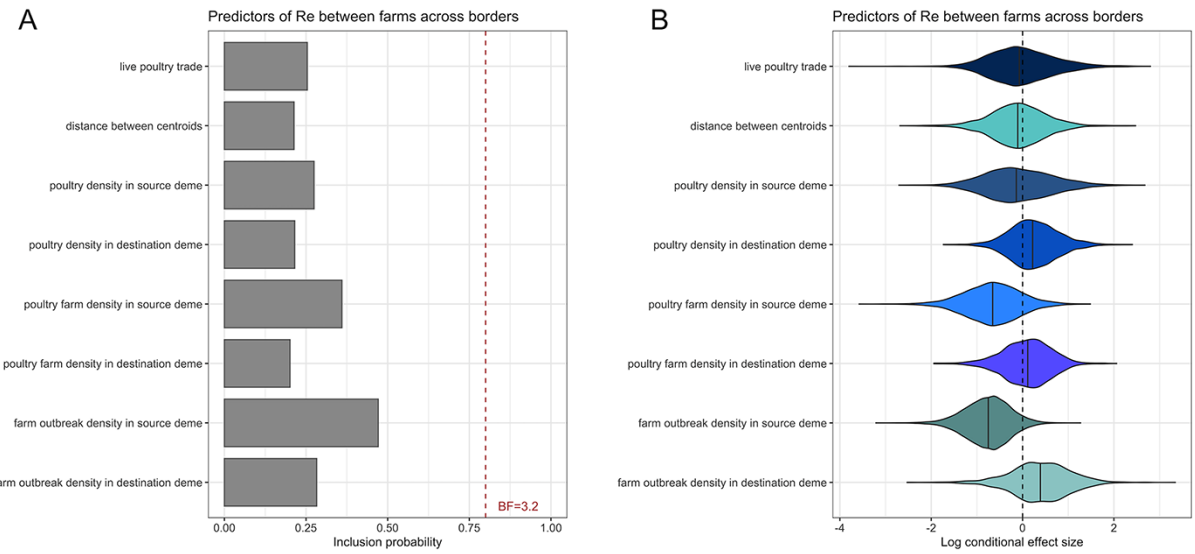

Figure 6. (A) Inclusion probability for predictors of the between-farm virus spread across borders. This represents the proportion of the posterior samples in which each predictor was included in the model. BFs were used to determine the contribution of each predictor in the GLM. BFs were calculated for each predictor to quantify which of the posterior and prior inclusion probabilities of the given predictor in the model is more likely. The cutoff for substantial contribution of a given predictor in the GLM was set at 3.2. (B) Log conditional effect sizes for predictors of the between-farm virus spread across borders. This represents the (log) contribution of each predictor when the corresponding predictor was included in the model ($\beta_i|\delta_I = 1$), where β_i is the coefficient and the binary indicator δ_I for each predictor i.

posterior and prior inclusion probabilities of the given predictor in the model is more likely. Figure 6B shows the log conditional effect size, which represents the log contribution of the given predictor when it was included in the model. None of the predictors were statistically supported to be associated with the spread of H5N8 virus between poultry farms across borders, illustrated by the low BF metric (<3.2) (Fig. 6A) and the similar distribution between the posterior coefficient estimates (Fig. 6B) and the prior (SI Appendix Table S2).

Discussion

In each country, we showed that the first introduction of H5N8 virus from wild birds to poultry farms likely occurred during autumn, which is in line with the timing of the arrival of migratory wild birds in Europe (BTO 2017). In addition, we estimated a delay of 3–16 weeks (depending on the country) between the inferred date of the first virus introduction and the date of the first officially reported poultry farm outbreak, likely illustrating the effectiveness of different surveillance strategies. The longest delay (16 weeks) was observed in the Czech Republic, where most outbreaks occurred in small-sized farms (<100 birds), while they mainly affected large-sized farms (>10,000 birds) in Germany, Hungary, and Poland (Napp et al. 2018). While a total of 442 poultry farm outbreaks and 372 wild bird cases in the four countries were officially reported (FAO 2021), we showed that these official numbers of detections could have been under-reported, especially in the wild bird population, likely due to challenges related to wildlife surveillance (Artois et al. 2009). High reporting rates of poultry farm outbreaks were found in Germany, Hungary, and Poland, likely linked to the high mortality rates of poultry following H5N8 virus infection, along with the active surveillance implemented around reported poultry farm outbreaks (EFSA et al. 2017). However, lower and again more delayed reporting rates were found for the poultry farms in the Czech Republic. These results suggest that the likelihood of reporting infected farms is likely associated with the characteristics of the farm. However, whether this is due to differences in farm size or other factors linked to the farm size (such as different farmers' knowledge, attitudes, and practices) needs further investigation.

Following the first virus introduction, we demonstrated that in Germany, Hungary, and Poland, the epidemic was dominated by local farm-to-farm transmission events. In Germany, local farm-to-farm transmission increased between February and May 2017, likely illustrating the cluster of turkey farm outbreaks which represented approximately 25 per cent of the total number of poultry farm outbreaks in the country (EFSA et al. 2017). In Hungary, a peak in the number of farm-to-farm transmission events was reported between October and November 2016, during which most outbreaks clustered in time and space (Napp et al. 2018). Moreover, the epidemic in these countries was also partly driven by wild bird-to-farm transmission, particularly in the middle of the epidemic, showing that the role of wild birds was not limited to the onset of the epidemic. These outcomes also emphasize that in-place biosecurity measures in Germany, Hungary, and Poland were sufficient to prevent continued incursions from farms across borders (such as a ban on international trade) (EFSA et al. 2017) but were less effective against local farm-to-farm and wild bird-to-farm transmission. Having more detailed knowledge of how poultry farms are connected with one another in those countries could help contain future outbreaks by disrupting the network of potential transmissions between poultry farms. Important efforts are also necessary to ensure that prevention strategies aiming at limiting the virus spread between wild birds and poultry (such as restriction of outdoor access and providing indoor feed and drinking water) (Artois et al. 2009) are implemented during high-risk periods. In addition, more sustainable strategies should be explored for poultry farms for which access to outdoor areas is part of the production specifications. The contribution of wild birds to poultry farm outbreaks was even more substantial in the Czech Republic, in which we showed that the epidemic was dominated by wild bird-to-farm transmission events. Accordingly, we also showed that the majority of farm-to-wild bird transmission events were from the Czech Republic. This provides evidence that

small-sized farms could be more exposed to virus transmission from wild birds than large commercial farms. Again, this could be explained by differences in farm size or other factors linked to it (such as different farming practices, access to outdoor areas, or biosecurity levels), which requires further attention. In wild birds, the epidemic was dominated by wild bird-to-wild bird transmission events. The number of wild bird-to-wild bird transmission events, however, has decreased drastically since February 2017, likely linked to a decrease in wild bird density with migration to warmer climates (Hill et al. 2016) and a decrease in virus survival in the environment due to the temperature-dependence of H5N8 virus transmission (EFSA et al. 2017).

We also attempted to uncover factors that could potentially predict the spread of H5N8 virus between farms across countries. However, none of the investigated predictors were identified as a supportive predictor of the viral spread. This could be explained by the low number of transmission events between poultry farms across countries (Fig. 5). This is in line with outbreak investigations on affected poultry farms in Europe, which showed that the likelihood of H5N8 virus introduction from one country to another via personnel contacts, trade of live poultry, feed, or poultry products was negligible (Lycett et al. 2016), although unreported cross-border activities could not be excluded. Using phylodynamic approaches, one previous study has found geographic proximity, sharing borders, and live poultry trade (when using time-dependent predictors) to be strong drivers of avian influenza (AI) virus spread between countries in Asia (Yang et al. 2019). The comparison of these previous GLM results to our study may not be appropriate due to differences in farming systems between Europe and Asia. In addition, our predictors ignore other potential drivers of virus spread, such as wild bird migration, different farming systems, and biosecurity levels among countries. For example, the scattered distribution of H5N8 sequences from wild birds among sequences from poultry farms of different countries on the MCC tree could support the possibility of wild birds' movements facilitating virus spread between poultry farms across countries (Fig. 1). It is also possible that transmission between countries is linked to the trade of poultry products or other cryptic means that were not tested in this study due to a lack of information. In the future, we recommend further investigation of predictors with a higher scale of temporal and spatial resolutions, which could allow for stronger contribution levels (Yang et al. 2019).

The 2016–7 epidemic of H5N8 virus in Europe remains, like other epidemics of AI viruses, epidemiologically complex as it involves multiple wild bird species that vary in spatial ecology and clinical disease severity. During the epidemic, the virus was detected in a large number of wild bird species, mainly those of the *Anseriformes* orders (ducks, geese, and swans), including mute swans (*Cygnus olor*), tufted ducks (*Aythya fuligula*), Whooper swans (*Cygnus cygnus*), Eurasian widgeons (*Mareca penelope*), and mallards (*Anas platyrhynchos*) (EFSA et al. 2017). Among these species, some can be mostly sedentary in given areas while partially or wholly migratory in others (BTO 2017), implying that some species can act as sentinels in some areas or long-distance vectors of H5N8 virus in others (Keawcharoen et al. 2008). Consequently, wild bird population structure may be much more complex than what was assumed in this study. For example, on the MCC tree, we observed H5N8 sequences from wild birds both within and between clusters of sequences from poultry farms of the same country, which could illustrate the presence of sedentary and migratory wild birds, respectively. Similarly, the virus was detected in several poultry species and farm types, which may play different roles in the virus spread due to discrepancies in virus infection susceptibility and farming practices (EFSA et al. 2017). Unfortunately, limited information on virus prevalence or epidemiology in various domestic and wild host species between countries makes it difficult to treat species separately, thereby necessitating the grouping used here.

Bayesian phylogeographic approaches (Lemey et al. 2009) are more common than structured phylodynamic approaches (like the MTBD model) to infer the transmission of lineages between different host species, their popularity being partly associated with their computational efficiency (Trovão et al. 2015; Lycett et al. 2016). However, one shortcoming of Bayesian phylogeographic approaches is the assumption of independence between the phylogeny and the transmission process, which can lead to loss of information (De Maio et al. 2015; Müller, Rasmussen, and Stadler 2017; Bloomfield et al. 2019). Another shortcoming is the assumption of proportionality between the sample sizes across subpopulations and the subpopulation sizes, which make it sensitive to biased sampling. Unlike Bayesian phylogeographic approaches, MTBD models explicitly integrate how lineages transmit within and between sub-populations while accounting for the sampling effort, making the estimations more robust to sampling bias (Kühnert et al. 2016; Scire et al. 2020). This has made it possible to infer transmission parameters, such as the effective reproduction number R_e, among poultry farms and wild birds based on pathogens' genome sequences. We expect those estimates being useful to parameterize predictive models of virus spread aiming at optimizing control strategies. We also inferred that the median farm-level infectious periods ranged from 7 to 14 days, suggesting that some countries were quicker at depopulating than others. This also emphasizes that a back-tracing window of approximately 2 weeks would be sufficient to capture the period during which a farm was infectious. Only one sequence was available per poultry farm, meaning that within-farm genetic diversity was not taken into account. However, this is a reasonable assumption due to the short period of the poultry farm outbreaks prior to detection and culling. More importantly, the present study demonstrates how relevant these models can be (i) to inform on the number of unreported infections, (ii) to reconstruct previous unobserved infections prior to the first officially reported infection, and (iii) to discriminate transmission events within a given host species from incursions across species that are more challenging using traditional wildlife and epidemiological methods (Blanchong et al. 2016; Guinat et al. 2021). Therefore, such phylodynamic tools can complement or even substitute for traditional epidemiological tools.

Phylodynamics provides one avenue for quantifying patterns and identifying drivers of infectious disease transmission dynamics at the wildlife–domestic animal interface, which is a fundamental challenge for veterinary epidemiology. We expect that our results will be valuable in better informing policy decision-making as means to reduce the impact of future epidemics of HPAI viruses.

Methods

Selection and alignment of sequences

H5N8 genome sequences (of HA segment) collected during winter 2016–7 from four severely affected European countries (Czech Republic, Germany, Hungary, and Poland) were downloaded from GISAID on 1 September 2020. Information about the sample collection and the sequencing techniques used to generate these sequences can be found in previous phylogenetic studies (Nagy et al. 2018; Pohlmann et al. 2018; Świętoń and Śmietanka 2018; Śmietanka et al. 2020). Only one sequence was available per poultry farm and per wild bird, meaning that the transmission

dynamics of H5N8 were inferred at the farm-to-farm, wild bird-to-farm, and farm-to-wild bird levels. Sequences were annotated with available sampling dates, locations and hosts, aligned using MAFFT v7 (Katoh and Standley 2013) and manually edited using AliView v1.26 (Larsson 2014). The dataset consisted of 190 poultry farm sequences and 130 wild bird sequences (SI Appendix Table S6 and Fig. S6).

Phylodynamic analysis

Multi-type birth–death model

The MTBD model was used as tree prior (Kühnert et al. 2016; Scire et al. 2020). Similar to compartmental models in epidemiology, the MTBD model involves the partitioning of the population into discrete subpopulations according to their properties, referred to here as deme. Sequences were organized into five demes, according to the host type and geographical location: 'poultry farms in Czech Republic', 'poultry farms in Germany', 'poultry farms in Hungary', 'poultry farms in Poland', and 'wild birds in the four countries' (SI Appendix Table S6). All sequences from wild birds were aggregated into one deme (not depending on the geographical location as for poultry farms) since it was assumed that the majority of sampled wild bird species (mainly mallards and swans) could move freely among countries (Atkinson et al. 2006). Under the MTBD model, infected hosts (poultry farms or wild birds) from a given deme could transmit the virus to another host from the same deme (with a parameter within-deme R_e), eventually become noninfectious due to recovery or death/depopulation (with a rate δ), be sequenced and sampled upon becoming non-infectious (with a proportion s, and thus are included into the dataset), or could transmit the virus to another host from another deme (with a parameter between-deme R_e). All transmissions become noninfectious and sampling processes were assumed to be deme-specific and constant through time, except for the within-deme R_e that was assumed constant across four time intervals, corresponding to the four phases of the epidemic (SI Appendix Fig. S1). It was assumed that once sampled, a given host could not be infected and sampled again since infected poultry farms were subject to culling following the confirmation of infection and sampling of wild birds was from a mortality event (EFSA et al. 2017).

The prior values and distributions of the MTBD model parameters are described in SI Appendix Table S2. The MTBD was combined with a HKY + Γ4 nucleotide substitution process with a relaxed molecular clock (Drummond et al. 2006) defined by a Log-normal(0.001, 1.25) prior (Beerens et al. 2017; Fusaro et al. 2017; Alarcon et al. 2018). The origin of the tree was given a Lognormal(–0.2, 0.2) distribution prior, corresponding to the median date 1 July 2016 (95 per cent HPD: 6 February 2016–19 October 2016) (Beerens et al. 2017; Fusaro et al. 2017; Świętoń and Śmietanka 2018) and assumed to be associated to the deme 'wild birds in the four countries', since the source of the first poultry farm outbreak in the four countries was likely attributed to infected migratory wild birds from Northern Eurasia (EFSA et al. 2017). All R_e parameters were given a Lognormal(0, 1) distribution prior (Iglesias et al. 2011; Grear et al. 2018; Andronico et al. 2019). The become non-infectious rate was given a Lognormal(52, 0.6) distribution prior (Grear et al. 2018; Leyson et al. 2019; Willgert et al. 2020). For each deme, the sampling proportion was given a uniform distribution prior with lower and upper bounds informed by the number of sequences and reported poultry farm outbreaks/wild bird cases (FAO 2021). Given the severity of the clinical signs affecting the majority of poultry combined with active surveillance around reported poultry farm outbreaks (EFSA et al. 2017), the number of unreported poultry farm outbreaks was considered relatively low in all countries. On the contrary, given the difficulty of catching and sampling wild birds, it was assumed that infected wild birds were significantly under-sampled, relative to poultry farms.

Predictors of H5N8 virus spread between poultry farms across borders

The MTBD model was extended with a GLM to inform the H5N8 virus spread between poultry farms across borders by 19 time-independent predictors (Lemey et al. 2014; Müller, Dudas, and Stadler 2019): the 2016 live poultry trade (FAOSTAT, 2016), the 2016 poultry density in the source and destination deme (FAOSTAT, 2016), the 2014 poultry farm density in the source and destination deme (EFSA et al. 2017), the 2017 farm outbreak density in the source and destination deme (FAO 2021), and the 2021 human density in the source and destination deme (Wikipedia 2021), whether two countries shared borders and the distance between countries' centroids. To account for potential missing predictors, we also included predictors to assess the virus spread from or to one individual country (SI Appendix Table S5). In this GLM parametrization, the between-deme R_e parameters act as the outcome to a log-linear function of the predictors. For each predictor i, the GLM parametrization also includes a regression coefficient β_i which quantifies the (log) contribution of the predictor and a binary indicator variable δ_i which quantifies the probability of the predictor to be included in the model (SI Appendix Table S2). To avoid collinearity among predictors, predictors were removed when the Pearson correlation exceeded >0.7 (SI Appendix Fig. S7). To reduce the effect of different predictors' magnitude, all non-binary predictors were log-transformed and standardized before inclusion in the GLM. BFs were used to determine the contribution of each predictor in the GLM (Kass and Raftery 1995; Lemey et al. 2014; Magee et al. 2015). BFs were calculated for each predictor to quantify which of the posterior and prior inclusion probabilities of the given predictor in the model ($\delta_i = 1$) is more likely. The cut-off of the substantial contribution of a given predictor in the GLM was set at 3.2 (Kass and Raftery 1995), implying that its posterior inclusion probability in the model was 3.2-fold more likely than its prior inclusion probability (0.50).

Inference of model parameters, structured trees, and epidemic trajectories

Phylodynamic analysis was implemented using the BDMM-Prime package (Vaughan, 2022) for BEAST v2.6.3 (Bouckaert et al. 2014) and the BEAGLE library (Ayres et al. 2012) to improve computational performance. Posterior phylogenetic trees and parameters of the nucleotide substitution model, of the relaxed clock model, of the MTBD model, and of the GLM were estimated using Markov chain Monte Carlo chains. All analyses were run for 40–50 million steps across three independent Markov chains and states were sampled every 10,000 steps. The first 10 per cent of steps from each analysis were discarded as burn-in before states from the chains were pooled using Log-Combiner v2.6.3 (Bouckaert et al. 2014). Convergence was assessed in Tracer v1.7 (Rambaut et al. 2018) by ensuring that the estimated sampling size values associated with the estimated parameters were all >200.

The structured phylogenetic trees (i.e. when phylogenetic trees are associated with a specific deme along their branches (Vaughan et al. 2014)) were inferred by applying a stochastic mapping algorithm (Freyman and Höhna 2019) implemented in BDMM-Prime (Vaughan, 2022) to a subsampled set of posterior phylogenetic

trees and model parameters ($n = 500$) generated by the MTBD analysis. The MCC tree was obtained from the structured trees in TreeAnnotator v2.6.3 (Bouckaert et al. 2014) and annotated using the ggtree package (Yu et al. 2017) in R v4.0.2 (Team 2013). The epidemic trajectories (i.e. the number of newly infected hosts per deme over time, due to within-deme and between-deme transmission (Vaughan et al. 2019)) were simulated from a subsampled set of posterior phylogenetic trees and model parameters ($n = 500$) generated by the MTBD analysis (Vaughan, 2022).

To test the robustness of the phylodynamic analysis with respect to changes in the R_e priors, a separate set of analyses were performed using broader and tighter priors on the within-deme R_e (Lognormal(0,2) or Lognormal(1,1)) and between-deme R_e (Unif(0,5)).

Data availability

All H5N8 genome sequences of HA segment are available in the GISAID database (https://www.gisaid.org). The prior values and distributions of the model parameters are described in SI Appendix Table S2. Details on the predictor data are available in SI Appendix Table S5. The BEAST 2 XML file used to perform the phylodynamic analysis, together with the accession numbers of the genome sequences and the R scripts are available from https://github.com/ClaireGuinat/h5n8_bdmm-prime.git.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

The authors are grateful to the cEvo group (ETH Zurich, Switzerland) for providing useful comments on this project. They also acknowledge the EFSA Public Access to documents Team for providing data on the poultry farm census in Europe. They also thank all submitting laboratories for sharing H5N8 genome sequences in the GISAID database. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 842621.

Conflict of interest: The authors declare no competing interests.

References

Alarcon, P. et al. (2018) 'Comparison of 2016–17 and Previous Epizootics of Highly Pathogenic Avian Influenza H5 Guangdong Lineage in Europe', *Emerging Infectious Diseases*, 24: 2270.

Anderson, R. M., and May, R. M. (1979) 'Population Biology of Infectious Diseases: Part I', *Nature*, 280: 361–7.

Andronico, A. et al. (2019) 'Highly Pathogenic Avian Influenza H5N8 in South-west France 2016–2017: A Modeling Study of Control Strategies', *Epidemics*, 28: 100340.

Artois, M. et al. (2009) 'Outbreaks of Highly Pathogenic Avian Influenza in Europe: The Risks Associated with Wild Birds', *Revue Scientifique Et Technique de l'OIE*, 28: 69.

Atkinson, P. W. et al. (2006) 'Urgent preliminary assessment of ornithological data relevant to the spread of avian influenze in Europe'. Wetlands International.

Ayres, D. L. et al. (2012) 'BEAGLE: An Application Programming Interface and High-performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 61: 170–3.

Beerens, N. et al. (2017) 'Multiple Reassorted Viruses as Cause of Highly Pathogenic Avian Influenza A(H5N8) Virus Epidemic, the Netherlands, 2016', *Emerging Infectious Diseases*, 23: 1974–81.

Blanchong, J. A. et al. (2016) 'Application of Genetics and Genomics to Wildlife Epidemiology', *The Journal of Wildlife Management*, 80: 593–608.

Bloomfield, S. et al. (2019) 'Investigation of the Validity of Two Bayesian Ancestral State Reconstruction Models for Estimating Salmonella Transmission during Outbreaks', *PLoS One*, 14: e0214169.

Bouckaert, R. et al. (2014) 'BEAST 2: A Software Platform for Bayesian Evolutionary Analysis', *PLoS Computational Biology*, 10: e1003537.

BTO. (2017), *British Trust for Ornithology (BTO)* <https://www.bto.org> accessed May 2021.

De Maio, N. et al. (2015) 'New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation', *PLoS Genetics*, 11: e1005421.

Drummond, A. J. et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

du Plessis, L., and Stadler, T. (2015) 'Getting to the Root of Epidemic Spread with Phylodynamic Analysis of Genomic Data', *Trends in Microbiology*, 23: 383–6.

Dudas, G. et al. (2018) 'MERS-CoV Spillover at the Camel-human Interface', *Elife*, 7: e31257.

EFSA, European Centre For Disease Prevention And Control, European Union Reference Laboratory For Avian Influenza et al. (2017) 'Avian Influenza Overview October 2016–August 2017', *EFSA Journal*, 15: e05018.

FAO. (2021), *Global Animal Disease Information System (Empres-i)* <https://empres-i.review.fao.org/#/> accessed Jan 2021.

FAOSTAT. (2016), *Production and Trade of Live Animals* <http://www.fao.org/faostat/en/#compare> accessed May 2021.

Faria, N. R. et al. (2018) 'Genomic and Epidemiological Monitoring of Yellow Fever Virus Transmission Potential', *Science*, 361: 894–9.

Freyman, W. A., and Höhna, S. (2019) 'Stochastic Character Mapping of State-dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-compatible Onagraceae Lineages', *Systematic Biology*, 68: 505–19.

Fusaro, A. et al. (2017) 'Genetic Diversity of Highly Pathogenic Avian Influenza A(H5N8/H5N5) Viruses in Italy, 2016–17', *Emerging Infectious Diseases*, 23: 1543–7.

Globig, A. et al. (2018) 'Highly Pathogenic Avian Influenza H5N8 Clade 2.3.4.4b In Germany in 2016/2017', *Frontiers in Veterinary Science*, 4: 240.

Grear, D. A. et al. (2018) 'Inferring Epidemiologic Dynamics from Viral Evolution: 2014-2015 Eurasian/North American Highly Pathogenic Avian Influenza Viruses Exceed Transmission Threshold, R0 = 1, in Wild Birds and Poultry in North America', *Evolutionary Applications*, 11: 547–57.

Guinat, C. et al. (2020a) 'Biosecurity Risk Factors for Highly Pathogenic Avian Influenza (H5N8) Virus Infection in Duck Farms, France', *Transboundary and Emerging Diseases*, 67: 2961–70.

——— et al. (2020b) 'Role of Live-Duck Movement Networks in Transmission of Avian Influenza, France, 2016–2017', *Emerging Infectious Diseases*, 26: 472–80.

——— et al. (2018) 'Exploring the Wind-Borne Spread of Highly Pathogenic Avian Influenza H5N8 during the 2016–2017 Epizootic in France', *Avian Diseases*, 63: 246–8.

——— et al. (2021) 'What Can Phylodynamics Bring to Animal Health Research?', *Trends in Ecology & Evolution*, 36: 837–47.

Hill, N. J. et al. (2016) 'Transmission of Influenza Reflects Seasonality of Wild Birds across the Annual Cycle', *Ecology Letters*, 19: 915–25.

Iglesias, I. et al. (2011) 'Reproductive Ratio for the Local Spread of Highly Pathogenic Avian Influenza in Wild Bird Populations of Europe, 2005–2008', *Epidemiology and Infection*, 139: 99–104.

Kass, R. E., and Raftery, A. E. (1995) 'Bayes Factors', *Journal of the American Statistical Association*, 90: 773–95.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Keawcharoen, J. et al. (2008) 'Wild Ducks as Long-distance Vectors of Highly Pathogenic Avian Influenza Virus (H5N1)', *Emerging Infectious Diseases*, 14: 600.

Kühnert, D. et al. (2016) 'Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data', *Molecular Biology and Evolution*, 33: 2102–16.

Larsson, A. (2014) 'AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets', *Bioinformatics*, 30: 3276–8.

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

——— et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLOS Computational Biology*, 5: e1000520.

Leyson, C. et al. (2019) 'Pathogenicity and Genomic Changes of a 2016 European H5N8 Highly Pathogenic Avian Influenza Virus (Clade 2.3.4.4) In Experimentally Infected Mallards and Chickens', *Virology*, 537: 172–85.

Lycett, S. et al. (2016) 'Global Consortium for H5N8 and Related Influenza Viruses. *Role for Migratory Wild Birds in the Global Spread of Avian Influenza H5N8*', *Science*, 354: 213–7.

Magee, D. et al. (2015) 'Combining Phylogeography and Spatial Epidemiology to Uncover Predictors of H5N1 Influenza A Virus Diffusion', *Archives of Virology*, 160: 215–24.

Mulatti, P. et al. (2018) 'Integration of Genetic and Epidemiological Data to Infer H5N8 HPAI Virus Transmission Dynamics during the 2016-2017 Epidemic in Italy', *Scientific Reports*, 8: 18037.

Müller, N. F., Dudas, G., and Stadler, T. (2019) 'Inferring Time-dependent Migration and Coalescence Patterns from Genetic Sequence and Predictor Data in Structured Populations', *Virus Evolution*, 5: vez030.

Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017) 'The Structured Coalescent and Its Approximations', *Molecular Biology and Evolution*, 34: 2970–81.

Nadeau, S. A. et al. (2021) 'The Origin and Early Spread of SARS-CoV-2 in Europe', *Proceedings of the National Academy of Sciences*, 118: 9.

Nagy, A. et al. (2018) 'Microevolution and Independent Incursions as Main Forces Shaping H5 Hemagglutinin Diversity during a H5N8/H5N5 Highly Pathogenic Avian Influenza Outbreak in Czech Republic in 2017', *Archives of Virology*, 163: 2219–24.

Napp, S. et al. (2018) 'Emergence and Spread of Highly Pathogenic Avian Influenza A(H5N8) in Europe in 2016-2017', *Transboundary and Emerging Diseases*, 65: 1217–26.

Pohlmann, A. et al. (2018) 'Swarm Incursions of Reassortants of Highly Pathogenic Avian Influenza Virus Strains H5N8 and H5N5, Clade 2.3.4.4 B, Germany, Winter 2016/17', *Scientific Reports*, 8: 1–6.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901.

Scire, J. et al. (2020), 'Phylodynamic Analysis of Genetic Sequencing Data from Structured Populations', *Viruses 2022*, 14: 1648.

Scoizec, A. et al. (2018) 'Airborne Detection of H5N8 Highly Pathogenic Avian Influenza Virus Genome in Poultry Farms, France', *Frontiers in Veterinary Science*, 5: 15.

Śmietanka, K. et al. (2020) 'Highly Pathogenic Avian Influenza H5N8 in Poland in 2019–2020', *Journal of Veterinary Research*, 64: 469.

Świętoń, E., and Śmietanka, K. (2018) 'Phylogenetic and Molecular Analysis of Highly Pathogenic Avian Influenza H5N8 and H5N5 Viruses Detected in Poland in 2016–2017', *Transboundary and Emerging Diseases*, 65: 1664–70.

Team, R. C. (2013) R: A Language and Environment for Statistical Computing.

Trovão, N. S. et al. (2015) 'Bayesian Inference Reveals Host-Specific Contributions to the Epidemic Expansion of Influenza A H5N1', *Molecular Biology and Evolution*, 32: 3264–75.

Vaughan, T. (2022), *BDMM-Prime Package* <https://github.com/tgvaughan/BDMM-Prime> accessed Mar 2022.

Vaughan, T. G. et al. (2014) 'Efficient Bayesian Inference under the Structured Coalescent', *Bioinformatics (Oxford, England)*, 30: 2272–9.

——— et al. (2019) 'Estimating Epidemic Incidence and Prevalence from Genomic Data', *Molecular Biology and Evolution*, 36: 1804–16.

Volz, E. M., Koelle, K., and Bedford, T. (2013) 'Viral Phylodynamics', *PLoS Computational Biology*, 9: e1002947.

Wikipedia. (2021), *World Population* <https://en.wikipedia.org/wiki/World_population> accessed Jul 2021.

Willgert, K. et al. (2020) 'Transmission of Highly Pathogenic Avian Influenza in the Nomadic Free-grazing Duck Production System in Viet Nam', *Scientific Reports*, 10: 1–11.

Yang, J. et al. (2019) 'Inferring Host Roles in Bayesian Phylodynamics of Global Avian Influenza A Virus H9N2', *Virology*, 538: 86–96.

Yu, G. et al. (2017) 'Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data', *Methods in Ecology and Evolution*, 8: 28–36.