Phylodynamic assessment of control measures for highly pathogenic avian influenza epidemics in France

Debapriyo Chakraborty^{1*}, Claire Guinat^{2,3}, Nicola F. Müller⁴, Francois-Xavier Briand⁵, Mathieu Andraud⁵, Axelle Scoizec⁵, Sophie Lebouquin⁵, Eric Niqueux⁵, Audrey Schmitz⁵, Beatrice Grasland⁵, Jean-Luc Guerin¹, Mathilde C. Paul¹ and Timothée Vergne¹

¹IHAP, Université de Toulouse, INRAE, ENVT, 23 Chemin des Capelles, 31300 Toulouse, France.

²Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

³Swiss Institute of Bioinformatics (SIB), Switzerland

⁴Vaccine and Infectious Disease, Fred Hutchinson Cancer Research Centre, Seattle, Washington State, USA

⁵The French Agency for Food, Environmental and Occupational Health & Safety (ANSES) Laboratory of Ploufragan-Plouzané-Niort, Ploufragan, France

*Corresponding author email: debapriyoc@gmail.com

Abstract

Phylodynamic methods have successfully been used to describe viral spread history but their applications for assessing specific control measures are rare. In 2016-17, France experienced a devastating epidemic of a highly pathogenic avian influenza virus (H5N8 clade 2.3.4.4b). Using 196 viral genomes, we conducted a phylodynamic analysis combined with generalised linear model and showed that the large-scale preventive culling of ducks significantly reduced the viral spread between *départements* (French administrative division). We also found that the virus likely spread more frequently between *départements* that shared borders, but the spread was not linked to duck transport between *départements*. Duck transport within *départements* increased the within-*département* transmission intensity, although the association was weak. Together, these results indicated that the virus spread in short-distances, either between adjacent *départements* or within *départements*. Results also suggested that the restrictions on duck transport within *départements* might not have stopped the viral spread completely. Overall, by testing specific hypothesis related to different control measures, we demonstrated that phylodynamics methods are capable of investigating the impacts of control measures on viral spread.

Introduction

During the winter of 2016, an epidemic of the highly pathogenic avian influenza (HPAI) virus occurred in France. It resulted in 484 infected poultry farms and approximately 7 million culled birds [1]. It remains critically important to learn as much as possible from this devastating epidemic, particularly as France once again experienced an HPAI outbreak in December 2020 [2]. The 2016-17 epidemic was caused by an H5N8 virus of the lineage 2.3.4.4b (A/Gs/Gd/1/96 clade) of Asian origin that was likely introduced by migratory birds [3]. The epidemic was largely concentrated in southwest France, encompassing the Occitanie and the Nouvelle-Aquitaine *régions* (the largest French territorial administrative division) [4]. The first incidence in poultry was recorded, using clinical symptoms, in a duck breeding farm in the eastern Tarn *département* (administrative division equivalent to county or district) on 25 November 2016 and the infection was confirmed—based on lab diagnosis—on 28 November (Fig. 1). Subsequently, the epidemic spread westwards [1], following the westerly distribution of poultry farms within the region. The last case of the epidemic was reported on 23 March 2017 [1].

Two major outbreak responses, namely preventive culling and stopping the transport of *foie gras* ducks between farms, were implemented during the epidemic. Culling was implemented locally in infected farms during the initial phase of the epidemic. However, these local measures proved to be insufficient in curbing the rapid spread of the virus. Consequently, there were three phases of large-scale culling. On 4 January 2017, all duck farms located in 150 communes across four *départements*, namely Gers, Landes, Pyrenees-Atlantiques and Hautes-Pyrenees, were notified to be preventively culled (Fig. 1). On 14 February, the number of communes under notification was doubled. On 21 February, the

notified areas were increased for the last time to include more than 500 communes in total. Additionally, following detection, ducks could not be moved from or to infected farms [5].

Epidemiological studies based on incidence data have provided important insights into the HPAI epidemic patterns [1] and the control measures [5,6]. Additionally, a recent phylogenetic study analysed 196 viral sequences that were identified in poultry farms and reported that the genotype that caused the large-sale epidemic in southwest France was associated with five geographic clusters that could be the consequence of a few long-distance transmission events followed by local spread [3]. However, the ecological and epidemiological drivers of these transmission patterns are still unknown. To fill this gap, we used a phylodynamic framework that we applied to the 2016-2017 epidemic data. Viral phylodynamics is the study of viral phylogenies and how they are shaped by potential interactions between epidemiological, immunological and evolutionary processes [7,8]. These methods can detect spatio-temporal patterns of epidemics based on viral genomic data, which represent a powerful source of information and are complimentary to epidemiological data [9]. Although phylodynamics methods are frequently used for reconstructing epidemic patterns, their application in assessing control measures is still rare [10]. Here, we used a phylodynamic framework to quantify the HPAI epidemic in southwest France and tested the effect of large-scale culling, along with other potential ecological and epidemiological drivers. Doing so, we underscore the versatility of phylodynamics methods in addressing both, specific control measure-based questions, as well as more broad ones regarding viral spread history.

Figure 1. The spatio-temporal dynamics of the 2016-17 H5N8 epidemic in southwest France.

The bars represent the daily number of infected farms identified in each *département*, with each *département* represented by a different colour. The number of infected farms represented those that were confirmed through PCR diagnostics during the epidemic period. The notification dates of the three culling phases are marked by three vertical lines. Inset. Location of southwest France with each study *département* marked with a specific colour corresponding to the source of samples.

Results

Viral phylogeny and spread history

To infer the phylogeny and the geographical spread of H5N8, we used a recently developed phylodynamic model called the marginal approximation of the structured coalescent (MASCOT) [11]. It models the coalescence of lineages within and the migration of lineages between geographical units and has been applied elsewhere to study the spread of

viral epidemics [12,13]. In our study, the geographical unit was the *département* as it was considered to provide a good compromise between the precision of the epidemiological inference and computational constraints, given the amount of data available. In Fig. 2A, we showed the evolutionary relationship between viral lineages from different *départements* in the form of a time-scaled summarised phylogenetic tree. Additionally, we show the most likely viral spread history of the lineages between *départements*, each of which is represented by a different colour. A change in colour across tree branches represents spread events between *départements*. We inferred Tarn to be the most likely source of all the viral lineages in southwest France (median posterior probability = 0.75, 95% credible interval; CI = 0.16-1). The median day of the emergence was estimated to be middle of November 2016 (95% CI = 27 Oct. - 22 Nov.). The clustering of basal sequences indicated a single introduction event in southwest France, followed by epidemic transmission (short terminal branches) mainly towards the west.

According to the inferred migration history (Fig. 2B), the virus mostly spread between neighbouring *départements*. The only exception to this pattern was migration from Tarn to Gers. In most cases, the virus migrated in one direction, from one *département* to other and did not come back to the source. However, in a few instances—between Gers and Landes and occasionally between Landes and Pyrénées-Atlantiques, we observed bidirectional spread of the virus. Gers was the most frequent source of spread and was responsible for spread into three other *départements*, namely Lot-et-Garonne (once), Landes (in multiple occasions) and Haute-Pyrenees (once). On the other hand, both Gers and Landes were the only *départements* where viruses were introduced from more than one *département*—namely Tarn (once) and Landes (in multiple occasions) into Gers and Gers (in multiple occasions) and Pyrénées -Atlantiques (at least twice) and Landes (in multiple occasions). On

the other end of the spectrum, Hautes-Pyrénées and Aveyron were the exceptional *départements* that received viruses in single events and never passed it forward to any other *départements*. Towards the end of the outbreak, lineages reached Pyrénées-Atlantiques in multiple occasions, all exclusively from Landes despite sharing its border with two other *départements* (Gers and Hautes-Pyrenees) as well.

The estimated viral effective population size N_e distribution for Landes—most affected in terms of case numbers—and Gers—second most affected—*départements* were proportional to the corresponding time series of case numbers (Figs. 2C and 2D).

Figure 2. Inferred epidemics dynamics based on structured coalescent population model. (A) A time-scaled maximum clade credibility (MCC) phylogenetic tree representing the evolutionary relationship between viral lineages. The colour of a branch indicates the inferred location (see legend) with the highest posterior support. A colour change on a

branch indicates a virus spread event. Numbers of all the nodes are shown in supplementary figure; (B) A schematic of the prominent (maximum probability of inferred location > 0.7) viral spread events between *départements*. (C) Estimates of viral effective population size (line in red; 95% highest posterior density; HPD in blue) and the time series of infected number of farms in Landes (black bar chart), the most affected *département*. (D) Estimates of viral population size (line in red; 95% HPD in Blue) and the time series of infected number of farms in Gers (black bar chart), the second most affected *département*.

Identifying the predictors of viral spread between *départements*

To identify the putative predictors of viral spread between *départements*, we used the MASCOT model where the migrations rates through time are inferred with the help of a generalized linear model (GLM) and a number of potential predictors [13]. The results are shown in Fig. 3. In Fig. 3A, we show the viral spread predictors and their level of support in the form of the Bayes factor (K), which is a Bayesian alternative to classical Null-Hypothesis Significance Testing [14,15]. Predictors with at least substantial support ($K > 3.2$) were included in the final GLM [16]. In Fig. 3B, we show the estimated coefficients, which represents the strength and direction (positive or negative) of each predictor's effect on the between-*département* viral spread. However, since all predictors are standardised, the strength of the effect size is not easily interpretable. The analysis demonstrated that sharing of borders ($K = 3.2$) as well as the period of culling of the epidemic (before and after 4 January) ($K = 10$) to be predictors of the observed spread patterns associated with viral spread. We inferred both predictors to positively predict spread, meaning that spread was greater between *départements* with shared borders and that spread was reduced after the 4 January culling initiation date.

Figure 3. The generalised linear models (GLMs) of viral spread between *départements*. (A)

The viral spread predictors and their level of support based on indicator probability, which

was estimated based on Bayesian Stochastic Search Variable Selection (BSSVS). Different

strengths of support are marked by vertical lines. Predictors whose indicators had at least

substantial support (Bayes factor, BF = 3.2) were included in the final GLM. (B) The boxplots

depict the strength and direction of each predictor's effect.

Identifying the predictors of viral effective population size N_e

To identify the putative predictors of viral effective size N_e within *départements*, we used

the MASCOT model where the N_e through time are defined as a generalized linear model

(GLM) [13]. In Fig. 4, we showed the results of a phylodynamic GLM analysis where we

investigated the relationship between viral effective population size N_e and a number of

potential predictors. In Fig. 4A, we show the N_e predictors and their level of support based

on indicator probability. The predictors whose indicators had at least substantial support

(Bayes factor, $K = 3.2$) were included in the final GLM. In Fig. 4B, we showed the estimated

coefficients, which represented the strength and direction of each predictor's effect. Based

on our inclusion criterion, the final GLM again included only two predictors, namely the case

numbers ($K > 100$, decisive) and the duck movement within *département* ($K > 3.2$).

Figure 4. The generalised linear models (GLMs) of viral effective population size N_e within *départements*. (A) The N_e predictors and their level of support based on indicator probability, which was estimated based on Bayesian Stochastic Search Variable Selection (BSSVS). Different strengths of support are marked by vertical lines. Predictors whose indicators had at least substantial support (Bayes factor, BF = 3.2) were included in the final GLM. (B) The boxplots depict the strength and direction of each predictor's effect.

Discussion

In 2016-17, France experienced a devastating epidemic of highly pathogenic avian influenza (HPAI) epidemic caused by an H5N8 subtype of the lineage 2.3.4.4b (A/Gs/Gd/1/96 clade). This outbreak caused major economic losses to the industry and contributed to the implementation of stricter farm biosecurity measures [6]. Unfortunately, the virus returned to France last winter causing yet another epidemic in the south-western part of the country highlighting the importance of understanding its transmission dynamics. One of those gaps has been assessing whether control measures were effective. Employing phylodynamic methods that combine genomic and epidemiological data, we showed that the large-scale preventive culling measure initiated on 4 January 2017 was successful in reducing viral spread between *départements*. At regional scale, we showed that viral spread between

départements occurred more frequently between *départements* that shared borders than between *départements* that were far apart, what is consistent with the geographical clusters identified in Briand et al. (2021). Our results could not find links between the viral spread between *départements* and duck transport. Within *départements*, we found that duck movements were positively but weakly associated with the viral effective population size. However, the number of infected farms was a powerful predictor of the effective population size. Additionally, our phylogeographic analysis estimated the date of viral emergence close to the date of the first detection, traced the origin of the epidemic to Tarn and exhibited detailed viral spread history between *départements*.

Preventive culling is a common control measure of viral epidemics of livestock. To reduce the epidemic size, it is often implemented in case of HPAI outbreaks because there is no treatment and vaccination is often not an option in a disease-free country like France. However, due to the elimination of potentially healthy birds and operational costs, it can be seen as an extremely expensive and controversial strategy. Therefore, it is often limited to infected farms and potentially-infected farms in the vicinity. Large-scale culling is usually implemented as a last resort to contain a rapidly spreading disease such as HPAI, after other less-expensive (e.g., local culling) and less-operationally challenging (e.g., ban on livestock movements) measures fail. Based on the HPAI epidemiologic data from the 2016-17 French outbreaks, a previous modelling study showed that a local culling strategy—culling of all poultry farms within 1 km of an infected farm—could have halved the epidemic size; but a large-scale culling strategy (up to 5 km) would only have had marginal benefits [6]. Based on this evidence and the fact that large-scale culling is operationally expensive and challenging to implement, a timely implementation of the culling of infected flocks was advocated. This finding has been echoed by other studies, both in case of HPAI [17] and foot and mouth

disease (FMD) [18,19]. However, culling could also be effective in reducing viral spread, which may not be contained locally. In that respect, our results complemented the epidemiological studies by focusing on viral spread and showed that large-scale culling is useful in containing regional spread of viruses—particularly for rapidly spreading viruses such as HPAI viruses—notwithstanding its impact on the number of infected farms. Together, these results can inform culling decision-making based on the type of pathogen and epidemic stage. For instance, it would be prudent to implement local-culling strategy for a slowly-spreading virus and delay large-scale culling until absolutely necessary. But, large-scale culling could be a relevant early response for a rapidly-spreading virus as far as containing the virus is concerned. Overall, culling decision-making requires a more detailed understanding of culling strategies and their holistic impact on the virus.

In our study, we also investigated the role of the two extensions of the large-scale preventive culling, which were successively implemented on 14 and 21 February 2017. Despite including more communes for culling, our results could not find evidence that these two culling phases reduced the transmission of H5N8 beyond the effect that the first culling phase on 4 January 2017 had.

Our results did not find evidence that the amount of duck movements between *départements* impacted the spread of the virus between *départements*. This finding—along with the evidence of between-*département* viral spread only between adjacent *départements*—suggests that most live-duck movements probably did not play any role in viral spread. This is consistent with observations that the duck movements that could have been responsible for the spread of the virus were quite limited and mostly occurred at the very early phases of the epidemic before stringent movement bans were implemented [5].

This could also be explained by the implementation of a PCR-test for all duck flocks that were about to be transported. One other explanation of this pattern could be that the ban on movements successfully disrupted the association between viral spread and long-distance duck movements. However, we did not consider the duck movements as a time varying predictor, which could have impacted our ability to detect duck movements as a predictor of viral transmission between *départements*. This hypothesis could be tested in the future with duck movement data collected over time (time-dependent variable).

Alternatively, our results could also mean that the virus did not spread through duck movements in the long-distance, but by some other mechanisms. Two major contenders for alternative mechanisms of viral spread are humans and farm equipment as fomites and wild birds as carriers. We did not find any significant association between human population density—which acted as a proxy for human activity—and viral spread. This suggested that human populations, which were not specifically involved in the duck-producing industry, did not play a major role in spreading the virus. An interesting development of this analysis would be to include equipment sharing data, although collection of this data is challenging as such records are probably not maintained methodically. Regarding the wild-bird related mechanisms, a study on Chinese poultry industry indicated that wild birds did not play a major role in spreading avian influenza viruses along the value chain across the country, in contrast to between-country spread for which wild birds play a prominent role [20]. This is likely to be the case in southwest France as well. Another potential mechanism could be wind-borne transmission, which is suspected in case of HPAI [21–23]. But, a previous study did not find any link between the direction of wind during the epidemic and the direction of viral spread, suggesting that the transmission was unlikely to be wind-borne [24]. However,

the study was limited to a specific period during the epidemic, hence the role of wind in spreading the virus is still an open question.

Our results indicated that the virus entered southwest France through Tarn and that this event represented a single source of the epidemic. This observation is likely to be robust because i) all the samples from Tarn are monophyletic and connect to the root, ii) terminal branches are short, iii) it is consistent with a recent study based on phylogenetic approaches [3]. Although Tarn itself is not considered to have high exposure risk from migratory birds, the single introduction pattern is likely to be explained by the European routes of migratory waterfowls because the estimated emergence interval coincided with the winter migration season (October-November) of many wild waterfowl species. These species are now known to carry and spread multiple avian influenza A viruses in poultry around the world, Still, we cannot completely reject the possibility of multiple introductions of a virus on the basis of the phylogeny alone [25,26]. Indeed, if the genomic sequences from distinct introductions are similar and in turn form monophyletic clusters, then the number of introductions could be underestimated. However, given avian influenza viruses are RNA viruses with high mutations rates, this alternative hypothesis is unlikely to hold. Our finding of a single introduction that acted as the source of the whole epidemic supports a previous phylogenetic study [3] and is consistent with outbreak investigations and epidemiological analyses [1,6]. Indeed, the index case in poultry was suspected and confirmed in a duck-breeding farm in Tarn, the day after that farm sent live-ducks to force-feeding farms in Gers and Lot-et Garonne. This event is likely to have triggered the viral spread [5].

Overall, by successfully analysing the impact of control measures, we showed the versatility of viral phylodynamic methods in providing both a broad understanding of

epidemics and the means to test hypotheses related to specific control measures. Finally, the methodology used in this study can be efficiently adapted to study the impact of control measures on many other viral epidemics.

Materials and Methods

Genomic data, bioinformatics and phylogenetic methods

196 HPAI H5N8 viruses were isolated amongst the 484 farms that experienced an outbreak between November 2016 and March 2017. The viral genomes were sequenced (one sequence per farm) at the French national reference laboratory for avian influenza (ANSES-Ploufragan, France). Each sequence was geocoded and associated with the *département* where the outbreak was located (Tarn, Aveyron, Lot-et-Garonne, Gers, Landes, Pyrénées-Atlantiques and Hautes-Pyrénées) along with the collection date. The whole concatenated virus genomic sequences were aligned using the MUSCLE multiple sequence alignment algorithm [27] implemented in MegaX software with default parameters [28]. Influenza viral genome is particularly prone to frequent genomic reassortments [29], which, if undetected, may provide spurious genomic signals. We therefore checked for the absence of recombination and reassortments in our aligned sequences using five different algorithms, namely BootScan, CHIMERA, MaxCHI, RDP and SisScan, implemented in RDP v4.1 [30]. None of the algorithms detected any recombination events. Therefore, we used all available sequences ($n = 196$) across the seven *départements* for downstream analyses.

We conducted an exploratory phylogenetic analysis, based on a quick maximum likelihood method, to ensure the geographic fidelity of the samples and that the samples

collected earlier were closer to the root of the phylogenetic tree. This analysis justified the use of the structured and more computationally intensive models.

Phylogenetic reconstruction of viral phylogeny and spread history

In our study, we used the GTR+Gamma₄ substitution model which was selected as the best model by the SMS algorithm in the Programme PhyML ver. 3 [31]. We also used a strict molecular clock as is common for avian influenza viruses[32,33] and also because the maximum likelihood tree root to tip divergence and sampling dates of the samples were strongly positively correlated . The phylogenetic temporal structure was assessed using Tempest ver. 1.5.3 [34].

Phylogenetic and generalised linear models

To infer the spread of H5N8, we used a recently developed phylodynamic model called the marginal approximation of the structured coalescent (MASCOT) [11]. This model was adapted in order to describe the coalescence of lineages within and the migration of lineages between *departments* and has been applied several times to study the spread of pathogens [12,13]. To identify putative predictors of viral spread, we used the MASCOT model where the migrations rates and effective population sizes through time are defined as a generalised linear model (GLM) [13].

Migration rate predictors

To investigate the effects of large-scale culling of ducks, we compared viral spread before and after preventive culling was notified by French government. The assumption was that if culling was effective then viral spread should be curtailed after the date of notification. To test this scenario, we created a predictor based on 4 January 2017. This was the date of the

first notification when preventive culling was implemented in 150 communes located at the border between Gers, Landes, Pyrenees-Atlantiques and Hautes-Pyrenees départements. We also created another predictor based on 14 February, when a second notification was issued to extend the culling to other communes. The last predictor of culling was based on 21 February, when the last notification was issued to include yet more communes under culling control measure. In Fig. 1, we showed the daily distribution of the number of outbreaks in each of the seven *départements* and marked the three culling notification dates.

Movement of ducks between different farms is a major feature of the *foie-gras* production in southwest France. Ducks, in large flocks of several thousands, are first raised in breeding farms for up to 12 weeks. Then, they are divided into small groups of hundreds and transported in batches to force-feeding farms across the region. Small batches are necessary because the process of force-feeding is labour-intensive and can optimally be done only in small batches. At the force-feeding-farms, the ducks are fattened for 12 days and then are sent off to slaughterhouse for harvest. As a result of this process, farms are continuously receiving and sending out shipments of ducks potentially facilitating the spread of the virus as well. To test any link between duck movement and viral spread between *départements*, we considered the total number of shipments exchanged between *départements*. This data was obtained from the professional organisation of fattening duck producers (CIFO) for the period between October 2016 and February 2017 [35]. This predictor was considered time-independent.

Density of livestock is often linked to pathogen spread [36–38] and density of poultry farms in southwest France were found to be associated the occurrence of outbreaks in the

case of the French HPAI 2016-17 epidemic [35]. To test the role of poultry density in driving viral spread, we considered the densities of duck and chicken farms per *département* as predictors. During epidemics in livestock, humans can spread pathogens between farms, either as fomites directly or through sharing of equipment [39,40]. Hence, we also considered human population per *département* as a separate predictor. These time-independent variables were computed from the French Directorate-General for Food (DGAI) and CIFOOG databases [35].

We also considered the total number of infected farms in a *département* to be a potential predictor of viral spread as *départements* with high number of cases could represent a major source of virus for other departments. Also, as the first case was detected in Tarn, we considered a potential predictor that represented Tarn to the most likely source of viral spread for all *départements*.

Finally, two time-independent spatial variables were considered as previous studies identified spatial patterns in the epidemic spread [1,35]. We tested the effect of the Euclidean distance between *départements* centroids as well as the border sharing between *départements* on the viral spread between departments.

Effective population size (N_e) predictors

In the case of Ebola virus epidemics, the number of detected cases was found to be highly predictive of the viral effective population size (N_e) [10,13,41]. In our analysis, we therefore used the number of infected farms in each *département* per day, which was smoothed using a 7-day moving average. However, if virus lineages are erroneously time-stamped due to unknown sampling delay, a GLM analysis may result in spurious association between case number and N_e [10]. So, for our second case number-based predictor, we

artificially incorporated a one-week delay in the case number distribution to check for any delay [13].

To test any link between duck movement and N_e within a *département*, we considered the total number of shipments exchanged between farms within a *département* as another potential predictor. Livestock density plays an important role in viral transmission and hence could be linked to N_e [42]. We used the densities of duck and chicken farms per *département* as potential predictors. Humans, as fomites, can also influence N_e and hence human population density was also used as a predictor. Both these predictors were time-independent. Also, all non-binary variables were log transformed and standardised so that their means and standard deviations equalled 0 and 1, respectively.

The modelling exercises and the Markov chain Monte Carlo (MCMC) simulations were conducted in BEAST v2.6.3 [43]. For posterior sampling, we used the coupled MCMC package [44] with 100 million iterations. The convergence and mixing of the MCMC chains were checked visually in Tracer v1.7.1 [45]. An analysis was considered successful if the effective sample size of each of the parameters was at least 200. The time-scaled phylogenetic tree with the most likely *département* of each lineage was created in FigTree v1.4.3 [46].

Selection of generalised linear model

Model selection requires extensive computation, particularly for models with many parameters. Hence, MASCOT provides an efficient model averaging approach, the Bayesian stochastic search variable selection (BSSVS), that is used to calculate a binary indicator variable [13,47]. This variable indicates if its predictor has contributed (1 = yes; 0 = No) to the GLM. BSSVS results in an estimate of the posterior inclusion probability or support for each

predictor. Each predictor is also associated with a coefficient [13]. The purpose of the coefficient is to express predictor's effect size and the value can range between $-\infty$ to ∞ . We also measured the confidence in the result based on Bayes factors (K). Based on convention [16], we decided the confidence to be not substantial ($1 > K < 3.2$), substantial ($3.2 > K < 10$), strong ($10 > K < 100$) or decisive ($K > 100$). In the final model, we included only the predictors with at least substantially confident result. These statistical analyses were conducted in R v4.0.3 [48]. All the data, including genetic data accession numbers, xml files and R scripts, are available in the following repository, https://github.com/dc27708/H5N8_MASCOT.

Acknowledgements

This work was financially supported by the FEDER/*Région Occitanie Recherche et Sociétés* 2018—AI-TRACK and the PREDYT project (*Fonds pour la Recherche sur l'Influenza Aviaire*). This study was also supported by the “Chair for Avian Biosecurity”, hosted by the National Veterinary College of Toulouse and funded by the Direction Generale de l'Alimentation, Ministère de l'Agriculture et de l'Alimentation, France. NFM is funded by the Swiss National Science Foundation (P2EZP3_191891). CG is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 842621. DC gratefully acknowledges the contribution of the INRAE MaIAGE unit in the form of free access to their MIGALE computer cluster. We also thank all the members of the EPIDESA team for their support and help.

Competing interests

The authors declare no competing interests.

References

1. Guinat C, Nicolas G, Vergne T, Bronner A, Durand B, Courcoul A, et al. Spatio-temporal patterns of highly pathogenic avian influenza virus subtype H5N8 spread, France, 2016 to 2017. *Eurosurveillance*. 2018;23.
2. Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Marangon S, Niqueux É, et al. Avian influenza overview December 2020 – February 2021. *EFSA J*. 2021;19: e06497. doi:<https://doi.org/10.2903/j.efsa.2021.6497>
3. Briand F-X, Niqueux E, Schmitz A, Martenot C, Cherbonnel M, Massin P, et al. Highly Pathogenic Avian Influenza A (H5N8) Virus Spread by Short-and Long-Range Transmission, France, 2016–17. *Emerg Infect Dis*. 2021;27: 508–516.
4. Bronner A, Niqueux E, Schmitz A, Le Bouquin S, Huneau-Salaün A, Guinat C, et al. Description de l'épisode d'influenza aviaire hautement pathogène en France en 2016-2017. *Bulletin épidémiologique, santé animale et alimentation*. Jul 2017;79: 13–17.
5. Guinat C, Durand B, Vergne T, Corre T, Rautureau S, Scoizec A, et al. Role of Live-Duck Movement Networks in Transmission of Avian Influenza, France, 2016–2017. *Emerg Infect Dis J*. 2020;26. doi:10.3201/eid2603.190412
6. Andronico A, Courcoul A, Bronner A, Scoizec A, Lebouquin-Leneveu S, Guinat C, et al. Highly pathogenic avian influenza H5N8 in south-west France 2016–2017: A modeling study of control strategies. *Epidemics*. 2019;28: 100340. doi:10.1016/j.epidem.2019.03.006
7. Volz EM, Koelle K, Bedford T. Viral Phylodynamics. *PLOS Comput Biol*. 2013;9: e1002947. doi:10.1371/journal.pcbi.1002947
8. Holmes EC, Grenfell BT. Discovering the Phylodynamics of RNA Viruses. *PLOS Comput Biol*. 2009;5: e1000505. doi:10.1371/journal.pcbi.1000505
9. Guinat C, Vergne T, Kocher A, Chakraborty D, Paul MC, Ducatez M, et al. What can phylodynamics bring to animal health research? *Trends Ecol Evol*. Accepted.
10. Dellicour S, Baele G, Dudas G, Faria NR, Pybus OG, Suchard MA, et al. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. *Nat Commun*. 2018;9: 2222. doi:10.1038/s41467-018-03763-2

11. Müller NF, Rasmussen D, Stadler T. MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*. 2018;34: 3843–3848.
12. Yang J, Müller NF, Bouckaert R, Xu B, Drummond AJ. Bayesian phylodynamics of avian influenza A virus H9N2 in Asia with time-dependent predictors of migration. *PLOS Comput Biol*. 2019;15: e1007189. doi:10.1371/journal.pcbi.1007189
13. Müller NF, Dudas G, Stadler T. Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations. *Virus Evol*. 2019;5: vez030.
14. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2: 6–10. doi:10.1038/s41562-017-0189-z
15. Keyzers C, Gazzola V, Wagenmakers E-J. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat Neurosci*. 2020;23: 788–799. doi:10.1038/s41593-020-0660-4
16. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90: 773–795.
17. Le Menach A, Vergu E, Grais RF, Smith DL, Flahault A. Key strategies for reducing spread of avian influenza among commercial poultry holdings: lessons for transmission to humans. *Proc R Soc B Biol Sci*. 2006;273: 2467–2475. doi:10.1098/rspb.2006.3609
18. Ferguson NM, Donnelly CA, Anderson RM. The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*. 2001;292: 1155–1160. doi:10.1126/science.1061020
19. Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*. 2001;294: 813–817. doi:10.1126/science.1065973
20. Yang Q, Zhao X, Lemey P, Suchard MA, Bi Y, Shi W, et al. Assessing the role of live poultry trade in community-structured transmission of avian influenza in China. *Proc Natl Acad Sci*. 2020;117: 5949–5954. doi:10.1073/pnas.1906954117
21. Zhao Y, Richardson B, Takle E, Chai L, Schmitt D, Xin H. Airborne transmission may have played a role in the spread of 2015 highly pathogenic avian influenza outbreaks in the United States. *Sci Rep*. 2019;9: 11755. doi:10.1038/s41598-019-47788-z
22. Ypma RJF, Jonges M, Bataille A, Stegeman A, Koch G, van Boven M, et al. Genetic Data Provide Evidence for Wind-Mediated Transmission of Highly Pathogenic Avian Influenza. *J Infect Dis*. 2013;207: 730–735. doi:10.1093/infdis/jis757
23. Scoizec A, Niqueux E, Thomas R, Daniel P, Schmitz A, Le Bouquin S. Airborne Detection of H5N8 Highly Pathogenic Avian Influenza Virus Genome in Poultry Farms, France. *Front Vet Sci*. 2018;5. doi:10.3389/fvets.2018.00015

- 494 24. Guinat C, Rouchy N, Camy F, Guérin JL, Paul MC. Exploring the Wind-Borne Spread of
495 Highly Pathogenic Avian Influenza H5N8 During the 2016–2017 Epizootic in France. *Avian*
496 *Dis.* 2018;63: 246–248. doi:10.1637/11881-042718-ResNote.1
- 497 25. Hill SC, Lee Y-J, Song B-M, Kang H-M, Lee E-K, Hanna A, et al. Wild waterfowl migration
498 and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza
499 in the Republic of Korea. *Infect Genet Evol.* 2015;34: 267–277.
- 500 26. da Silva Filipe A, Shepherd JG, Williams T, Hughes J, Aranday-Cortes E, Asamaphan P, et
501 al. Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland
502 Europe into Scotland. *Nat Microbiol.* 2021;6: 112–122. doi:10.1038/s41564-020-00838-z
- 503 27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
504 throughput. *Nucleic Acids Res.* 2004;32: 1792–1797.
- 505 28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics
506 analysis across computing platforms. *Mol Biol Evol.* 2018;35: 1547–1549.
- 507 29. Lycett SJ, Duchatel F, Digard P. A brief history of bird flu. *Philos Trans R Soc B.* 2019;374:
508 20180257.
- 509 30. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of
510 recombination patterns in virus genomes. *Virus Evol.* 2015;1.
- 511 31. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms
512 and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance
513 of PhyML 3.0. *Syst Biol.* 2010;59: 307–321. doi:10.1093/sysbio/syq010
- 514 32. Fourment M, Holmes EC. Avian influenza virus exhibits distinct evolutionary dynamics in
515 wild birds and poultry. *BMC Evol Biol.* 2015;15: 120. doi:10.1186/s12862-015-0410-5
- 516 33. Si Y-J, Lee IW, Kim E-H, Kim Y-I, Kwon H-I, Park S-J, et al. Genetic characterisation of
517 novel, highly pathogenic avian influenza (HPAI) H5N6 viruses isolated in birds, South
518 Korea, November 2016. *Eurosurveillance.* 2017;22: 30434. doi:10.2807/1560-
519 7917.ES.2017.22.1.30434
- 520 34. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of
521 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016;2:
522 vew007.
- 523 35. Guinat C, Artois J, Bronner A, Guérin JL, Gilbert M, Paul MC. Duck production systems
524 and highly pathogenic avian influenza H5N8 in France, 2016–2017. *Sci Rep.* 2019;9: 6177.
525 doi:10.1038/s41598-019-42607-x
- 526 36. Cantrell DL, Groner ML, Ben-Horin T, Grant J, Revie CW. Modeling Pathogen Dispersal in
527 Marine Fish and Shellfish. *Trends Parasitol.* 2020;36: 239–249.
528 doi:10.1016/j.pt.2019.12.013
- 529 37. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock.
530 *Philos Trans R Soc B Biol Sci.* 2015;370: 20140107. doi:10.1098/rstb.2014.0107

38. Tildesley MJ, Brand S, Brooks Pollock E, Bradbury NV, Werkman M, Keeling MJ. The role of movement restrictions in limiting the economic impact of livestock infections. *Nat Sustain.* 2019;2: 834–840. doi:10.1038/s41893-019-0356-5
39. Mansley LM. The challenge of FMD control in the 2001 UK FMD epidemic. *Animal Production in Europe: The Way Forward in a Changing World “in-Between” congress of the International Society for Animal Hygiene Saint-Malo: Citeseer. Citeseer; 2004.*
40. van Andel M, Tildesley MJ, Gates MC. Challenges and opportunities for using national animal datasets to support foot-and-mouth disease control. *Transbound Emerg Dis.* 2020.
41. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature.* 2017;544: 309–315. doi:10.1038/nature22040
42. Meadows AJ, Mundt CC, Keeling MJ, Tildesley MJ. Disentangling the influence of livestock vs. farm density on livestock disease epidemics. *Ecosphere.* 2018;9: e02294. doi:<https://doi.org/10.1002/ecs2.2294>
43. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10: e1003537.
44. Müller NF, Bouckaert RR. Adaptive parallel tempering for BEAST 2. *bioRxiv.* 2020; 603514. doi:10.1101/603514
45. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67: 901.
46. Rambaut A. FigTree. Edinburgh; 2016. Available: <http://tree.bio.ed.ac.uk/software/figtree/>
47. Lemey P, Suchard M, Rambaut A. Reconstructing the initial global spread of a human influenza pandemic. *PLoS Curr.* 2009;1. doi:10.1371/currents.RRN1031
48. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020. Available: <https://www.R-project.org/>