

1 **Title:**

2 Disentangling the role of poultry farms and wild birds in the spread of highly pathogenic  
3 avian influenza virus H5N8 in Europe

4

5 **Authors:** Claire Guinat<sup>1,2\*</sup>, Cecilia Valenzuela Agui<sup>1,2</sup>, Timothy G. Vaughan<sup>1,2</sup>, Jérémie  
6 Scire<sup>1,2</sup>, Anne Pohlmann<sup>3</sup>, Christoph Staubach<sup>3</sup>, Jacqueline King<sup>3</sup>, Edyta Swieton<sup>4</sup>, Adam  
7 Dan<sup>5</sup>, Lenka Cernikova<sup>6</sup>, Mariette F. Ducatez<sup>7</sup>, Tanja Stadler<sup>1,2</sup>

8

9 **Affiliations:**

10 <sup>1</sup>Department of Biosystems Science and Engineering, ETH Zurich, Switzerland

11 <sup>2</sup>Swiss Institute of Bioinformatics (SIB), Switzerland

12 <sup>3</sup>Friedrich-Loeffler-Institut, Germany

13 <sup>4</sup>National Veterinary Research Institute, Poland

14 <sup>5</sup>Danam Vet Molbiol, Kozseg, Hungary

15 <sup>6</sup>State Veterinary Institute, Prague, Czech Republic

16 <sup>7</sup>INRAE-ENVIT, France

17

18 **\*Corresponding author**

19 Dr. Claire Guinat

20 DVM, MSc, PhD

21 Computational Evolution group

22 Mattenstrasse 26

23 4058 Basel

24 Switzerland

25 [claire.guinat@bsse.ethz.ch](mailto:claire.guinat@bsse.ethz.ch)

26

27

28 **Abstract (250 words) 246**

29       Recent outbreaks of highly pathogenic avian influenza H5N8 virus in Europe have  
30 caused severe damage to animal health, wildlife conservation and livestock economic  
31 sustainability. While epidemiological and phylogenetic studies have generated important  
32 clues about the virus spread in Europe, they remained opaque to the specific role of poultry  
33 farms and wild birds. Using a phylodynamic framework, we inferred the H5N8 virus  
34 transmission dynamics among poultry farms and wild birds in four severely affected  
35 countries and investigated drivers of spread between farms across borders during the 2016-17  
36 epidemic. Based on existing genetic data, we showed that the virus was likely introduced into  
37 poultry farms during the autumn, in line with the timing of arrival of migratory wild birds.  
38 Then, transmission was mainly driven by farm-to-farm transmission in Germany, Hungary

39 and Poland, suggesting that better understanding of how infected farms are connected in  
40 those countries would greatly help control efforts. In contrast, the epidemic was dominated  
41 by wild bird-to-farm transmission in Czech Republic, meaning that more sustainable  
42 prevention strategies should be developed to reduce virus exposure from wild birds. We  
43 inferred effective reproduction number  $R_e$  estimates among poultry farms and wild birds. We  
44 expect those estimates being useful to parameterize predictive models of virus spread aiming  
45 at optimising control strategies. None of the investigated predictors related to live poultry  
46 trade, poultry census and geographic proximity were identified as supportive predictors of the  
47 viral spread between farms across borders, suggesting that other drivers should be considered  
48 in future studies.

49

50 **Significance statement (120 words) 118**

51 In winter 2016-17, Europe was severely hit by an unprecedented epidemic of highly  
52 pathogenic avian influenza (HPAI) H5N8 virus, causing significant impact on animal health,  
53 wildlife conservation and livestock economic sustainability. By applying phylodynamic tools  
54 to H5N8 sequence data collected from poultry farms and wild birds during the epidemic, we  
55 quantified how effectively the first infections were detected, how fast the virus spread, how  
56 many infections were missed and how many transmission events occurred at the wildlife-  
57 domestic interface. Also, we investigated predictors of the virus spread between farms across  
58 borders. These results are crucial to better understand the virus transmission dynamics, with  
59 the view to inform policy decision-making and reduce the impact of future epidemics of  
60 HPAI viruses.

61

62 **Keywords:** phylodynamics, Europe, highly pathogenic avian influenza, H5N8, virus  
63 transmission and evolution, Bayesian inference

64

65 **Main text 4646**

66

67 **Introduction**

68

69           Since the beginning of the 21<sup>st</sup> century, the highly pathogenic avian influenza (HPAI)

70 H5N8 virus (clade 2.3.4.4b) represents one of the most serious threats to animal health,

71 wildlife conservation and livestock economic sustainability. In June 2016, the virus was

72 detected in wild birds in regions of Central Asia (at the Ubsu-Nur and Qinghai lakes, known

73 as migration stop-overs) and subsequently spread to other Asian countries and Europe (1). By

74 the end of 2017, the virus had caused one of the most severe epidemic in Europe, in terms of

75 number of poultry outbreaks, wild bird cases and affected countries (1). Most of poultry

76 outbreaks occurred in France (37.8%), followed by Hungary (21.5%), Germany (8.5%),

77 Poland (5.8%) and Czech Republic (3.9%) (1). While no human cases were observed, the

78 control strategies that were implemented in the affected countries resulted in the culling of

79 several million poultry, causing devastating socio-economic impacts for the poultry industry.

80           The emergence of H5N8 virus in Europe was likely attributable to infected migratory

81 wild birds from Northern Eurasia, leading to occasional or multiple viral incursions into

82 poultry farms (2–7). After emergence, farm-to-farm transmission was likely the main driver

83 of the epidemic, with contact with infected poultry and contaminated fomites, such as

84 vehicles or equipment, being a major risk factor for farm infection (1–3, 6, 8, 9). In a number

85 of cases, high poultry density and substantial gaps in farm biosecurity were also identified as

86 potential risk factors for farm infection (1, 3, 6, 10). The possibility of airborne transmission

87 between poultry farms was also suggested, without being conclusively demonstrated (11, 12).

88           While such epidemiological and phylogenetic studies have generated important clues

89 about the H5N8 virus transmission patterns in Europe, they remained opaque to the specific

90 role of poultry farms and wild birds in the disease spread. In particular, understanding the

91 viral transmission dynamics among these two subpopulations is crucial to determine which of  
92 these two has the greatest potential to drive the viral transmission during epidemics, which, in  
93 turn, represents critical information to better target control strategies. When appropriate  
94 pathogen genetic and epidemiological data is collected, phylodynamic methods can fill this  
95 critical gap (13–15). By fitting population dynamic models to genetic sequences collected  
96 during epidemics, these tools aim at quantifying disease transmission dynamics and have  
97 been particularly used to study the spread of infectious diseases in structured populations, be  
98 they stratified by time, species or geography (16–18). Importantly, birth-death model-based  
99 approaches (19) explicitly allow for the direct estimation of key epidemiological parameters,  
100 such as the effective reproduction number  $R_e$  (which captures the number of secondary  
101 infections generated at any time during an epidemic in a partially immune population) (20),  
102 while taking into account the sampling effort.

103         Using a phylodynamic framework, this study aimed at disentangling the role of  
104 poultry farms and wild birds in the spread of H5N8 in Europe during the 2016-2017  
105 epidemic. We fitted a phylodynamic model with geographical and host structure to H5N8  
106 genome sequences of the HA segment collected from both host types (190 from poultry farms  
107 and 130 from wild birds) in four severely affected European countries (Czech Republic,  
108 Germany, Hungary and Poland) to: (i) estimate the early patterns of virus spread, (ii) infer the  
109 number of unreported infections, (iii) provide  $R_e$  estimates, (iv) discriminate the number of  
110 new infections arising from local transmission versus importation events and (v) identify  
111 factors driving the virus spread between farms across borders.

112

## 113 **Results**

114         We fitted a multi-type birth-death (MTBD) model to the aligned sequences (19, 21) to  
115 co-infer epidemiological parameters, along with the underlying structured phylogenetic trees

116 (22) and epidemic trajectories (23) (see Materials and Methods). The model was structured  
117 according to the host type and geographical location, resulting into five demes: ‘poultry farms  
118 in Czech Republic’, ‘poultry farms in Germany’, ‘poultry farms in Hungary’, ‘poultry farms  
119 in Poland’ and ‘wild birds in the four countries’. This latter was assumed to represent the  
120 epidemic origin (24) and all transmission, become non-infectious and sampling processes  
121 were assumed to be deme-specific and constant through time (except for the within-deme  $R_e$   
122 that was estimated across four-time intervals) (19, 21).

### 123 *Early patterns of H5N8 virus spread in Europe*

124 The maximum clade credibility (MCC) tree reconstructed using the MTBD model is  
125 shown in [Figure 1](#). Sequences from poultry farms of the same country for Germany, Hungary  
126 and Poland were generally clustered together in the tree, while sequences from poultry farms  
127 for Czech Republic were more scattered in the tree as were wild birds’ sequences. From the  
128 epidemic trajectories, we extracted the first transmission events and summarized them over  
129 time to inform the early patterns of virus spread. [Figure 2](#) shows the temporal distribution of  
130 inferred dates of the first imported (i.e. from another deme), local (i.e. within-deme) and  
131 exported (i.e. to another deme) outbreak/case per deme together with the first officially  
132 reported outbreak/case for comparison (25). Overall, the inferred dates of the first imported  
133 poultry farm outbreak were before the date of the first officially reported outbreak in each  
134 deme, with a higher delay in Czech Republic deme (median: 115 days, 95% High Posterior  
135 Density (HPD): 70 – 158), i.e. approximately 16 weeks) compared to the others (from  
136 median: 21 days, 95% HPD: 5 – 54 to median: 46 days, 95% HPD: 5 – 106, i.e.  
137 approximately 3 to 7 weeks) ([SI Appendix Table S1](#)). The first events of local virus  
138 transmission and exportation also took place before the first poultry farm outbreak was  
139 officially reported in each deme and occurred very rapidly after the first virus introduction ([SI](#)  
140 [Appendix Table S1](#)).

## 141 *Number of unreported H5N8 infections*

142 From the epidemic trajectories, we extracted the number of poultry farms and wild  
143 birds that became non-infectious (following death, culling or recovery of the poultry  
144 flock/wild bird) and summarized it over time. This number was used to inform the number of  
145 outbreaks/cases that could have been missed during the epidemic. [Figure 3](#) represents the  
146 temporal distribution of the inferred cumulative number of no-longer infectious  
147 outbreaks/cases per deme together with the cumulative number of officially reported  
148 outbreaks/cases for comparison (25). The cumulative number of officially reported poultry  
149 farm outbreaks (94 for Germany, 240 for Hungary and 65 for Poland) were within the  
150 inferred 95% HPD (median: 80, 95% HPD: 26-1196 for Germany, median: 401, 95% HPD:  
151 142-1126 for Hungary and median: 96, 95% HPD: 34-398 for Poland). More discrepancies  
152 were observed for poultry farms in Czech Republic and wild birds in the countries, the  
153 cumulative number of officially reported outbreaks/cases (43 for Czech Republic and 372 for  
154 wild birds) being outside the inferred 95% HPD (median: 134, 95% HPD: 54 – 356 for Czech  
155 Republic and median: 4062, 95% HPD: 1138-14569 for wild birds).

## 156 *Key epidemiological parameters of H5N8 virus spread*

157 [Figures 4A](#) and [4B](#) illustrate the posterior distributions for within-deme and between-  
158 deme  $R_e$  respectively, together with the prior ([SI Appendix Table S2](#)) for comparison. The  
159 within-deme  $R_e$  was estimated across four-time intervals, corresponding to the four phases of  
160 the epidemic ([SI Appendix Figure S1](#)). Note that only one sequence was available per poultry  
161 farm and per wild bird, meaning that the within-deme  $R_e$  represents the farm-to-farm/wild  
162 bird-to-wild bird virus transmission and the between-deme  $R_e$  represents the cross-species  
163 and cross-country virus transmission (i.e. farm-to-wild bird/wild bird-to-farm/farm-to-farm  
164 across countries). For most demes, the median within-deme  $R_e$  posteriors were above or close  
165 to 1 during the first time period, while a decrease was generally observed throughout the

166 subsequent time periods (Figures 4A, SI Appendix Table S3). However, these  $R_e$  estimates  
167 slightly increased again during the fourth time period (Feb – May 2017) in Germany,  
168 Hungary and Poland. The highest median  $R_e$  estimates were observed between poultry farms  
169 in Hungary and between wild birds in the four countries during the first time period (Oct –  
170 Nov 2016). Overall, the between-deme  $R_e$  estimates were much lower than the within-deme  
171  $R_e$  estimates, with extremely low values (median within the range of  $10^{-3}$  –  $10^{-2}$ ) for those  
172 representing farm-to-farm across countries and wild bird-to-farm transmission (Figure 4B, SI  
173 Appendix Table S3). Slightly higher values were found for those representing farm-to-wild  
174 bird transmission (median of 0.1 – 0.4). One exception was found for the between-deme  $R_e$   
175 estimate representing farm-to-wild bird transmission in Czech Republic, with a median of 4.6  
176 (95% HPD: 0.9-9.2). The infectious period was also inferred for each deme, with the highest  
177 median found for poultry farms in Czech Republic (median: 14 days, 95% HPD: 7 – 23) and  
178 wild birds in the four countries (median: 14 days, 95% HPD: 11 - 19) (SI Appendix Figure S2  
179 and Table S3). The infectious period was slightly lower for poultry farms in Germany  
180 (median: 10, 95%HPD: 6 – 16), Hungary (median: 8 days, 95% HPD: 4 – 14) and Poland  
181 (median: 7 days, 95%HPD: 4 – 11).

### 182 ***Number of local H5N8 transmission versus importation events***

183 From the epidemic trajectories, we can discriminate the number of poultry farm  
184 outbreaks and wild bird cases arising from local virus transmission (i.e. farm-to-farm/wild  
185 bird-to-wild bird) versus virus importation (i.e. farm-to-wild bird/wild bird-to-farm/farm-to-  
186 farm across countries) events. Figure 5 illustrates the temporal distribution of the inferred  
187 median number of local transmission and importation events per deme. In Germany, Hungary  
188 and Poland, the epidemic was dominated by local virus transmission events between poultry  
189 farms (median: 109, 95% HPD: 61 – 5750, median: 316, 95% HPD: 144 – 888 and median:  
190 77, 95% HPD: 32 – 367, respectively) with an increase around March 2017, November 2016

191 and December 2016, respectively (SI Appendix Table S4). In Czech Republic, the epidemic  
192 was dominated by importations from wild birds (median: 115, 95% HPD: 57 – 230). For all  
193 countries, an increase in the number of importation events from wild birds were observed  
194 around January – February 2017. The epidemic in wild birds was also dominated by local  
195 transmission between wild birds (median: 3,075, 95% HPD: 807 – 8,575) and the highest  
196 number of importations were coming from poultry farms in Czech Republic (median: 972,  
197 95% HPD: 77 – 5,569).

### 198 *Predictors of H5N8 virus spread between poultry farms across borders*

199 Alongside inferring the transmission dynamics of H5N8, potential drivers of virus  
200 spread between poultry farms across the four countries were investigated by quantifying the  
201 corresponding  $R_e$  parameter in the MTBD model with a generalized linear model (GLM) (26,  
202 27) (see Materials and Methods). There were eight predictors included in the model: the 2016  
203 live poultry trade (28), the 2016 poultry density in the source and destination deme (28), the  
204 2014 poultry farm density in the source and destination deme (24), the 2017 farm outbreak  
205 density in the source and destination deme (25) and the distance between countries' centroids  
206 (SI Appendix Table S5). The highest number of poultry was moved from Czech Republic (22  
207 million) and Germany (20 million) to Poland. The highest poultry density was reported in  
208 Poland (600 birds/km<sup>2</sup>) while the highest poultry farm density was found in Germany (0.2  
209 farm/km<sup>2</sup>). The highest poultry farm outbreak density was reported in Hungary (0.003  
210 outbreak/km<sup>2</sup>). Countries centroids were 382 to 790 km apart. Figure 6A shows, for each  
211 predictor, the inclusion probability which represents the proportion of the posterior samples  
212 in which the given predictor was included in the model and the Bayes Factor (BF) which  
213 quantifies which of the posterior and prior inclusion probabilities of the given predictor in the  
214 model is more likely. Figure 6B shows the log conditional effect size which represents the log  
215 contribution of the given predictor when it was included in the model. None of the predictors

216 were statistically supported to be associated with the spread of H5N8 virus between poultry  
217 farms across borders, illustrated by the low BF metric (<3.2) (Figure 6A) and the similar  
218 distribution between the posterior coefficient estimates (Figure 6B) and the prior (SI  
219 Appendix Table S2).

220

## 221 Discussion

222 The epidemic trajectories showed that, in each country, the first introduction of H5N8  
223 virus from wild birds to poultry farms likely occurred during autumn, which is in line with  
224 the timing of arrival of migratory wild birds in Europe (29). Also, the epidemic trajectories  
225 indicated that there was a delay of 3 to 16 weeks (depending of the country) between the  
226 inferred date of the first virus introduction and the date of the first officially reported poultry  
227 farm outbreak, likely illustrating different surveillance strategies' effectiveness. The longest  
228 delay (16 weeks) was observed in Czech Republic, where most outbreaks occurred in small  
229 size farms (< 100 birds), while they mainly affected large size farms (> 10,000 birds) in  
230 Germany, Hungary and Poland (1). While a total of 442 poultry farm outbreaks and 372 wild  
231 bird cases in the four countries were officially reported (25), the epidemic trajectories showed  
232 that these numbers could have been under-estimated, especially in the wild bird population,  
233 likely due to challenges related to wildlife surveillance (30). High reporting rates of poultry  
234 farm outbreaks were found in Germany, Hungary and Poland, likely linked to the high  
235 mortality rates of poultry following H5N8 virus infection, along with the active surveillance  
236 implemented around reported poultry farm outbreaks (24). However, lower and again more  
237 delayed reporting rates were found for the poultry farms in Czech Republic. These results  
238 suggest that the likelihood of reporting infected farms is likely associated with the  
239 characteristics of the farm. However, whether this results of differences in size or other

240 factors linked to the farm size (such as different farmers' knowledge, attitudes and practices)  
241 needs further investigation.

242         Following the first virus introduction, the epidemic trajectories demonstrated that in  
243 Germany, Hungary and Poland, the epidemic was dominated by local farm-to-farm  
244 transmission events. In Germany, local farm-to-farm transmission increased between  
245 February and May 2017, likely illustrating the cluster of turkey farm outbreaks which  
246 represented approximately 25% of the total number of poultry farm outbreaks in the country  
247 (24). In Hungary, a peak of in the number of farm-to-farm transmission events was reported  
248 between October and November 2016, during which most outbreaks clustered in time and  
249 space (1). Moreover, the epidemic in these countries was also partly driven by wild bird-to-  
250 farm transmission (in particular in the middle of the epidemic) showing that the role of wild  
251 birds was likely greater than expected and was not limited to the onset of the epidemic. These  
252 outcomes also emphasize that in-place biosecurity measures in Germany, Hungary and  
253 Poland were sufficient to prevent continued incursions from farms across borders (such as  
254 ban of international trade) (24), but were less effective against local farm-to-farm and wild  
255 bird-to-farm transmission. Having more detailed knowledge of how poultry farms are  
256 connected with one another in those countries could help containing future outbreaks by  
257 disrupting the network of potential transmissions between poultry farms. Important efforts are  
258 also necessary to ensure that prevention strategies aiming at limiting the virus spread between  
259 wild birds and poultry (such as restriction of outdoor access and providing indoor feed and  
260 drinking water) (30) are implemented during high-risk periods. Also, more sustainable  
261 strategies should be explored for poultry farms for which access to outdoor areas is part of the  
262 production specifications. The contribution of wild birds to poultry farms outbreaks was even  
263 more substantial in Czech Republic, in which the epidemic trajectories showed that the  
264 epidemic was dominated by wild bird-to-farm transmission events. Accordingly, they also

265 showed that the majority of farm-to-wild bird transmission events were from Czech Republic.  
266 This provides evidence that small size farms could be more exposed to virus transmission  
267 from wild birds than large commercial farms. Again, this could be explained by differences in  
268 farm size or other factors linked to the farm size (such as different farming practices – access  
269 to outdoor area – or biosecurity levels) which requires further attention. In wild birds, the  
270 epidemic was dominated by wild bird-to-wild bird transmission events. The number of wild  
271 bird-to-wild bird transmission events however decreased drastically from February 2017,  
272 likely linked to the decrease in wild bird density with migration to warmer climates (31) and  
273 the decrease in virus survival in the environment due to temperature-dependence of H5N8  
274 virus transmission (24).

275         We also attempted to uncover factors that could potentially predict the spread of  
276 H5N8 virus between farms across countries. However, none of the investigated predictors  
277 were identified as supportive predictor of the viral spread. This is in line with outbreak  
278 investigations on affected poultry farms in Europe, which showed that the likelihood of  
279 H5N8 virus introduction from one country to another via personnel contacts, trade of live  
280 poultry, feed, or poultry products was negligible (7), although unreported cross-border  
281 activities could not be excluded. Using phylodynamic approaches, one previous study has  
282 found geographic proximity, sharing borders and live poultry trade (when using time-  
283 dependent predictors) to be strong drivers of AI virus spread between countries in Asia (32).  
284 The comparison of this previous GLM results to our study may not be appropriate due to  
285 differences in farming systems between Europe and Asia. Also, our predictors ignore other  
286 potential drivers of virus spread, such as wild bird migration, different farming systems and  
287 biosecurity levels among countries. For example, the scattered distribution of H5N8  
288 sequences from wild birds among sequences from poultry farms of different countries on the  
289 MCC tree could support the possibility of wild birds' movements facilitating virus spread

290 between poultry farms across countries. It is also possible that transmission between  
291 countries is linked to trade of poultry products or other cryptic means that were not tested in  
292 this study due to lack of information. In the future, we recommend further investigation of  
293 predictors with a higher scale of temporal and spatial resolutions, which could allow for  
294 stronger contribution levels (32).

295         The 2016-2017 epidemic of H5N8 virus in Europe remains, like other epidemics of AI  
296 viruses, epidemiologically complex as it involved multiple wild bird species that vary in  
297 spatial ecology and clinical disease severity. During the epidemic, the virus was detected in a  
298 large number of wild bird species, mainly those of the *Anseriformes* orders (ducks, geese,  
299 swans), including mute swans (*Cygnus olor*), tufted ducks (*Aythya fuligula*), Whooper swans  
300 (*Cygnus cygnus*), Eurasian widgeons (*Mareca penelope*) and mallards (*Anas platyrhynchos*)  
301 (24). Among these species, some can be mostly sedentary in given areas while partially or  
302 wholly migratory in others (29), meaning that some species can act as sentinels in some areas  
303 or long-distance vectors of H5N8 virus in others (33). Consequently, wild bird population  
304 structure may be much more complex than what was assumed in this study. For example, on  
305 the MCC tree, we observed H5N8 sequences from wild birds both within and between  
306 clusters of sequences from poultry farms of the same country, which could illustrate the  
307 presence of sedentary and migratory wild birds, respectively. Similarly, the virus was  
308 detected in several poultry species and farm types, which may play different roles in the virus  
309 spread due to discrepancies in virus infection susceptibility and farming practices (1, 24).  
310 Unfortunately, limited information on virus prevalence or epidemiology in various domestic  
311 and wild host species between countries makes it difficult to treat species separately, thereby  
312 necessitating the grouping used here.

313         Bayesian phylogeographic approaches (34) are more common than structured  
314 phylodynamic approaches (like the MTBD model) to infer the transmission of lineages

315 between different host species, their popularity being partly associated to their computational  
316 efficiency (7, 35). However, one shortcoming of Bayesian phylogeographic approaches is the  
317 assumption of independence between the phylogeny and the transmission process, which can  
318 lead to loss of information (36–38). Another shortcoming is the assumption of proportionality  
319 between the sample sizes across sub-populations and the subpopulation sizes, which make it  
320 sensitive to biased sampling. Unlike Bayesian phylogeographic approaches, MTBD models  
321 explicitly integrate how lineages transmit within and between sub-populations while  
322 accounting for the sampling effort, making the estimations more robust to sampling bias (19,  
323 21). This has made it possible to infer transmission parameters, such as the effective  
324 reproduction number  $R_e$ , among poultry farms and wild birds based on pathogen genome  
325 sequences. We expect those estimates being useful to parameterize predictive models of virus  
326 spread aiming at optimising control strategies. We also inferred that the median farm-level  
327 infectious periods ranged from 7 to 14 days, suggesting that some countries were quicker at  
328 depopulating than others. This also emphasizes that a back-tracing window of approximately  
329 2 weeks would be sufficient to capture the period during which a farm was infectious. Only  
330 one sequence was available per poultry farm, meaning that within-farm genetic diversity was  
331 not taken into account. However, this is a reasonable assumption due to the short-period of  
332 the poultry farm outbreaks prior to detection and culling. More importantly, the present study  
333 demonstrates how relevant these models can be (i) to inform on the number of unreported  
334 infections, (ii) to reconstruct previous unobserved infections prior to the first officially  
335 reported infection and (iii) to discriminate transmission events within a given host species  
336 from incursions across species, that are more challenging using traditional wildlife and  
337 epidemiological methods (15, 39). Therefore, such phylodynamic tools can complement or  
338 even substitute for traditional epidemiological tools.

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

344

## 345 **Methods**

### 346 *Selection and alignment of sequences*

347           All H5N8 genome sequences of HA segment collected during winter 2016-2017 from  
348 four severely affected European countries (Czech Republic, Germany, Hungary and Poland)  
349 were downloaded from GISAID on Sept 1<sup>st</sup>, 2020. Only one sequence was available per  
350 poultry farm and per wild bird, meaning that the transmission dynamics of H5N8 were  
351 inferred at the farm-to-farm, wild bird-to-farm and farm-to-wild bird levels. Selected  
352 sequences were annotated with available sampling dates, locations and hosts, aligned using  
353 MAFFT v7 (40) and manually edited using AliView v1.26 (41). The final dataset consisted of  
354 190 genome sequences from infected poultry farms and 130 from infected wild birds ([SI](#)  
355 [Appendix Table S6 and Figure S3](#)).

### 356 *Phylogenetic analysis*

#### 357 *Multi-type birth-death model*

358           The multi-type birth-death (MTBD) model was fitted to the sequence alignment (19,  
359 21). Under this model, infected hosts could transmit the virus to another host from the same  
360 discrete subpopulation (with a parameter within-deme  $R_e$ ), referred to here as deme,  
361 eventually become non-infectious due to recovery or death/depopulation (with a rate  $\delta$ ), be  
362 sequenced and sampled upon becoming non-infectious (with a proportion  $s$ , and thus are  
363 included into the dataset) or could transmit the virus to another host from another deme (with

364 a parameter between-deme  $R_e$ ). All transmission, become non-infectious and sampling  
365 processes are assumed to be deme-specific and constant through time, except for the within-  
366 deme  $R_e$  that was estimated across four-time intervals, corresponding to the four phases of the  
367 epidemic (SI Appendix Figure S1).

368 Under this model, sequences were organized into five demes, according to the host  
369 type and geographical location: ‘poultry farms in Czech Republic’, ‘poultry farms in  
370 Germany’, ‘poultry farms in Hungary’, ‘poultry farms in Poland’ and ‘wild birds in the four  
371 countries’ (SI Appendix Table S6). All sequences from wild birds were aggregated into one  
372 deme (not depending on the geographical location as for poultry farms) since it was assumed  
373 that the majority of sampled wild bird species (mainly mallards and swans) could move freely  
374 among countries (42). It was assumed that once sampled, a given host could not be infected  
375 and sampled again since infected poultry farms were subject to culling following the  
376 confirmation of infection and sampling of wild birds was from a mortality event (24).

377 The prior values and distributions of the model parameters are described in SI  
378 Appendix Table S2. The MTBD was specified as the tree prior using a HKY +  $\Gamma_4$  nucleotide  
379 substitution process with a relaxed molecular clock (43) defined by a Lognormal(0.001, 1.25)  
380 prior (4, 5, 44). The origin of the tree was given a Lognormal(-0.2,0.2) prior, corresponding  
381 to the median date 1<sup>st</sup> of July 2016 (95% HPD: 6<sup>th</sup> of February 2016 – 19<sup>th</sup> of October 2016)  
382 (2, 4, 5) and assumed to be associated to the deme ‘wild birds in the four countries’, since the  
383 source of the first poultry farm outbreak in the four countries was likely attributed to infected  
384 migratory wild birds from Northern Eurasia (24). All  $R_e$  parameters were given a  
385 Lognormal(0,1) prior (9, 45, 46). The become non-infectious rate was given a  
386 Lognormal(52,0.6) prior (46–48). For each deme, the sampling proportion was given a  
387 Uniform distribution prior with lower and upper bounds informed by the number of  
388 sequences and reported poultry farm outbreaks/wild bird cases (25). Given the severity of the

389 clinical signs affecting the majority of poultry combined with active surveillance around  
390 reported poultry farm outbreaks (24), the number of unreported poultry farm outbreaks was  
391 considered relatively low in all countries. On the contrary, given the difficulty of catching  
392 and sampling wild birds, it was assumed that infected wild birds were significantly under-  
393 sampled, relative to poultry farms.

#### 394 *Predictors of H5N8 virus spread between poultry farms across borders*

395 The MTBD model was extended with a generalized linear model (GLM) to inform the  
396 H5N8 virus spread between poultry farms across borders by 19 time-independent predictors  
397 (26, 27): the 2016 live poultry trade (28), the 2016 poultry density in the source and  
398 destination deme (28), the 2014 poultry farm density in the source and destination deme (24),  
399 the 2017 farm outbreak density in the source and destination deme (25), the 2021 human  
400 density in the source and destination deme (49), whether two countries shared borders and the  
401 distance between countries' centroids. To account for potential missing predictors, we also  
402 included predictors to assess the virus spread from or to one individual country (SI Appendix  
403 Table S5). In this GLM parametrization, the between-deme  $R_e$  parameters act as the outcome  
404 to a log-linear function of the predictors. For each predictor  $i$ , the GLM parametrization also  
405 includes a regression coefficient  $\beta$  which quantifies the (log) contribution of the predictor  
406 and a binary indicator variable  $\delta_i$  which quantifies the probability of the predictor to be  
407 included in the model (SI Appendix Table S2). To avoid collinearity among predictors,  
408 predictors were removed when the Pearson correlation exceeded  $> 0.7$  (SI Appendix Figure  
409 S4). To reduce the effect of different predictors' magnitude, all non-binary predictors were  
410 log-transformed and standardized before inclusion in the GLM. Bayes Factors (BF) were  
411 used to determine the contribution of each predictor in the GLM (26, 50, 51). BF were  
412 calculated for each predictor to quantify which of the posterior and prior inclusion  
413 probabilities of the given predictor in the model ( $\delta_i = 1$ ) is more likely. The cutoff for

414 substantial contribution of a given predictor in the GLM was set at 3.2 (51), meaning that its  
415 posterior inclusion probability in the model was 3.2-fold more likely than its prior inclusion  
416 probability (0.50).

#### 417 *Inference of MTBD model parameters, structured trees and epidemic trajectories*

418         Phylogenetic analysis was implemented using the BDMM-Prime package (52) for  
419 BEAST v2.6.3 (53) and the BEAGLE library (54) to improve computational performance.  
420 All analyses were run for 40-50 million steps across three independent Markov chains  
421 (MCMC) and states were sampled every 10,000 steps. The first 10% of steps from each  
422 analysis were discarded as burn-in before states from the chains were pooled using Log-  
423 Combiner v2.6.3 (53). Convergence was assessed in Tracer v1.7 (55) by ensuring that the  
424 estimated sampling size (ESS) values associated with the estimated parameters were all >  
425 200.

426         The structured trees (i.e. when phylogenetic trees are associated with a specific deme  
427 along their branches (22)) were inferred by applying a stochastic mapping algorithm (56)  
428 implemented in BDMM-Prime (52) to a subsampled set of posterior structured trees and  
429 model parameters (n=500) generated by the MTBD analysis. The MCC tree was obtained  
430 from the structured trees in TreeAnnotator v2.6.3 (53), and annotated using the ggtree  
431 package (57) in R v4.0.2 (58). For each set of posterior model parameters set and associated  
432 structured tree, an epidemic trajectory (i.e. corresponding to the sequence of transmission,  
433 become non-infectious and sampling events that occur throughout a given epidemic (23)) was  
434 drawn from the distribution of such trajectories conditional on the tree and parameters (52).

435

#### 436 **Data availability**

437 All H5N8 genome sequences of HA segment are available on the GISAID database  
438 (<https://www.gisaid.org>). The prior values and distributions of the model parameters are

439 described in [SI Appendix Table S2](#). Details on the predictor data are available in [SI Appendix](#)  
440 [Table S5](#). The BEAST 2 XML file used to perform the phylodynamic analysis, together with  
441 the R scripts are available from [https://github.com/ClaireGuinat/h5n8\\_bdmm-prime.git](https://github.com/ClaireGuinat/h5n8_bdmm-prime.git).

442

443

#### 444 **Acknowledgements**

445 The authors are grateful to the cEvo group (ETH Zurich, Switzerland) for providing  
446 useful comments on this project. They also acknowledge the EFSA Public Access to  
447 documents Team for providing data on the poultry farm census in Europe. This project has  
448 received funding from the European Union's Horizon 2020 research and innovation  
449 programme under the Marie Skłodowska-Curie grant agreement No 842621.

450

451 **Competing interests:** The authors declare no competing interests.

452

#### 453 **References**

- 454 1. S. Napp, N. Majó, R. Sánchez-González, J. Vergara-Alert, Emergence and spread of  
455 highly pathogenic avian influenza A(H5N8) in Europe in 2016-2017. *Transbound.*  
456 *Emerg. Dis.* **65**, 1217–1226 (2018).
- 457 2. E. Świętoń, K. Śmietanka, Phylogenetic and molecular analysis of highly pathogenic  
458 avian influenza H5N8 and H5N5 viruses detected in Poland in 2016–2017. *Transbound.*  
459 *Emerg. Dis.* **65**, 1664–1670 (2018).
- 460 3. A. Globig, *et al.*, Highly Pathogenic Avian Influenza H5N8 Clade 2.3.4.4b in Germany  
461 in 2016/2017. *Front. Vet. Sci.* **4** (2018).
- 462 4. A. Fusaro, *et al.*, Genetic Diversity of Highly Pathogenic Avian Influenza  
463 A(H5N8/H5N5) Viruses in Italy, 2016–17. *Emerg. Infect. Dis.* **23**, 1543–1547 (2017).
- 464 5. N. Beerens, *et al.*, Multiple Reassorted Viruses as Cause of Highly Pathogenic Avian  
465 Influenza A(H5N8) Virus Epidemic, the Netherlands, 2016. *Emerg. Infect. Dis.* **23**,  
466 1974–1981 (2017).
- 467 6. P. Mulatti, *et al.*, Integration of genetic and epidemiological data to infer H5N8 HPAI  
468 virus transmission dynamics during the 2016-2017 epidemic in Italy. *Sci. Rep.* **8**, 18037  
469 (2018).

- 470 7. S. Lycett, *et al.*, Global Consortium for H5N8 and Related Influenza Viruses. *Role*  
471 *Migr. Wild Birds Glob. Spread Avian Influenza H5N8 Sci.* **354**, 213–217 (2016).
- 472 8. C. Guinat, *et al.*, Role of Live-Duck Movement Networks in Transmission of Avian  
473 Influenza, France, 2016–2017. *Emerg. Infect. Dis.* **26**, 472–480 (2020).
- 474 9. A. Andronico, *et al.*, Highly pathogenic avian influenza H5N8 in south-west France  
475 2016–2017: A modeling study of control strategies. *Epidemics* **28**, 100340 (2019).
- 476 10. C. Guinat, *et al.*, Biosecurity risk factors for highly pathogenic avian influenza (H5N8)  
477 virus infection in duck farms, France. *Transbound. Emerg. Dis.* **67**, 2961–2970 (2020).
- 478 11. C. Guinat, N. Rouchy, F. Camy, J. L. Guérin, M. C. Paul, Exploring the Wind-Borne  
479 Spread of Highly Pathogenic Avian Influenza H5N8 During the 2016–2017 Epizootic in  
480 France. *Avian Dis.* **63**, 246–248 (2018).
- 481 12. A. Scoizec, *et al.*, Airborne Detection of H5N8 Highly Pathogenic Avian Influenza  
482 Virus Genome in Poultry Farms, France. *Front. Vet. Sci.* **5** (2018).
- 483 13. E. M. Volz, K. Koelle, T. Bedford, Viral phylodynamics. *PLoS Comput Biol* **9**,  
484 e1002947 (2013).
- 485 14. L. du Plessis, T. Stadler, Getting to the root of epidemic spread with phylodynamic  
486 analysis of genomic data. *Trends Microbiol.* **23**, 383–386 (2015).
- 487 15. C. Guinat, *et al.*, What can phylodynamics bring to animal health research? *Trends Ecol.*  
488 *Evol.* (2021).
- 489 16. G. Dudas, L. M. Carvalho, A. Rambaut, T. Bedford, MERS-CoV spillover at the camel-  
490 human interface. *Elife* **7**, e31257 (2018).
- 491 17. N. R. Faria, *et al.*, Genomic and epidemiological monitoring of yellow fever virus  
492 transmission potential. *Science* **361**, 894–899 (2018).
- 493 18. S. A. Nadeau, T. G. Vaughan, J. Scire, J. S. Huisman, T. Stadler, The origin and early  
494 spread of SARS-CoV-2 in Europe. *Proc. Natl. Acad. Sci.* **118** (2021).
- 495 19. D. Kühnert, T. Stadler, T. G. Vaughan, A. J. Drummond, Phylodynamics with  
496 migration: a computational framework to quantify population structure from genomic  
497 data. *Mol. Biol. Evol.* **33**, 2102–2116 (2016).
- 498 20. R. M. Anderson, R. M. May, Population biology of infectious diseases: Part I. *Nature*  
499 **280**, 361–367 (1979).
- 500 21. J. Scire, J. Barido-Sottani, D. Kühnert, T. G. Vaughan, T. Stadler, Improved multi-type  
501 birth-death phylodynamic inference in BEAST 2. *BioRxiv* (2020).
- 502 22. T. G. Vaughan, D. Kühnert, A. Poppinga, D. Welch, A. J. Drummond, Efficient Bayesian  
503 inference under the structured coalescent. *Bioinforma. Oxf. Engl.* **30**, 2272–2279 (2014).
- 504 23. T. G. Vaughan, *et al.*, Estimating epidemic incidence and prevalence from genomic  
505 data. *Mol. Biol. Evol.* **36**, 1804–1816 (2019).

- 506 24. EFSA, *et al.*, Avian influenza overview October 2016–August 2017. *EFSA J.* **15**,  
507 e05018 (2017).
- 508 25. FAO, Global Animal Disease Information System (Empres-i) [Available at:  
509 <https://empres-i.review.fao.org/#/> (accessed in January 2021)] (2021).
- 510 26. P. Lemey, *et al.*, Unifying viral genetics and human transportation data to predict the  
511 global transmission dynamics of human influenza H3N2. *PloS Pathog* **10**, e1003932  
512 (2014).
- 513 27. N. F. Müller, G. Dudas, T. Stadler, Inferring time-dependent migration and coalescence  
514 patterns from genetic sequence and predictor data in structured populations. *Virus Evol.*  
515 **5**, vez030 (2019).
- 516 28. FAOSTAT, FAOSTAT - production and trade of live animals [Available at:  
517 <http://www.fao.org/faostat/en/#compare> (Accessed in May 2021)] (2021).
- 518 29. BTO, British Trust for Ornithology (BTO) <https://www.bto.org> [Accessed in May 2021]  
519 (2017).
- 520 30. M. Artois, *et al.*, Outbreaks of highly pathogenic avian influenza in Europe: the risks  
521 associated with wild birds. *Rev. Sci. Tech.* **28**, 69 (2009).
- 522 31. N. J. Hill, *et al.*, Transmission of influenza reflects seasonality of wild birds across the  
523 annual cycle. *Ecol. Lett.* **19**, 915–925 (2016).
- 524 32. J. Yang, D. Xie, Z. Nie, B. Xu, A. J. Drummond, Inferring host roles in bayesian  
525 phylodynamics of global avian influenza A virus H9N2. *Virology* **538**, 86–96 (2019).
- 526 33. J. Keawcharoen, *et al.*, Wild ducks as long-distance vectors of highly pathogenic avian  
527 influenza virus (H5N1). *Emerg. Infect. Dis.* **14**, 600 (2008).
- 528 34. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian Phylogeography  
529 Finds Its Roots. *PLOS Comput. Biol.* **5**, e1000520 (2009).
- 530 35. N. S. Trovão, M. A. Suchard, G. Baele, M. Gilbert, P. Lemey, Bayesian Inference  
531 Reveals Host-Specific Contributions to the Epidemic Expansion of Influenza A H5N1.  
532 *Mol. Biol. Evol.* **32**, 3264–3275 (2015).
- 533 36. N. De Maio, C.-H. Wu, K. M. O'Reilly, D. Wilson, New Routes to Phylogeography: A  
534 Bayesian Structured Coalescent Approximation. *PLoS Genet.* **11**, e1005421 (2015).
- 535 37. S. Bloomfield, *et al.*, Investigation of the validity of two Bayesian ancestral state  
536 reconstruction models for estimating Salmonella transmission during outbreaks. *Plos*  
537 *One* **14**, e0214169 (2019).
- 538 38. N. F. Müller, D. A. Rasmussen, T. Stadler, The Structured Coalescent and Its  
539 Approximations. *Mol. Biol. Evol.* **34**, 2970–2981 (2017).
- 540 39. J. A. Blanchong, S. J. Robinson, M. D. Samuel, J. T. Foster, Application of genetics and  
541 genomics to wildlife epidemiology. *J. Wildl. Manag.* **80**, 593–608 (2016).

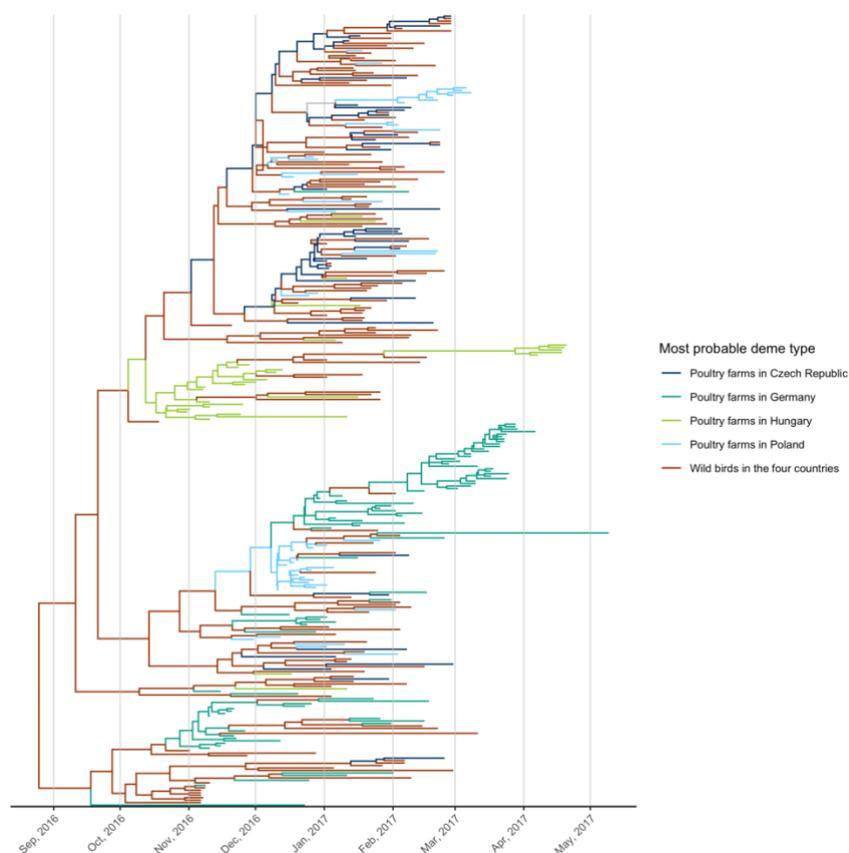
- 542 40. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7:  
543 Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- 544 41. A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large  
545 datasets. *Bioinformatics* **30**, 3276–3278 (2014).
- 546 42. P. W. Atkinson, *et al.*, “Urgent preliminary assessment of ornithological data relevant to  
547 the spread of Avian Influenze in Europe” (Wetlands International, 2006).
- 548 43. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and  
549 dating with confidence. *PLoS Biol.* **4**, e88 (2006).
- 550 44. P. Alarcon, *et al.*, Comparison of 2016–17 and previous epizootics of highly pathogenic  
551 avian influenza H5 Guangdong lineage in Europe. *Emerg. Infect. Dis.* **24**, 2270 (2018).
- 552 45. I. Iglesias, *et al.*, Reproductive ratio for the local spread of highly pathogenic avian  
553 influenza in wild bird populations of Europe, 2005–2008. *Epidemiol. Infect.* **139**, 99–  
554 104 (2011).
- 555 46. D. A. Grear, J. S. Hall, R. J. Dusek, H. S. Ip, Inferring epidemiologic dynamics from  
556 viral evolution: 2014–2015 Eurasian/North American highly pathogenic avian influenza  
557 viruses exceed transmission threshold,  $R_0 = 1$ , in wild birds and poultry in North  
558 America. *Evol. Appl.* **11**, 547–557 (2018).
- 559 47. C. Leyson, *et al.*, Pathogenicity and genomic changes of a 2016 European H5N8 highly  
560 pathogenic avian influenza virus (clade 2.3. 4.4) in experimentally infected mallards and  
561 chickens. *Virology* **537**, 172–185 (2019).
- 562 48. K. Willgert, *et al.*, Transmission of highly pathogenic avian influenza in the nomadic  
563 free-grazing duck production system in Viet Nam. *Sci. Rep.* **10**, 1–11 (2020).
- 564 49. Wikipedia, World population [Available at:  
565 [https://en.wikipedia.org/wiki/World\\_population](https://en.wikipedia.org/wiki/World_population) (Accessed in July 2021)] (2021).
- 566 50. D. Magee, R. Beard, M. A. Suchard, P. Lemey, M. Scotch, Combining phylogeography  
567 and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion.  
568 *Arch. Virol.* **160**, 215–224 (2015).
- 569 51. R. E. Kass, A. E. Raftery, Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
- 570 52. T. Vaughan, BDMM Prime (In prep.).
- 571 53. R. Bouckaert, *et al.*, BEAST 2: a software platform for Bayesian evolutionary analysis.  
572 *PLoS Comput Biol* **10**, e1003537 (2014).
- 573 54. D. L. Ayres, *et al.*, BEAGLE: an application programming interface and high-  
574 performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173  
575 (2012).
- 576 55. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior  
577 summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901 (2018).

- 578 56. W. A. Freyman, S. Höhna, Stochastic character mapping of state-dependent  
579 diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae  
580 lineages. *Syst. Biol.* **68**, 505–519 (2019).
- 581 57. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree: an R package for visualization  
582 and annotation of phylogenetic trees with their covariates and other associated data.  
583 *Methods Ecol. Evol.* **8**, 28–36 (2017).
- 584 58. R. C. Team, R: A language and environment for statistical computing (2013).

585

586

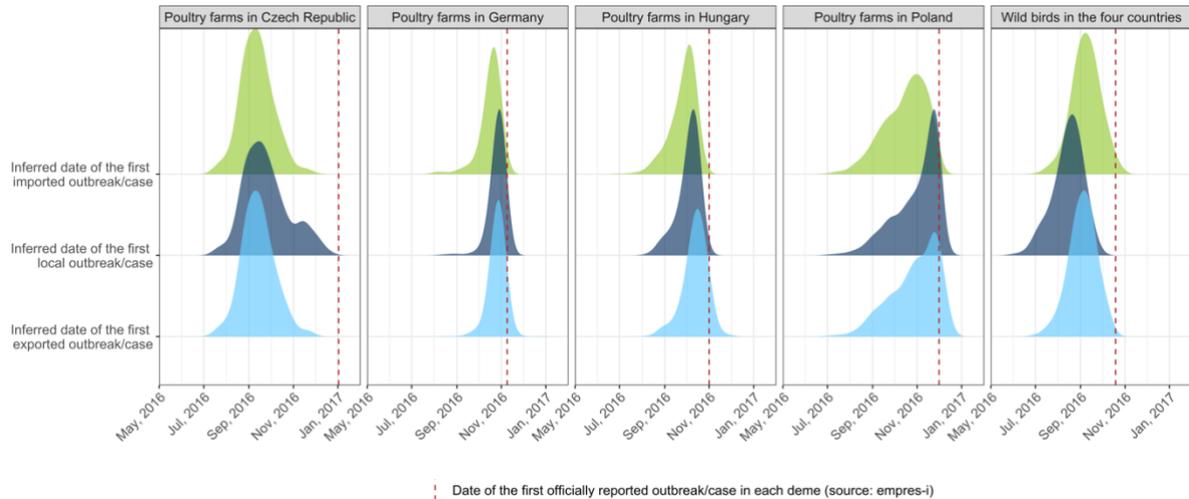
## 587 **Figures**



588

589 **Figure 1.** Time-scaled maximum clade credibility phylogenetic tree of the HA segment of  
590 HPAI H5N8 virus sequenced from poultry farms and wild birds during the 2016-2017  
591 epidemic in Czech Republic, Germany, Hungary and Poland. The color of the tree branches  
592 shows the deme type with the highest probability (see legend) and uncertainty in deme type  
593 assignment is shown in grey. There is evidence for virus spread among neighbouring poultry

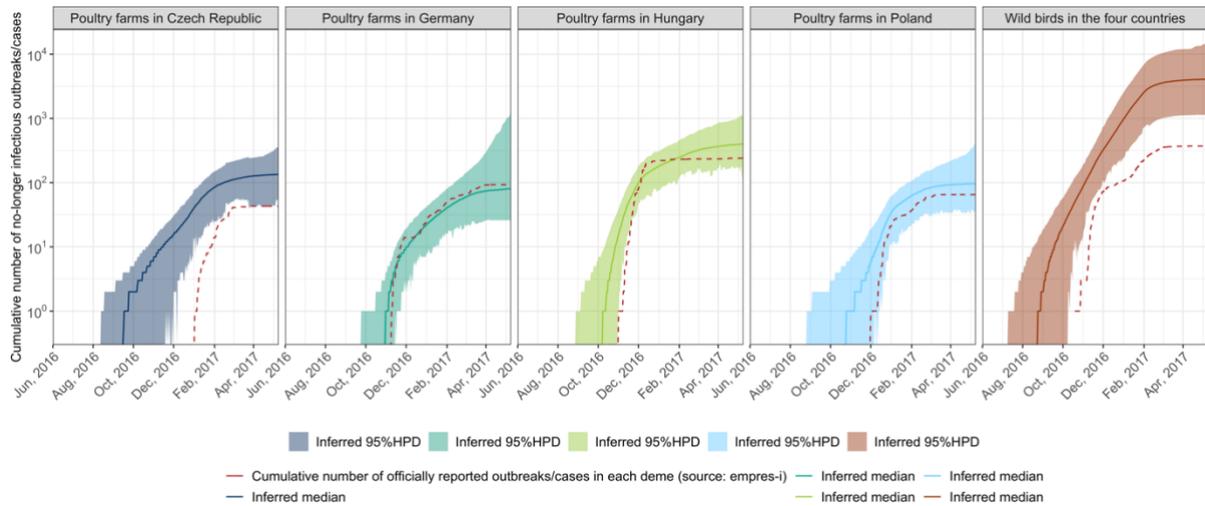
594 farms illustrated by the presence of clusters of H5N8 sequences from poultry farms of the  
595 same country (mainly Germany, Hungary and Poland) in the tree, with the possibility of wild  
596 birds' movements facilitating virus spread between poultry farms across countries, illustrated  
597 by the dispersal distribution of H5N8 sequences from wild birds.  
598



599

600 **Figure 2.** Temporal distribution of the inferred date of the first imported, local and  
601 exported outbreak/case per deme. The red dashed line represents the first officially reported  
602 outbreak/case per deme for comparison. In this graph, for each trajectory and each deme, we  
603 extracted the date of first imported, local and exported outbreak/case and summarized them  
604 over time in a probability density.

605



606

607 **Figure 3.** Temporal distribution of the inferred cumulative number of no longer infectious

608 outbreaks/cases per deme. The solid line represents the median inferred, the colored areas

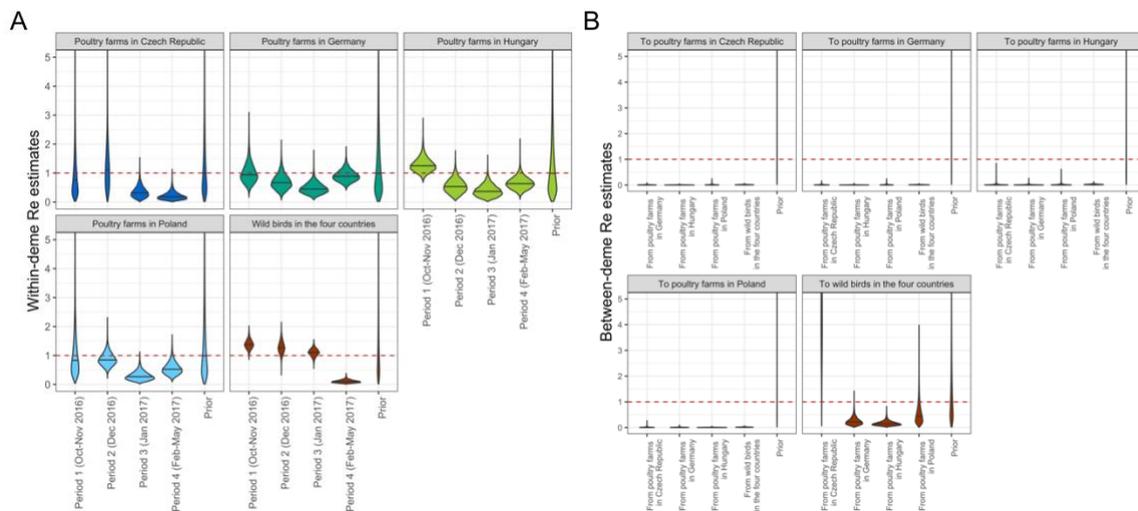
609 represent the 95% HPD. The red dashed line represents the cumulative number of officially

610 reported outbreaks/cases in log scale (25). In this graph, for each trajectory and each deme,

611 we extracted the cumulative number of become non-infectious events and summarized them

612 over time in log scale.

613



614

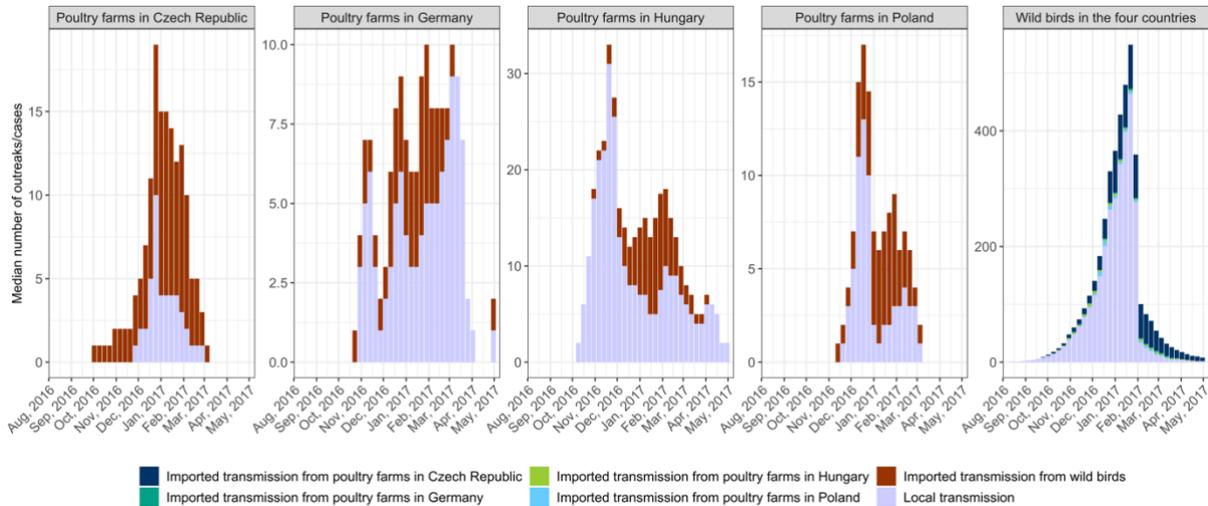
615 **Figure 4.** A) Posterior distributions for the within-deme  $R_e$  values across four-time intervals.

616 Solid horizontal lines represent median values and the dashed red line represents the

617 threshold between epidemic growth and fade out. B) Posterior distributions for the between-

618 deme  $R_e$  values. Solid horizontal lines represent median values and the dashed red line  
 619 represents the threshold between epidemic growth and fade out.

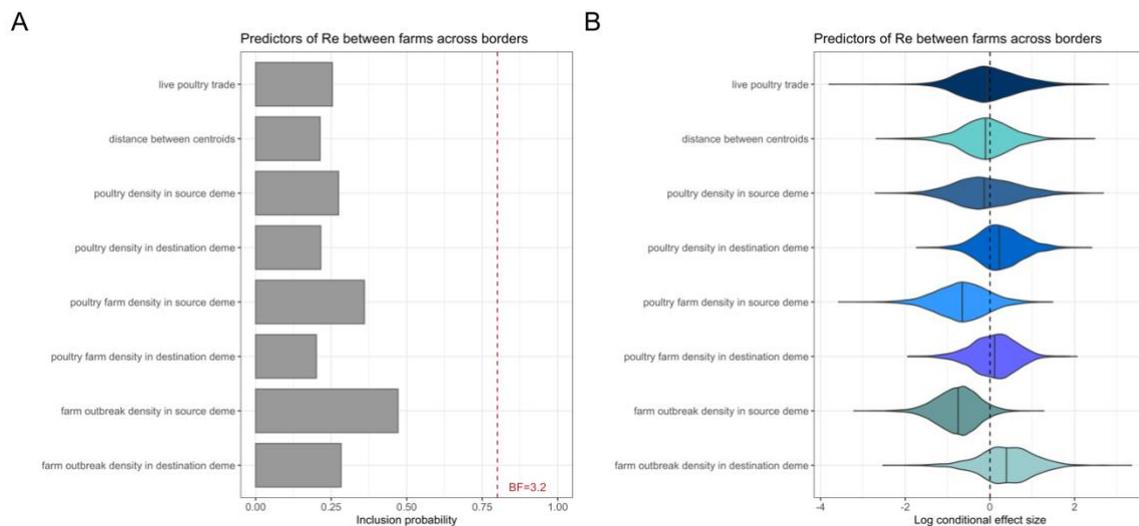
620



621

622 **Figure 5.** Temporal distribution of the inferred median number of poultry farm outbreaks and  
 623 wild bird cases arising from local transmission and importation events per deme. In this  
 624 graph, for each trajectory and each deme, we computed the median number of within-deme  
 625 and between-deme transmission events over time.

626



627

628 **Figure 6.** A) Inclusion probability for predictors of the between-farm H5N8 virus spread  
 629 across borders. This represents the proportion of the posterior samples in which each

630 predictor was included in the model. Bayes Factors (BF) were used to determine the  
631 contribution of each predictor in the GLM. BF were calculated for each predictor to quantify  
632 which of the posterior and prior inclusion probabilities of the given predictor in the model is  
633 more likely. The cutoff for substantial contribution of a given predictor in the GLM was set at  
634 3.2 (51). B) Log conditional effect sizes for predictors of the between-farm H5N8 virus  
635 spread across borders. This represents the (log) contribution of each predictor when the  
636 corresponding predictor was included in the model ( $\beta_i | \delta_i = 1$ ), where  $\beta_i$  is the coefficient and  
637 the binary indicator  $\delta_i$  for each predictor  $i$ .  
638