1 TITLE: Agricultural and geographic factors shaped the North American 2015 highly pathogenic
2 avian influenza H5N2 outbreak
3 SHORT TITLE: Agricultural and geographic factors of the 2015 HPAI US Outbreak
4 AUTHORs: Joseph T. Hicks[1], Dong-Hun Lee[2], Venkata R. Duvuuri[1], Mia Kim Torchetti[3], David
5 E Swayne[4], Justin Bahl[1,5]
6
7 AUTHOR AFFILIATIONS: [1]Center for Ecology of Infectious Diseases, Department of
8 Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, Georgia,
9 United States; [2]Department of Pathobiology and Veterinary Science, the University of
10 Connecticut, Storrs, Connecticut, United States; [3]U.S. Department of Agriculture, Ames, Iowa,
11 United States; [4]Exotic and Emerging Avian Viral Diseases Research Unit, U.S. National Poultry
12 Research Center, Agricultural Research Service, U.S. Department of Agriculture, Athens,
13 Georgia, United States; [5]Duke-NUS Graduate Medical School, Singapore.
14 CORRESPONDING AUTHORS:
15 Justin Bahl, justin.bahl@uga.edu
16 KEYWORDS: highly pathogenic avian influenza, phylogenetics, phylodynamcs, avian disease,
17 Bayesian inference, coalescent models, discrete trait evolution, generalized linear model
18

ABSTRACT

The 2014 – 2015 highly pathogenic avian influenza (HPAI) H5NX outbreak represents the largest and most expensive HPAI outbreak in the United States to date. Despite extensive traditional and molecular epidemiological studies, factors associated with the spread of HPAI among midwestern poultry premises remain unclear. To better understand the dynamics of this outbreak, 182 full genome HPAI H5N2 sequences isolated from commercial layer chicken and turkey production premises were analyzed using evolutionary models modified to incorporate epidemiological and geographic information. Epidemiological compartmental models constructed in a phylogenetic framework provided evidence that poultry type acted as a barrier to the transmission of virus among midwestern poultry farms. Furthermore, after initial introduction, a continuous external source of virus was not needed to explain the propagation of HPAI cases within the commercial poultry industries. Discrete trait diffusion models indicated that within state viral transitions occurred more frequently than inter-state transitions. Distance, road density and proportion of water coverage were all supported as associated with viral transition between county groups (Bayes Factor > 3.0). Together these findings indicate that the midwestern poultry industries were not a single homogenous population, but rather, the outbreak was shaped by poultry sectors and geographic factors.

AUTHOR SUMMARY

The highly pathogenic avian influenza outbreak among poultry farms in the midwestern United States appears to be influenced by agricultural and geographic factors. After initial introduction of the virus into the poultry industries, no further introductions (such as from a wild bird reservoir) were necessary to explain the continuation of the outbreak from March to June 2015. Additionally, evidence suggests that proximity increases the chances of viral movement between

two locations. While many hypotheses have been proposed to explain the transmission of virus among poultry farms, the support for road density as an important driver of viral movement suggests human-mediated viral transportation played a key role in the spread of the highly pathogenic H5N2 outbreak in North America.

INTRODUCTION

In 2014, a novel reassortant highly pathogenic avian influenza (HPAI) H5N8 virus of the hemagglutinin (HA) clade 2.3.4.4 was identified in South Korean poultry and wild birds and quickly spread to other Asian countries and Europe (1–3). By the end of 2014, both the Eurasian H5N8 virus and its reassortant H5N2 containing Eurasian- and North American-origin gene segments, were reported in western Canada and the United States (4–6). The ensuing 2014-2015 North American HPAI outbreak marked the largest and most expensive HPAI outbreak in the United States to date (7). In late November 2014, commercial poultry flocks in British Columbia, Canada were reported to be infected with the novel reassortant HPAI H5N2 (5), soon followed by HPAI H5N8 isolation within wild birds in the United States Pacific Northwest (4). Over the next several months, sporadic infections arose in wild and domestic birds, including both commercial production and backyard poultry operations. In March 2015, a drastic increase of HPAI H5N2 cases was observed within domestic poultry in the Midwestern United States. By the end of the outbreak in June 2015, over 50.4 billion poultry died or were culled due to the outbreak, costing the US government over $850 million, the poultry industries an estimated $700 million to $1 billion and had a negative $3.3 billion impact on the economy (7–9).

Risk factors that explain the continued transmission of HPAI between domestic poultry facilities remain unclear. For instance, previous analyses have provided conflicting evidence as to the role of wild birds in the propagation of the outbreak within the midwestern poultry industries. Despite frequent reports of wild birds on the grounds and within barns of HPAI-positive turkey premises (10), a case-control study found no significant difference in exposure to wild birds between positive turkey premises and matched controls (11). Similarly, one phylodynamic analysis found no evidence of continued HPAI introductions into the midwestern

poultry industries (12), but other models have suggested multiple introductions (13,14). Geographic and environmental variables, such as human population, agricultural, climatological, and ecological measures, may help explain farm-to-farm transmission observed within the poultry industries. For example, proximity between midwestern poultry premises has been implicated as an important risk factor for HPAI infection (11,13). Although it has been suggested that poultry production type did not affect outbreak transmission (12), this has not been formally tested. Despite extensive molecular epidemiological studies, such environmental and ecological covariates of viral spread during this outbreak have not been investigated.

Direct epidemiological links between most poultry premises have not been established (15), limiting the ability to investigate risk factors that facilitated HPAI transmission among poultry farms. The incorporation of pathogen genetic sequence data into epidemiological investigations can elucidate network connections between infectious entities, be that individual hosts or populations, such as poultry farms. One approach is viral phylodynamic modeling, ie. the integration of epidemiological and evolutionary models to explore viral ecological dynamics. Based on the assumption that viral epidemiology and evolution occur on the same time scale, viral phylodynamic modeling can reveal underlying population structure and epidemiological parameters. Recent incorporation of generalized linear models (GLM), a family of commonly used regression methods, into Bayesian phylogenetic frameworks have enabled investigation into the impact of ecological factors on the geographic diffusion of viral pathogens (16,17). Through such an approach, factors associated with HPAI movement within United States poultry industries can be identified, informing future control efforts.

In this study, we integrated epidemiological and ecological parameters with genomic sequence data collected contemporaneously with the midwestern poultry industries HPAI

92 outbreak to formally test outstanding hypotheses. Whole genome HPAI H5N2 sequence data
93 isolated from layer chicken and turkey premises were analyzed using evolutionary model-based
94 techniques. First, we developed population models to test the importance of poultry sector
95 divisions (i.e. layer chicken vs turkey industries) and external viral introductions from an
96 unsampled avian population in the propagation of the outbreak. Second, we evaluated ecological
97 predictors of geographic diffusion of virus among midwestern counties to help identify
98 environmental and human variables associated with viral transmission. Together, these analyses
99 use information that accumulated within the HPAI H5N2 genome during the outbreak to help
100 decipher higher-order patterns of viral dispersal among commercial poultry farms.

101 RESULTS

102 *HPAI H5N2 Evolution within Domestic Poultry*

103 182 full genome HPAI H5N2 genetic sequences, each representing a single commercial
104 poultry farm operation across 49 counties in six states (Iowa, Minnesota, Nebraska, North
105 Dakota, South Dakota, and Wisconsin), were included in the present analysis. The sequences
106 were isolated from samples collected between March 25 and June 15, 2015 from positive turkey
107 premises (72.5%) and layer chicken farms (27.5%). Two molecular clock assumptions and three
108 "traditional" coalescent models (i.e., constant population, exponential growth, and extended
109 Bayesian skyline plot [EBSP]) were compared with marginal likelihood estimation (MLE) to
110 evaluate the underlying population and evolutionary dynamics of the 2015 HPAI outbreak. The
111 highly flexible EBSP coalescent with a strict molecular clock assumption had the best fit for the
112 included sequence data (log(MLE) = -25884.06). The supported molecular clock assumption
113 varied depending on the coalescent model employed. Statistical support between the relaxed and
114 strict molecular clock assumptions was ambivalent for the constant and exponential coalescent

115 models (log Bayes factor of relaxed compared to strict molecular clock ($logBF_{R-S}$) = 0.06 and -
116 0.3, respectively; Fig 1A). Stronger evidence for the strict clock was observed when the EBSP
117 coalescent was implemented ($logBF_{R-S}$ = -4.45). Similarity between molecular clocks was also
118 demonstrated by the limited impact of the molecular clock assumptions on phylogenetic tree
119 parameter estimates such as evolutionary rate and time to the most recent common ancestor
120 (TMRCA; Fig 1B and 1C). For example, under the EBSP coalescent, the relaxed mean clock rate
121 was $6.84x10^{-3}$ substitutions per site per year (95% highest posterior density (HPD): $6.09x10^{-3}$ -
122 $7.59x10^{-3}$) compared to the strict clock rate estimate of $6.77x10^{-3}$ substitutions per site per year
123 (95% HPD: $5.98x10^{-3}$ - $7.56x10^{-3}$). In contrast, selection of the coalescent model influenced the
124 TMRCA of the included sequences. Under the strict molecular clock, EBSP coalescent models
125 estimated the TMRCA as March 1, 2015 (95% HPD: February 16 to March 10, 2015) while the
126 remaining traditional coalescent models had a TMRCA at least two weeks earlier (Fig 1C).
127
128 Fig 1. Evolutionary history of HPAI H5N2 isolated from commercial poultry premises, 2015.
129 (A) Bayes factor (BF) tests between molecular clock and coalescent evolutionary models.
130 For each coalescent model (exponential growth [Expo] and extended Bayesian skyline
131 plot [EBSP]), BF was calculated using the constant coalescent model as reference (Const,
132 indicated with asterisk) under the same molecular clock model. Two horizontal gray
133 reference lines denote log(BF) = 1 and log(BF) = 5, which represent support and very
134 strong support, respectively, for improved fit over the reference. (B) Molecular clock rate
135 (substitutions per site per year) comparison between molecular clock and coalescent
136 evolutionary models. (C) The estimated time of the most recent common ancestor
137 (TMRCA; decimal year) compared between molecular clock and coalescent evolutionary

models. (D) Maximum clade credibility tree representing the ancestral reconstruction of poultry industry (layer chicken vs. turkey) across the evolutionary history of the outbreak. The ancestral reconstruction assumed an EBSP coalescent and strict molecular clock evolutionary model. Tree branches are colored based on the most probable poultry industry of the descendant node. Thin gray node bars represent the 95% highest posterior density (HPD) of the node height (i.e., the time at which that ancestor is estimated to have existed).

HPAI H5N2 Host Dispersion and Population Dynamics

To explore the extent of viral dispersal between poultry industries, multiple phylogenetic-based methods were performed: the structured coalescent, the discrete trait diffusion model, and epidemiologic compartmental model-based coalescent. Each of these methods estimate a different approximation for the dispersal of virus between populations. The structured coalescent treats layer chicken premises and turkey premises as separate population demes between which virus was allowed to "migrate," and thus estimates a *migration rate* between the two demes. In contrast, discrete trait diffusion models treat the trait of interest (here, poultry industry) as a characteristic that evolves over time, inferring a *transition rate*, analogous to a nucleotide substitution model. Finally, compartmental models enable the calculation of *transmission rates* between the two poultry compartments. Although all approximate the amount of viral dispersal among the poultry industries, each measure is calculated differently with unique assumptions and so are referred to by a particular term. All methods estimated that viral dispersal from layer chicken premises to turkey premises occurred more frequently than from turkey premises to layer chicken (Supplemental Table S1). In the structured coalescent, the migration rate from

161 layer chicken to turkey premises was much greater than the reverse (migration rate from
162 chickens to turkeys: 12.6, 95% HPD: 6.2 – 18.7; migration rate from turkeys to chickens: 0.7,
163 95% HPD: 0.00001 – 2.2). The transition rates between the poultry industries estimated from the
164 discrete trait diffusion model were much more similar to each other (transition rate from
165 chickens to turkeys: 1.4, 95% HPD: 0.04 – 3.9; transition rate from turkeys to chickens: 0.3, 95%
166 HPD: 0.003 – 0.9). These models suggest the dispersion of virus between poultry industries was
167 not symmetrical, potentially indicating poultry type played a role in the outbreak dynamics.
168 To formally test this hypothesis, we used epidemiological compartmental model
169 equations to describe the coalescent process (18). Four competing scenarios were constructed
170 (Fig 2A). Models 1 and 2 described a homogenous poultry population that differed by the
171 presence of a continuous external viral source in Model 2. In contrast, Models 3 and 4 described
172 a host population stratified by poultry production system, again differing based on an external
173 viral source in Model 4. It should be noted that due to the sampling scheme of genetic sequences
174 (one HPAI whole genome sequence per infected premises), the epidemiologic unit of interest
175 was the premises (or farm), and not the individual bird. That is, findings of the compartmental
176 models should be interpreted on the farm-to-farm scale, not the dynamics of transmission
177 between individual birds. Akaike's information criteria for Markov chain Monte Carlo (AICM)
178 calculated from the posterior sample of structured tree likelihood estimates revealed that Model 3
179 provided the best fit for the data under both strict and relaxed molecular clock assumptions
180 (AICM under strict clock = 330.1; under relaxed clock = 376.3; Fig 2B, Supplemental Table S2).
181 This suggests the midwestern portion of the 2015 HPAI outbreak was isolated from external
182 sources but most likely structured by poultry production system. Four transmission rates were
183 estimated for Model 3 to describe the interaction between the layer chicken and turkey

184 populations: two within-poultry system rates (β_T and β_C) and two between-poultry system rates
185 (β_{TC} and β_{CT}). The model estimated the transmission rates within the turkey production system to
186 be highest (β_T = 11.6, 95%HPD: 2.0 – 22.0), followed by transmission rates from chicken farms
187 to turkey farms (β_{CT} = 4.9, 95% HPD: 0.6 – 9.6). The lowest transmission rate was estimated
188 from turkey farms to chicken farms (β_{TC} = 0.1, 95% HPD: 0.02 – 0.22). This is similar to the
189 results of the structured coalescent model and discrete trait model described above
190 (Supplemental Table S1). Infectious period of a farm also varied substantially between the two
191 production systems. A HPAI-positive turkey premises was estimated to remain infectious for 5.7
192 days (95% HPD: 4.3 – 10.5), whereas layer chicken premises were estimated to remain
193 infectious for 32.1 days (95% HPD: 22.4 – 49.3; Fig 2C).
194
195 Fig 2. Comparison of hypothesized HPAI H5N2 epidemiological compartmental models. (A)
196 Each compartmental model represents a Susceptible-Infectious-Removed (SIR) model
197 with varied population heterogeneity: 1) a single, closed, homogenous population, 2) a
198 single, homogenous population with a continual external source of virus (U), 3) a closed
199 population, stratified by poultry system (turkeys (T) and layer chickens (C)), and 4) the
200 stratified population with a continual external source of virus. (B) Compartmental model
201 fit for the midwestern highly pathogenic avian influenza (HPAI) H5N2 outbreak, 2015.
202 Akaike's information criteria for Markov chain Monte Carlo (AICM) calculated based on
203 the posterior distribution of the structured tree likelihood was used to evaluate the relative
204 model fit for the four assessed compartmental models under differing molecular clock
205 assumptions. Under both molecular clocks, Model 3 provided the best model fit. (C)
206 Estimated infectious period of layer chicken and turkey farms during the 2015 midwestern

207 highly pathogenic avian influenza (HPAI) H5N2 outbreak. During model specification, an
208 informative prior was provided for the Bayesian process. This prior probability
209 distribution was based on the reported average time from HPAI confirmation to
210 depopulation plus 5 days to allow for delay between infection and HPAI confirmation.
211 Model 3 estimated the infectious period for layer chickens to be longer than expected
212 given the prior information.
213
214 *Ecologic Predictors of HPAI H5N2 Geographic Diffusion*
215 Using the posterior distribution of phylogenetic trees estimated under the EBSP
216 coalescent and strict molecular clock assumptions, discrete trait diffusion models were estimated
217 to describe the geographic dispersal of HPAI H5N2 throughout the midwestern United States.
218 County of origin was used as the basis to categorize the 182 sequences. Counties were grouped
219 based on their state and whether sequences within the county exclusively originated from
220 commercial turkey premises. For example, Iowan counties with only turkey cases were grouped
221 separately from Iowan counties which had at least one layer chicken case. County groups with
222 only turkey cases are henceforth referred to as turkey-exclusive while county groups with at least
223 one layer chicken case are referred to as mixed poultry. The complete ancestral reconstruction of
224 the midwestern outbreak is shown in Fig 3A. The three largest transition rates were observed
225 between county groups within the same state, particularly Minnesota and Iowa (Fig 3B;
226 Supplemental Table S3). The most frequent transitions occurred from Minnesota mixed poultry
227 counties to Minnesota turkey-exclusive counties (median rate: 3.3 transitions per year; 95% HPD
228 0.7 – 6.4; BF = 490.6). In Minnesota, the reverse rate (i.e., from turkey-exclusive counties to
229 mixed poultry counties) was also decisively supported with a relatively high transition rate (2.3

transitions per year; 95% HPD: 0.6 – 4.6; BF = 2,007.1). The second most frequent transition had the highest statistical support and occurred from Iowan mixed poultry counties to Iowan turkey-exclusive counties (3.3 transitions per year; 95% HPD 1.4 – 5.7; BF = 28,139.6). Three inter-state transitions were also decisively supported, but less frequent. These transitions were estimated from Iowa mixed-poultry counties to Minnesota turkey-exclusive counties (0.9 transitions per year; 95% HPD 0.2 – 2.2; BF = 14,068.3), from Minnesota turkey-exclusive counties to Wisconsin turkey-exclusive counties (0.74 transitions per year; 95% HPD 0.1 – 1.8; BF = 202.3) and from Wisconsin turkey-exclusive counties to Iowan mixed-poultry counties (0.9 transitions per year; 95% HPD 0.01 – 2.6; BF = 134.8). All supported transition rates (BF > 3.0) were found either within a state or between states that share borders, except for a single weakly supported rate from South Dakota turkey counties to Wisconsin mixed poultry counties (0.6 transitions per year; 95% HPD 0.0002 – 2.1; BF = 3.2). This suggests geographic distance influences the dispersal of HPAI H5N2 among midwestern counties.

Fig 3. Discrete trait diffusion model of HPAI H5N2 among midwestern county groups. (A) Maximum clade credibility tree representing the ancestral reconstruction of county groups across the evolutionary history of the outbreak. The ancestral reconstruction was based on an EBSP coalescent and strict molecular clock evolutionary model. Tree branches are colored based on the most probable county group of the descendant node. Thin gray node bars represent the 95% highest posterior density (HPD) of the node height (i.e., the time at which that ancestor is estimated to have existed). (B) Diffusion rate summary among county groups. County groups were defined based on state and composition of host type within the county. Counties with only turkey cases (turkey exclusive; T) were grouped

253 separately from counties with at least one layer chicken case (mixed poultry; C). Arrows
254 represent transition rates with strong support (Bayes factor > 10) with arrow thickness
255 proportional to the magnitude of transition rate. (C) Conditional effect size of
256 environmental and geographic covariates within the generalized linear model (GLM).
257 Conditional effect size represents the effect size of the variable coefficient given inclusion
258 in the GLM. Supported covariates (Bayes factor > 3) are bolded. Covariates are ordered
259 by Bayes factor. The dashed gray line represents a conditional effect size of 0, signifying
260 little impact of the covariate on viral dispersal.
261
262 The discrete trait diffusion model was extended with a GLM that assessed the impact of
263 distance and other environmental variables on the transition rates among the defined county
264 groups. County characteristics for the 9 modeled variables are summarized in Table 1. On
265 average, county centers were 266 km apart, ranging from 30 to 862 km. HPAI-positive counties
266 had a higher density of layer chicken farms (0.02 farms/km^2) than turkey farms (0.004
267 farms/km^2). Counties also had a broad range of human population density ranging from about 1
268 to 58 people/km^2. Of the 9 variables included in the GLM, three were statistically supported to
269 be associated with diffusion of HPAI H5N2 among county groups (Fig 3C, Supplemental Table
270 S4). Distance between county group centroid was decisively supported to be negatively
271 associated with transition between two groups (log conditional effect size = -1.0; 95% HPD -1.2,
272 -0.8; BF = 216,262.9). In other words, viral transitions are less likely between county groups that
273 are separated by a greater distance. Road density of the origin county group was positively
274 associated with viral dispersion (log conditional effect size = 1.2; 95% HPD 0.6 – 1.7; BF =
275 42.8). That is, county groups with a higher density of roads were associated with higher

dispersion rates to other county groups. The proportion of the destination county group covered with water was only weakly supported for inclusion in the GLM (log conditional effect size = 0.6; 95% HPD 0.2 – 0.9; BF = 3.9).

Table 1. Demographic and geographic characteristics of the 49 United States counties with HPAI-positive commercial poultry premises during the H5N2 outbreak, 2015.

	Mean	Standard Deviation	Minimum	Maximum
Distance between counties (km)	265.95	153.45	30.18	861.99
Layer Chicken Farm Density (farms/km^2)	0.02	0.02	0.001	0.09
Turkey Farm Density (farms/km^2)	0.004	0.004	3.97×10^{-6}	0.01
Human Population Density (humans/km^2)	12.53	11.21	1.39	58.07
Road Density (km/km^2)	1.89	0.32	1.18	2.78
Water Coverage (%)	1.56	2.21	0.02	11.34
Important Bird Area (%)	4.23	7.60	0.0	30.96
Agricultural Land Use (%)	78.07	13.28	32.46	90.62
Frozen days	19.82	4.93	12	38

DISCUSSION

Our exploration of population models to describe the 2015 midwestern United States HPAI H5N2 outbreak provides evidence that upon entering the midwestern poultry industries, no further viral introductions from outside sources were needed to explain the observed epidemiological trajectory. Furthermore, the statistical support for a stratified poultry population suggests that poultry industries should not be considered a homogenous host population for viral pathogens. This is also supported by the discrete trait diffusion analyses, which demonstrate that geographic factors influence viral dispersion among counties, indicating heterogeneity among geographic locations. Multiple factors including poultry production system barriers and

293 geographic characteristics appear to have influenced the course of the outbreak within poultry
294 industries.

295 In our analysis, the EBSP coalescent model had better support than the other traditional
296 coalescent models in terms of model fit. This is most likely a reflection of EBSP's flexibility, i.e.
297 the piece-wise nature of this method, which facilitates the identification of complex population
298 changes. Coalescent theory has been a popular technique to infer population demographics
299 underlying viral outbreaks (19–28). By relating effective population size to the rate at which
300 phylogenetic lineages converge backwards in time, the coalescent has become a powerful tool to
301 infer demographic changes even in the face of incomplete sampling. Traditionally, the estimation
302 of the coalescent process required rigid prior assumptions in the form of simplistic mathematical
303 growth functions (e.g., constant population size or exponential growth). To better reflect
304 biological reality, methods have been developed that incorporate more flexibility than a one to
305 two parameter mathematical function (29,30). For instance, the EBSP assumes demographic
306 changes follow a smoothed, piece-wise, linear function whose change points are inferred from
307 the sequence data (31). To date, mathematical methods to incorporate population structure into
308 EBSP coalescent models have not been developed even though population structure has been
309 observed to confound EBSP estimates (32).

310 Despite the flexibility of EBSP, compartmental-based coalescent models are worth
311 assessing as they allow for direct incorporation and hypothesis testing of specific population
312 structures. Rather than the non-parametric, piece-wise approach of EBSP, the prior mathematical
313 functions assumed are ordinary differential equations (ODEs) constructed from the specification
314 of epidemiological compartmental models. It is the parameters of these ODEs that are fit during
315 the Markov chain Monte Carlo (MCMC) process. Among the four analyzed compartmental

316 models, we found that the closed, stratified population provided the best fit for the sequence data,
317 suggesting layer chickens and turkeys represented two separate host populations that interacted
318 with each other, but did not receive virus from a continuous external source. Interestingly, when
319 only observing the single homogenous population models (Models 1 and 2), the inclusion of an
320 external viral source (Model 2) improves model fit compared to the closed population model
321 (Model 1). Once the population structure of poultry type is included (Models 3 and 4), the closed
322 population model provides a better fit than that with continual viral introductions. This
323 observation underlines the importance of including population heterogeneity within evolutionary
324 demographic models to explain observed viral diversity and population dynamics.
325 To help improve identifiability of the remaining parameters within the compartmental
326 model, expected prior distributions for the infectious period of affected premises were specified
327 based on reported USDA data (7). Despite the informative assumption, the infectious period of
328 layer chicken farms was estimated to be longer than expected. In our model, we assumed a 5-day
329 period between the onset of infectivity of the farm and reporting of HPAI infection. Delays in the
330 identification and/or reporting of HPAI infection could result in infectious periods that begin
331 well before the assumed 5 days. Continued infectivity beyond the completion of flock
332 depopulation is another likely contributor to prolonged infectious periods. Although commercial
333 poultry depopulation occurred on average 6.4 days after National Veterinary Services Laboratory
334 (NVSL) HPAI confirmation, premises were not considered to be virus-free until, on average,
335 87.7 days following confirmation (7). In either case, our models suggest layer chicken farms
336 remained infectious for much longer than turkey farms, potentially explaining why the
337 transmission rate from chicken farms to turkey farms was higher than its counterpart. In fact,
338 regardless of the model (i.e., structured coalescent, discrete trait diffusion model, or

339 compartmental model), layer chicken farms played a more central role to viral transmission than
340 turkey farms during the outbreak. This may seem contradictory to experimental evidence that
341 demonstrated the HPAI H5N2 virus had longer mean death times in turkeys (5 – 6 days)
342 compared to chickens (2 – 3 days (33). However, such experimental infections only describe
343 transmission information on the individual bird scale, rather than the farm-to-farm transmission
344 scale captured in this analysis. Although it may be that individual turkeys survive longer, in
345 practice turkey premises were quicker to be depopulated, resulting in a shorter farm-level
346 infectious period compared to chicken farms. Because intervention (i.e. depopulation) was
347 performed on the farm level, individual-level infectious periods alone are not adequate to
348 describe the overall observed outbreak dynamics.

349 The implementation of a GLM into a Bayesian discrete trait analysis has been previously
350 applied to HPAI in China (34) and Egypt (35), providing evidence that environmental,
351 agricultural and anthropogenic factors influence viral movement. Due to differences in social,
352 governmental and agricultural systems, the generalization of these previous GLM results to other
353 countries may not be appropriate. Instead, these studies provide a framework to identify
354 epidemiological covariates of the North American HPAI H5N2 outbreak. Our results indicate
355 that distance and road density are key factors that influenced the geographic spread of HPAI
356 H5N2 among midwestern counties in the spring of 2015. A recent spatial modeling analysis
357 revealed that HPAI spread among Minnesota poultry premises was heavily distance-dependent
358 during the 2015 outbreak (13). Our results support this claim by providing evidence that the
359 frequency of shared viral diversity increases as the distance between two counties decreases.
360 Risk of infection due to proximity can also be observed in our discrete trait diffusion model in
361 which within-state HPAI spread was much more frequent than inter-state spread. HPAI

362 movement between states may explain Bonney, et al.'s finding that distance-independent
363 transmissions significantly improved the fit of their transmission kernel model (13). Although
364 the causal relationship between the supported covariates and viral dispersal cannot be determined
365 from our analysis, the statistical support for road density within the GLM may provide evidence
366 for the relative importance of anthropogenic movement of virus. High road density may correlate
367 with better logistic connectivity between farms, increasing the likelihood that an infected
368 premises will export virus to nearby farms and counties. Road density has been associated with
369 HPAI H5N1 outbreaks in Bangladesh (36), Thailand (37), Romania (38), Indonesia (39,40), and
370 Nigeria (41), although high road density in these countries may reflect greater human population
371 density and, therefore, a higher likelihood of case detection (42). Intensive commercial poultry
372 surveillance during the 2015 outbreak and the lack of support for human population density as a
373 covariate within our model suggest that the statistical support for road density in the dispersal of
374 HPAI among midwestern counties may not merely be an artifact of sampling bias or
375 confounding. The third variable associated with HPAI dispersal was the proportion of a
376 destination county group covered by surface water. In other words, counties with a larger
377 proportion of surface water received virus more frequently compared to those with less surface
378 water. Surface water resources have been associated with HPAI dispersal and prevalence in
379 China and may signify movement of virus by migrating waterfowl that stopover in lakes, rivers
380 and wetlands (34,43). In our analysis, this variable was only weakly supported and had a
381 relatively small effect size. Additionally, other variables that represent potential migratory
382 stopover habitats, such as Important Bird Areas and agricultural land, were not supported within
383 the model, suggesting that if wild birds contributed to HPAI dispersal within the Midwest, their
384 role was limited. This further supports previous studies, which indicated that the midwestern

385 portion of the outbreak was driven by inter-farm transmission (11,12,14). Several mechanisms
386 have been proposed to explain HPAI transmission between farms during the 2015 outbreak,
387 including equipment sharing, personnel overlap, and aerosolization.

388 Due to the restricted number of sequences in the presented analysis, the number of
389 variables and demographic scenarios that could be modelled was limited. This also affected the
390 resolution of the geographic covariates that could be included within the GLM. Ideally, the
391 environmental and agricultural characteristics of each individual farm or county would be
392 evaluated as predictor for HPAI spread; however, the individual transition rates between 182
393 farms or even 49 counties would be impossible to accurately estimate from the 182 sequences of
394 this dataset. For this reason, sequences were categorized into county groups, resulting in a
395 manageable transition rate matrix as well as permitting the summarization of environmental
396 characteristics across a few counties rather than across an entire state.

397 Despite these limitations, our results present several implications for future HPAI
398 surveillance and control in the United States. While wild birds may provide a means of viral
399 dispersal across large distances and initial introduction into an area, evidence suggests the HPAI
400 outbreak within the midwestern poultry industries could be maintained without continued
401 introductions. In this sense, in-place biosecurity efforts may have been enough to prevent
402 continued viral introductions from outside sources (including wild birds, backyard poultry
403 flocks, or long-distance movement from other geographic regions), but were ineffective against
404 local farm-to-farm transmission. For instance, it has been suggested that biosecurity factors could
405 explain the lack of HPAI cases within the broiler chicken industry in the Midwest (44). The
406 association of distance between and road density of county groups with HPAI dispersal suggests
407 human transportation modes may have played an important role in dispersal of HPAI across the

Midwest. A better understanding of how HPAI-positive farms are logistically connected would greatly aid surveillance and control efforts. With the knowledge of how these farms share personnel and equipment, future outbreaks could be contained by disruption of the transportation network.

METHODS

Dataset

Whole genome HPAI H5N2 sequences collected, isolated and sequenced by the United States Department of Agriculture (USDA) during the 2014 – 2015 North American HPAI outbreak served as the basis for the analyzed data set. Full description of their collection and sequencing has been reported elsewhere (14). A subset of this sequence data was selected to better investigate the farm-to-farm transmission dynamics of the midwestern portion of the HPAI H5N2 outbreak. This subset was defined by the following inclusion criteria: 1) sequences isolated from commercial domestic poultry samples and 2) membership of the sequence in a phylogenetically distinct group, as determined by maximum likelihood estimation by Lee, et al (14). These viruses represented midwestern HPAI-positive poultry premises from the latter part of the outbreak, which was defined by a rapid increase in incidence within the midwestern poultry industries. As within-farm epidemiological dynamics were not of interest in this analysis, only one viral sequence per positive poultry premises was included. Viruses isolated from backyard poultry operations and wild birds were not included due to the incongruency in surveillance and sampling between these populations and the domestic poultry industries. A full list of the included sequence names and accession numbers are provided in Supplemental Table S5.

Coalescent Model Comparison

Coalescent theory provides the statistical framework to estimate population changes over time from genetic sequence data. To investigate the population dynamics of the midwestern poultry portion of the outbreak, various coalescent population model prior assumptions were implemented and compared in BEAST2 (45). Using ModelFinder (46) as implemented in the IQ-TREE software package (http://www.iqtree.org/), the Kimura three parameter (K3P; i.e., one transition rate and 2 transversion rates) model (47) with unequal base frequencies and a gamma distribution of rate variation among sites (48) was determined as the best fit nucleotide substitution model and was used for each BEAST2 model. All coalescent models were separately estimated under both strict and lognormally distributed, uncorrelated, relaxed molecular clock assumptions. For each BEAST2 model, at least three independent MCMC runs of 50 million chain length were initiated from random starting trees. Convergence was assessed in Tracer v1.5, ensuring an effective sample size (ESS) > 200 for each estimated parameter. If ESS < 200, the discarded burn-in fraction was increased or more MCMC runs were performed. Three "traditional" coalescent models (i.e., constant population, exponential growth, and EBSP (31)) were performed to investigate demographic dynamics. Model fit was compared among the coalescent and molecular clock models with path sampling to calculate the marginal likelihood estimate (MLE) (49). Estimating the marginal likelihood enables the calculation of a Bayes Factor (BF), which is a ratio of two marginal likelihoods. A log(BF) > 5 indicates very strong statistical support for one model over the other (50). Viral dispersion between poultry industries (layer chicken vs. turkey) was initially estimated with a simple discrete trait diffusion model as well as a structured coalescent (25). The EBSP coalescent model was used as the tree prior for the discrete trait diffusion model. Both viral dispersion models were performed under both strict and relaxed molecular clock, as above.

454 A recently developed structured coalescent-based BEAST2 package (PhyDyn) was used
455 to investigate more complex pathogen population scenarios by specifying epidemiological
456 compartmental models (18). Four alternative compartmental models were assessed to investigate
457 the presence of population structure by poultry type (layer chicken vs. turkey) and continual viral
458 introductions from an unknown source population. Each compartmental model was a
459 Susceptible-Infectious-Removed (SIR) model with varied population heterogeneity (Fig 2A): 1)
460 a single, closed, homogenous population, 2) a closed population, stratified by poultry system, 3)
461 a single, homogenous population with a continual external source of virus, and 4) a stratified
462 population with a continual external source of virus. By including models with an external viral
463 source, the models test whether this aspect of the outbreak was insulated or involved repeated
464 introductions of HPAI from wild birds, backyard poultry, or undetected HPAI-positive premises.
465 Since marginal likelihood estimation via path sampling has not yet been developed for the
466 PhyDyn package, Akaike Information Criterion for MCMC (AICM) (51) was used to assess
467 model fit and was calculated from the posterior MCMC sample of the structured tree likelihood
468 with the R package, aicm (https://rdrr.io/cran/geiger/man/aicm.html).

469 *Discrete trait diffusion models*

470 To estimate the impact of environmental variables on the geographic diffusion of HPAI
471 between midwestern counties, a discrete trait diffusion model was constructed and further
472 extended with a generalized linear model (GLM) in BEAST v1.10 (52). Discrete trait diffusion
473 models are a phylogeographic technique in which each analyzed genetic sequence is assigned an
474 observed characteristic trait that is assumed to have changed across the viral evolutionary history
475 in a continuous time Markov chain process (53). Transition rates among these observed traits can
476 then be inferred. In this analysis, the discrete character trait definition was based on the United

States county in which the HPAI-positive poultry premises was located. Counties were then categorized by state and whether the county's sequences exclusively originated from turkey production premises. In contrast to the simplified discrete trait model performed parallel to the structured coalescent model above, this model enables geographic dispersion of the HPAI virus to be estimated.

The geographic discrete trait diffusion model was extended with a GLM to assess the impact of environmental covariates on the viral transition rates among county categories. In this approach, viral diffusion rates among discrete geographic regions act as the outcome to a log-linear combination of environmental variables, regression coefficients and indicator variables (17). Environmental and anthropogenic variables were selected based on previous indication of their importance to avian influenza risk (42). Layer chicken farm density and turkey farm density were calculated from USDA 2012 census data (https://quickstats.nass.usda.gov/) divided by the land area of the county group. Human population density and proportion of county covered in water was obtained from United States census data (https://factfinder.census.gov/). The remaining variables were summarized per county group using ArcGIS Pro. Geographic distance was calculated as the linear distance between county group centroid. Road density was estimated as the total length of road per county divided by the total county group area. Proportion of county designated as an important bird area (IBA) was calculated using the publicly available Audubon Important Bird Areas and Conservation Priorities data (54). Proportion of the county group used for agriculture (i.e., covered by pasture, hay or cultivated crops) was obtained from the United States Geological Survey National Land Cover Database created in 2011 and amended in 2014 (55). The number of frozen days was calculated from daily freeze-thaw satellite data from March 1 to June 15, 2015 (56,57). A frozen day was defined as a day in which more than half of the

county group area had a temperature measured as below 0 C. All covariate measures were log-transformed and standardized before inclusion in the GLM.

The discrete trait diffusion models were applied to the empirical distribution of phylogenetic trees from the best fitting evolutionary model. For each diffusion model, three independent MCMC runs of 1 million steps in length were performed, sampling every 100 steps. Convergence was assessed in Tracer v1.5, ensuring ESS > 200 for each estimated parameter. Removing the first 10% of each run as burn-in and re-sampling every 300 steps, log and tree files were combined using LogCombiner in the BEAST v1.10 software suite. Statistical support for transition rates in the discrete trait diffusion model and the covariate coefficients of the GLM were inferred using Bayesian stochastic search variable selection (BSSVS). Briefly, for each estimated parameter, an indicator variable (I) is stochastically turned on (I = 1) or off (I = 0) at each step of the MCMC (16,53). The posterior distribution of indicator values can be used to calculate a Bayes factor (BF), indicating the level of statistical support for the inclusion of that parameter in the model. BF support was defined in the following categories: no support (BF < 3.0), substantial support (3.0 ≤ BF < 10.0), strong support (10.0 ≤ BF < 30.0), very strong support (30.0 ≤ BF < 100.0), and decisive support (BF ≥ 100.0). Median transition rates, median conditional coefficients, 95% highest posterior density (HPD) and BF were calculated using personalized Python scripts.

ACKNOWLEDGEMENTS

We would like to thank Alex Heri for her critical review of this manuscript.

REFERENCES

1. Lee YJ, Kang HM, Lee EK, Song BM, Jeong J, Kwon YK, et al. Novel reassortant influenza A(H5N8) viruses, South Korea, 2014. Emerg Infect Dis. 2014 Jun;20(6):1087–

9.

2. Dalby AR, Iqbal M. The European and Japanese outbreaks of H5N8 derive from a single source population providing evidence for the dispersal along the long distance bird migratory flyways [Internet]. Vol. 3, PeerJ. 2015. p. e934. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25945320

3. Harder T, Maurer-Stroh S, Pohlmann A, Starick E, Höreth-Böntgen D, Albrecht K, et al. Influenza A(H5N8) Virus Similar to Strain in Korea Causing Highly Pathogenic Avian Influenza in Germany. Emerg Infect Dis [Internet]. 2015 May;21(5):860–3. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25897703

4. Ip HS, Torchetti MK, Crespo R, Kohrs P, DeBruyn P, Mansfield KG, et al. Novel Eurasian highly pathogenic avian influenza A H5 viruses in wild birds, Washington, USA, 2014. Emerg Infect Dis. 2015 May;21(5):886–90.

5. Pasick J, Berhane Y, Joseph T, Bowes V, Hisanaga T, Handel K, et al. Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014. Sci Rep. 2015 Mar;5:9484.

6. Lee DH, Bahl J, Torchetti MK, Killian ML, Ip HS, DeLiberto TJ, et al. Highly Pathogenic Avian Influenza Viruses and Generation of Novel Reassortants, United States, 2014-2015. Emerg Infect Dis. 2016 Jul;22(7):1283–5.

7. APHIS U. Final Report for the 2014–2015 Outbreak of Highly Pathogenic Avian Influenza (HPAI) in the United States. 2016 Aug.

8. Johansson RC, Preston WP, Seitzinger AH. Government Spending to Control Highly Pathogenic Avian Influenza. Choices. 2016 Jan;31(2).

9. Swayne DE, Hill RE, Clifford J. Safe application of regionalization for trade in poultry

546 and poultry products during highly pathogenic avian influenza outbreaks in the USA.
547 Avian Pathol [Internet]. Taylor & Francis; 2017 Mar 4 [cited 2019 Apr 23];46(2):125–30.
548 Available from: https://www.tandfonline.com/doi/full/10.1080/03079457.2016.1257775
549 10. Dargatz D, Beam A, Wainwright S, McCluskey B. Case Series of Turkey Farms from the
550 H5N2 Highly Pathogenic Avian Influenza Outbreak in the United States During 2015.
551 Avian Dis. 2016 Jun;60(2):467–72.
552 11. Wells SJ, Kromm MM, VanBeusekom ET, Sorley EJ, Sundaram ME, VanderWaal K, et
553 al. Epidemiologic Investigation of Highly Pathogenic H5N2 Avian Influenza Among
554 Upper Midwest U.S. Turkey Farms, 2015. Avian Dis. 2017 Jun;61(2):198–204.
555 12. Grear DA, Hall JS, Dusek RJ, Ip HS. Inferring epidemiologic dynamics from viral
556 evolution: 2014-2015 Eurasian/North American highly pathogenic avian influenza viruses
557 exceed transmission threshold, R0 = 1, in wild birds and poultry in North America. Evol
558 Appl. 2017 Dec;11(4):547–57.
559 13. Bonney PJ, Malladi S, Boender GJ, Weaver JT, Ssematimba A, Halvorson DA, et al.
560 Spatial transmission of H5N2 highly pathogenic avian influenza between Minnesota
561 poultry premises during the 2015 outbreak. PLoS One [Internet]. 2018;13(9):e0204262.
562 Available from: https://www.ncbi.nlm.nih.gov/pubmed/30240402
563 14. Lee D-H, Torchetti MK, Hicks J, Killian ML, Bahl J, Pantin-Jackwood M, et al.
564 Transmission dynamics of highly pathogenic avian influenzvirus a(H5Nx) clade 2.3.4.4,
565 North America, 2014–2015. Emerg Infect Dis. 2018;24(10).
566 15. USDA APHIS. Epidemiologic and Other Analyses of HPAI-Affected Poultry Flocks:
567 September 9, 2015 Report. 2015.
568 16. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying Viral

569 Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of
570 Human Influenza H3N2. PLOS Pathog [Internet]. 2014;10(2):e1003932. Available from:
571 http://dx.doi.org/10.1371%2Fjournal.ppat.1003932
572 17. Baele G, A. MS, Rambaut A, Lemey P. Emerging Concepts of Data Integration in
573 Pathogen Phylodynamics. Syst Biol. 2016;66(1):e65.
574 18. Volz EM, Siveroni I. Bayesian phylodynamic inference with complex models. PLoS
575 Comput Biol [Internet]. 2018 Nov;14(11):e1006546. Available from:
576 https://www.ncbi.nlm.nih.gov/pubmed/30422979
577 19. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-
578 human interface. Vol. 7, eLife. 2018.
579 20. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating Mutation Parameters,
580 Population History and Genealogy Simultaneously From Temporally Spaced Sequence
581 Data. Genetics [Internet]. 2002;161(3):1307–20. Available from:
582 http://tmclibrary.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV1NS_QwEB5
583 UELyIr5-
584 rr5CTJ3fpZ5IePMiiiDdB0Vto86HCtrvY7mH99c6kqbwrnt5j27S0mcxknuTpMwBpMonG
585 P2JCjlhIW5zbrEuESbRx2qtNSWe4kYZw4_Q6nb7kD_fieQMGFhmRLNeoipPm_c3TLU
586 O_fvh6axNHuJFOmSv0fJy15EU3n8-uFrW_
587 21. Dearlove B, Wilson DJ. Coalescent inference for infectious disease: meta-analysis of
588 hepatitis C. Philos Trans R Soc Lond B Biol Sci [Internet]. 2013 Mar 19 [cited 2019 Apr
589 15];368(1614):20120314. Available from:
590 http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2012.0314
591 22. Biek R, Drummond AJ, Poss M. A virus reveals population structure and recent

592 demographic history of its carnivore host. Science. 2006 Jan;311(5760):538–41.
593 23. Möller S, du Plessis L, Stadler T. Impact of the tree prior on estimating clock rates during
594 epidemic outbreaks. Proc Natl Acad Sci [Internet]. 2018;115(16):4200–5. Available from:
595 http://www.pnas.org/content/115/16/4200.abstract
596 24. Streicker DG, Altizer SM, Velasco-Villa A, Rupprecht CE. Variable evolutionary routes
597 to host establishment across repeated rabies virus host shifts among bats. Proc Natl Acad
598 Sci U S A [Internet]. 2012 Nov 27 [cited 2019 Apr 15];109(48):19715–20. Available
599 from: http://www.pnas.org/cgi/doi/10.1073/pnas.1203456109
600 25. Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ. Efficient Bayesian
601 inference under the structured coalescent. Bioinformatics [Internet]. 2014 Aug 15 [cited
602 2019 Feb 11];30(16):2272–9. Available from:
603 http://www.ncbi.nlm.nih.gov/pubmed/24753484
604 26. Aldunate F, Gámbaro F, Fajardo A, Soñora M, Cristina J. Evidence of increasing
605 diversification of Zika virus strains isolated in the American continent. J Med Virol
606 [Internet]. John Wiley & Sons, Ltd; 2017 Dec 1 [cited 2019 Apr 15];89(12):2059–63.
607 Available from: http://doi.wiley.com/10.1002/jmv.24910
608 27. Alkhamis MA, Perez AM, Murtaugh MP, Wang X, Morrison RB. Applications of
609 Bayesian Phylodynamic Methods in a Recent U.S. Porcine Reproductive and Respiratory
610 Syndrome Virus Outbreak. Front Microbiol [Internet]. 2016 Feb 2 [cited 2019 Apr
611 15];7:67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26870024
612 28. Carrington CVF, Foster JE, Pybus OG, Bennett SN, Holmes EC. Invasion and
613 Maintenance of Dengue Virus Type 2 and Type 4 in the Americas. J Virol [Internet]. 2005
614 Dec 1 [cited 2019 Apr 15];79(23):14680–7. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/16282468

29. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences [Internet]. Vol. 22, Molecular biology and evolution. 2005. p. 1185–92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15703244

30. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008 Jul;25(7):1459–71.

31. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci. BMC Evol Biol [Internet]. BioMed Central; 2008 Oct 23 [cited 2019 Feb 11];8(1):289. Available from: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-8-289

32. Heller R, Chikhi L, Siegismund HR. The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. Mailund T, editor. PLoS One [Internet]. Public Library of Science; 2013 May 7 [cited 2019 Apr 14];8(5):e62992. Available from: https://dx.plos.org/10.1371/journal.pone.0062992

33. Lee DH, Bertran K, Kwon JH, Swayne DE. Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. J Vet Sci. 2017 Aug;18(S1):269–80.

34. Lu L, Brown AJL, Lycett SJ. Quantifying predictors for the spatial diffusion of avian influenza virus in China. BMC Evol Biol [Internet]. 2017;17(1):16. Available from: https://doi.org/10.1186/s12862-016-0845-3

35. Magee D, Beard R, Suchard MA, Lemey P, Scotch M. Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion. Arch

638 Virol. Austria; 2015 Jan;160(1):215–24.

639 36. Loth L, Gilbert M, Osmani MG, Kalam AM, Xiao X. Risk factors and clusters of Highly
640 Pathogenic Avian Influenza H5N1 outbreaks in Bangladesh. Prev Vet Med [Internet].
641 NIH Public Access; 2010 Aug 1 [cited 2019 Mar 26];96(1–2):104–13. Available from:
642 http://www.ncbi.nlm.nih.gov/pubmed/20554337

643 37. Paul M, Tavornpanich S, Abrial D, Gasqui P, Charras-Garrido M, Thanapongtharm W, et
644 al. Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1:
645 prospects from a spatial-based model. Vet Res [Internet]. BioMed Central; 2010 [cited
646 2019 Mar 26];41(3):28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20003910

647 38. Ward MP, Maftei D, Apostu C, Suru A. Environmental and anthropogenic risk factors for
648 highly pathogenic avian influenza subtype H5N1 outbreaks in Romania, 2005--2006. Vet
649 Res Commun [Internet]. 2008 Dec 5 [cited 2019 Mar 26];32(8):627–34. Available from:
650 http://link.springer.com/10.1007/s11259-008-9064-8

651 39. Loth L, Gilbert M, Wu J, Czarnecki C, Hidayat M, Xiao X. Identifying risk factors of
652 highly pathogenic avian influenza (H5N1 subtype) in Indonesia. Prev Vet Med [Internet].
653 2011 Oct 1 [cited 2019 Mar 26];102(1):50–8. Available from:
654 https://linkinghub.elsevier.com/retrieve/pii/S016758771100208X

655 40. Yupiana Y, de Vlas SJ, Adnan NM, Richardus JH. Risk factors of poultry outbreaks and
656 human cases of H5N1 avian influenza virus infection in West Java Province, Indonesia.
657 Int J Infect Dis [Internet]. 2010 Sep [cited 2019 Mar 26];14(9):e800–5. Available from:
658 https://linkinghub.elsevier.com/retrieve/pii/S1201971210024070

659 41. Rivas AL, Chowell G, Schwager SJ, Fasina FO, Hoogesteijn AL, Smith SD, et al. Lessons
660 from Nigeria: the role of roads in the geo-temporal progression of avian influenza (H5N1)

661 virus. Epidemiol Infect [Internet]. Cambridge University Press; 2010 Feb 5 [cited 2019
662 Mar 27];138(02):192. Available from:
663 http://www.journals.cambridge.org/abstract_S0950268809990495
664 42. Gilbert M, Pfeiffer DU. Risk factor modelling of the spatio-temporal patterns of highly
665 pathogenic avian influenza (HPAIV) H5N1: A review. Spat Spatiotemporal Epidemiol
666 [Internet]. 2012 Sep;3(3):173–83. Available from:
667 https://www.sciencedirect.com/science/article/pii/S1877584512000032
668 43. Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, et al. Spatial distribution and
669 risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China [Internet]. Vol. 7,
670 PLoS Pathogens. 2011. p. e1001308. Available from:
671 https://www.ncbi.nlm.nih.gov/pubmed/21408202
672 44. Bertran K, Lee D-H, Balzli C, Pantin-Jackwood MJ, Spackman E, Swayne DE. Age is not
673 a determinant factor in susceptibility of broilers to H5N2 clade 2.3.4.4 high pathogenicity
674 avian influenza virus. Vet Res [Internet]. BioMed Central; 2016 Dec 21 [cited 2019 Apr
675 25];47(1):116. Available from:
676 http://veterinaryresearch.biomedcentral.com/articles/10.1186/s13567-016-0401-6
677 45. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a
678 software platform for Bayesian evolutionary analysis. PLoS Comput Biol [Internet].
679 2014;10(4):e1003537. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24722319
680 46. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
681 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017
682 Jun;14(6):587–9.
683 47. Kimura M. Estimation of evolutionary distances between homologous nucleotide

684 sequences. Proc Natl Acad Sci U S A [Internet]. National Academy of Sciences; 1981 Jan
685 1 [cited 2019 Feb 11];78(1):454–8. Available from:
686 http://www.ncbi.nlm.nih.gov/pubmed/6165991
687 48. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable
688 rates over sites: Approximate methods. J Mol Evol [Internet]. Springer-Verlag; 1994 Sep
689 [cited 2019 Feb 11];39(3):306–14. Available from:
690 http://link.springer.com/10.1007/BF00160154
691 49. Lartillot N, Philippe H. Computing Bayes Factors Using Thermodynamic Integration. Syst
692 Biol [Internet]. 2006 Apr;55(2):195–207. Available from:
693 https://www.jstor.org/stable/20142916
694 50. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc [Internet]. 1995 Jun;90(430):773–
695 95. Available from: https://www.jstor.org/stable/2291091
696 51. Raftery AE, Newton M, Satagopan J, Krivitsky P, Raftery G. Estimating the integrated
697 likelihood via posterior simulation using the harmonic mean identity [Internet]. 2007
698 [cited 2019 Feb 11]. Available from:
699 https://www.scienceopen.com/document?vid=ba5eccc7-2c10-4ec4-b8aa-319939c45f1b
700 52. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian
701 phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol [Internet].
702 2018;4(1):vey016. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29942656
703 53. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its
704 roots. PLoS Comput Biol [Internet]. 2009 Sep;5(9):e1000520. Available from:
705 http://www.ncbi.nlm.nih.gov/pubmed/19779555
706 54. National Audubon Society. Important Bird Areas Database, Boundary Digital Data Set.

2018.

55. U.S. Geological Survey. NLCD 2011 Land Cover (2011 Edition, amended 2014) [Internet]. U.S. Geological Survey; 2014. Available from: https://www.sciencebase.gov/catalog/item/5a1c31b9e4b09fc93dd63982

56. Kim Y, Kimball JS, Glassy J, McDonald KC. No TitleMEaSUREs Northern Hemisphere Polar EASE-Grid 2.0 Daily 6 km Land Freeze/Thaw Status from AMSR-E and AMSR2, Version 1. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center; 2018.

57. Kim Y, Kimball JS, Glassy J, Du J. An extended global Earth system data record on daily landscape freeze–thaw status determined from satellite passive microwave remote sensing. Earth Syst Sci Data [Internet]. 2017 Feb 16 [cited 2019 Mar 25];9(1):133–47. Available from: https://www.earth-syst-sci-data.net/9/133/2017/

721 SUPPLEMENTAL MATERIAL CAPTIONS

722 Table S1. Estimates of viral transmission between poultry industries during the 2015 highly
723 pathogenic avian influenza virus H5N2 outbreak within the midwestern United States.

724 Table S2. Akaike's information criteria for Markov chain Monte Carlo (AICM) for the
725 epidemiological compartment-based coalescent models.

726 Table S3. Discrete trait diffusion matrix of the midwestern highly pathogenic avian influenza
727 (HPAI) H5N2 outbreak, 2015. Median rates and associated 95% highest posterior density
728 intervals (in brackets) are presented in each cell. The diffusion model is asymmetrical, and
729 therefore, rates have directionality from a source county group (indicated on the left) to a
730 sink county group (indicated across the top). County groups were defined by state (IA -
731 Iowa, MN - Minnesota, ND - North Dakota, NE - Nebraska, SD - South Dakota, WI -
732 Wisconsin) and composition of poultry type (T - turkey exclusive, CM - layer chicken
733 exclusive and mixed poultry). Rates are colored by the level of Bayes factor support. Gray
734 rates represent no support.

735 Table S4. Generalized linear model (GLM) conditional effect sizes and statistical support for
736 agricultural and geographic covariates of the dispersal of highly pathogenic avian
737 influenza (HPAI) H5N2 among midwestern county groups. Conditional effect size and
738 95% highest posterior density (HPD) were calculated based on the estimated GLM
739 coefficients given the Bayesian stochastic search variable selection (BSSVS) indicator =
740 1. The posterior probability (PP) refers to the proportion of Markov chain Monte Carlo
741 (MCMC) samples in which the BSSVS indicator = 1. Bayes factor (BF) > 3.0 indicates
742 statistical support for the inclusion of the covariate within the GLM.

743 Table S5. Accession number and names of 182 included HPAI H5N2 full genome sequences.

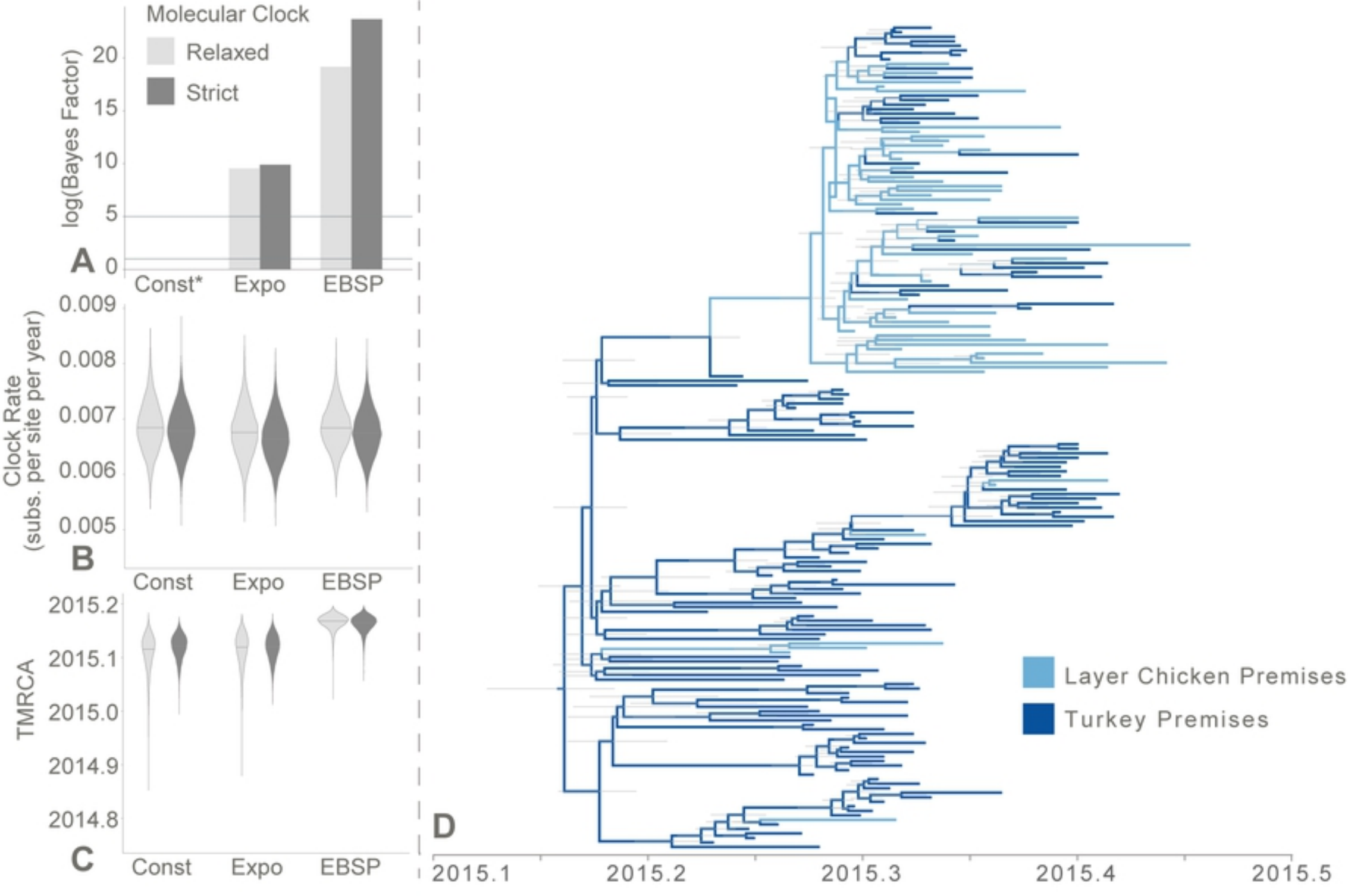
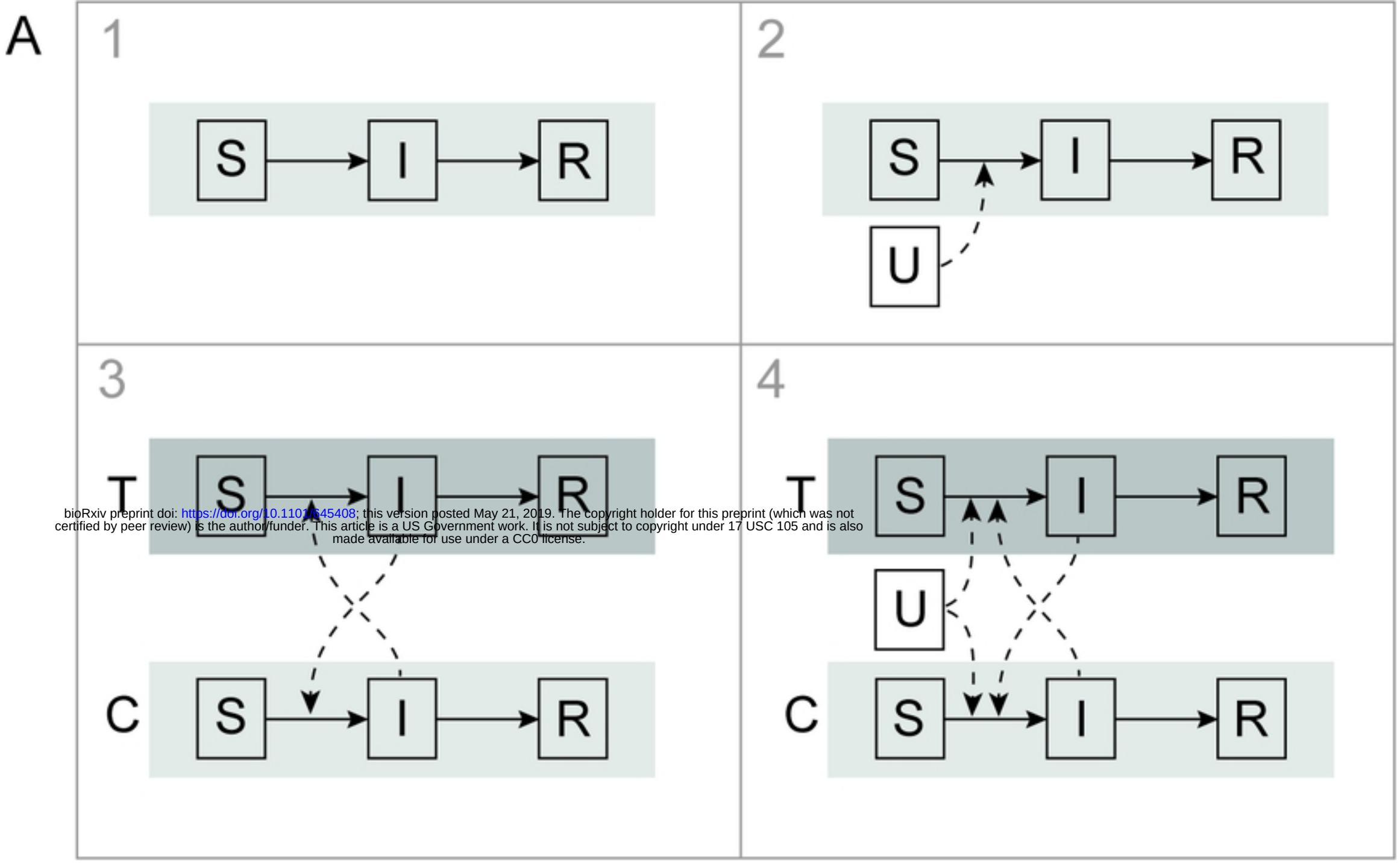
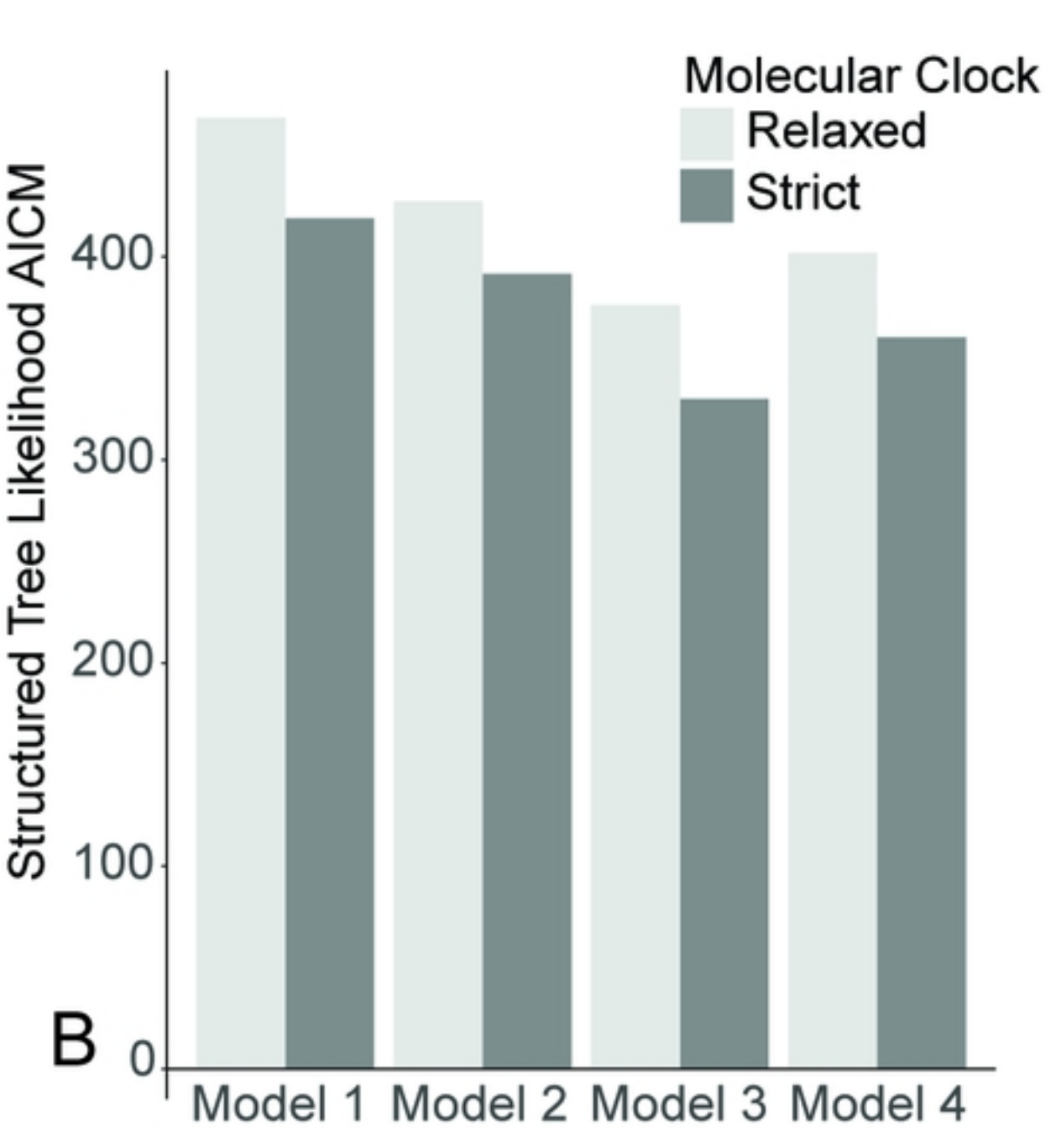

Molecular Clock
Relaxed
Strict
log(Bayes Factor)
20
15
10
5
0
A
Const*
Expo
EBSP
Clock Rate
(subs. per site per year)
0.009
0.008
0.007
0.006
0.005
B
Const
Expo
EBSP
TMRCA
2015.2
2015.1
2015.0
2014.9
2014.8
C
Const
Expo
EBSP
D
2015.1
2015.2
2015.3
2015.4
2015.5
Layer Chicken Premises
Turkey Premises

Figure1





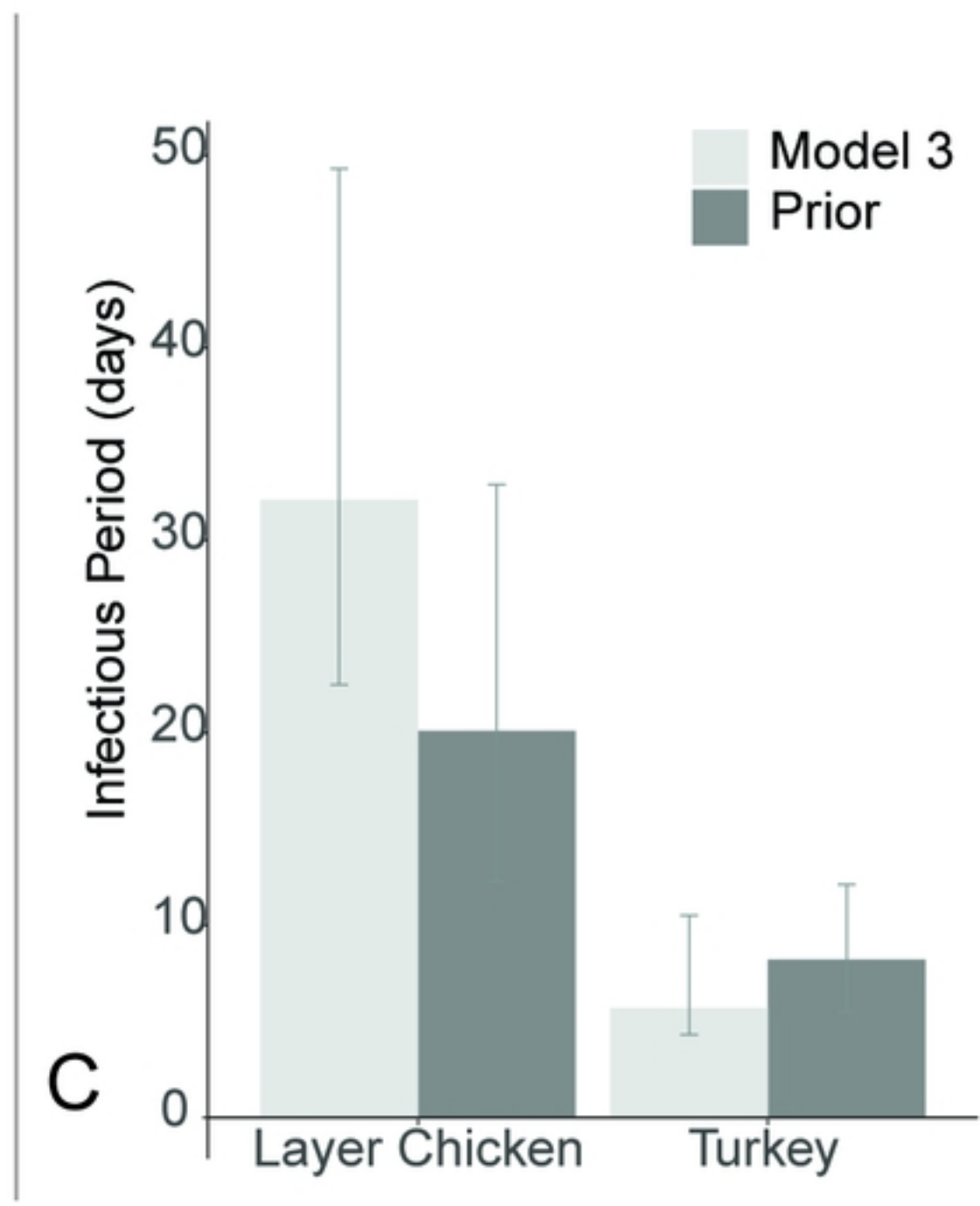


Figure2

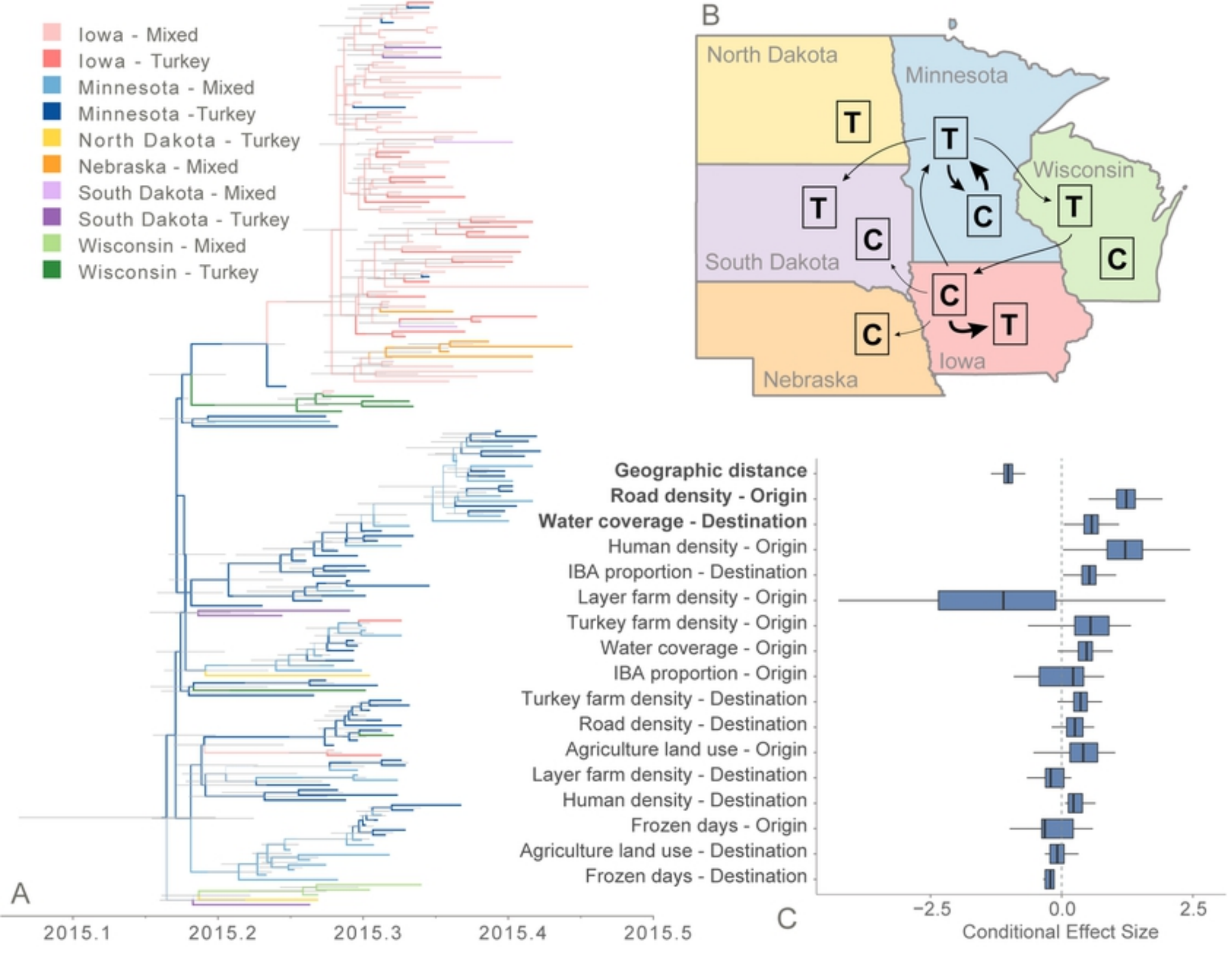

Iowa - Mixed
Iowa - Turkey
Minnesota - Mixed
Minnesota -Turkey
North Dakota - Turkey
Nebraska - Mixed
South Dakota - Mixed
South Dakota - Turkey
Wisconsin - Mixed
Wisconsin - Turkey
A
2015.1
2015.2
2015.3
2015.4
2015.5
B
North Dakota
Minnesota
Wisconsin
South Dakota
Iowa
Nebraska
T
C
Geographic distance
Road density - Origin
Water coverage - Destination
Human density - Origin
IBA proportion - Destination
Layer farm density - Origin
Turkey farm density - Origin
Water coverage - Origin
IBA proportion - Origin
Turkey farm density - Destination
Road density - Destination
Agriculture land use - Origin
Layer farm density - Destination
Human density - Destination
Frozen days - Origin
Agriculture land use - Destination
Frozen days - Destination
C
−2.5
0.0
2.5
Conditional Effect Size

Figure3