Author/s:
Hayama, Y;Firestone, SM;Stevenson, MA;Yamamoto, T;Nishi, T;Shimizu, Y;Tsutsui, T

Title:
Reconstructing a transmission network and identifying risk factors of secondary transmissions in the 2010 foot-and-mouth disease outbreak in Japan

Date:
2019-09-01

Citation:
Hayama, Y., Firestone, S. M., Stevenson, M. A., Yamamoto, T., Nishi, T., Shimizu, Y. & Tsutsui, T. (2019). Reconstructing a transmission network and identifying risk factors of secondary transmissions in the 2010 foot-and-mouth disease outbreak in Japan. TRANSBOUNDARY AND EMERGING DISEASES, 66 (5), pp.2074-2086. https://doi.org/10.1111/tbed.13256.

Persistent Link:
https://hdl.handle.net/11343/285967

DR. YOKO HAYAMA (Orcid ID : 0000-0003-4814-0945)

Article type : Original Article

Reconstructing a transmission network and identifying risk factors of secondary transmissions in the 2010 foot-and-mouth disease outbreak in Japan

Running title: FMD Transmission network and risk factors

Yoko Hayama*[1], Simon M. Firestone[2], Mark A. Stevenson[2], Takehisa Yamamoto[1], Tatsuya Nishi[3], Yumiko Shimizu[1], Toshiyuki Tsutsui[1]

[1]Viral Disease and Epidemiology Research Division, National Institute of Animal Health, National Agriculture Research Organization, Tsukuba, Ibaraki, Japan

[2]Asia-Pacific Centre for Animal Health, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[3] Exotic Disease Research Station, National Institute of Animal Health, National Agriculture and Food Research Organization, Japan

[Corresponding author information]

Yoko Hayama

Viral Disease and Epidemiology Research Division, National Institute of Animal Health, National

Agriculture and Food Research Organization, 3-1-5 Kannondai, Tsukuba, Ibaraki 305-0856, Japan

Tel/Fax: +81-29-838-7769, E-mail: hayama@affrc.go.jp

Summary

Research aimed at understanding transmission networks, representing a network of "who infected whom" for an infectious disease outbreak, have been actively conducted in recent years. Transmission network models incorporating epidemiological and genetic data are valuable for elucidating disease transmission pathways. In this study, we reconstructed the transmission network of the foot-and-mouth disease (FMD) epidemic in Japan in 2010, and explored farm-level risk factors associated with increased risk of secondary transmission.

A published, systematic Bayesian transmission network model was applied to epidemiological data of 292 infected farms and whole genome sequence data of 104 of the infected farms. This model can make inferences for known infected farms even lacking genetic data. After estimating the consensus network, the accuracy of the network was examined by comparison with epidemiological data. Then, risk factors inferred to have been sources of secondary transmission were explored using zero-inflated Poisson regression model.

As far as we are aware, this study represents the largest FMD outbreak transmission network to be published by such means combining epidemiological and genetic data. The consensus network reasonably generated the epidemiological links, which were estimated from the actual epidemiological investigation. Among 292 farms, 101 farms (35%) were inferred to have been the sources of secondary transmission, and amongst these farms, the median number of secondary cases was 2 (min:1 – max:18) farms. The farm-type (small and large -sized pig farms), the number of days from onset to notification, and the number of susceptible farms within a 1-kilometer radius were

significantly associated with secondary transmission.

Transmission network modeling enabled inference of the connections between infected farms during the FMD epidemic and identified important factors for controlling the risk of secondary transmission. This study demonstrated that the predominant susceptible species held on a farm, farm size, and animal density were associated with increased onwards transmission.

Keywords

transmission network, secondary transmission, zero-inflated Poisson model, foot-and-mouth disease, Japan

Introduction

Foot-and-mouth disease (FMD) is a highly contagious and economically important disease of cloven-hoofed animals (OIE, 2017). An outbreak of FMD in a previously free country has a major economic and social impact, and results in the loss of export markets, causing severe disruption to domestic markets. In Japan, a large-scale FMD epidemic by serotype O virus occurred in 2010 in Miyazaki Prefecture in the southern part of Japan (Muroga et al., 2012) (Fig. 1). Due to the rapid spread of the disease, emergency vaccination was conducted to contain it. Ultimately, a total of 292 farms were affected, and almost 290,000 animals, including both infected and vaccinated animals, were culled (vaccination-to-kill policy). The epidemic caused severe economic damage to livestock and related industries in the affected region (Hayama et al., 2017).

When an outbreak of an infectious disease such as FMD occurs, an epidemiological investigation is conducted to identify potential transmission pathways through contact-tracing of animals, people,

and vehicles from infected farms. However, in practice, obtaining detailed and accurate contact-tracing data requires great effort and is sometimes difficult to acquire immediately after an outbreak. Field investigations are resource-intensive and may result in a high degree of uncertainty with regard to the source of infection and timing of unobserved transmission (Jewell et al., 2009). Additionally, it is uncertain how efficiently the transmission occurs via movements of animals, people, and vehicles between farms.

Modelling the transmission network of infectious disease outbreaks has been a highly active research area in recent years. Transmission network model can be used to infer the network of "who infected whom" during an infectious disease epidemic (Cottam et al., 2008; De Maio et al., 2016; Jombart et al, 2014; Klinkenberg et al, 2017; Lau et al., 2015; Morelli et al., 2012; Ympa et al., 2013). This approach enables us to infer the transmission network by incorporating epidemiological and genetic data. A detailed review and benchmarking study of recently published transmission network models is available (Firestone et al., 2019). Estimation of the transmission network for an infectious disease outbreak enables us to understand the mechanism of disease transmission, which complements the epidemiological investigations and helps us to better understand the features of spread. Transmission networks are particularly valuable for quantifying the number of secondary transmissions produced by a single primary case. Identification of outliers that caused high numbers of secondary cases (i.e., superspreading events) and analysis for the factors and mechanisms behind superspreading events is essential to the establishment of effective control measures for future outbreaks.

FMD virus (FMDV), from the genus *Aphthovirus* and family *Picornaviridae*, is a single-stranded positive-sense RNA virus with an approximately 8.4-kilobase (kb) genome (Carrillo et al., 2005). As with other RNA viruses, the FMDV genome is highly mutable (mutation rate of 10^{-3} to 10^{-5} per

96 genomic replication). Thus, due to the low fidelity of RNA replication producing sufficient genetic
97 variation, FMD outbreaks have been studied to reconstruct the transmission networks at the farm
98 level (Drake and Holland, 1999; Cottam et al., 2008; Kao et al., 2014).
99
100 In the current study, we first reconstructed a transmission network of the 2010 FMD epidemic in
101 Japan using a transmission network modeling method that combines available epidemiological and
102 genetic data. Next, farm-level risk factors associated with increased risk of secondary transmission
103 were explored from the estimated transmission network.
104
105 **Materials and methods**
106 Epidemiological data
107 In this study, on the assumption that all infected farms had been detected during the epidemic,
108 epidemiological data from all 292 infected farms detected during the 2010 FMD epidemic in Japan
109 (207 cattle farms, 84 pig farms, and 1 goat farm) were used for the reconstruction of the transmission
110 network. These data contained the geographical location of farms, farming information such as
111 species and number of animals per farm, as well as disease-related information including estimated
112 date of the first clinical onset, and dates of notification and completion of culling on detected farms.
113 The date of the first clinical onset on detected farms was estimated from the ages of lesions on
114 infected animals, the results of serological testing (antibody titers), and information from farmers, as
115 described by Hayama et al. (2012). Serological testing was performed at the National Institute of
116 Animal Health (NIAH) using liquid-phase blocking ELISA (The Pirbright Institute, U.K.) against
117 the FMD type-O virus, according to the manufacturer's instructions.
118
119 Genetic data

During the epidemic, clinical samples were submitted by Miyazaki Prefectural Government to NIAH for diagnosis of FMD. Then, 104 virus isolates were obtained from the samples collected from 104 farms, as described previously (Nishi et al., 2017). The isolated viruses were identified as type-O, belonging to genotype Mya-98 of Southeast Asia (SEA) topotype, based on VP1 region sequencing (Valdazo-Gonzalez et al., 2013). The nucleotide sequences of 104 viruses showed a pairwise identity of more than 99.56%, without any genetic deletions or insertions (Nishi et al., 2017). In this study, a dataset composed of the full length of the L-fragment sequences (7667 bases) from the 104 FMDVs was used (Genbank accessions: LC149617-LC149720 (Nishi et al., 2017)).

Transmission network model

In selecting the transmission network model, the following two points were considered to estimate the transmission network of FMD epidemic in Japan: 1) the model makes use of farm distance data; 2) the model can handle observed infected farms missing genetic data. For estimation of the FMD epidemic, the first point was necessary to reconstruct the spatial pattern of transmission network, and the second was essential because not all infected farms in this epidemic had genetic data available.

Through a review of comparisons between the features of several transmission network models (De Maio et al., 2016; Firestone et al., 2019), a systematic Bayesian transmission network model published by Lau et al. (Lau et al., 2015) was considered as the most appropriate method in this study. This model is a Bayesian Markov Chain Monte Carlo (MCMC) framework that simultaneously and explicitly infers the transmission network and unobserved transmitted sequences by systematically integrating the available epidemiological and genetic data. This model allows the implementation of the farm distance data_using a spatial kernel, and can handle observed infected cases that are missing genetic data. Most available models have similar features and infer based on

144 similar input data. They do however differ markedly in their formulations and assumptions, thus the
145 need to implement the most appropriate for the purpose, based on consideration of the model
146 formulation and assumptions, and performance in any available comparison studies. Thus, Lau's
147 model was chosen to reconstruct the transmission network and analyze the spatial and temporal
148 spread of this disease during the FMD epidemic despite lack of genomic data for some of affected
149 farms. This model has recently been demonstrated to be the most accurate for inferring the
150 transmission network and timing of exposures based on a benchmarking study that evaluated the
151 reproducibility of transmission networks using simulated FMD outbreaks (Firestone et al., 2019).
152
153 In this model, the underlying epidemic process was incorporated through the spatial-temporal
154 stochastic SEIR model, where farms have states of susceptible (S), exposed (E), infectious (I), and
155 Removed (R). The spatial relationship K between an infectious farm i and susceptible farm j was
156 represented by the power law kernel function, which allows occasional long-range transmissions of
157 infection with frequent short-range transmissions.
158 $$K(d_{ij};\kappa) = \frac{1}{1 + d_{ij}^{\kappa}}$$
159 where d_{ij} is the Euclidean distance between farms and κ is the power of the kernel to be inferred. The
160 overall transmission rate of farm j to become infected by any other farm i was determined based on
161 the term $\alpha + \beta\sum_i K(d_{ij};\kappa)$, where α and β are the primary and secondary transmission rates,
162 respectively. Here, the primary transmission rate α is the rate of causing primary cases (i.e. initial
163 infections from external sources), and the secondary transmission rate β is the rate of infection from
164 the surrounding infectious farms.
165
166 A nucleotide sequence is composed of four nucleotide bases: adenine (A), guanine (G), cytosine (C),
167 and thymine (T) for DNA; whereas A, G, C and uracil (U) instead of T for RNA. These nucleotide

bases are classified as pyrimidines (A and G) and purines (T(U) and C). Transitions are interchanges between bases in the same category (A-G or C-T(U)), and transversions are interchanges between bases from the different categories. Specifically, the process of nucleotide substitutions was modeled by a continuous-time Markov process, adopting the two-parameter Kimura model (Yang, 2006), which allows different rates of transition (μ_1) and transversion (μ_2). Additionally, in this transmission network model, a universal master sequence G_M was assumed to be an origin sequence for stochastically generating the primary sequences (i.e. actual sequences passed to primary cases) (Lau et al., 2015). Specifically, for each farm inferred to be an index, each nucleotide base of the 'primary sequence' assumed to have initiated infection in animals on this farm was assumed to have a probability p (the variation parameter) of differing from the base at the corresponding site in G_M. Those nucleotide bases that differed from the corresponding site in G_M, were drawn uniformly from the three possible alternatives. The G_M, whose initial value was arbitrarily set as the first observed sequence among all observed samples, was also updated simultaneously in the MCMC algorithm. By applying the Bayesian integration framework to all infected farms, the complete set of transmitted sequences, the transmission network, and all exposure times are simultaneously inferred.

In this study, all unobserved parameters were given uninformative uniform priors and the following inputs were given initial values (supplementary file, S. Table 1): date of exposure for each farm, the source of infection for each farm (initially proposed randomly based on farms likely infectious at time of exposure), the universal master sequence (G_M), the sequence on each farm at each transmission event (including farms for which genetic data were unavailable), the spatial transmission kernel parameter (κ), the mean and variance of the latent period (*lat*), the mean of the periods between day of onset and the last day of culling (approximately the farm-level infectious period, denoted as c), primary and secondary transmission rates (α and β), the rates of transitions and

192 transversions (μ_1 and μ_2), and the variation parameter (p) representing the proportion of variant sites
193 in each primary infection compared to the G_M sequence.
194
195 Lau's systematic Bayesian model was compiled from C++ source code. We ran four chains for
196 550,000 MCMC iterations. For each chain, the first 50,000 iterations were discarded as burn-in
197 based on assessment of convergence with Gelman and Rubin's shrink factor (Gelman and Rubin,
198 1992) and thinning the chains by 100 based on assessment of autocorrelation (Plummer et al., 2006).
199 Thus, a total of 20,000 samples were obtained.
200
201 After running the MCMC iterations, the consensus transmission network was estimated considering
202 model support, which is the posterior probability density for each link that supports the potential
203 links of a target farm and a source farm. Specifically, in the consensus network, for each target farm,
204 the source was considered as that with the highest posterior support. Additionally, we examined
205 whether the level of posterior support depends on whether or not genetic data are available from the
206 source or target farm, or both. Then posterior support was compared for the links based on the
207 different availability of genetic data between source and target farms, and the multiple comparisons
208 test was performed using the Wilcoxon signed rank test with Bonferroni correction. Moreover,
209 posterior distributions of model parameters were examined. To compare the spatial and temporal
210 pattern of the consensus network and the observed data, the spatial distribution of the disease was
211 plotted based on the estimated exposure date, and the temporal trend of the number of exposed farms
212 per day was calculated using the estimated exposure date.
213
214 Analysis of factors associated with secondary transmission
215 The number of secondary transmissions per infected farm was calculated from the consensus

network. Subsequently, a zero-inflated Poisson (ZIP) regression model (Dohoo et al., 2009) was used to determine the associations between farm-level factors and the count of secondary transmissions originating from each infected farm. Because there were many farms that did not cause secondary transmission during the epidemic, ZIP models, which deal with a large number of zero counts by simultaneously fitting a logistic-regression model and a Poisson model, are highly appropriate for analyzing the factors associated with secondary transmission. ZIP models enable us to examine the factors from two aspects: 1) what factors were associated with whether or not a farm caused any secondary transmissions; and 2) for farms that infected others, what factors were associated with the number of secondary transmissions. The count of secondary transmissions was the outcome variable in the Poisson part of the model. The association between an explanatory variable and the outcome variable was described in terms of a count ratio (CR). The outcome in the logistic part of the ZIP-regression model is expressed as the probability of a zero count; therefore, coefficients (including odds ratios (OR)) would have the opposite interpretation to what they would have in a usual logistic regression (Dohoo et al., 2009). Therefore, to aid interpretation, in the current study, all ORs were transformed into the inverse values, representing the likelihood of causing secondary transmissions.

The explanatory variables for each infected farm were animal species, farm size, the number of nearby susceptible farms, the number of days from clinical onset to notification, days from notification to end of culling, and days from clinical onset to end of culling. Animal species and farm size were classified as either small-sized cattle farm, large-sized cattle farm, small-sized pig farm, or large-sized pig farm, with consideration of the median number of animals on the infected farms for each species (i.e., the cut-off value of small-sized farms was less than 50 animals and less than 1,000 animals for cattle and pig farms, respectively). The number of susceptible farms located

240 within 100 m, 500 m, and 1 km of each infected farm was calculated. A susceptible farm is defined
241 as a farm that had not been exposed during the infectious period of the infected farm under
242 consideration. The explanatory variables were first examined using a univariable ZIP regression
243 model. Only variables with a p-value less than 0.1 and with a pairwise correlation coefficient of less
244 than 0.5 were used in the multivariable models. The final model was constructed using a manual
245 backward and forward variable selection approach. The goodness-of-fit of the estimated models was
246 compared using Akaike's information criterion (AIC). The model with the smallest AIC was deemed
247 to be the best-fitting model. Goodness-of-fit between the ZIP model and ordinary Poisson model
248 were evaluated by Vuong's test (Vuong, 1989), which is used for determining the best model.
249
250 Validation of the results of multivariable analysis
251 In the analysis of the factors associated with secondary transmission, the number of secondary cases
252 derived from the consensus network was used. While the consensus network includes the most likely
253 link for each target farm (source farm with the highest model support among the potential sources), a
254 target farm has the possibility of linking with other potential sources. Therefore, the effect of the
255 other possible networks on the results of the multivariable analysis was validated using a
256 pseudo-bootstrapping approach. Specifically, based on the values of support of the possible source
257 farms on each target farm, one possible source was randomly selected, and then a possible network
258 (not the consensus network) was constructed. A total of 100 possible networks were used for the
259 multivariable analysis, and the distributions of each coefficient were compared with the result of the
260 consensus network.
261
262 Analysis of the outputs from the transmission network model and all statistical analyses were
263 performed using R version 3.4.3 (R Core Team, 2017). All maps were made using ArcGIS 10 (ESRI,

264 Redlands CA).

265

266 **Results**

267 Consensus network

268 The consensus transmission network inferred for the FMD epidemic in 2010 in Miyazaki Prefecture
269 of Japan is presented in Fig. 2 in a hierarchical layout. This figure shows that some infected farms
270 caused more secondary transmissions than others. Among 292 infected farms, 101 farms (35%) were
271 inferred to have caused secondary transmission events. In the early phase of the epidemic, 12 farms
272 were estimated to be infected before the implementation of movement restriction; seven of these
273 farms caused secondary transmissions (58%). After the implementation of movement restriction, 94
274 of 280 farms (34%) caused secondary transmissions. Demonstrating the effect of movement
275 restrictions, there was a 42% reduction in the risk of onwards transmission for infected farms
276 detected after movement restrictions, compared with those farms detected before (relative risk =
277 0.58; 95% confidence interval: 0.35, 0.95; p-value = 0.11). The distribution of the number of
278 secondary cases is shown in Fig. 3. The median number of secondary cases was 2 (min:1 – max:18)
279 farms.

280

281 The consensus network corresponded well with the spatial and temporal dynamics observed during
282 the epidemic (Fig. 4, Fig. 5, and S. Fig 1, supplementary file). A cattle farm (Case 6) was estimated
283 as a primary case (i.e. source of the epidemic), and the transmission from this primary case to the
284 first farm notified (the index cattle farm, Case 1) was reconstructed with strong support (100%). The
285 long-distance transmission from Case 7 to Case 9 (both cattle farms) was also reconstructed with
286 strong support (100%).

287

The median level of model confidence (support) for links in the consensus network was 0.16 (5^{th} – 95^{th} percentile: 0.04 – 0.80). Fig. 6 shows a comparison of the support for links in the consensus network for the combinations of source and target farms where genetic data were available or not available. The support of the links where genetic data were available for both the inferred source and target farms ("aa", n = 36) was statistically significantly higher than that of the following: links where genetic data were available for source farms but not for target farms ("an", n = 84), links where genetic data were available for target farms but not for source farms ("na", n = 68), and links where genetic data were not available for source and target farms ("nn", n = 104) (all comparisons $p < 0.001$). The support of the links "na" was also higher than that of links "an" and "nn" (both comparisons $p < 0.001$).

Posterior distributions of model parameters such as the rates of transitions and transversions, the parameter of the spatial kernel, the mean and variance of the latent period, and the mean period from day of onset to the last day of culling are described in Table 1. The median and 95% credible intervals (CrI) of the transition rate and transversion rate per site per day were 1.71×10^{-5} (95% CrI: $1.50 \times 10^{-5} – 1.96 \times 10^{-5}$) and 0.12×10^{-5} (95% CrI: $0.090 \times 10^{-5} – 0.16 \times 10^{-5}$), respectively. Based on these values, the between-farm level values of the total mutation rate ($\mu_1 + 2 \times \mu_2$) and the ratio of transitions to transversion (μ_1/μ_2) were estimated to be 1.95×10^{-5} per site per day and 13.9, respectively. The mean latent period was estimated as 4.8 days (95% CrI: 3.6 – 6.0 days) and the mean period from onset of infectiousness to removal was 15.2 days (95% CrI: 13.6 – 17.1 days). The mean period between the exposure time to notification was 5.8 days (95% CrI: 2.4 – 13.4 days). The earliest date of exposure was estimated to be 25.2 days (95% CrI: 24.1 – 26.0 days) prior to the first confirmed case confirmation (20^{th} April 2010).

312 Factors associated with the number of secondary transmissions

313 In the univariable analysis, the following variables showed a relationship with the number of
314 secondary transmissions (p-values less than 0.1): farm type, number of susceptible farms within 1
315 km of a source farm, number of days from clinical onset to notification, and number of days from
316 notification to end of culling (S. Table 2, supplementary file). The multivariable analysis (Table 2)
317 showed that small-sized pig farms and large-sized pig farms were significantly associated with
318 increased counts of secondary cases. The CRs of small-sized pig farms and large-sized pig farms
319 with reference to small-sized cattle farms were 1.92 ($p = 0.005$), and 2.60 ($p < 0.001$), respectively.
320 Moreover, each additional day of delay from clinical onset to notification was associated with an
321 9.0% increase in the number of secondary transmissions from a farm (CR 1.09; $p < 0.001$). Days
322 from clinical onset to notification (inverse OR 1.32; $p < 0.001$) and the number of susceptible farms
323 within 1 km (inverse OR 1.06; $p = 0.003$) were also identified as risk factors for causing secondary
324 transmission. The Vuong's test produced a value for V of 4.27, which was significant ($p < 0.001$)
325 and more than 1.96, indicating that ZIP model was preferred over the ordinal Poisson model (Dohoo
326 et al., 2009).
327
328 Validation of the multivariable analysis results

329 A comparison of the results of the multivariable analysis using the consensus network and that of the
330 other possible networks is shown in S. Fig 2 (supplementary file). In the count component of model
331 outputs, CRs of the variables in the consensus network (except for that of number of susceptible
332 farms within 1 km) were included in the 95% CrI of those in the possible networks. The variables
333 that were significant risk factors in the consensus network were also indicated risk factors in the
334 possible networks (i.e. the 2.5th percentiles were larger than 1). The count ratio of the number of
335 susceptible farms within 1 km in the consensus network was slightly lower than that of the possible

336 networks, and this variable was a risk factor in the possible networks. In the logistic component of
337 model outputs, the inverse ORs of variables in consensus network (except for that of days from
338 clinical onset to notification) were included in the 95% CrI of those in the possible networks. The
339 inverse OR of days from clinical onset to notification was a little higher in the consensus network
340 than in the possible networks. Similar to that in the logistic component of model outputs, variables
341 that were significant risk factors in the consensus network were also indicated as risk factors in the
342 possible networks.
343
344 **Discussion**
345 The transmission network for the 2010 FMD epidemic in Miyazaki Prefecture of Japan was
346 reconstructed by implementing Lau's systematic Bayesian model (Lau et al., 2015) combining
347 epidemiological and genetic data. As far as the authors are aware, this is the largest outbreak
348 transmission network for FMD to as yet be inferred by such means combining epidemiological and
349 genetic data, with most other studies inferring either very small subsets of large outbreaks or
350 considerably smaller simulated epidemics (Lau et al., 2015; De Maio et al., 2016; Cottam et al.,
351 2008; Morelli et al., 2012; Klinkenberg et al., 2017). The estimated consensus network corresponded
352 well with the spatial and temporal disease transmission observed during the epidemic. It also
353 demonstrated that at least 12 farms were already infected before the movement restriction, and the
354 introduction of the disease was estimated 25.2 days (95% CrI: 24.1 – 26.0 days) prior to the first
355 confirmed case (20th April 2010). Among 292 infected farms, 101 of the farms (35%) were most
356 likely to have caused secondary transmissions. There was a large reduction in the risk of onwards
357 transmission after movement restrictions (though it was not statistically significant), and secondary
358 transmission in the early phase of the outbreak likely led to the large-scale FMD epidemic.
359

The transmission from the primary case (Case 6) to the index case (Case 1) was inferred with strong support in this study. According to epidemiological investigations during the initial phase of the epidemic, the FMDV was suspected to have been initially introduced into Case 6 farm, housing cattle and water buffalo (*Bubalus arnee*), located 600 m from the index case, and the virus was suspected to have been introduced at the end of March (Muroga et al., 2012). The source of infection of Case 6 remains unknown and was likely outside of Japan. In addition, long-distance transmission (Case 7 to Case 9), a distance of more than 70 km, was inferred with strong support. The epidemiological investigation revealed that an animal transportation truck was shared between these two farms and was used to ship cattle to the slaughter house before imposing movement restriction (Muroga et al., 2012). In attempts with several other published transmission network models (data not shown), no model succeeded in accurately inferring this long-distance transmission. It could be interpreted that the estimated transmission network by Lau's model succeeded in reasonably reconstructing the epidemiological links observed during the actual outbreak. Moreover, this study assumed that all infected farms had been detected during the epidemic. During this FMD epidemic, because the infected animals showed typical clinical signs of FMD, this allowed the farmers and veterinarians to find and notify the susceptible cases immediately. Following the detection of an infected farm, clinical surveillance and serological surveillance were conducted several times for the susceptible animals on the farms within the movement restriction zone based on the FMD control guideline (MAFF, 2004). Thus, the possibility of missing cases during the epidemic would be minimal, and the effect of missing cases on the inference made by this study would not be substantial.

In a comparative study of transmission network models using simulated FMD outbreak data (Firestone et al., 2019), Lau's systematic Bayesian model was found to be the most accurate for the

purpose of our investigation. Lau's model has been developed to facilitate systematic integration of epidemiological and genetic data, jointly inferring the transmission network and genome of the transmitted virus, even when data on a subset of the infected population is unavailable (Lau et al., 2015). This approach appears to provide a highly reasonable inference of the transmission network for this outbreak, although the levels of model support were lower than those reported for similarly sized simulated outbreaks (Firestone et al., 2019), or in the original model development and verification studies (Lau et al., 2015). This difference in model support between the current study and the previous studies may be either due to the simulations in previous studies not capturing the full extent of the complexity of the real system, or the lower availability of genetic data in this real setting.

For future improvement of the transmission network modelling, several considerations should be noted. First, in the current Lau's model, the transmission process was simplified to stabilize the MCMC estimation for transmission networks; that is, the transmission process implemented in the spatial-temporal SEIR model only depends on the distance between farms and does not include the effect of animal species or farm size. Considering the characteristics of FMD, incorporating terms representing different effects of animal species and farm size could make the model fit in a more biologically plausible manner. Additionally, because Lau's model assumes a dominant strain on an exposure at any time point, the within-host dynamics of the pathogens has not been considered (Lau et al., 2015). Thus, more sophisticated genomic evolution models including within-host dynamics could be considered. Finally, because this model does not explicitly assume the process of back-mutation, the effect of back-mutation on the estimation of transmission networks has not been analyzed in detail. In the case of a virus with a high mutation rate, such as FMDV, it might be better to perform detailed analysis of how back-mutation could affect the estimation of transmission

408 networks. For application to real outbreak data, these improvements and further comparative
409 evaluations of transmission network modeling methods could increase the accuracy of inferring the
410 transmission network.
411
412 Lau's systematic Bayesian model has the power to estimate a transmission network under the
413 condition of incomplete genetic data. However, it should be noted that the absence of genetic data
414 could introduce potential bias in estimating the transmission network and result in low accuracy of
415 the network. To gain higher accuracy in reconstructing transmission network, it is desirable to
416 increase available genetic data. In using this model, high accuracy has been demonstrated even when
417 50% of farms were missing genetic data (Firestone et al., 2019). In the current study, the highest
418 support (median 0.61, 5^{th} – 95^{th} percentile: 0.13 – 1.00) was observed for the links whose genetic
419 data was available in both source and target farms. During the FMD epidemic in 2010 in Japan, the
420 virus genome was obtained from 36% (104/292 farms) of the infected farms. In future outbreaks, it
421 is important to collect samples from as many suspicious/infected farms and as many animals per
422 farm as possible. Under such intense sampling and with appropriate field investigation and
423 epidemiological resources it may be possible to infer the transmission network in near real-time, thus
424 actively informing decision making. Such data collections will also allow more comprehensive
425 retrospective analyses, thus better informing preparedness for future events.
426
427 Farm-level risk factors associated with increased risk of secondary transmission were assessed based
428 on the inferred consensus transmission network. As far as the validation study, the effect of the
429 possible networks on the results of the multivariable analysis seems small. The results showed that
430 the farm type (host species and farm size) and farm density were significantly associated with
431 secondary transmission. When using small-sized cattle farms as a reference, pig farms generated

more secondary transmissions, and large-sized pig farms generated more secondary cases than small-sized pig farms. Pig farms, which have a high level of virus shedding (Alexandersen et al., 2003), are thought to have played an important role in disease dissemination during the epidemic, a result consistent with previous findings of the epidemiological studies of this epidemic (Nishiura et al., 2014; Hayama et al., 2013). Once an infection occurs in a large-sized farm, more animals get infections and shed a larger volume of virus. In general, as farm size increases, related persons or vehicles will also move more frequently onto or off the farms. Therefore, the large-sized farms are more likely to cause secondary transmissions compared to the small-sized farms. Moreover, infected farms surrounded by many susceptible farms within a range of 1 km posed a higher risk of causing secondary transmissions. Although this was not a significant risk factor associated with increasing the number of secondary cases in the consensus network, it should be noticed that this showed a potential association with increasing in the number of secondary cases based on the results of the validation study using the other possible networks. As pointed out in previous studies, areas densely populated with livestock are at a higher risk for large-scale epidemics than sparsely populated areas (Hayama et al., 2016; Backer et al., 2012; Durr et al., 2014; Le et al., 2005; Bessell et al., 2010). In such areas, transmission between farms within short distances of one another via multiple potential modes (termed "local spread") is considered highly probable because of the close proximity of farms, regardless of farm type. The analysis of the estimated transmission network revealed risk factors associated with secondary transmissions, namely, the chance of causing secondary transmission depends on the farm density, and the chance of causing onward transmission depends on the farm type (pig farms, especially, large-sized farms). For livestock areas with these risk factors, improved biosecurity measures should be encouraged to prevent the introduction of the disease, and better preparedness, such as efficient and rapid development of both human and material resources with appropriate organizational logistical support as well as a stockpile of the vaccine for emergency use,

456 will be required to achieve prompt and efficient containment of FMD outbreaks.

457

458 The number of days from clinical onset to notification was directly associated with causing
459 secondary transmission as well as increasing the number of secondary transmissions. This means
460 that the longer the period from clinical onset to notification posed a higher risk of causing further
461 transmissions. Although this factor seems to be slightly over-estimated in the consensus network, the
462 importance as a risk factor was not changed in the analysis of the possible networks. When animals
463 are infected by FMDV, clinical onset precedes virus-shedding (Alexandersen et al., 2003; Fukai et
464 al., 2011; Onozato et al., 2014). Thus, secondary transmission is more likely to occur when
465 notification is delayed, and such delays can potentially cause major economic and other impacts in
466 FMD outbreaks (Carpenter et al., 2011). Our results highlight the importance of raising awareness of
467 FMD and, from a risk-management perspective, of encouraging early notification of suspected cases.

468

469 Phylogenetic analysis has been widely used to investigate the evolutionary history of pathogens.
470 Genome sequences provide key insights into pathogen origins, as well as their evolution during an
471 epidemic (Holmes et al., 2016). However, it is not sufficient to estimate a transmission network
472 relying on the genetic data alone. We should consider that the transmission network and the
473 phylogenetic tree of an outbreak are different. As mentioned by Ympa et al. (2013), while the timing
474 of nodes in the transmission network corresponds to transmission times, the timing of internal nodes
475 of the phylogenetic tree corresponds to coalescent events that take place prior to transmission. Since
476 phylogenetic trees may be topologically dissimilar to transmission networks, especially under highly
477 dense sampling, interpreting phylogenetic proximity as epidemiological linkage can be dangerously
478 misleading (Worby et al., 2016). Transmission network modeling, which combines an evolutionary
479 process model using pathogen genetic data and a transmission dynamics model using

480 epidemiological data, is therefore a desirable approach to infer the potential linkages between
481 infected cases during an outbreak.
482
483 In conclusion, transmission network modeling was applied to real-world FMD epidemiological data,
484 and was found to be a useful method for tracing the spread of disease during the 2010 FMD
485 epidemic in Japan. Analyzing the secondary transmission from the estimated transmission network
486 also enabled us to understand the risk factors associated with the transmission of the disease. These
487 approaches provide valuable insights for understanding the mechanism of disease transmission and
488 for identifying the farm or areas at risk of FMD spread from perspectives of real-time analysis and
489 retrospective analysis, which will be useful for establishing effective prevention and control
490 strategies for future FMD outbreaks.
491
492 **Acknowledgements**
493 We are grateful to Dr. Max S. Y. Lau for his advice regarding implementation of the transmission
494 network analysis. This research was partially supported by an Australian Research Council DECRA
495 (DE160100477) and by the Japanese Ministry of Agriculture, Forestry and Fisheries (Management
496 Technologies for the Risk of Introduction of Livestock Infectious Diseases and Their Wildlife-borne
497 Spread in Japan, FY2018-2022).
498
499
500 **Conflict of interest**
501 The authors declare no conflicts of interest.
502
503 **References**

Alexandersen, S., Quan, M., Murphy, C., Knight, J., Zhang, Z. (2003). Studies of quantitative parameters of virus excretion and transmission in pigs and cattle experimentally infected with foot-and-mouth disease virus. *Journal of Comparative Pathology*, **129**, 268–282.

Alexandersen, S., Zhang, Z., Donaldson, A.I., Garland, A.J. (2003). The pathogenesis and diagnosis of foot-and-mouth disease. *Journal of Comparative Pathology*, **129**, 1–36.

Backer, J.A., Hagenaars, T.J., Nodelijk, G., van, Roermund H.J. (2012). Vaccination against foot-and-mouth disease I: epidemiological consequences. *Preventive Veterinary Medicine*, **107**, 27–40.

Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E. (2010). Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain. *International Journal of Infectious Diseases*, **14**, e210–215.

Carpenter, T.E., O'Brien, J.M., Hagerman, A.D., McCarl, B.A. (2011). Epidemic and economic impacts of delayed detection of foot-and-mouth disease: a case study of a simulated outbreak in California. *Journal of Veterinary Diagnostic Investigation*, **23**, 26–33.

Carrillo, C., Tulman, E.R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G.F., Rock, D.L. (2005). Comparative genomics of foot-and-mouth disease virus. *Journal of Virology*, **79**, 6487–6504.

Cottam, E.M., Thebaud, G., Wadsworth, J., Gloster, J., Mansley, L., Paton, D.J., King, D.P., Haydon, D.T. (2008). Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proceedings of the Royal Society: Biological Sciences*, **275**, 887–895.

De Maio, N., Wu, C.-H., Wilson D. J. (2016). SCOTTI: Efficient reconstruction of transmission within outbreaks with the structured coalescent. *PLoS Computational Biology*, **12**, e1005130.

Dohoo, I., Martin, W., Stryhn, H. (2009). Modelling Count and Rate Data. In Veterinary epidemiology research (2nd edition). VER Inc. Prince Edward Island, pp. 445–466.

528 Drake, J.W., Holland, J.J. (1999). Mutation rates among RNA viruses. *Proceedings of the National*
529 *Academy of Science U S A*, **96**, 13910–13913.
530 Durr, S., Fasel-Clemenz, C., Thur, B., Schwermer, H., Doherr, M.G., Dohna, H.Z., Carpenter, T.E.,
531 Perler, L., Hadorn, D.C. (2014). Evaluation of the benefit of emergency vaccination in a
532 foot-and-mouth disease free country with low livestock density. *Preventive Veterinary Medicine*,
533 **113**, 34–46.
534 Firestone, S.M., Hayama, Y., Bradhurst, R.A., Yamamoto, T., Tsutsui, T., Stevenson, M.A. (2019).
535 Reconstructing outbreaks: a methods comparison of transmission network models. *Scientific Reports,*
536 **9**, 4809.
537 Fukai, K., Morioka, K., Yoshida, K. (2011). An experimental infection in pigs using a
538 foot-and-mouth disease virus isolated from the 2010 epidemic in Japan. *The Journal of Veterinary*
539 *Medical Science*, **73**, 1207–1210.
540 Gelman, A., Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences.
541 *Statistical Science*, **7**, 457–472.
542 Hayama, Y., Muroga, N., Nishida, T., Kobayashi, S., Tsutsui, T. (2012). Risk factors for local spread
543 of foot-and-mouth disease, 2010 epidemic in Japan. *Research in Veterinary Science*, **93**, 631–635.
544 Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N., Tsutsui, T. (2013). Mathematical model of
545 the 2010 foot-and-mouth disease epidemic in Japan and evaluation of control measures. *Preventive*
546 *Veterinary Medicine*, **112**, 183–193.
547 Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N., Tsutsui, T. (2016). Potential impact of
548 species and livestock density on the epidemic size and effectiveness of control measures for
549 foot-and-mouth disease in Japan. *The Journal of Veterinary Medical Science*, **78**, 13–22.
550 Hayama, Y., Osada, Y., Oushiki, D., Tsutsui, T. (2017). An economic assessment of foot and mouth
551 disease in Japan. *Review Scientifique et Techique*, **36**, 207–215.

552 Holmes, E.C., Dudas, G., Rambaut, A., Andersen, K.G. (2016). The evolution of Ebola virus:
553 Insights from the 2013-2016 epidemic. *Nature*, **538**, 193–200.
554 Jewell, C.P., Keeling, M.J., Roberts, G.O. (2009). Predicting undetected infections during the 2007
555 foot-and-mouth disease outbreak. *Journal of the Royal Society Interface*, **6**, 1145–1151.
556 Jombart, T., Cori, A., Didelot, X., Cauchemez, S., Fraser, C., Ferguson, N. (2014) Bayesian
557 reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS*
558 *Computational Biology*, **10**, e1003457.
559 Kao, R.R., Haydon, D.T., Lycett, S.J., Murcia, P.R. (2014). Supersize me: how whole-genome
560 sequencing and big data are transforming epidemiology. *Trends in Microbiology*, **22**, 282–291.
561 Klinkenberg, D., Backer, J. A., Didelot, X., Colijn, C., Wallinga, J. (2017) Simultaneous inference
562 of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Computational Biology,*
563 **13**, e1005495.
564 Lau, M.S., Marion, G., Streftaris, G., Gibson, G. (2015). A systematic Bayesian integration of
565 epidemiological and genetic data. *PLoS Computational Biology*, **11**, e1004633.
566 Le Menach A., Legrand, J., Grais, R.F., Viboud, C., Valleron, A.J., Flahault, A. (2005). Modeling
567 spatial and temporal transmission of foot-and-mouth disease in France: identification of high-risk
568 areas. *Veterinary Research*, **36**, 699–712.
569 Ministry of Agriculture Forestry and Fisheries (MAFF). (2004). Guideline on FMD Prevention and
570 control (in Japanese).
571 http://www.maff.go.jp/j/syouan/douei/katiku_yobo/k_fmd/pdf/160401_fmd_guide.pdf
572 Morelli, M.J., Thebaud, G., Chadoeuf, J., King, D.P., Haydon, D.T., Soubeyrand, S (2012). A
573 bayesian inference framework to reconstruct transmission trees using epidemiological and genetic
574 data. *PLoS Computational Biology*, **8**, e1002768.
575 Muroga, N., Hayama, Y., Yamamoto, T., Kurogi, A., Tsuda, T., Tsutsui, T. (2012). The

foot-and-mouth disease epidemic in Japan, 2010. *The Journal of Veterinary Medical Science*, **74**, 399–404.

Nishi, T., Yamada, M., Fukai, K., Shimada, N., Morioka, K., Yoshida, K., Sakamoto, K., Kanno, T., Yamakawa, M. (2017). Genome variability of foot-and-mouth disease virus during the short period of the 2010 epidemic in Japan. *Veterinary Microbiology*, **199**, 62–67.

Nishiura, H., Ejima, K., Mizumoto, K., Nakaoka, S., Inaba, H., Imoto, S., Yamaguchi, R., Saito, M.M. (2014). Cost-effective length and timing of school closure during an influenza pandemic depend on the severity. *Theoretical Biology and Medical Modelling*, **11**, 5.

Onozato, H., Fukai, K., Kitano, R., Yamazoe, R., Morioka, K., Yamada, M., Ohashi, S., Yoshida, K., Kanno, T. (2014). Experimental infection of cattle and goats with a foot-and-mouth disease virus isolate from the 2010 epidemic in Japan. *Archives of Virology*, **159**, 2901–2908.

Plummer, M., Best, N., Cowles, K., Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.

The World Organisation for Animal Health (OIE). (2017). Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. http://www.oie.int/en/standard-setting/terrestrial-manual/access-online/.

Valdazo-Gonzalez, B., Timina, A., Scherbakov, A., Abdul-Hamid, N.F., Knowles, N.J., King, D.P. (2013). Multiple introductions of serotype O foot-and-mouth disease viruses into East Asia in 2010-2011. *Veterinary Research*, **44**, 76.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307–333.

Worby, C.J., O'Neill, P.D., Kypraios, T., Robotham, J.V., De, Angelis D., Cartwright, E.J., Peacock, S.J., Cooper, B.S. (2016). Reconstructing transmission trees for communicable diseases using densely sampled genetic data. *The Annals of Applied Statistics*, **10**, 395–417.

Yang, Z. (2006). Computational molecular evolution. Oxford University Press.

600 Ypma, R.J., van, Ballegooijen W.M., Wallinga, J. (2013). Relating phylogenetic trees to
601 transmission trees of infectious disease outbreaks. *Genetics*, **195**, 1055–1062.
602
603 **Figure Legends**
604 Fig. 1 The epidemic curve and epidemic area of the 2010 FMD epidemic in Japan.
605
606 Fig. 2 A hierarchical layout of a consensus network during the 2010 FMD epidemic in Japan. The
607 orange colored edges indicate the transmissions from Case 6 (source of the epidemic) to Case 1 and
608 Case 7 to Case 9. The generation of transmission is indicated vertically from top to bottom.
609
610 Fig. 3 Frequency distribution of the number of secondary transmissions.
611
612 Fig. 4 Spatial pattern of the reconstructed consensus network. (A) shows the network before the
613 movement restriction implemented on April 20 in 2010. “C6-C1” and “C7-C9” indicate transmission
614 from Case 6 to Case 1 and Case 7 to Case 9, respectively. (B) shows the long-distance transmission
615 from Case 7 to Case 9 before the implementation of the movement restriction. (C) shows the
616 network at the start of emergency vaccination on May 22 in 2010. (D) shows the network at the end
617 of the epidemic on July 4 in 2010.
618
619 Fig. 5 Temporal dynamics of the exposure time. The red line shows the number of exposed farms
620 per day based on the observed data, assuming that infected farms were exposed 5 days prior to
621 clinical onset date, considering the estimated latent period of 4.8 days. The black line shows the
622 number of exposed farms per day based on the median value of the estimated exposure time on each
623 infected farm. The gray lines show the number of exposed farm per day based on the exposure time

624 on each infected farm from all 20,000 MCMC samples.

625

626 Fig. 6 Comparison of support of the consensus network with or without the genetic data from source
627 and target farms. “aa”, genetic data were available from both source and target farms; “an”, genetic
628 data were only available from the source farm; “na”, genetic data were only available from target
629 farms; “nn”, genetic data were unavailable from both the source and target farms.

Table 1 Posterior median estimates and 95% credible intervals for the parameters

Parameter		Median	95% CrI
α	primary transmission rate	5.86×10^{-5}	$(2.10 \times 10^{-6}–3.13\times10^{-4})$
β	secondary transmission rate	0.45	(0.38–0.55)
mean(lat)	mean of latent period	4.84	(3.68–6.08)
var(lat)	variance of latent period	13.62	(6.20–25.42)
c	mean period from infection to removal	15.25	(13.63–17.11)
κ	parameter of spatial kernel	1.65	(1.42–1.93)
μ_1	transition rate	1.71×10^{-5}	$(1.50 \times 10^{-5}–1.96 \times 10^{-5})$
μ_2	transversion rate	0.12×10^{-5}	$(0.09 \times 10^{-5}–0.16 \times 10^{-5})$
p	variation parameter	0.52	(0.05–0.98)

CrI: Credible Interval

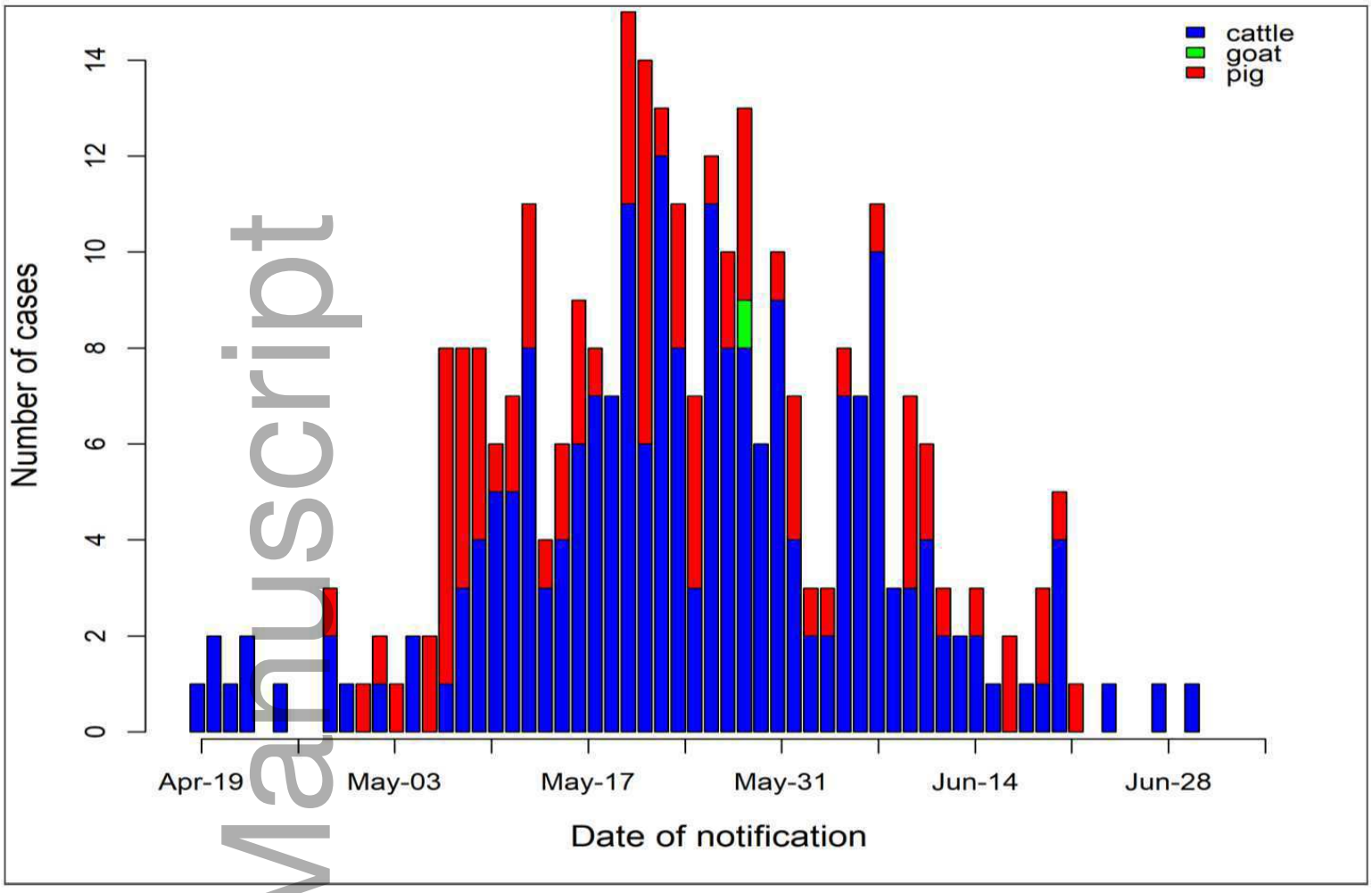
Table 2 Multivariable zero-inflated Poisson regression model for association between secondary transmission and farm characteristics

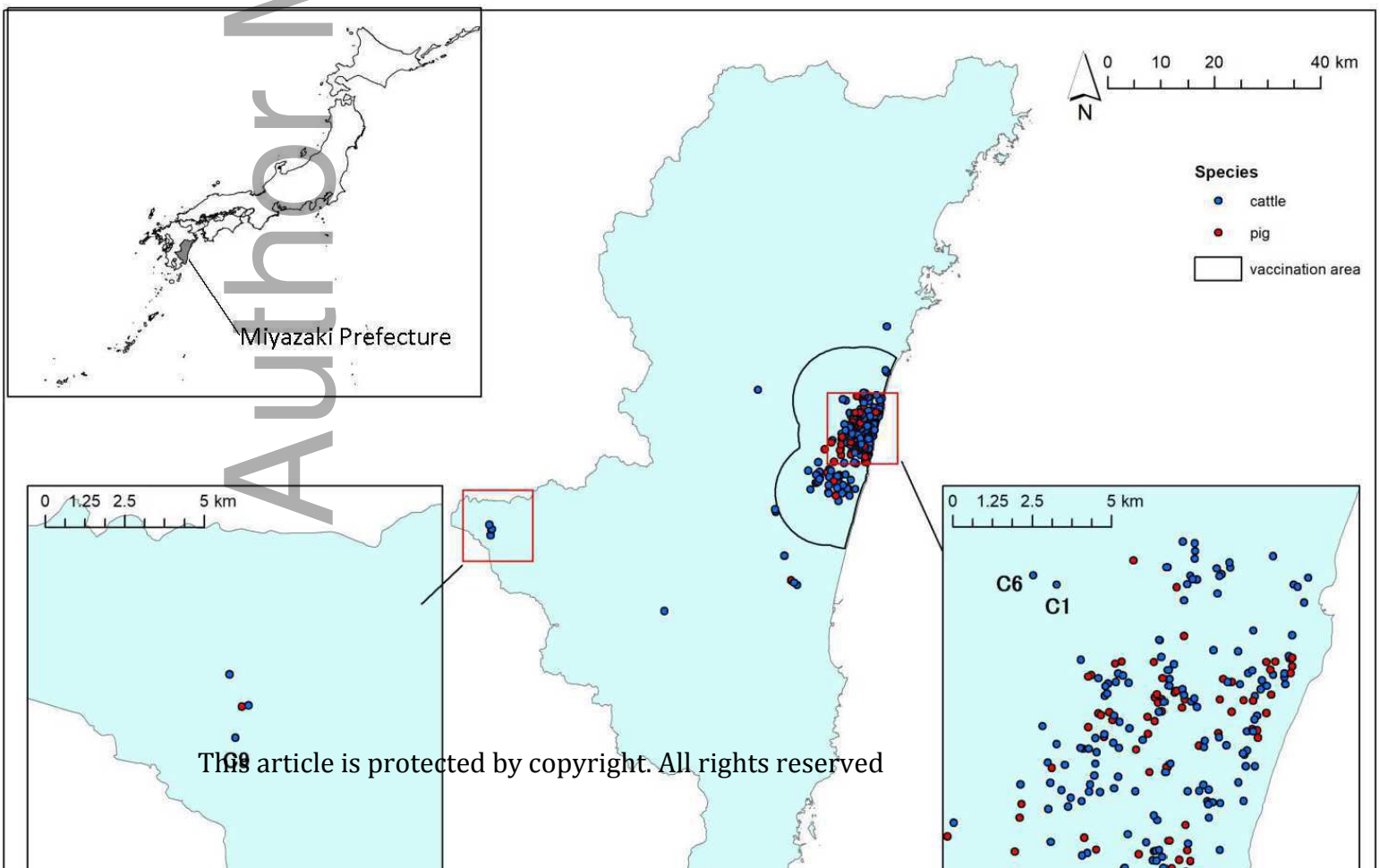
Variable	Levels	CR[a]	95% Confidence interval	p-value
Count part				
Species and farm size	small-sized cattle farm	1	–	–
	large-sized cattle farm	1.23	0.84, 1.82	0.289
	small-sized pig farm	1.92	1.22, 3.02	0.005
	large-sized pig farm	2.60	1.70, 3.98	< 0.001
Days from clinical onset to notification		1.09	1.06, 1.12	< 0.001
Number of susceptible farms within 1 km		1.00	0.99, 1.02	0.812
Logistic part		inverse OR[b]		
Species and farm size	small-sized cattle farm	1	–	–
	large-sized cattle farm	1.82	0.83, 3.99	0.132
	small-sized pig farm	0.77	0.29, 2.04	0.605
	large-sized pig farm	1.34	0.54, 3.37	0.528

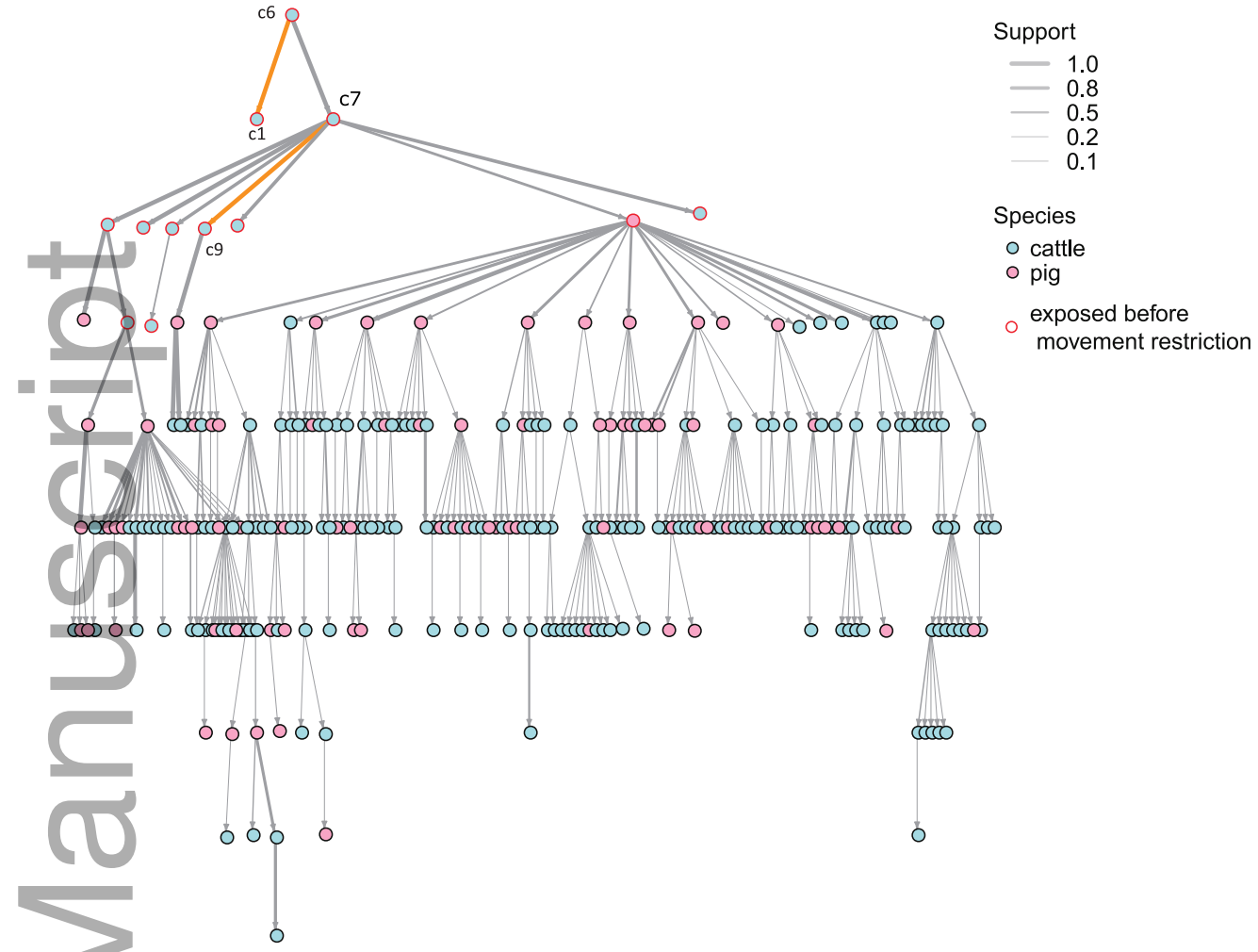
Days from clinical onset to notification	1.32	1.14, 1.55	< 0.001
Number of susceptible farms within 1 km	1.06	1.02, 1.11	0.003

[a] Count ratio estimated from count part of the model. CR>1 means that the factor is positively associated with the number of secondary transmissions.

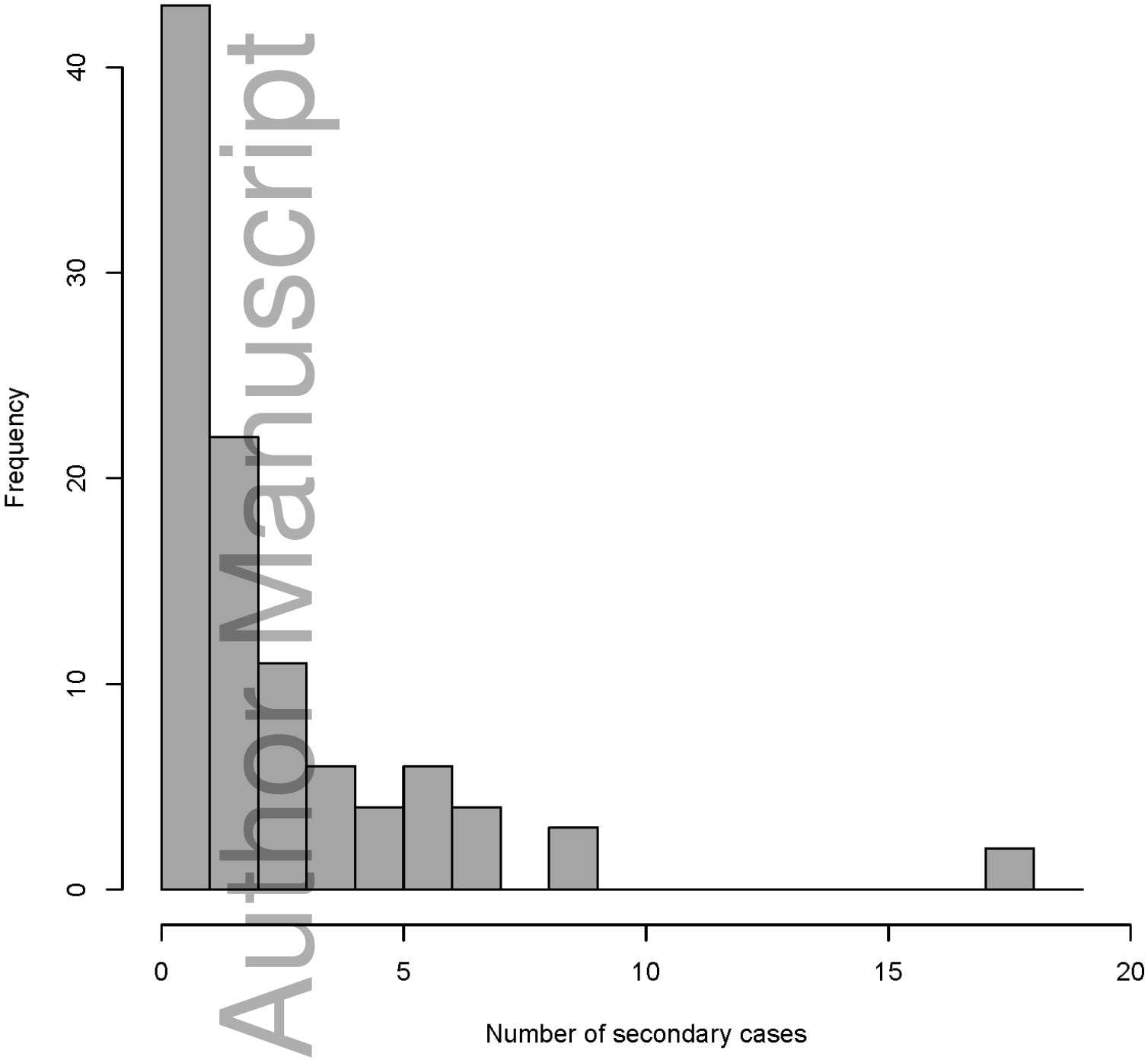
[b] Inverse value of odds ratio estimated from zero-inflation part of the model. Because the outcome in the logistic part of the ZIP-regression model was probability of a zero count, ORs were transformed into the inverse values. Inverse OR >1 means that the factor is positively associated with causing secondary transmission.



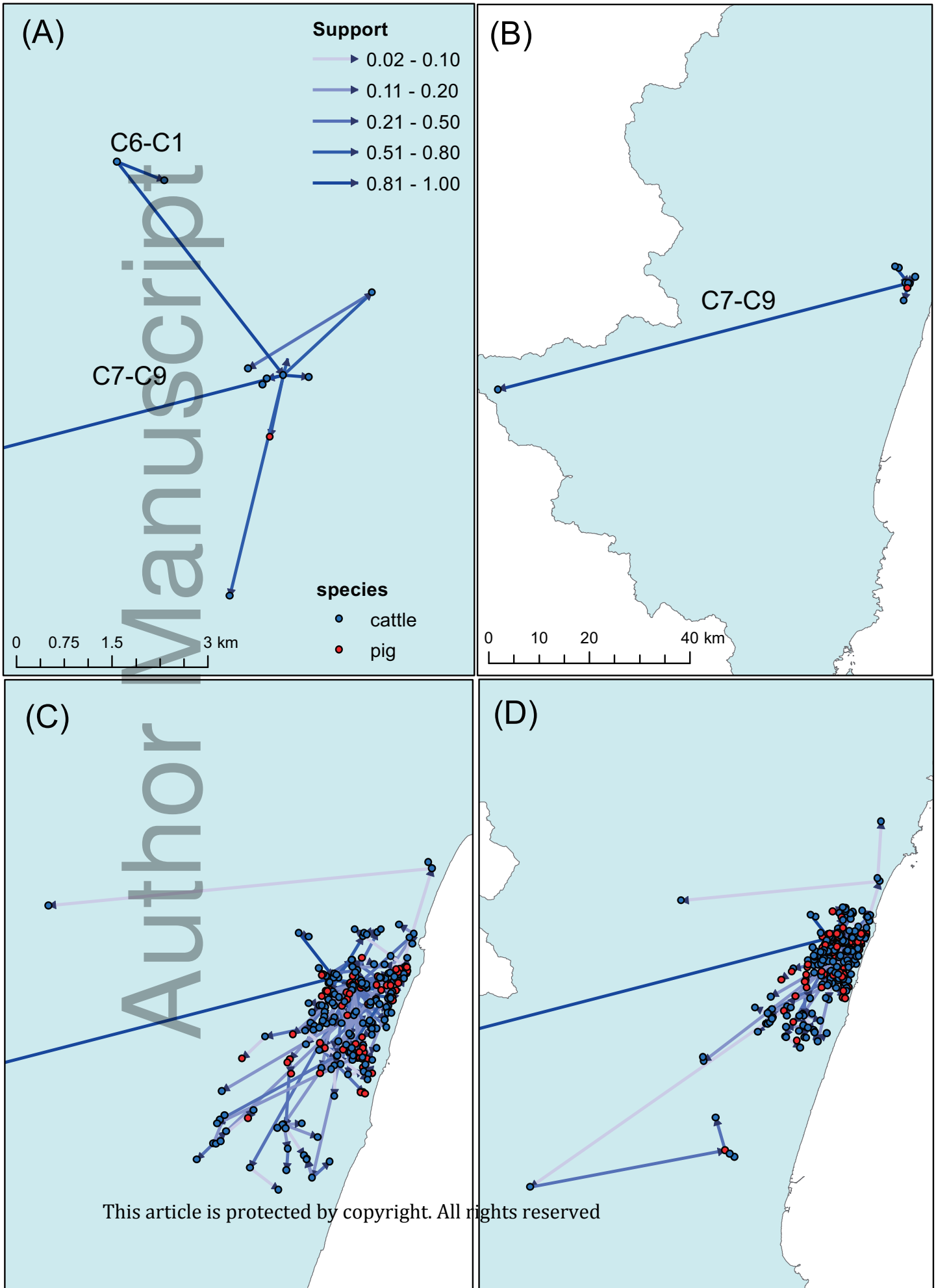


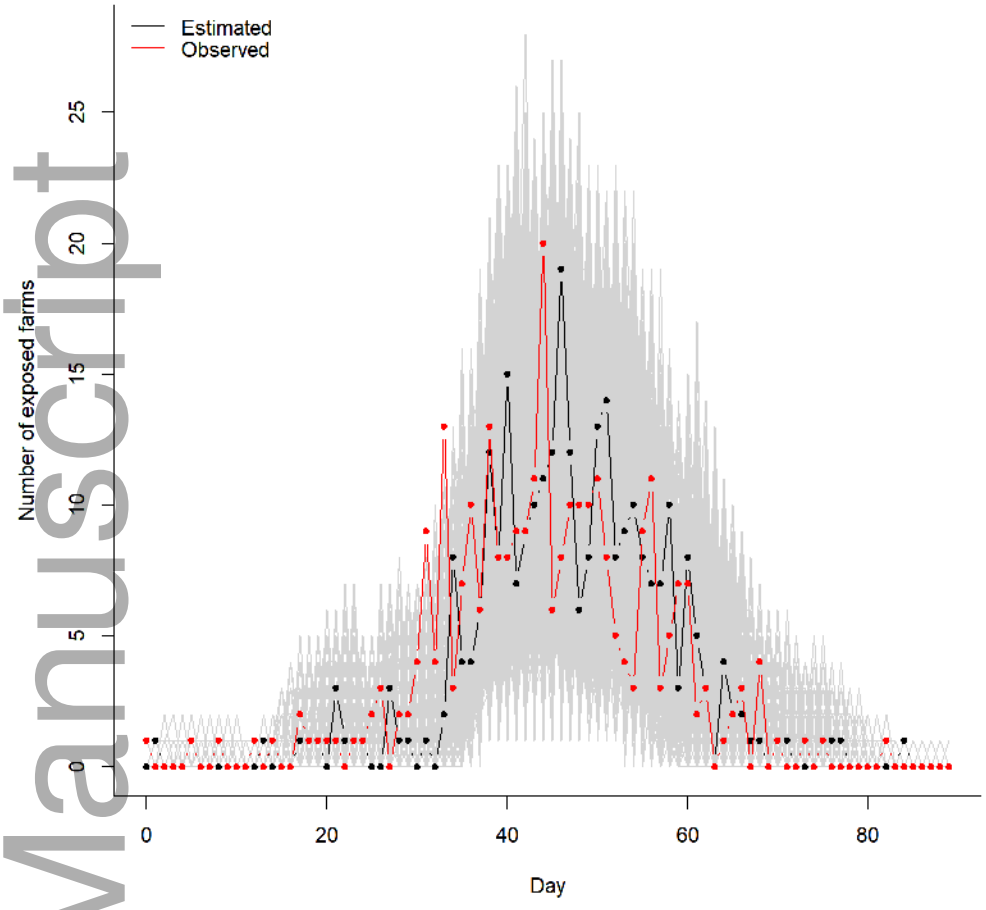


tbed_13256_f2.eps

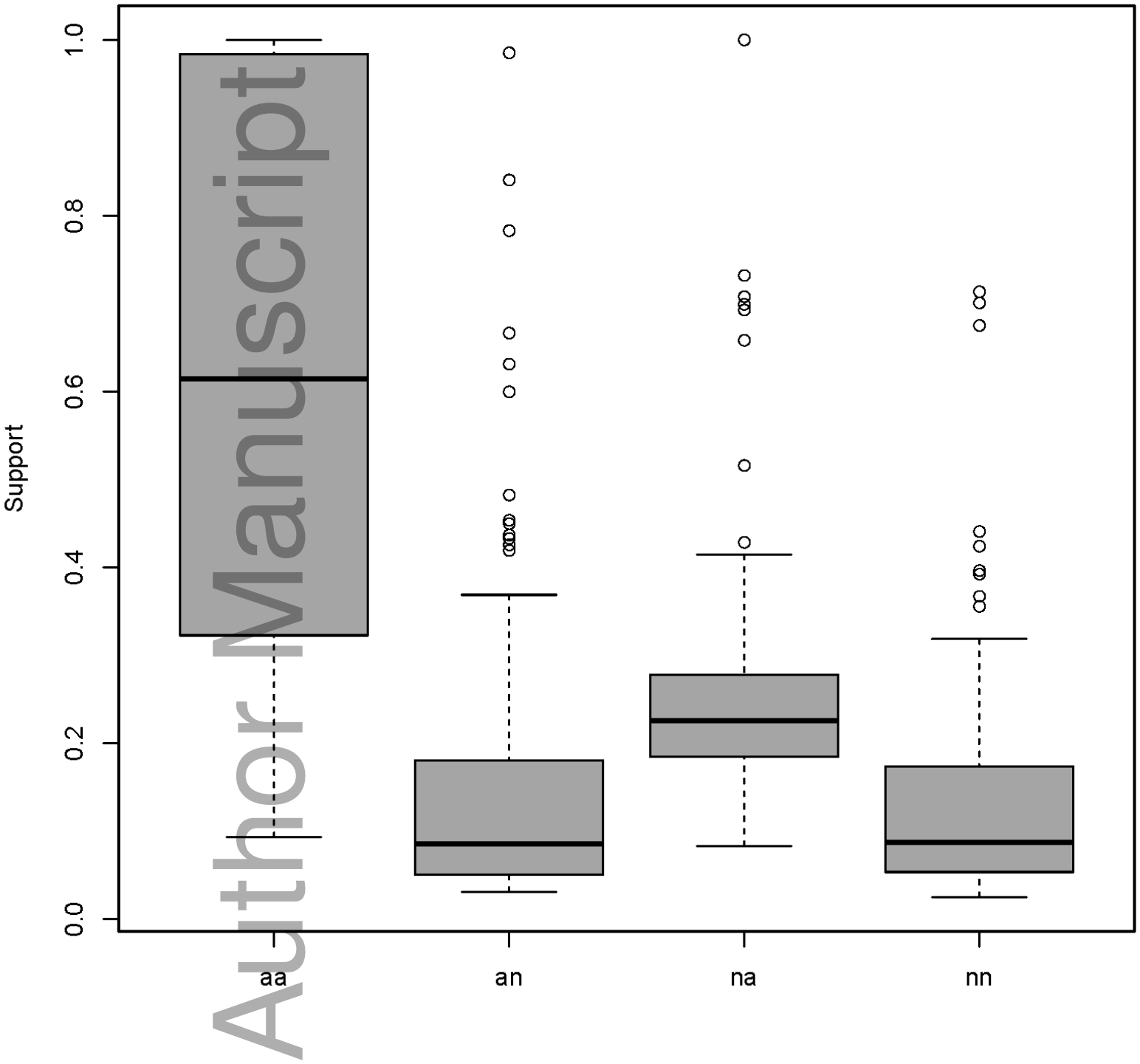


tbed_13256_f3.eps





tbed_13256_f5.tif



tbed_13256_f6.eps