SUPPLEMENT: Probabilistic reconstruction of measles transmission clusters in the United States between 2001 and 2016 using routine case reports.

1. Description of the US dataset

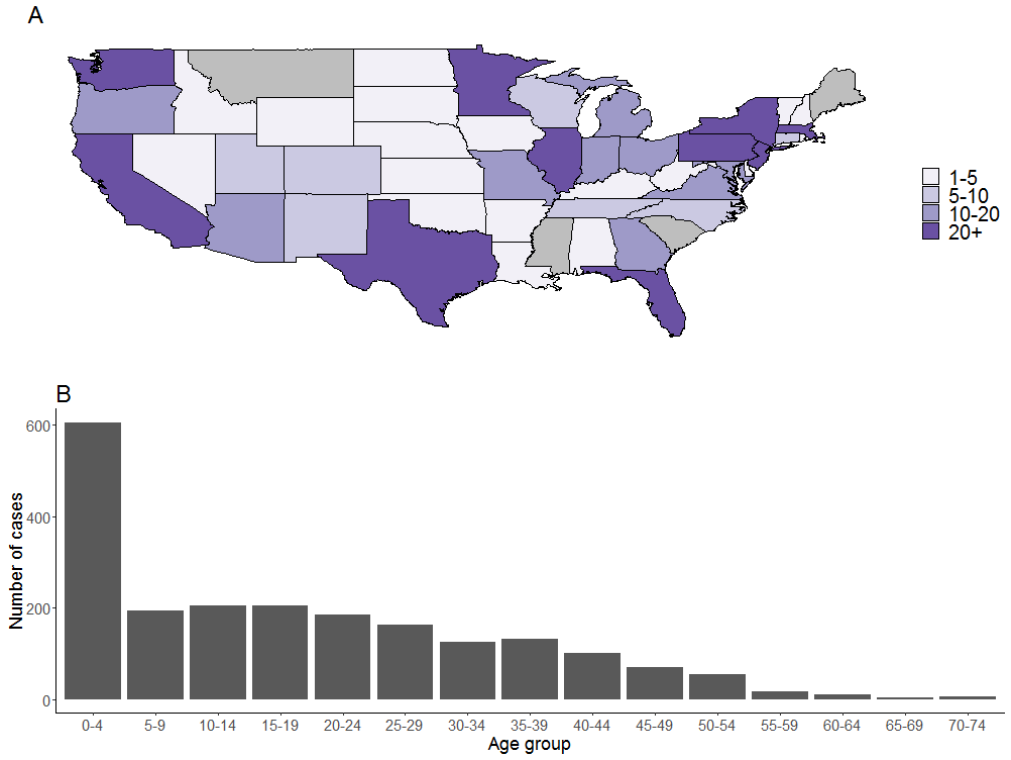

Figure S1: Description of the cases and clusters reported in the United States between 2001 and 2016. Panel A: number of clusters reported per state in the contact tracing investigations (5 clusters cover several states) and Panel B: Age distribution of the cases. Data from CDC's National Notifiable Disease Surveillance System. Hawaii and Alaska and not shown in Panel A.

2. Evaluation cluster matching

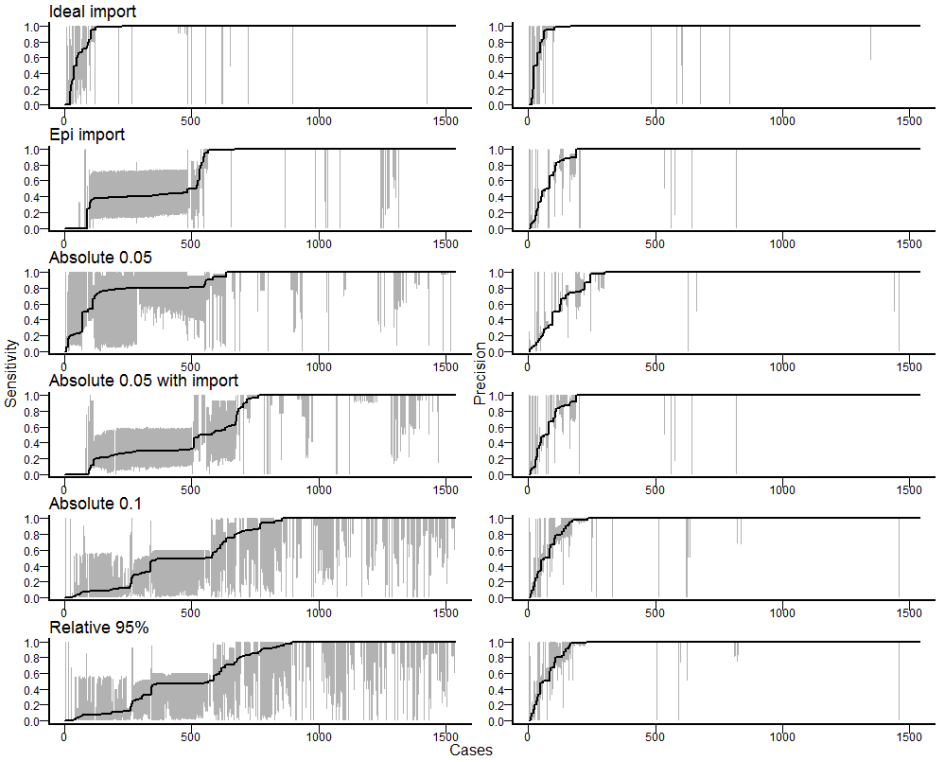

Figure S2: Confidence interval of the sensitivity (right panels) and precision (left panels) of the clusters of each case in the different runs. Grey areas represent the 95% credibility intervals, and the black line represents the median values of sensitivity or precision across all iterations.

3. Posterior distribution and convergence

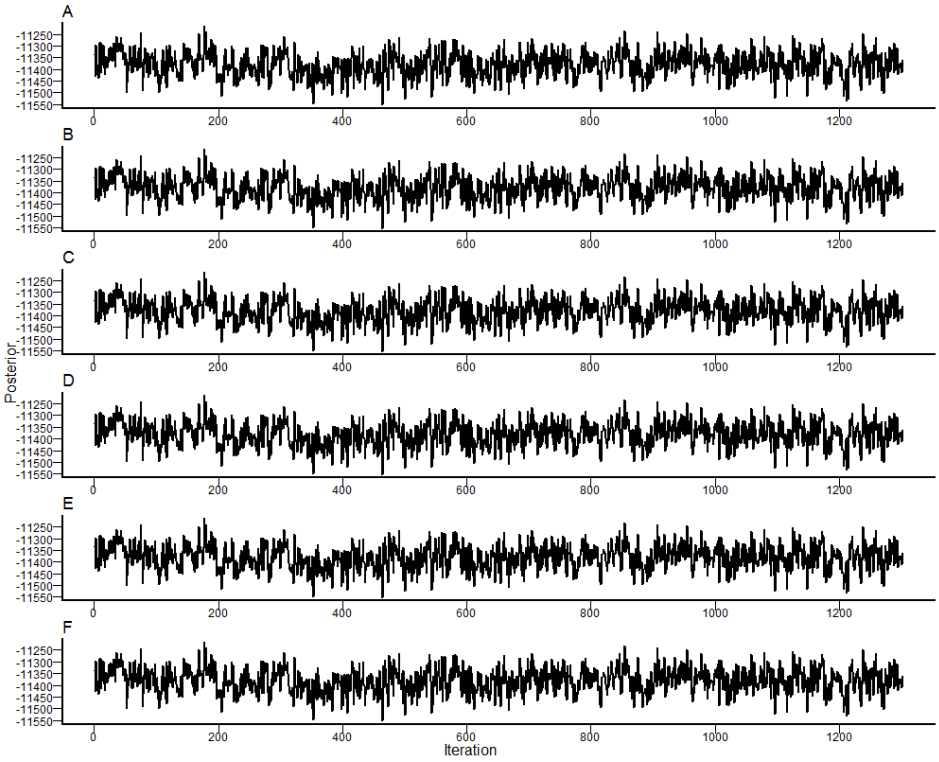

Figure S3: Trace of the posterior value of the MCMC runs after removing the burnin section and thinning. A: Ideal imports; B: Epi imports; C: Absolute threshold, $k = 0.05$ D: Absolute threshold, $k = 0.05$ with prior information on imports; E: Absolute threshold, $k = 0.1$; F: Relative threshold upper $\lambda = 97.5\%$ quantile.

4. Clusters stratified by state

a. Scenario 1

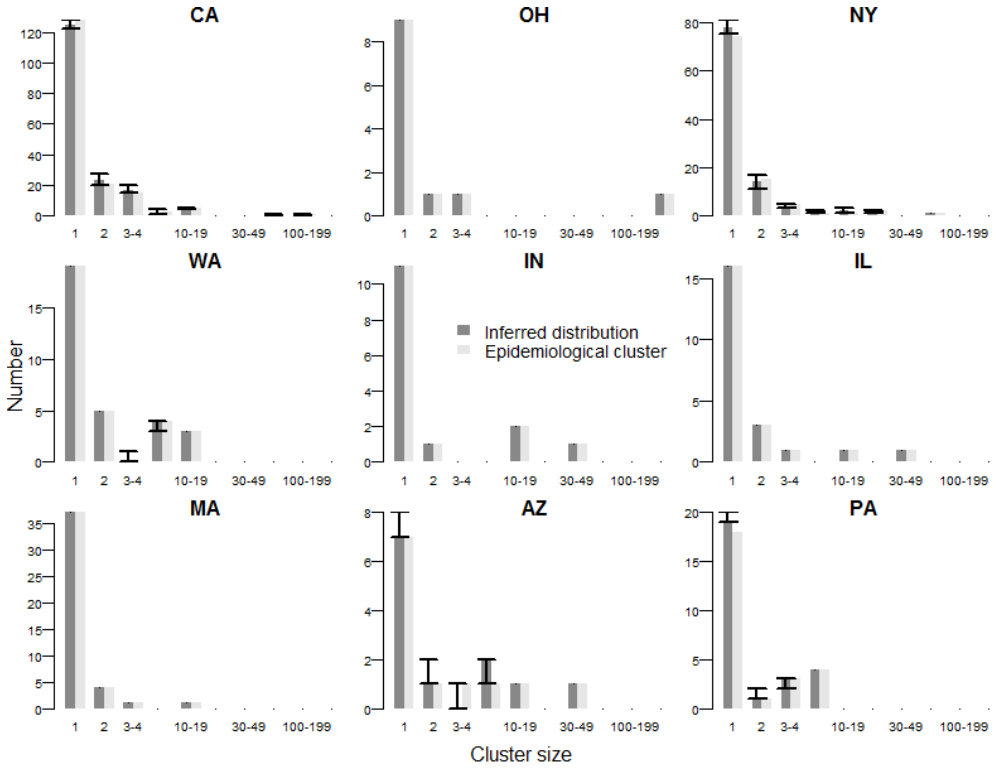

Figure S4: Comparison of the inferred and reference cluster size distributions in the nine states with most cases declared between 2001 and 2016 in scenario 1.

b. Scenario 2

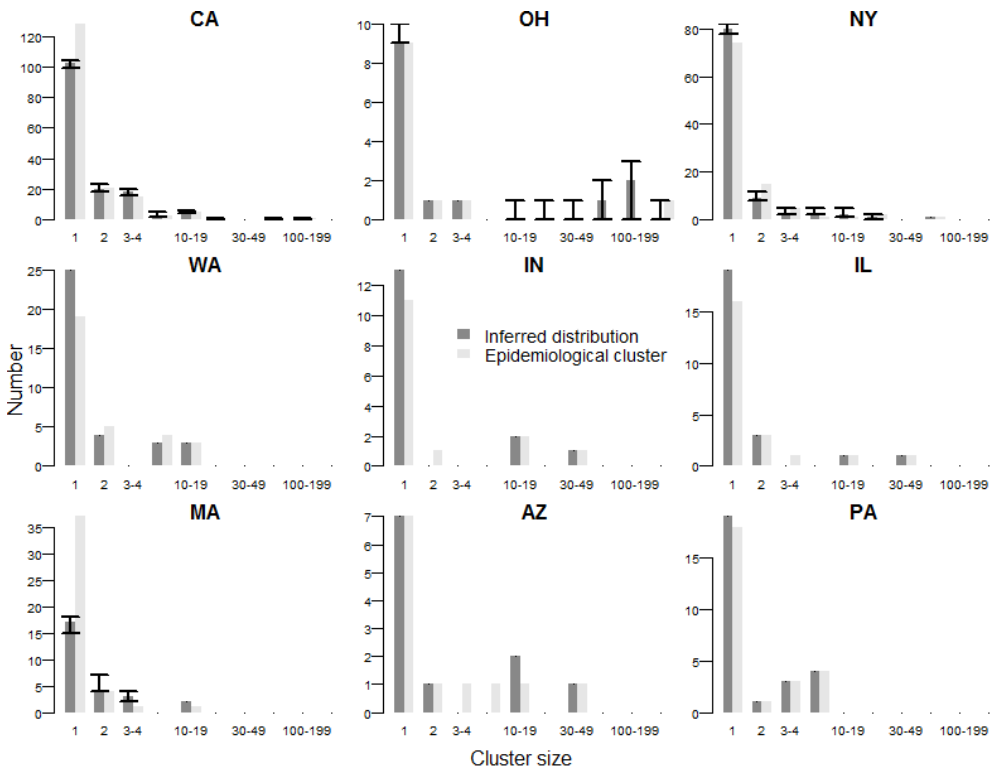

Figure S5: Comparison of the inferred and reference cluster size distributions in the nine states with most cases declared between 2001 and 2016 in scenario 2.

c. Scenario 3

i. Absolute threshold $k = 0.05$

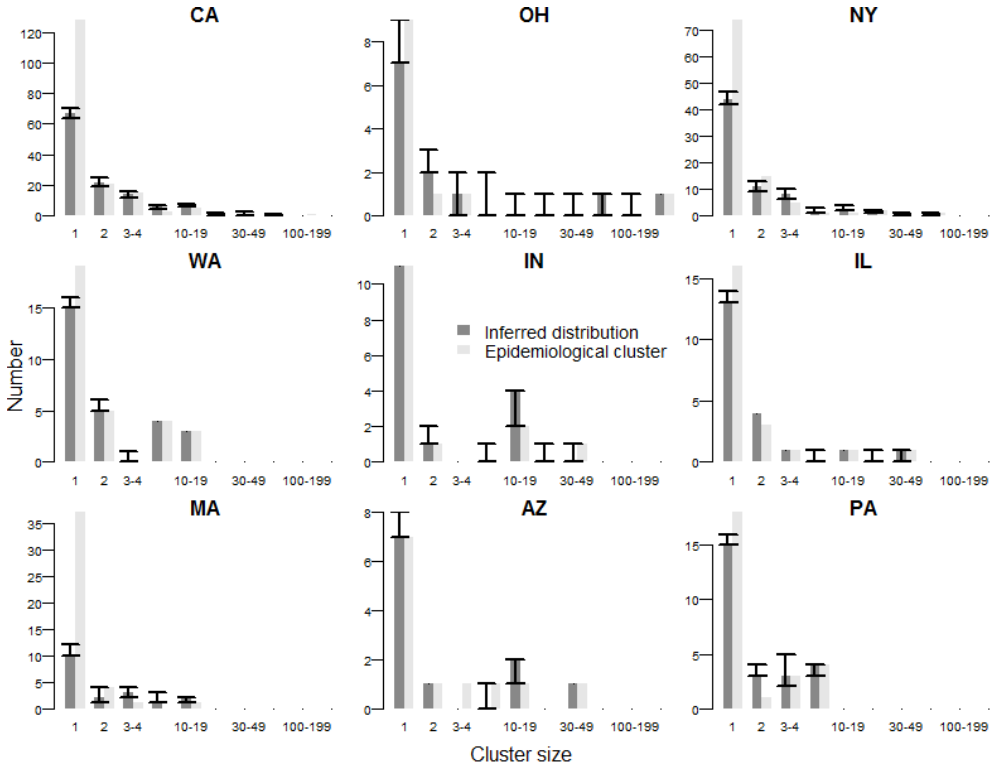

Figure S6: Comparison of the inferred and reference cluster size distributions in the nine states with most cases declared between 2001 and 2016 in scenario 3, with an absolute threshold $k = 0.05$ and no prior information on the import status of cases.

ii. Relative threshold $k = 95\%$

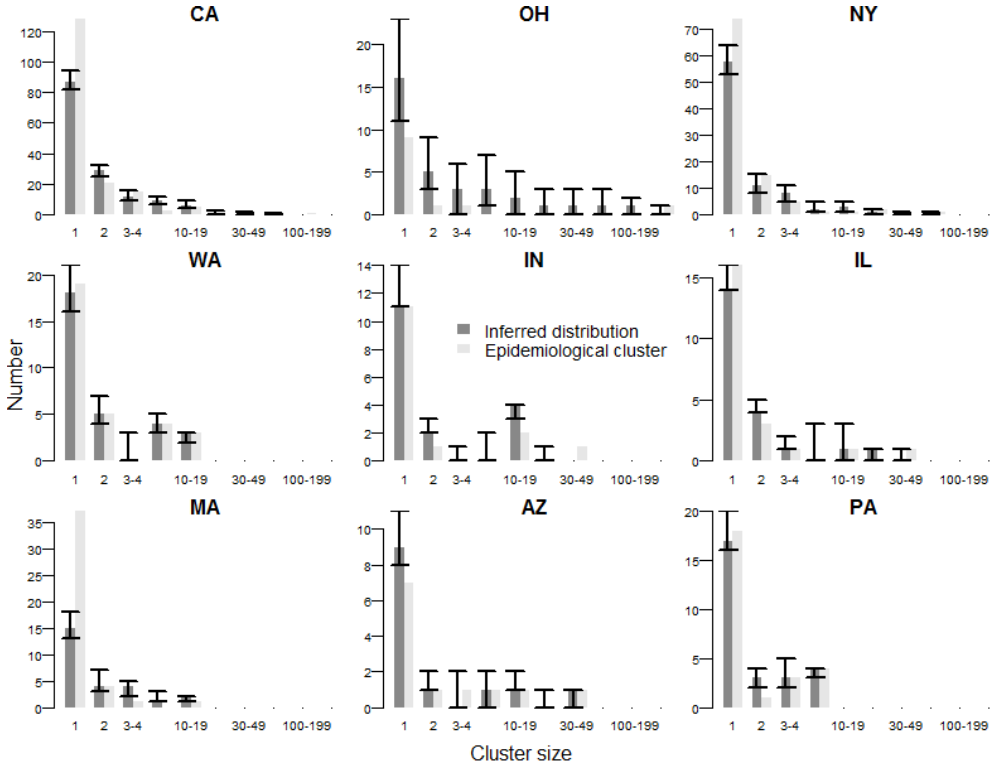

Figure S7: Comparison of the inferred and reference cluster size distributions in the nine states with most cases declared between 2001 and 2016 in scenario 3, with a relative threshold $k = 95\%$ and no prior information on the import status of cases.

iii. Epidemiological imports, and absolute threshold $k = 0.05$

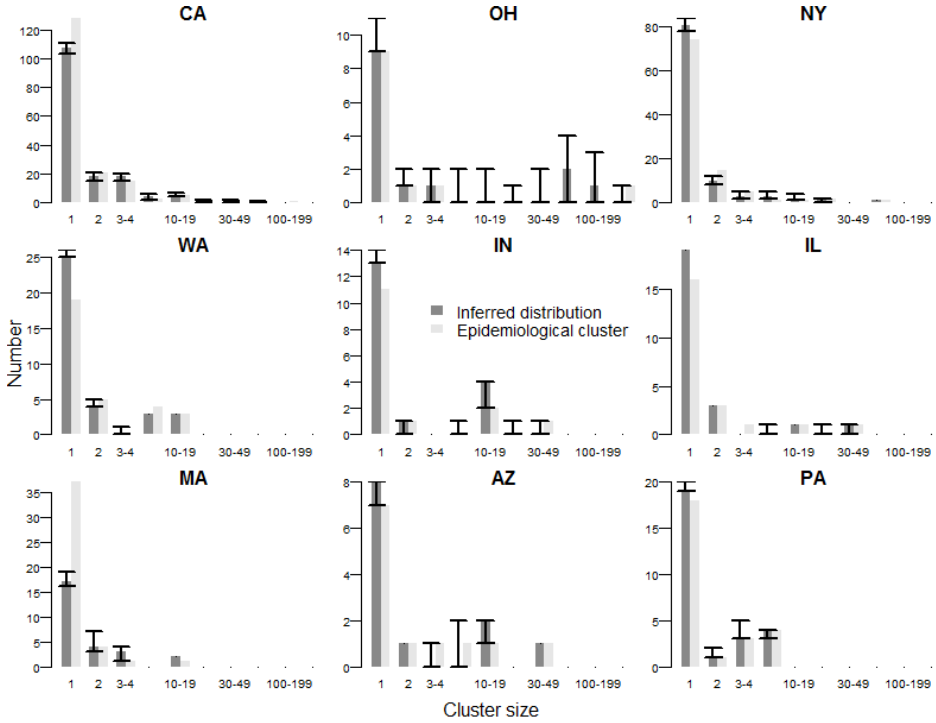

Figure S8: Comparison of the inferred and reference cluster size distributions in the nine states with most cases declared between 2001 and 2016 in scenario 3, with an absolute threshold $k = 0.05$ and using the import status distribution from the contact tracing investigations as prior information.

5. Parameter estimates

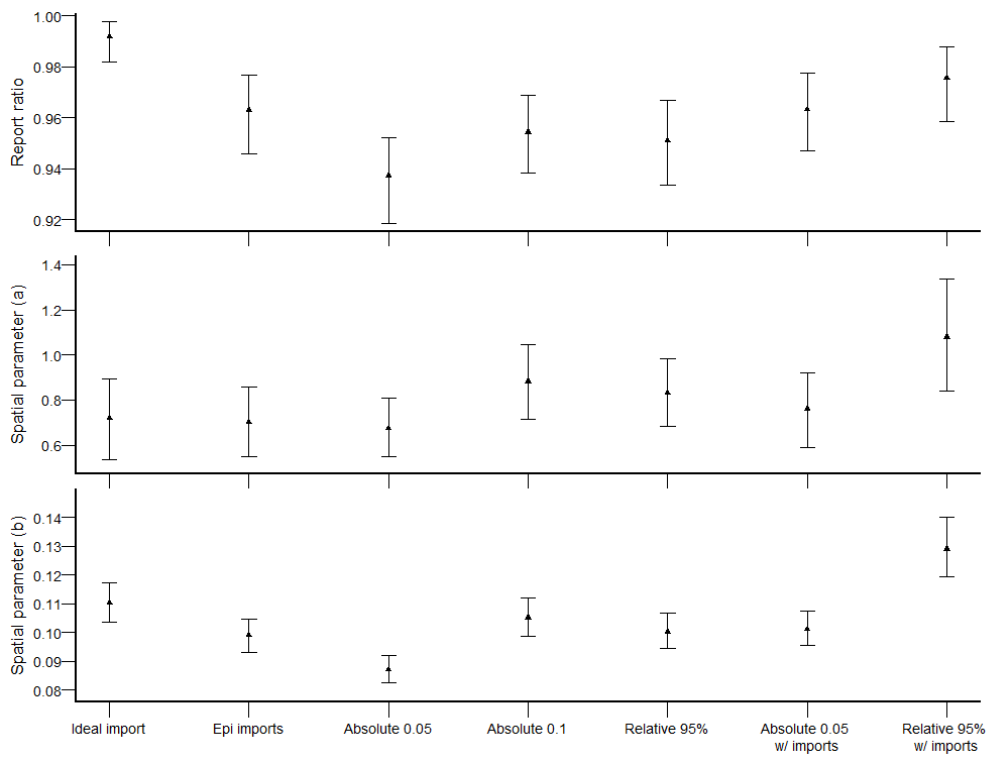

Figure S9: Estimation of A) the report ratio ρ, B) the spatial parameter a and C) the spatial parameter b in each scenario. The dots represent the median estimate, and the arrows correspond to the 95% credibility interval. The estimates were obtained after burnin and thinning.

6. Distance between transmission

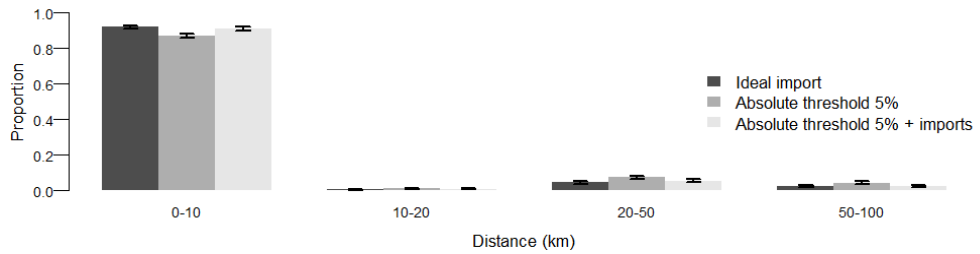

Figure S10: Distribution of the distance between connected cases in scenario 1, scenario 3 without prior and scenario 3 with prior information. Arrows correspond to the 95% credibility interval.

7. Impact of different components of likelihood

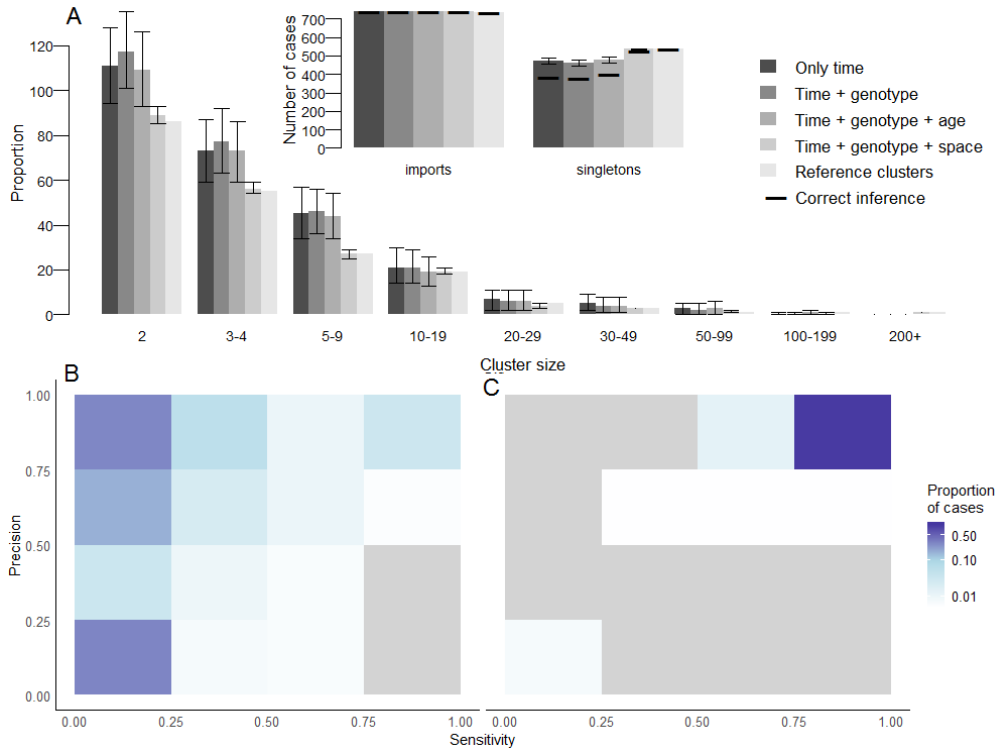

Figure S11: Description of transmission clusters inferred excluding certain components of the likelihood. In this example, the import status of import was not inferred, and taken from the ancestor in each cluster (Scenario 1). Panel A: Cluster size distribution using 1) only the time component of likelihood; 2) time and genotype; 3) time genotype and age; 4) time, genotype and space; compared to the reference clusters (lighgrey). Arrows represent the 95% credibility intervals of each estimate. Only clusters containing at least 2 cases are represented. Insert: Number of imports and number of isolated cases (singletons). For each scenario, the horizontal dark line represents the number of imports that are also imports in the reference clusters, same for singletons. Panel B: Heatmap representing the precision and sensitivity of the clusters for each case when only time and genotype are used to infer the transmission clusters, the cases were classified in a category depending on the proportion of their reference cluster that were inferred in the same cluster (x-axis) and the proportion of mismatches in the inferred cluster. Panel C: Same for time, genotype and space.

8. Number of secondary transmissions, overall and per state

a. Overall

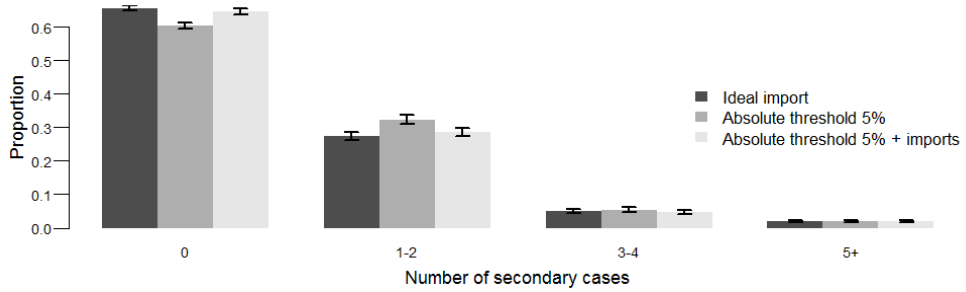

Figure S12: Distribution of the number of secondary cases caused by each case in scenario 1, scenario 3 without prior and scenario 3 with prior information. Arrows correspond to the 95% credibility interval. Cases were classified in groups of transmitter (no further transmission; 1-2 subsequent cases; 3-4 subsequent cases and more than 5 subsequent cases).

b. Maps per state

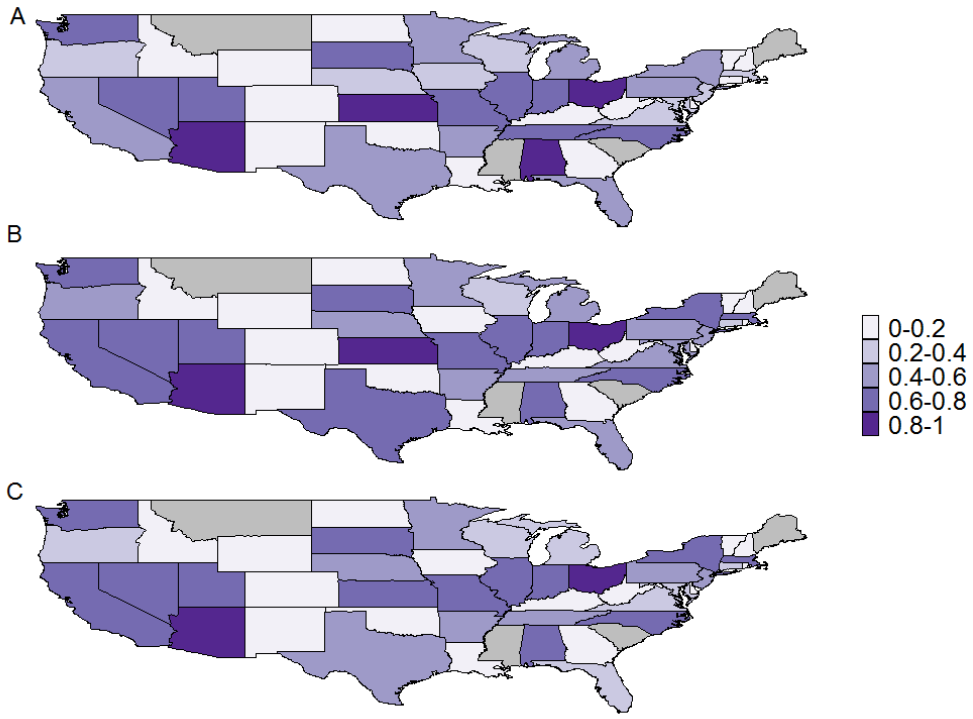

Figure S13: Average number of secondary cases caused by each case stratified by state in A) scenario 1, B) scenario 3 without prior and C) scenario 3 with prior information.

9. Impact of the proportion of genotype reported. Inference on simulated data

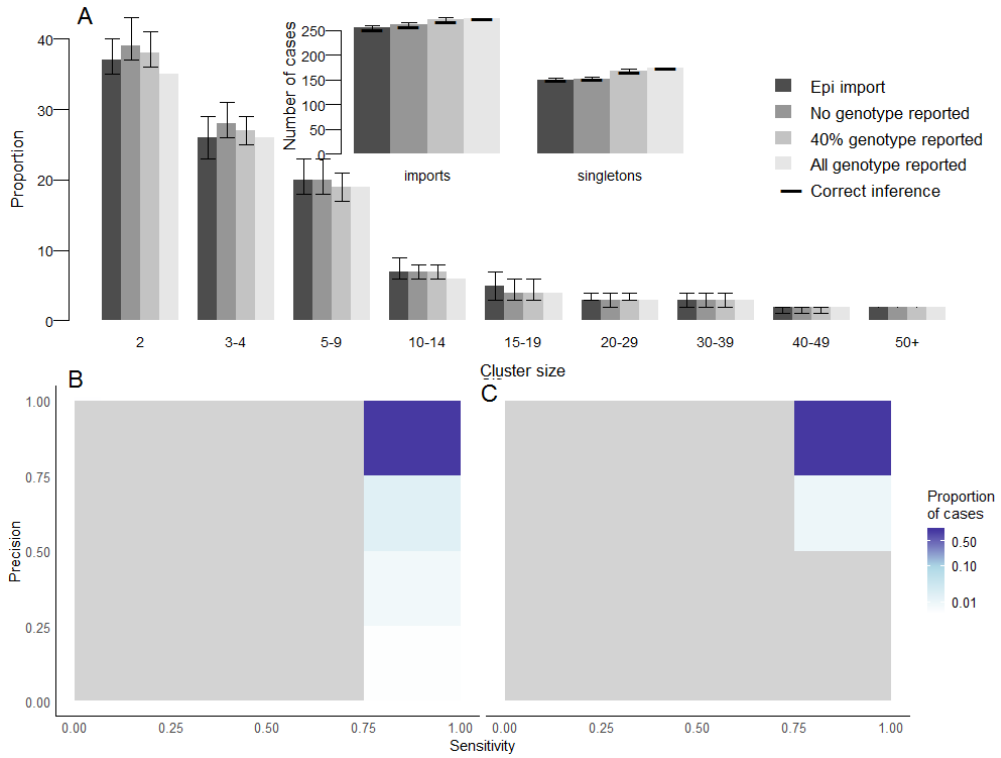

Figure S14: Description of transmission clusters inferred using simulated data (toy_outbreak_long in the *o2geosocial* package)*, depending on the proportion of genotyped cases in the data. Panel A: Cluster size distribution when 1) none of the cases was genotyped, 2) 40% of the cases were genotyped (similar to the US dataset) 3) All the cases were genotyped; compared to the reference clusters (lighgrey). Arrows represent the 95% credibility intervals of each estimate. Only clusters containing at least 2 cases are represented. Insert: Number of imports and number of isolated cases (singletons). For each scenario, the horizontal dark line represents the number of imports that are also imports in the reference clusters, same for singletons. Panel B: Heatmap representing the precision and sensitivity of the clusters for each case when no genotype was reported are used to infer the transmission clusters, the cases were classified in a category depending on the proportion of their reference cluster that were inferred in the same cluster (x-axis) and the proportion of mismatches in the inferred cluster. Panel C: Same when all cases were genotyped.*