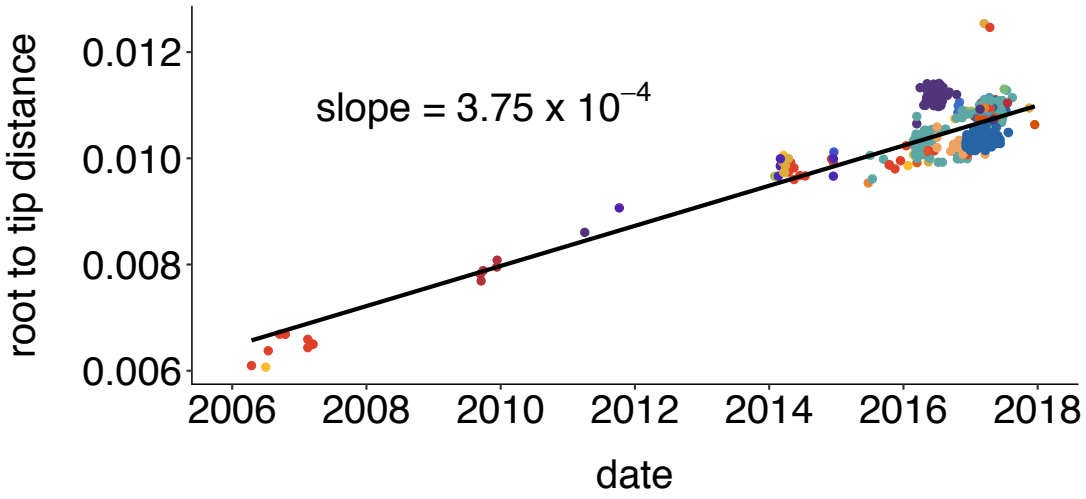
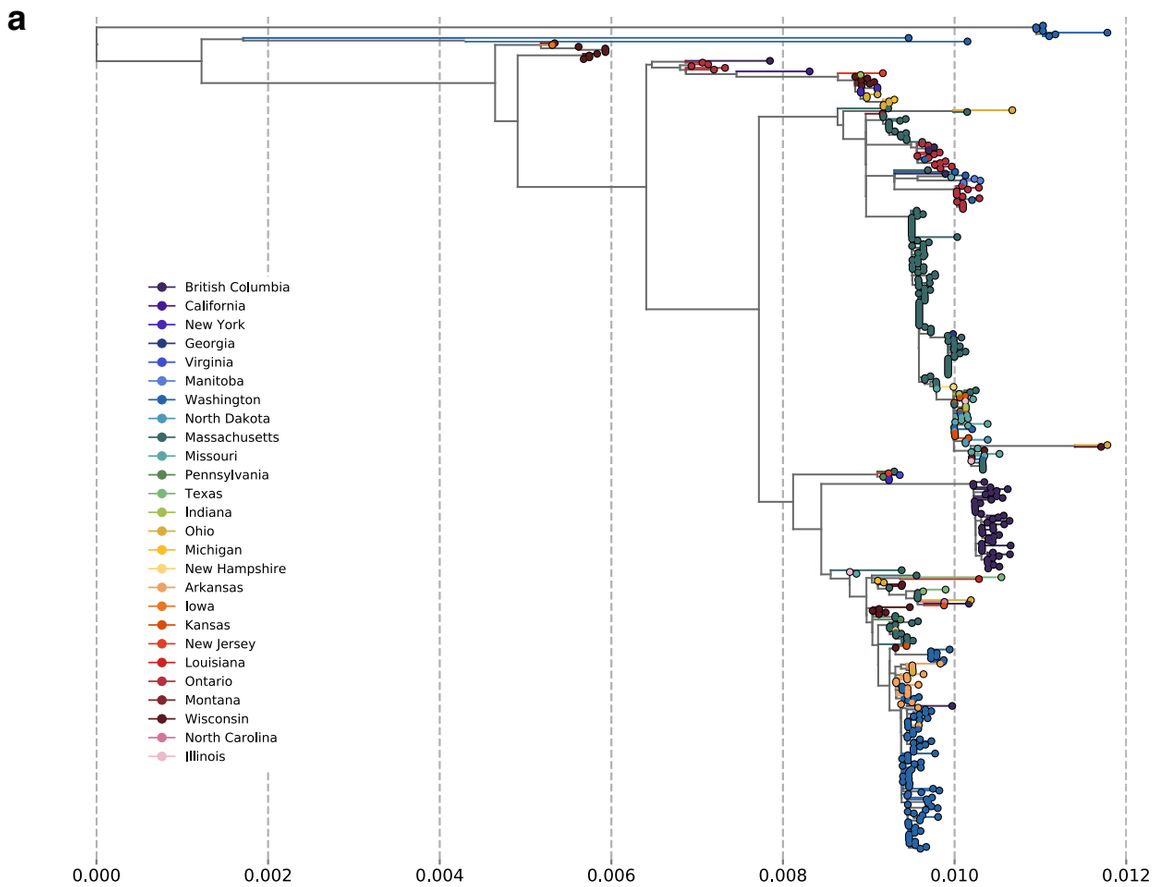
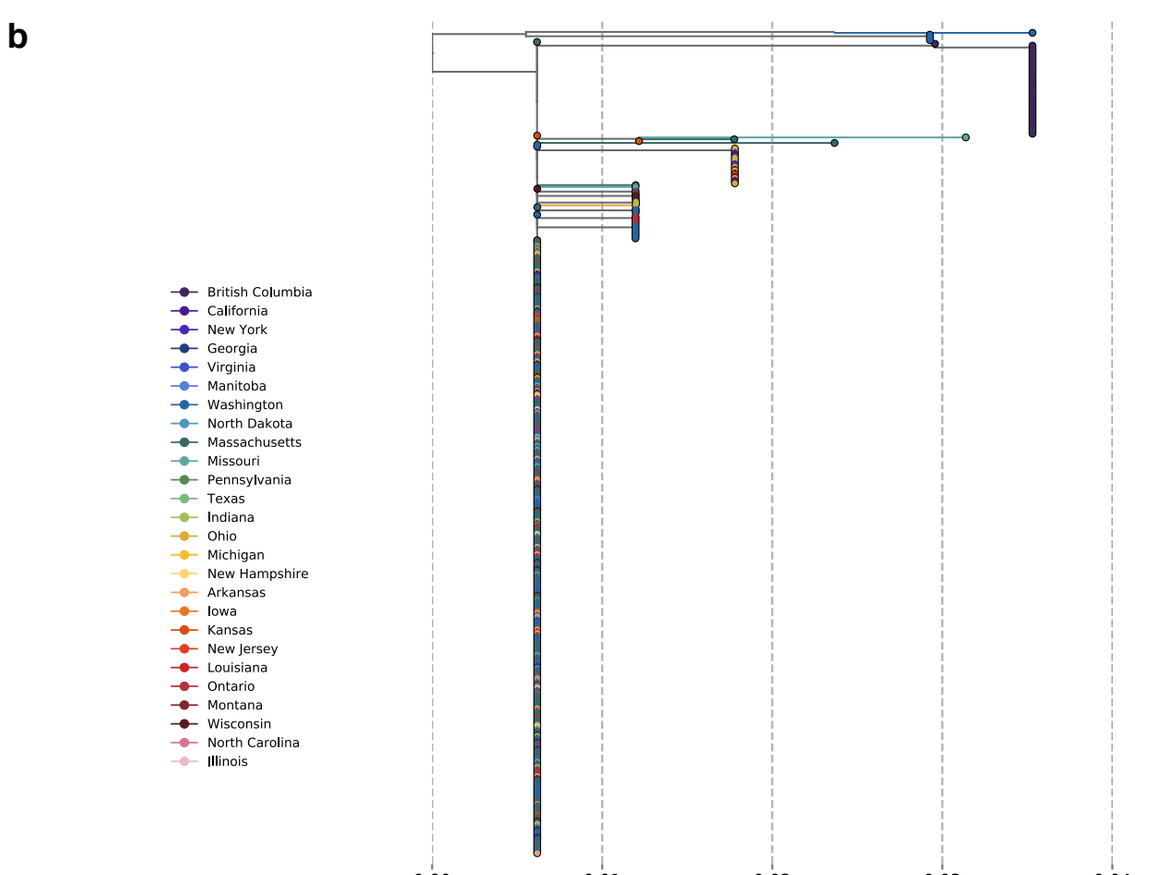


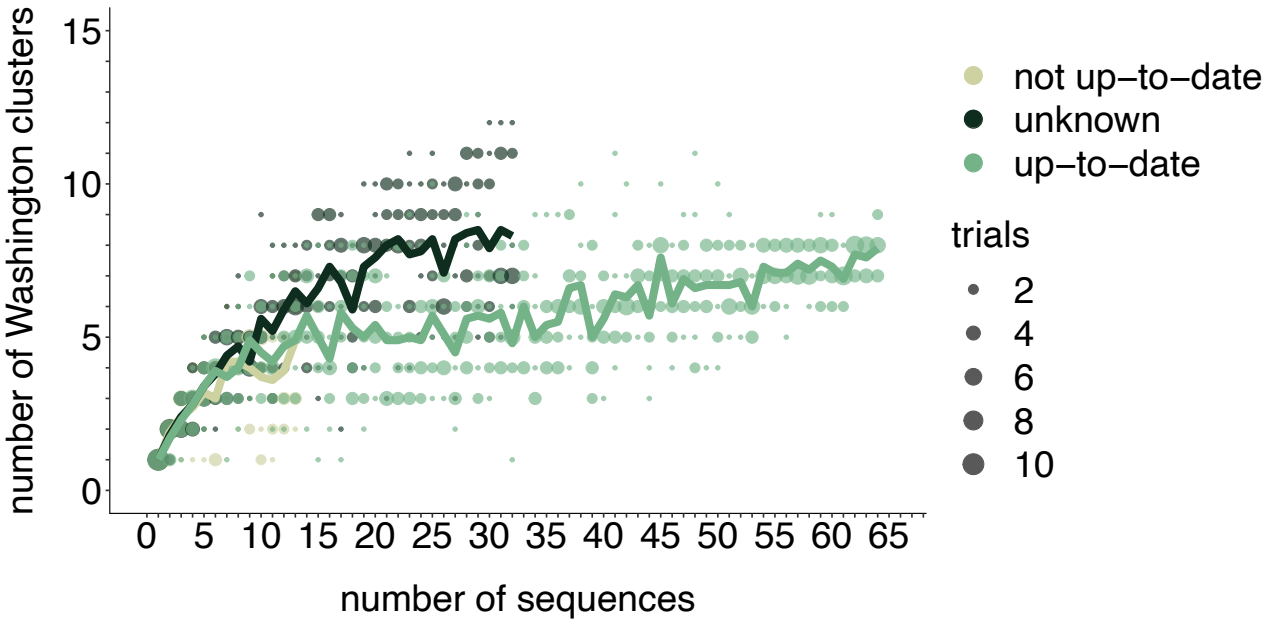
Supplemental Figure 1: Mumps genomes accumulate mutations linearly over time
We inferred a maximum likelihood phylogeny using IQTREE for all available complete mumps genomes of genotype G, sampled from North America between 2006 and 2018. We inferred the root-to-tip distance with TempEst and plot the root to tip divergence vs. sample collection date. Color represents geographic location (either Canadian province or US state), with colors corresponding to those in Figure 1. We infer that mumps genomes accumulate mutations at a rate of 3.75 x 10^{-4} substitutions per site per year.



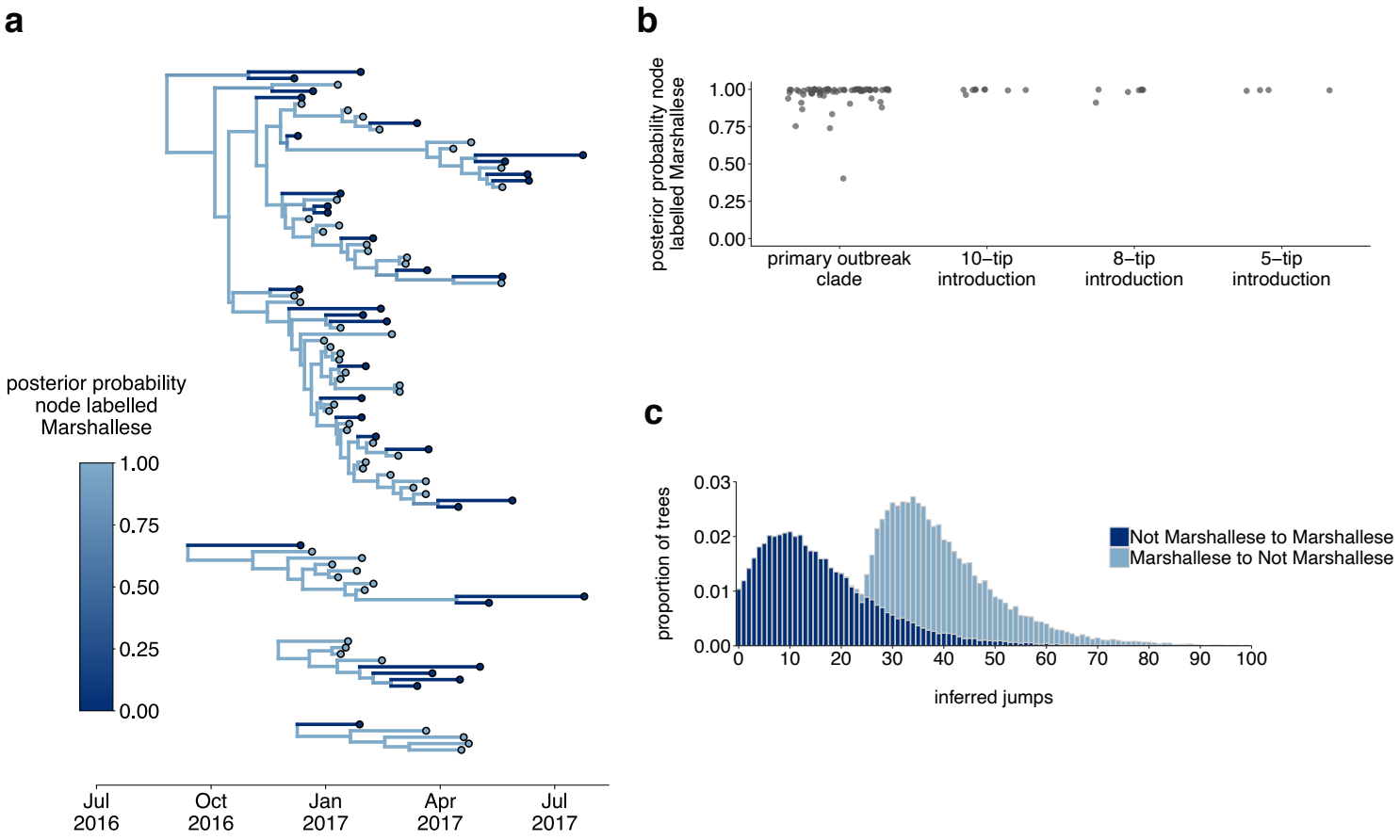


Supplemental Figure 2: SH gene sequences are inadequate for fine-scale resolution of mumps tranmission

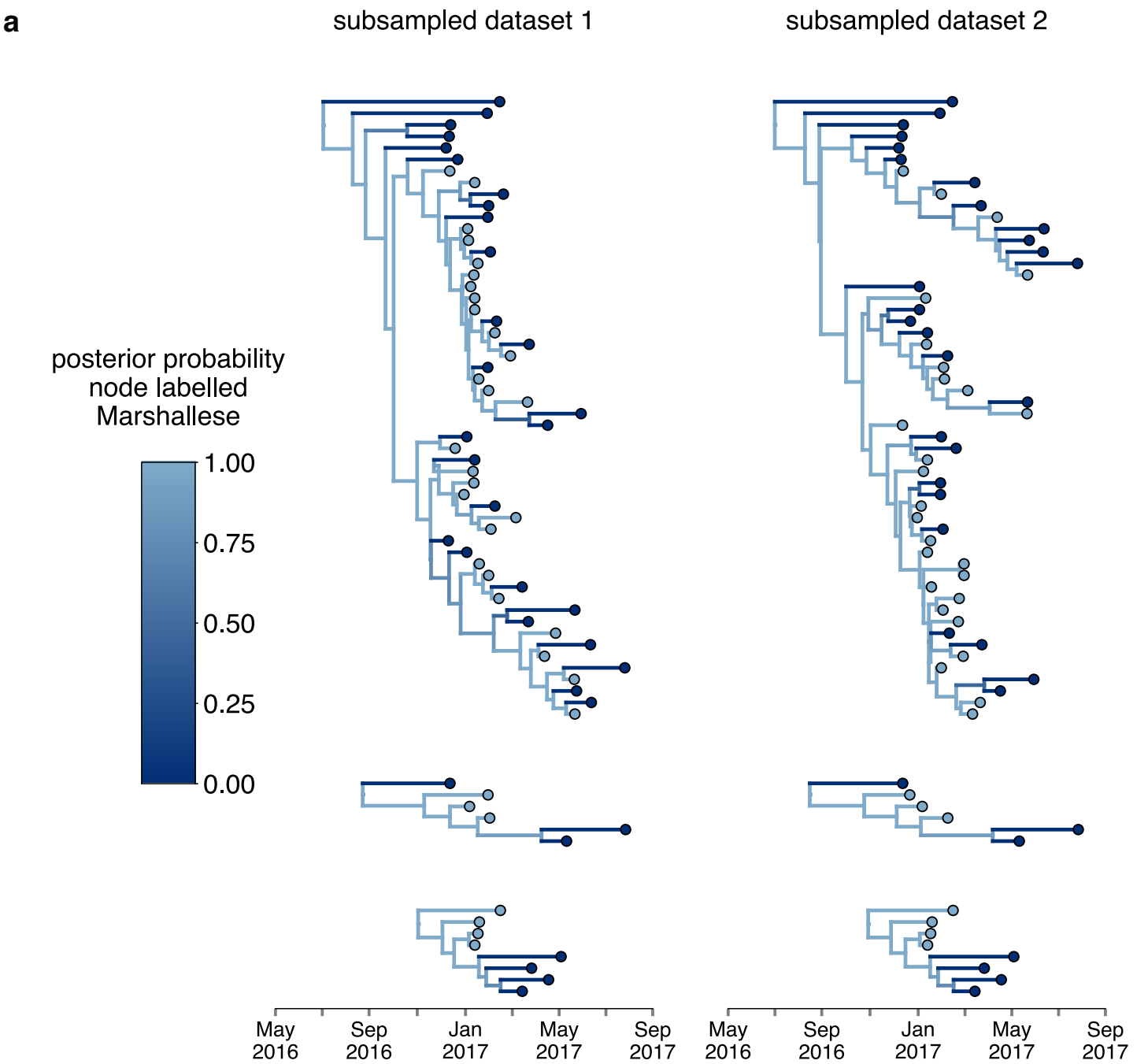
a. We inferred a maximum likelihood phylogeny using IQTREE for all available complete mumps genomes of genotype G, sampled from North America between 2006 and 2018. Color represents geographic location (either Canadian province or US state), and the x-axis displays divergence in substitutions per site per year. **b.** To compare whether similar results would have been obtained if we had only sequenced the SH gene, we truncated our sequences to include only the coding region for SH and again inferred a maximum likelihood phylogeny using the same procedure as in a. The vast majority of North American mumps sequences are identical and form a single polytomy, suggesting that SH sequencing alone provides limited resolution for inferring geographic spread.



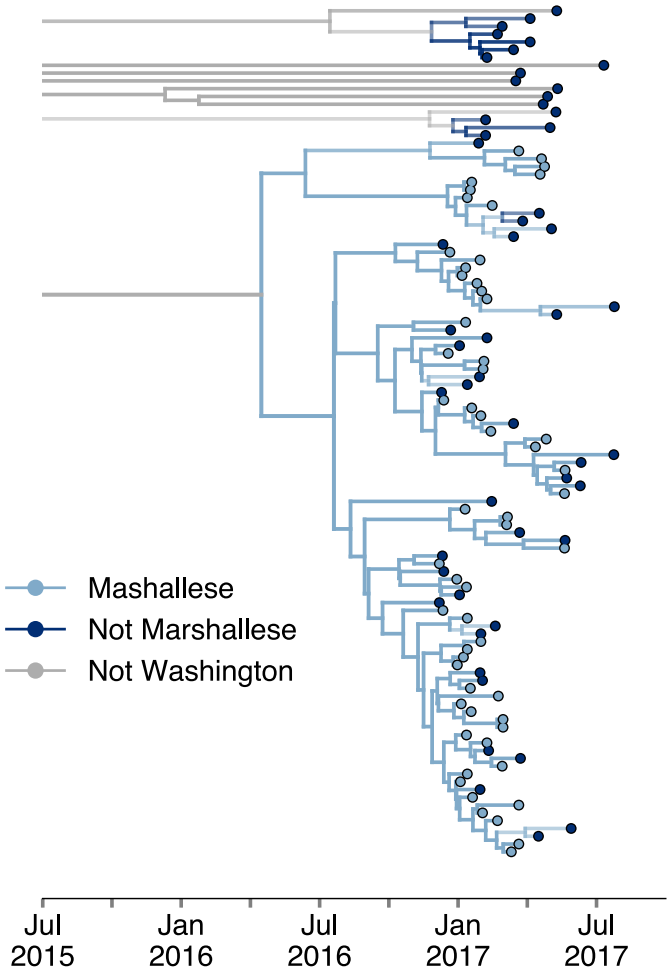
Supplemental Figure 3: Rarefaction results by vaccination status
We repeated the rarefaction analysis shown in Figure 4b for vaccination status. We separated all Washington tips and classified them by vaccination status into up-to-date, not up-to-date, or unknown vaccination status. We then performed a rarefaction analysis and plot the number of inferred Washington clusters (y-axis) as a function of the number of sequences included in the analysis (x-axis). Dark green represents unknown vaccination status, light green represents not up-to-date, and green represents up-to-date. The majority of sequences in our dataset were derived from individuals who were up-to-date for mumps vaccine. Each dot represents the number of trials in which that number of clusters was inferred, and the solid line represents the mean across trials.



Supplemental Figure 4: Inferences are similar under a higher migration rate prior
The results are shown for the exact same analyses displayed in Figure 5, except inferred under a model with a higher migration rate prior (10 instead of 1). **a.** Using the 4 Washington clusters that had a mixture of Marshallese and non-Marshallese cases, we inferred phylogenies using a structured coalescent model. Each group of sequences shared a clock model, migration model, and substitution model, but each topology was inferred separately, allowing us to incorporate information from all 4 clusters into the migration estimation. For each cluster, the maximum clade credibility tree is shown, where the color of each internal node represents the posterior probability that the node is Marshallese. **b.** For each internal node shown in panel **a**, we plot the posterior probability of that node being Marshallese. Across all 4 clusters, almost every internal node is inferred as Marshallese with high probability. **c.** The posterior distribution of the number of "jumps" or transmission events from Marshallese to not Marshallese (light blue) and not Marshallese to Marshallese (dark blue) inferred for the primary outbreak clade.

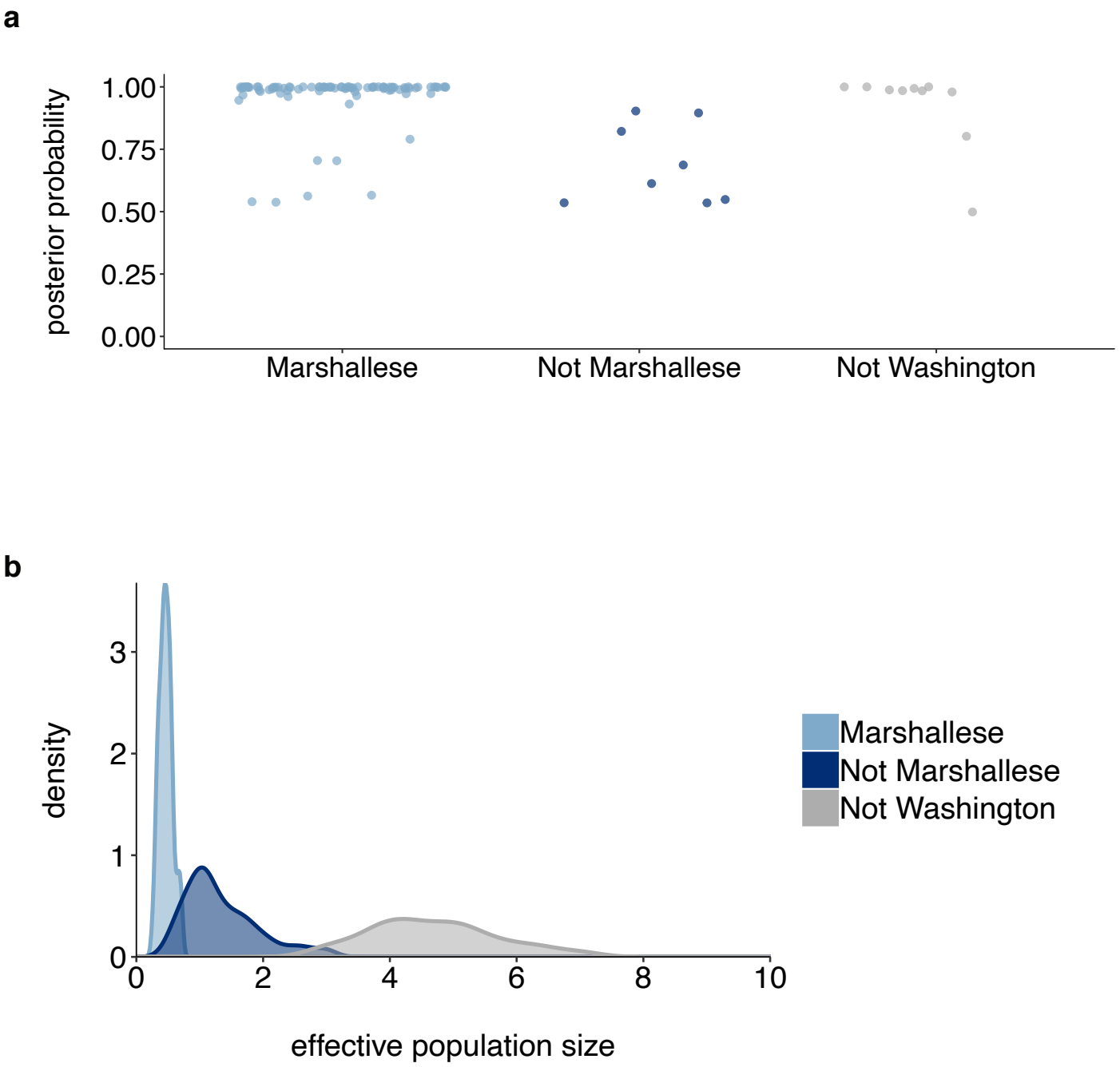


Supplemental Figure 5: Structured coalesent analyses are robust to sampling differences
To ensure that our results were robust to differences in sampling of Marshallese and non-Marshallese tips within the clusters used for this analysis, we subsampled our dataset 3 independent times, and ran 3 independent chains per unique subsampling. In each subsampled dataset, the number of Marshallese tips was randomly subsampled to be equal to the number of non-Marshallese tips in each of the 4 clusters. We then ran each of these subsampled datasets with the exact same model as run with the full dataset. In subsampled datasets 1 and 2, 2 out of 3 chains converged, and results were combined and displayed here. In the 3rd subsampled dataset, none of the 3 chains converged, so those results are not shown. **a.** For each subsampled dataseet, we plot the inferred maximum clade credibility tree from the combined tree outputs from the 2 converged chains. The color of each tip represents whether that sample was derived from a Marshallese or non-Marshallese case, and the color of the internal node represents the posterior probability of that internal node being Marshallese. **b.** For each tree shown in **a**, the posterior probability that each internal node is labelled as Marshallese is shown. The number of the subsampled dataset is shown on the x-axis and the posterior probability is shown on the y-axis.



Supplemental Figure 6: Including all Washington sequences recovers majority of transmission in Marshallese

To ensure that excluding non-Marshallese clusters did not skew our findings, we inferred a single tree using all Washington sequences. We performed a structured coalescent analysis specifying 3 groups: Marshallese, not Marshallese, and not Washington. Each internal node is colored by its most probable group, with its opacity specifying the posterior probability of being in that group (fully opaque being probability = 1, fully transparent being probability = 0).



Supplemental Figure 7: Posterior probabilities of internal node states
a. For the tree shown in Supplemental Figure 6, each internal node is plotted. For each intenral node, its color and placement on the x-axis represents its inferred most probable group (Marshallese, Not Marshallese, or Not Washington). The posterior probability of being labelled its most probable group is shown on the y-axis. We recover moderate support for a small number of non-Marshallese internal nodes, while the vast majority of internal nodes remain inferred as Marshallese. **b.** The 95% highest posterior density intervals of the inferred effective population sizes for Marshallese, non-Marshallese, and not Washington demes.

Supplemental Table 1: Mumps cases by age group

Age group	Mumps case count	2016-2017 average population in Washington[1]	Rate of cases per 100,000 individuals	Percentage of total cases
0 - 4	34	450,847	7.5	3.8%
5 - 9	104	462,951	22.5	11.7%
10 - 14	198	451,485	43.9	22.3%
15 - 19	214	455,612	47.0	24.1%
20 - 39	256	1,980,004	12.9	28.8%
40 - 64	80	2,348,529.5	3.4	10.0%
65+	2	1,097,571	0.2	0.22%
Unknown	1	NA	NA	0.11%

Supplemental Table 2: Outbreak characteristic and dataset composition

	counts in outbreak (%)	Count in dataset (%)
Up-to-date vaccination	574 (65%)	64 (58%)
Not up-to-date vaccination	86 (9.7%)	13 (12%)
Unknown vaccination status	229 (26%)	33 (30%)
Marshallese	465 (52%)	57 (52%)
Not Marshallese	424 (48%)	53 (48%)
Total	**889**	**110**

Supplemental Table 3: Logistic regression results of probability that phylogeny nodes has descendants

Predictor variable	Estimated coefficient (standard error)	Odds ratio (95% CI)	p-value
Not up-to-date	-0.77 (0.69)	0.46 (0.11, 1.70)	0.26
Vaccination status unknown	0.79 (0.80)	2.21 (0.59, 9.20)	0.32
Age	-1.2 (1.44)	0.30 (0.016, 4.70)	0.41
Community status	1.17 (0.43)	3.21 (1.39, 7.69)	0.0073

We evaluated the impact of vaccination status, age, and community membership on the probability that the phylogenetic node had descendants in the tree. Coefficients represent the increase in log odds of having descendants for the given predictor variable. Coefficients were exponentiated to produce odds ratios. We evaluated the impacts of having an unknown vaccination status, having a vaccination status that is not up-to-date, and being Marshallese as binary predictor variables. Age was evaluated as a continuous variable normalized such that values fall between 0 and 1.

Supplemental References

1. Estimates of April 1 population by age, sex, race and Hispanic origin. https://www.ofm.wa.gov/washington-data-research/population-demographics/population-estimates/estimates-april-1-population-age-sex-race-and-hispanic-origin.