

1 **Repeated introductions and intensive community transmission**
2 **fueled a mumps virus outbreak in Washington State**

3

4 Louise H. Moncla^{*1#}, Allison Black^{1,2#}, Chas DeBolt³, Misty Lang³, Nicholas R. Graff³, Ailyn C.
5 Pérez-Osorio³, Nicola F. Müller¹, Dirk Haselow⁴, Scott Lindquist³, Trevor Bedford^{*1,2}

6

7 1. Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle,
8 Washington, United States

9 2. Department of Epidemiology, University of Washington, Seattle, Washington, United States

10 3. Office of Communicable Disease Epidemiology, Washington State Department of Health,
11 Shoreline, Washington, United States

12 4. Arkansas Department of Health, Little Rock, Arkansas, United States

13

14 [#] these authors contributed equally to this work

15 * corresponding author; lhmoncla@gmail.com and trevor@bedford.io

16

17 Summary: Using genomic epidemiology to characterize a mumps virus outbreak in Washington
18 State

19

20 **Abstract**

21 In 2016/2017, Washington State experienced a mumps outbreak despite high childhood
22 vaccination rates, with cases more frequently detected among school-aged children and
23 members of the Marshallese community. We sequenced 166 mumps virus genomes collected in
24 Washington and other US states, and traced mumps introductions and transmission within
25 Washington. We uncover that mumps was introduced into Washington approximately 13 times,

26 primarily from Arkansas, sparking multiple co-circulating transmission chains. Although age and
27 vaccination status may have impacted transmission, our dataset could not quantify their precise
28 effects. Instead, the outbreak in Washington was overwhelmingly sustained by transmission
29 within the Marshallese community. Our findings underscore the utility of genomic data to clarify
30 epidemiologic factors driving transmission, and pinpoint contact networks as critical for mumps
31 transmission. These results imply that contact structures and historic disparities may leave
32 populations at increased risk for respiratory virus disease even when a vaccine is effective and
33 widely used.

34

35 **Introduction**

36 In 2016 and 2017, mumps virus swept the United States in the country's largest outbreak since
37 the pre-vaccine era (CDCMMWR, 2019). Washington State was heavily affected, reporting 889
38 confirmed and probable cases. Longitudinal studies (Davidkin et al., 2008), epidemiologic
39 outbreak investigations (Cardemil et al., 2017), and epidemic models (Lewnard and Grad, 2018)
40 suggest that mumps vaccine-induced immunity wanes over 13-30 years, consistent with the
41 preponderance of young adult cases in recent outbreaks. Like with other recent mumps
42 outbreaks, most Washington cases in 2016/17 were vaccinated. Unusually though, while most
43 US outbreaks in 2016/2017 were associated with university settings (Albertson et al., 2016;
44 Bonwitt et al., 2017; Donahue et al., 2017; Golwalkar et al., 2018; Shah et al., 2018; Wohl et al.,
45 2020), incidence in Washington was highest among children aged 10-18 years, younger than
46 expected given waning immunity. The outbreak was also peculiar in that approximately 52% of
47 the total cases were Marshallese, an ethnic community that comprises ~0.3% of Washington's
48 population. These same phenomena were also observed in Arkansas. Of the 2,954 confirmed
49 and probable Arkansas cases, 57% were Marshallese, and 57% of cases were children aged 5-
50 17 (Fields et al., 2019). Amongst infected school-aged children in Arkansas and Washington,
51 >90% had previously received 2 doses of MMR vaccine (Fields et al., 2019). The high

52 proportion of vaccinated cases, younger-than-expected age at infection, disproportionate impact
53 on the Marshallese community, and epidemiologic link to Arkansas suggest that factors beyond
54 waning immunity are necessary to explain mumps transmission during this outbreak in
55 Washington.

56

57 The US and the Marshall Islands are closely linked through a history that continues to impact
58 the health of US-residing Marshallese to this day. Between 1947 and 1986, the United States
59 occupied the Republic of Marshall Islands and detonated the equivalent of >7000 Hiroshima
60 size nuclear bombs as part of its nuclear testing program (Barker, 2012). The effects were
61 devastating, precipitating widespread environmental destruction, nuclear contamination, and
62 dire health consequences (Hallgren et al., 2015; Niedenthal, 1997; Palafox et al., 2007; Simon,
63 1997; Takahashi et al., 1997). Marshallese individuals inhabiting the targeted atolls were forcibly
64 moved to other islands, and many were exposed to nuclear fallout (Abella et al., 2019) that
65 persists on the Islands today (Bordner et al., 2016). Significant concern remains within the
66 community regarding long-term health impacts of nuclear exposure and its potential impacts on
67 immune function. Marshallese individuals living on and off the Islands experience significant
68 health disparities including a higher burden from infectious diseases and chronic health
69 conditions (Adams et al., 1986; Wong et al., 1979; Yamada et al., 2004). Compounding these
70 disparities, from 1996 to 2020 (Hirono, 2019), Marshallese individuals were specifically
71 excluded from Medicaid eligibility despite legal residency in the US permitted under the
72 Compact of Free Association (COFA) Treaty. As a result, many US-residing Marshallese are
73 uninsured, with poor access to healthcare (McElfish et al., 2015). Marshallese households are
74 more likely to be multigenerational and tend to be larger on average (Harris and Jones, 2005;
75 US Census Bureau, n.d.), potentially increasing the number and intensity of interactions among
76 individuals. These factors combined mean that Marshallese individuals may be at increased risk
77 of respiratory virus infection.

78

79 Clarifying the determinants of infectious disease transmission is important for prioritizing
80 prevention and mitigation resources. However, sampling bias presents a persistent challenge
81 for elucidating source-sink dynamics from genomic data (De Maio et al., 2015; Dudas et al.,
82 2018; Frost et al., 2015; Kühnert et al., 2011; Lemey et al., 2020; Stack et al., 2010), which may
83 undermine the utility of genomic epidemiological studies in some situations. Here, we formulate
84 a set of genomic epidemiological approaches that are robust to sampling frame and apply them
85 to investigate patterns of mumps transmission in Washington. We sequenced 110 mumps viral
86 genomes obtained from specimens collected from laboratory-confirmed mumps cases in
87 Washington State and another 56 from other US states collected between 2006 and 2018. We
88 employ a novel application of phylogeographic methods to detailed epidemiologic data on age,
89 vaccination status, and community membership, and develop a new statistic for quantifying
90 transmission in the tree. By combining these phylodynamic approaches with community health
91 advocate interviews that contextualize our results, we provide a framework for investigating viral
92 transmission dynamics that is sensitive to community health priorities and readily applicable to
93 other viral pathogens.

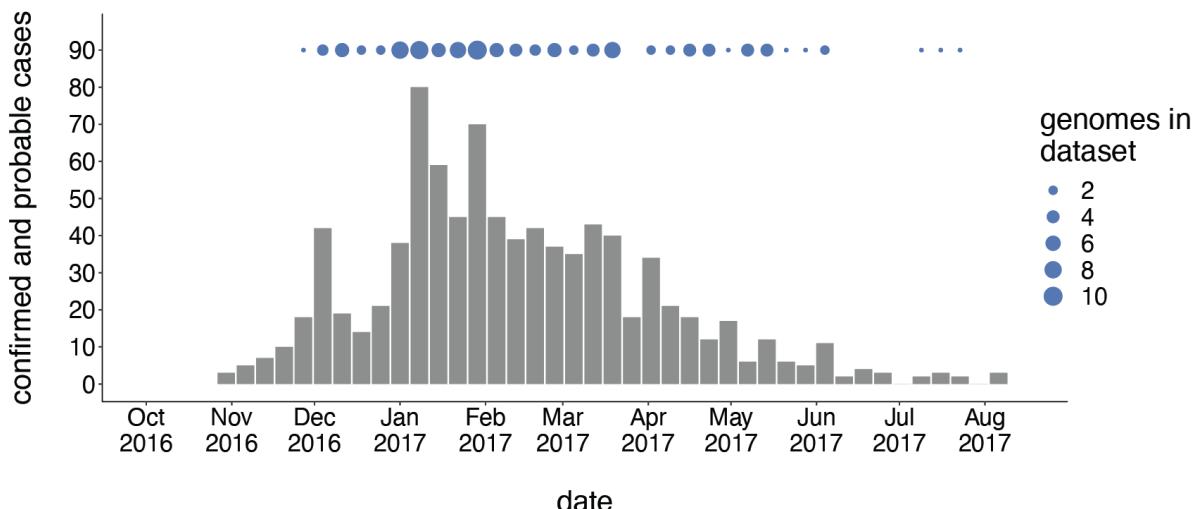
94

95 **Results**

96 **Outbreak characteristics and dataset composition**

97 We generated genome sequences for 110 PCR-positive mumps samples collected throughout
98 Washington State during 2016/2017, and 56 samples collected in Wisconsin, Ohio, Missouri,
99 Alabama, and North Carolina between 2006 and 2018 (**Supplementary File 1a**). The
100 Washington State outbreak began in October 2016, and peaked in winter of 2017, culminating
101 in 889 confirmed and probable cases across Washington (**Figure 1**). Individuals aged <1 to 64

102 years were affected, but incidence was highest among children aged 10-14 (44.9 cases per
103 100,000) and 15-19 (47.0 per 100,000) (**Supplementary File 1b**). Among outbreak cases



104

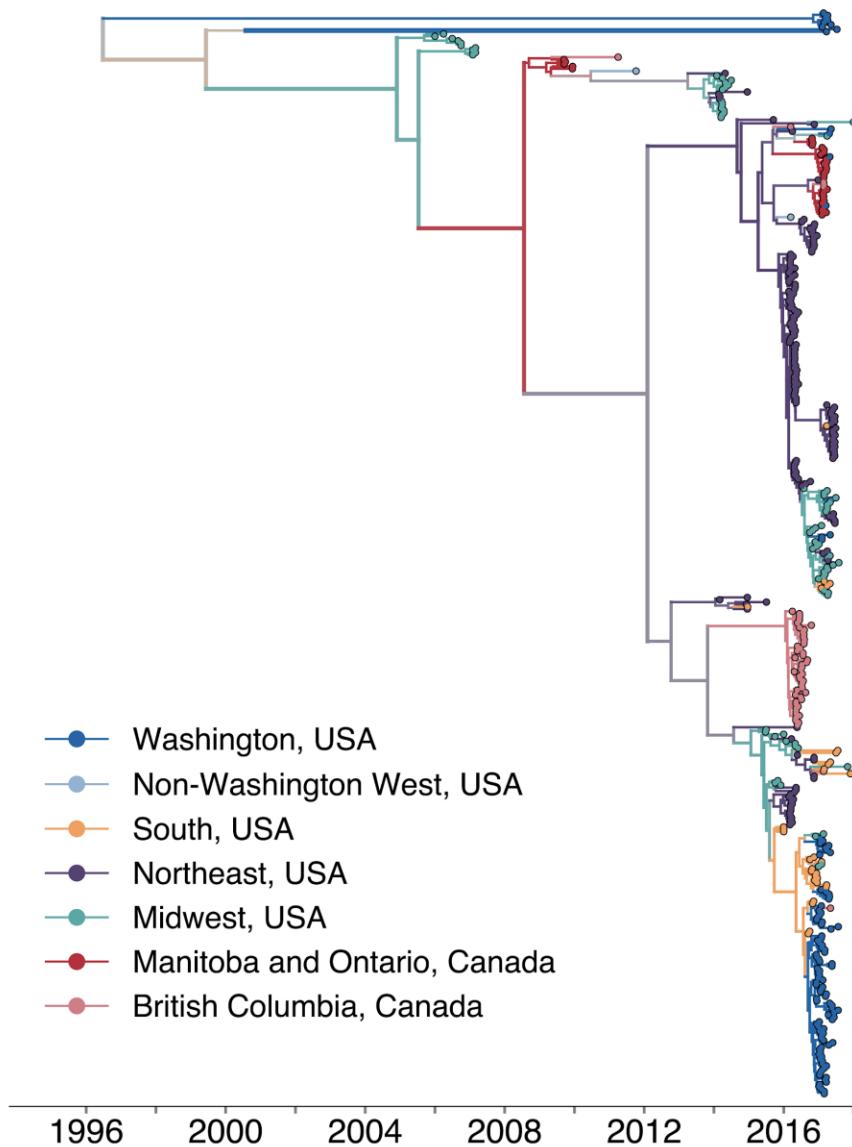
105 **Figure 1: Genomic sampling covers the duration of the outbreak.**

106 The first mumps case in Washington was reported on October 30, 2016, and case counts peaked in
107 winter of 2017. Here we show recorded numbers of confirmed and probable cases by epidemiologic (epi)
108 week. Blue dots above the epidemiologic curve represent the number of Washington genome sequences
109 sampled from viruses collected during that epi week.
110

111 5-19 years of age, 91% of individuals were considered up-to-date on mumps vaccine. Adults in
112 the age group most likely to be parents of school aged children (20-39 years old) were infected
113 at a rate of only 12.9 cases per 100,000, but comprised a significant proportion (29%) of total
114 cases (**Supplementary File 1b**). While Marshallese individuals comprise only ~0.3% of
115 Washington's total population, they accounted for 52% of reported mumps cases
116 (**Supplementary File 1c**). Among Marshallese cases aged 5-19, 93% were up-to-date on
117 vaccination, suggesting that this over-representation is not attributable to poor vaccine
118 coverage.
119

120 **Outbreaks across North America are related**

121 We combined our sequence data with publicly available full genome sequences sampled from
122 North America between 2006 and 2018, and built a time-resolved phylogeny, inferring migration
123 history among 27 US states and Canadian provinces (**Figure 2**, **Figure 2-figure supplement 1**)



124

125 **Figure 2: North American mumps outbreaks are related**
126 We combined all publicly available North American mumps genomes and built a time-resolved phylogeny.
127 We inferred geographic transmission history between each US state and Canadian province using a
128 discrete trait model, but have grouped these locations into regions for plotting purposes. A tree colored by
129 the full geographic transmission history across all 27 locations is shown in Figure 2 - figure supplement 2.
130 Here, we display the maximum clade credibility tree, where color represents geographic location. We
131 grouped US states by geography as follows: non-Washington West include California and Montana;
132 Midwest USA includes North Dakota, Kansas, Missouri, Iowa, Wisconsin, Indiana, Michigan, Ohio, and
133 Illinois; South USA includes North Carolina, Alabama, Virginia, Georgia, Texas, Arkansas, and Louisiana;

134 Northeast USA includes New York, Massachusetts, Pennsylvania, New Hampshire, and New Jersey.
135 Canadian provinces are also grouped by geographic area. The x-axis represents the collection date (for
136 tips), or the inferred time to the most recent common ancestor (for internal nodes). The internal node
137 coloring represents the sum of the posterior probabilities for each inferred geographic division within the
138 most probable region. For example, since we group Manitoba and Ontario into the same Canadian
139 region, if a node was inferred with highest probability to circulate in Manitoba, then the node would be
140 colored red to represent that Canadian region. The opacity of the color then corresponds to the sum of
141 the probabilities that the node circulated in Manitoba or that the node circulated in Ontario. The posterior
142 probability is expressed by the color gradient, where increasingly grey tone represents decreasing
143 certainty of the inferred geographic state. The ancestral state at the root was poorly resolved, and is
144 therefore colored mostly grey.

145

146 **Figure 2-figure supplement 2).** Sequences from samples collected between 2006 and 2014
147 clustered with other North American mumps viruses sampled from the same times. Nine
148 Washington sequences were highly divergent from other North American genotype G viruses,
149 with a time to the most recent common ancestor (TMRCA) of ~22 years (**Figure 2**, blue tips with
150 long branches clustered towards the top of the tree). To place these genomes in context, we
151 built a divergence tree using all publicly available global full genome mumps sequences (**Figure**
152 **2-figure supplement 3)**. Seven of these divergent Washington sequences cluster with viruses
153 sampled from New Zealand (**Figure 2-figure supplement 3**), suggesting they could be travel-
154 related. The other 2 sequences cluster with other divergent genotype G viruses sampled from
155 geographically disparate locations (**Figure 2-figure supplement 3**). The remaining Washington
156 sequences nest within the diversity of other North American viruses, and descend from the
157 same mumps lineage that has circulated in North America since 2006 (**Figure 2**). We observe
158 substantial geographic mixing along the tree. While viruses from the northeast (teal tips and
159 branches) seeded outbreaks in the Northeast and Midwest, we also infer transmission from the
160 Northeast to Southern US states and British Columbia. Despite the close geographic proximity
161 between British Columbia and Washington, most British Columbia sequences form a distinct
162 cluster on a long branch (**Figure 2**), suggesting seeding from an unsampled location. Although
163 viruses from Washington are scattered throughout the phylogeny, most cluster within a clade of
164 viruses sampled in Arkansas (**Figure 2**).

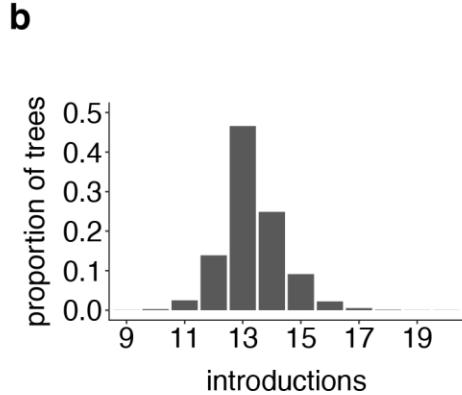
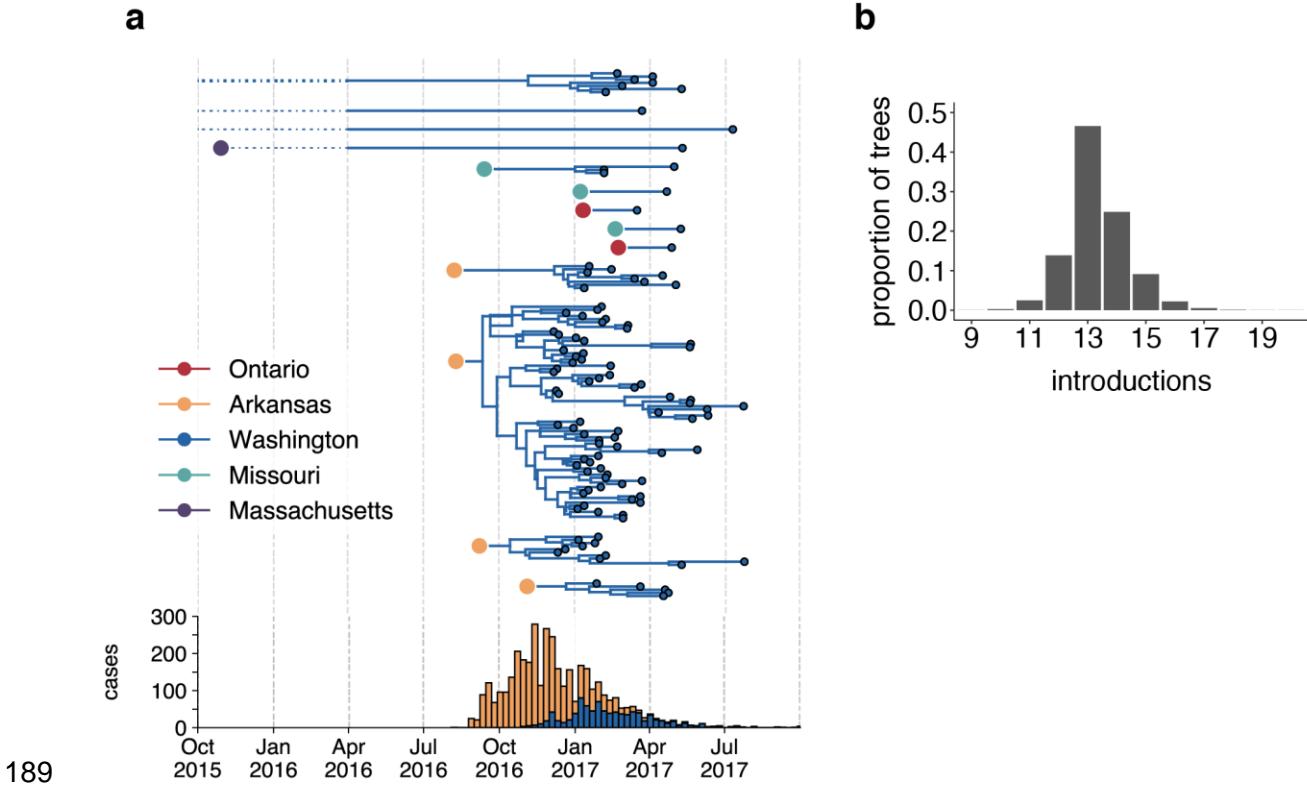
165

166 Mumps is classified into 12 genotypes (labelled A-N, excluding E and M) based on its SH gene
167 sequence. There is some evidence that mumps genotypes are geographically associated
168 (Nomenclature and Des virus, n.d.), and the vast majority of mumps viruses circulating in North
169 America since 2006 have been genotype G viruses. Although most samples in our dataset are
170 also genotype G, we did sequence three viruses that group in different genogroups. One
171 sample from Wisconsin in 2006 grouped with genotype A viruses, another sample from
172 Wisconsin in 2015 grouped with genotype H viruses, and one sample from Washington in 2017
173 grouped with genotype K viruses (**Figure 2-figure supplement 3**). The Washington genotype K
174 virus (Washington.USA/9.17/FH94/K) is closely related to a genotype K mumps virus collected
175 during a mumps outbreak in Massachusetts from an individual who reported international travel
176 (Wohl et al., 2020)(**Figure 2-figure supplement 3**). These divergent, non-genotype G genomes
177 were excluded from further phylogenetic analysis.

178

179 **Mumps was introduced into Washington multiple independent times**

180 Estimating the number and timing of viral introductions is important for estimating epidemiologic
181 parameters and evaluating public health surveillance systems, but detecting these dynamics
182 may be challenging with case count data alone (Faria et al., 2017; Grubaugh et al., 2017). The
183 Washington Department of Health had identified a single potential index case infected in
184 October of 2016. To determine whether the genomic data similarly supported a single
185 introduction of mumps to Washington state, we separated each introduction inferred in the
186 maximum clade credibility tree and plotted each as its own transmission chain (**Figure 3a**). We
187 enumerated the number of transitions into Washington in each tree in the posterior set, and
188 plotted the distribution of Washington introductions consistent with the phylogeny (**Figure 3b**).



202 Genomic data show that mumps was introduced into Washington State approximately 13

203 independent times (95% highest posterior density, HPD: 12 - 15), from geographically disparate

204 locations (**Figure 3**). In addition to the nine highly divergent Washington tips (**Figure 2-figure**

205 **supplement 3**), we detect one introduction from Massachusetts that descends from a long

206 branch. Prior to being sampled in Washington, this lineage was last inferred to circulate in

207 Massachusetts in late 2015. Thus rather than representing a direct introduction from

208 Massachusetts to Washington, this lineage likely moved through other geographic locations that

209 lack genomic sampling. We infer introductions from Ontario and Missouri that each lead to 1-3
210 sampled cases (**Figure 3b**), suggesting limited onward transmission following these
211 introductions. In contrast, 4 introductions from Arkansas account for 92/110 sequenced cases,
212 suggesting that these introductions led to more sustained chains of transmission following
213 introduction (**Figure 3b**). We refer to the largest cluster as the “primary outbreak clade,” and
214 infer its introduction from Arkansas to Washington around August of 2016 (August 7, 2016, 95%
215 HPD: July 11, 2016 to September 19, 2016, **Figure 3b**), 3.5 months before Washington’s first
216 reported case. These data reveal that what had appeared to be a single outbreak based on
217 case surveillance data was in fact a series of multiple introductions, primarily from Arkansas,
218 sparking overlapping and co-circulating transmission chains.

219

220 **SH gene sequencing is insufficient for fine-grained geographic inference**

221 Mumps virus surveillance and genotyping relies on the SH gene (Centers for Disease Control
222 and Prevention, 2019a), a short, 316 bp gene that is simple and rapid to sequence. To
223 determine whether SH gene sequencing would have produced similar results, we built a
224 divergence tree using our set of North American full genomes (**Figure 2-figure supplement 4**),
225 then truncated that data to include only SH gene sequences (**Figure 2-figure supplement 5**).
226 Almost all North American SH genes were identical, resulting in a single, large polytomy (**Figure**
227 **2-figure supplement 5**). This indicates that SH sequences lack sufficient resolution to elucidate
228 fine-grained patterns of geographic spread, consistent with previous findings (Gouma et al.,
229 2016; Wohl et al., 2020).

230

231 **Quantifying differences in transmission patterns within Washington**

232 In both Arkansas and Washington, Marshallese individuals comprised over 50% of mumps
233 cases, despite accounting for a much lower proportion of the population in both states.
234 Phylogenetic reconstruction links the outbreaks in Washington and Arkansas, placing most

235 sampled mumps genomes in Washington as descendant from Arkansas. We sought to
236 investigate how mumps transmission may have differed within Marshallese and non-
237 Marshallese communities within the same outbreak. Phylogenetic trees reflect the transmission
238 process and can be used to quantify differences in transmission patterns among population
239 groups. If transmission rates were distinct between Marshallese and non-Marshallese mumps
240 cases, we would expect the following: 1. Sequences from the high-transmitting group should be
241 more frequently detected upstream in transmission chains. 2. Introductions seeded into the
242 high-transmitting group should result in larger and more diverse clades in the tree. 3. The
243 internal nodes of the phylogeny should be predominantly composed by members of the high-
244 transmitting group, while members of the low-transmitting group should primarily be found at
245 terminal nodes, since less propagated transmission will cause the lineage to die out.

246

247 **Marshallese cases are enriched upstream in transmission chains**

248 We developed a transmission metric to quantify whether Marshallese cases were enriched at
249 the beginnings of successful transmission chains. We traverse the full genome divergence
250 phylogeny (**Figure 2-figure supplement 4**) from root to tip. When we encounter a tip that lies
251 on an internal node, we enumerate the number of tips that descend from its parent node. We
252 then classify each tip in the phylogeny as either a “basal tip” (i.e., there are tips detected
253 downstream of that tip) or a “terminal tip” (there are not tips detected downstream), and
254 compare the proportion of basal and terminal tips among groups (**Figure 4a**, see Methods for
255 more details). Given our sampling proportion (110 sequences/889 total cases, ~12%), we do not
256 expect to have captured true parent/child infection pairs. Rather, we expect to have
257 preferentially sampled long, successful transmission chains within the state. This allowed us to
258 test whether community membership, vaccination status, and age were associated with
259 sustained transmission via logistic regression(see Methods for details and statistical model).
260 While those with unknown vaccination status were more likely to be basal in the tree than those

261 with known up-to-date vaccination, the confidence interval could not exclude a null or positive
262 association between vaccination status and basal/terminal status (**Table 1**). Having an age of at
263 least 20 years predicted a mean lowered odds of being basal in the tree, but a wide range of
264 effects are plausible given our sample. Resolving the precise effects of vaccination status and
265 age would likely require a larger dataset. However, we do find evidence for community status as
266 a strong predictor for being basal on the phylogeny. Marshallese cases were significantly more
267 likely to be basal than non-Marshallese cases (odds ratio = 3.2, p = 0.00725, **Table 1**). While
268 only 27% (14/52) of non-Marshallese tips were ancestral to downstream samples, 56% (32/57)
269 of Marshallese tips were ancestral in lineages with sampled propagated transmission
270 (**Supplementary File 1d**). These results suggest that community membership was a significant
271 determinant of sustained transmission while controlling for vaccination status and age.

272

273 **Table 1: Associations between basal tip position in the phylogeny and possible**
274 **predictors of transmission.**

Predictor variable	Estimated coefficient (standard error)	Odds ratio (95% CI)	p-value
Not up-to-date	-0.76 (0.69)	0.47 (0.11, 1.73)	0.27
Vaccination status unknown	0.72 (0.77)	2.04 (0.47, 10.15)	0.35
Age ≥ 20 years	-0.38 (0.51)	0.69 (0.25, 1.86)	0.46
Community status	1.21 (0.42)	3.36 (1.49, 7.91)	0.0042

275 We evaluated the impact of vaccination status, age, and community membership on the probability that a
276 sampled virus was basal in the tree. Coefficients represent the increase in the log odds of being basal in
277 the tree for each given predictor variable while controlling for the others. Coefficients were exponentiated
278 to produce odds ratios. We evaluated the impacts of having an unknown vaccination status, having a
279 vaccination status that is not up-to-date, having an age of at least 20 years, and being Marshallese as
280 binary predictor variables.

281

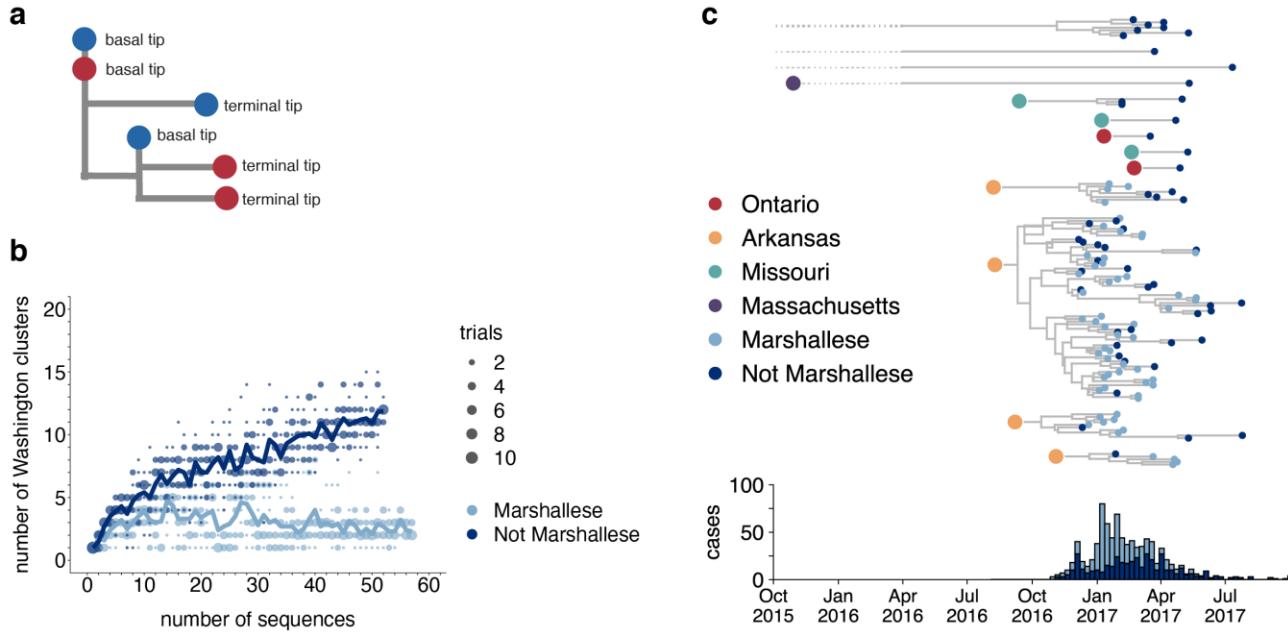
282 **Longer transmission chains are associated with community status**

283 In the absence of recombination, closely-linked infections will cluster together on the tree, while
284 unrelated infections should fall disparately on the tree, forming multiple smaller clusters. We

285 inferred the number of Washington-associated clades in the tree as a function of whether
286 sampled infections came from Marshallese or non-Marshallese individuals. Using the full North
287 American phylogeny, we removed all Washington sequences and separated them into viruses
288 sampled from cases noted as Marshallese or non-Marshallese. Then, separately for each
289 group, we added sequences back into the tree one by one, until all sequences for that group
290 had been added. For each number of sequences, we performed 10 independent trials (see
291 Methods for complete details), and at each step, we enumerated the number of inferred
292 Washington clusters in the phylogeny. For comparison, we also grouped tips by vaccination
293 status and repeated this analysis.

294

295 For viruses sampled from non-Marshallese individuals, the number of inferred clusters
296 increases linearly as tips are added to the tree (**Figure 4b**). This suggests that these infections
297 are not closely related, and are therefore not part of sustained transmission chains (**Figure 4b**).
298 In contrast, the number of inferred clusters for Marshallese tips stabilizes after ~10 tips are
299 added, even as almost 50 more sequences are added to the tree. This pattern likely arises
300 because many Marshallese infections are part of the same long transmission chain, such that
301 newly added tips nest within existing clusters. We do not observe similar differences among
302 vaccination groups (**Figure 4-figure supplement 1**). These findings are consistent with distinct
303 patterns of transmission among Marshallese versus non-Marshallese cases: transmission
304 among Marshallese individuals resulted in a small number of large clusters, while transmission
305 among non-Marshallese individuals are generally the result of disparate introductions that
306 generate shorter transmission chains.



307

308 **Figure 4: Marshalllese individuals sustain longer transmission chains**

309 **a.** A schematic for quantifying tips that lie “upstream” in transmission chains. For tips that lie on an
 310 internal node, meaning that they have a branch length separating them from their parent internal node of
 311 less than one mutation , we infer the number of child tips that descend from that tip’s parental node. For
 312 each tip in the example tree, its classification as either a “basal tip” or a “terminal tip” is annotated
 313 alongside it. All tips that have a nonzero branch length are annotated as terminal tips. We can then
 314 compare whether sequences of particular groups (here, blue vs. red) are more likely to be basal or
 315 terminal via logistic regression. **b.** We separated all Washington tips and classified them into Marshalllese
 316 and not Marshalllese. We then performed a rarefaction analysis and plotted the number of inferred
 317 Washington clusters (y-axis) as a function of the number of sequences included in the analysis (x-axis).
 318 Dark blue represents not Marshalllese sequences, and light blue represents Marshalllese sequences.
 319 Each dot represents the number of trials in which that number of clusters was inferred, and the solid line
 320 represents the mean across trials. **c.** The exploded tree as shown in Figure 3a is shown, but tips are now
 321 colored by whether they represent Marshalllese or non-Marshalllese cases. For reference, the number of
 322 Washington cases (y-axis) is plotted over time (x-axis), where bar color represents whether those cases
 323 were Marshalllese or not.

324

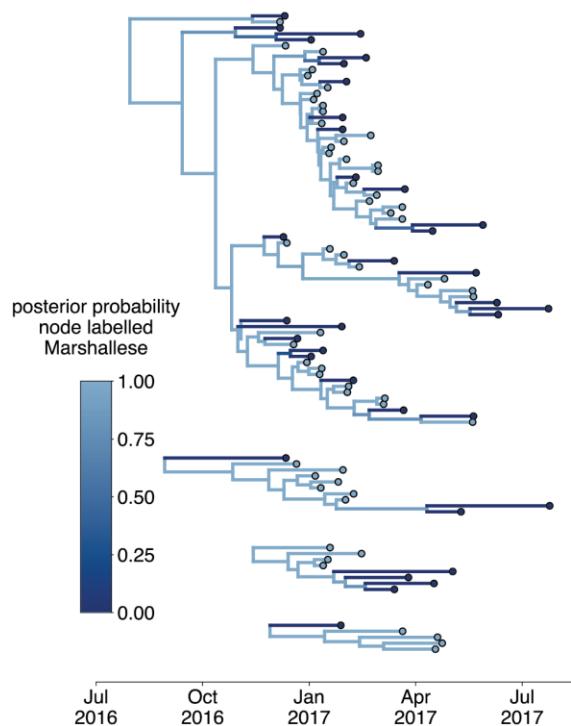
325 We next separated each Washington introduction and colored each tip by community
 326 membership. Every introduction that was not seeded from Arkansas led to exclusively non-
 327 Marshalllese infections, while introductions from Arkansas defined lineages that circulated for
 328 longer and were enriched with Marshalllese tips (**Figure 4c**). The primary outbreak clade is
 329 particularly enriched, containing 43 Marshalllese tips and 26 non-Marshalllese tips, hinting that
 330 transmission chains are longer when Marshalllese cases are present in a cluster.

331

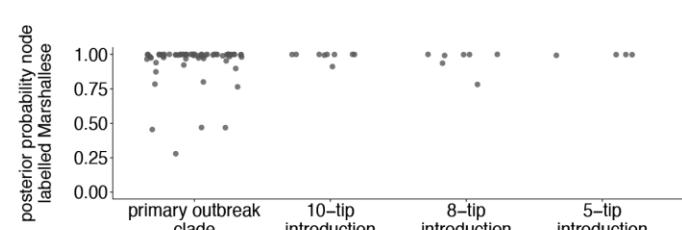
332 **Mumps transmitted efficiently within the Marshallese community**

333 Internal nodes on a phylogeny represent ancestors to subsequently sampled tips, while terminal
 334 nodes represent viral infections that did not give rise to sampled progeny. If the mumps
 335 outbreak were primarily sustained by transmission within one group, the backbone of the
 336 phylogeny and the majority of internal nodes should be inferred as circulating in that group. We
 337 selected the 4 introductions that contained both Marshallese and non-Marshallese tips (**Figure**
 338 **4c**, the 4 Arkansas introductions), and reconstructed ancestral states along the phylogeny and
 339 migration/transmission rates between Marshallese and non-Marshallese groups using a
 340 structured coalescent model.

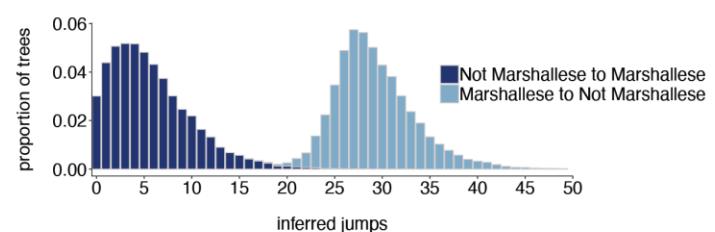
a



b



c



341

Jul 2016 Oct 2016 Jan 2017 Apr 2017 Jul 2017

342 **Figure 5: The Washington outbreak was sustained by transmission in the Marshallese community**

343 **a.** Using the 4 Washington clusters that had a mixture of Marshallese and non-Marshallese cases, we
 344 inferred phylogenies using a structured coalescent model. Each group of sequences shared a clock
 345 model, migration model, and substitution model, but each topology was inferred separately, allowing us to
 346 incorporate information from all 4 clusters into the migration estimation. For each cluster, the maximum
 347 clade credibility tree is shown, where the color of each internal node represents the posterior probability
 348 that the node is Marshallese. **b.** For each internal node shown in panel a, we plot the posterior probability
 349 of that node being Marshallese. Across all 4 clusters, 74 out of 88 internal nodes (84%) are inferred as
 350 Marshallese with a posterior probability of at least 0.95. **c.** The posterior distribution of the number of

351 “jumps” or transmission events from Marshallese to not Marshallese (light blue) and not Marshallese to
352 Marshallese (dark blue) inferred for the primary outbreak clade.
353

354 74/88 internal nodes were inferred to circulate within the Marshallese community with posterior
355 probability of at least 0.95 (**Figure 5a, b**). Movement of a lineage from the Marshallese deme
356 into the non-Marshallese deme subsequently caused the lineage to die out quickly (**Figure 5a**,
357 dark blue branches). This suggests that transmission was overwhelmingly maintained within the
358 Marshallese community, and that infections seeded into the non-Marshallese community did not
359 sustain prolonged transmission chains. We estimate substantially more transmission from
360 Marshallese to non-Marshallese groups than the opposite: within the primary outbreak clade, we
361 estimate 29 transmission events from Marshallese to non-Marshallese groups (95% HPD: 21,
362 37), and only 6 (95% HPD: 0, 14) from non-Marshallese to Marshallese groups (**Figure 5d**).
363 This strongly suggests that transmission predominantly occurred in one direction: transmission
364 events leading to non-Marshallese infections usually died out, and did not typically re-seed
365 circulation within the Marshallese community. These results hold true regardless of migration
366 rate prior (**Figure 5-figure supplement 1**).

367

368 To ensure that our results were not driven by unequal sampling within the analyzed clades, we
369 generated 3 datasets in which the number of Marshallese and non-Marshallese tips were
370 subsampled to be equal. For each of these 3 subsampled datasets, we ran 3 independent
371 chains under the same model described above. Chains converged for 2 of the 3 subsampled
372 datasets. In the converged chains, we recover very similar tree topologies (**Figure 5-figure**
373 **supplement 2a**) with equivalent phylogenetic reconstructions of lineage circulation within
374 Marshallese and non-Marshallese demes. We also recovered maximum clade credibility trees in
375 which the vast majority of the internal nodes are inferred to circulate within the Marshallese
376 deme (**Figure 5-figure supplement 2a,b**), confirming that our findings are robust to sampling,

377 consistent with past observations of model performance (De Maio et al., 2015; Dudas et al.,
378 2018; Vaughan et al., 2014).

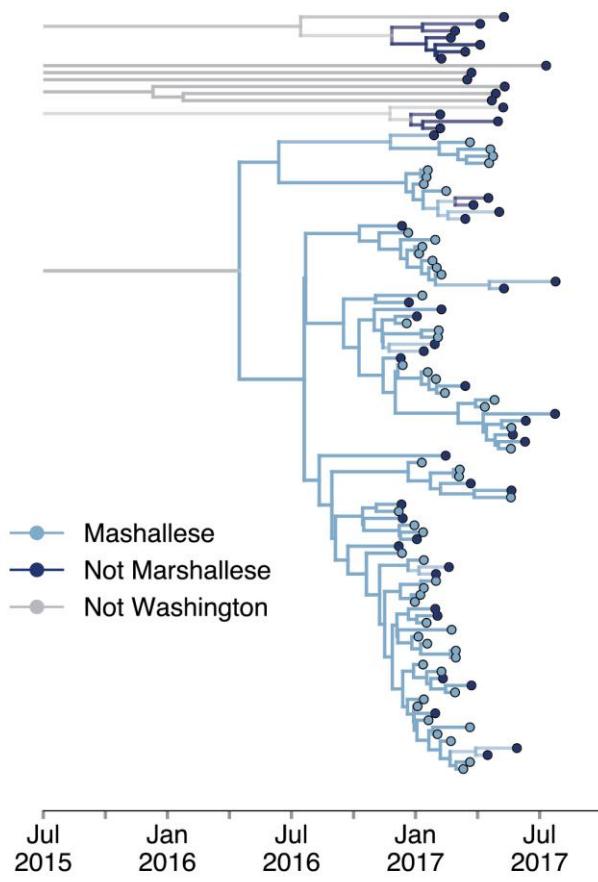
379

380 The above structured coalescent model requires both groups to be present in each cluster,
381 which meant that we had to exclude several small Washington introductions composed entirely
382 of non-Marshalllese tips (**Figure 4c**). To assess whether our findings would change if we
383 analyzed all sequenced samples, we performed an additional analysis incorporating all
384 Washington genotype G sequences in our dataset and estimated a single tree using an
385 approximate structured coalescent model (Müller et al., 2018). All Washington sequences were
386 annotated as either Marshalllese or not Marshalllese. To provide a “source” population for the
387 extensive diversity among our disparate Washington introductions, we also specified a third,
388 unsampled deme, for which migration was only allowed to proceed outward. As above, we
389 inferred very few non-Marshalllese internal nodes (**Figure 6 and Figure 6-figure supplement**
390 **1**). All internal nodes in the primary outbreak group are inferred as Marshalllese with high
391 probability, while non-Marshalllese cases are present as terminal nodes. We recovered support
392 for a single non-Marshalllese cluster, indicating limited sustained transmission in the non-
393 Marshalllese population.

394

395 Structured coalescent models infer the effective population size (N_e) for each group, which
396 reflects the number of infections necessary to generate the observed genetic diversity.
397 Differences in N_e can result from different transmission rates or different numbers of infected
398 individuals (Volz, 2012), and can therefore approximate differences in disease frequency
399 between groups. While the total number of Marshalllese and non-Marshalllese cases reported
400 through the public health surveillance system were similar (**Supplementary File 1b**), we
401 estimate that N_e for the non-Marshalllese group is approximately 3 times higher than that of the
402 Marshalllese group. Assuming the same number of infected individuals in each group, lower N_e 's

403 suggest higher transmission rates (Volz, 2012), suggesting more transmission within the
404 Marshallese deme. Taken together, our results suggest that the outbreak was primarily
405 sustained by transmission within the Marshallese community. While we do observe spillover into
406 the non-Marshallese community, transmission was generally not as successful there, resulting
407 in short, terminal transmission chains.



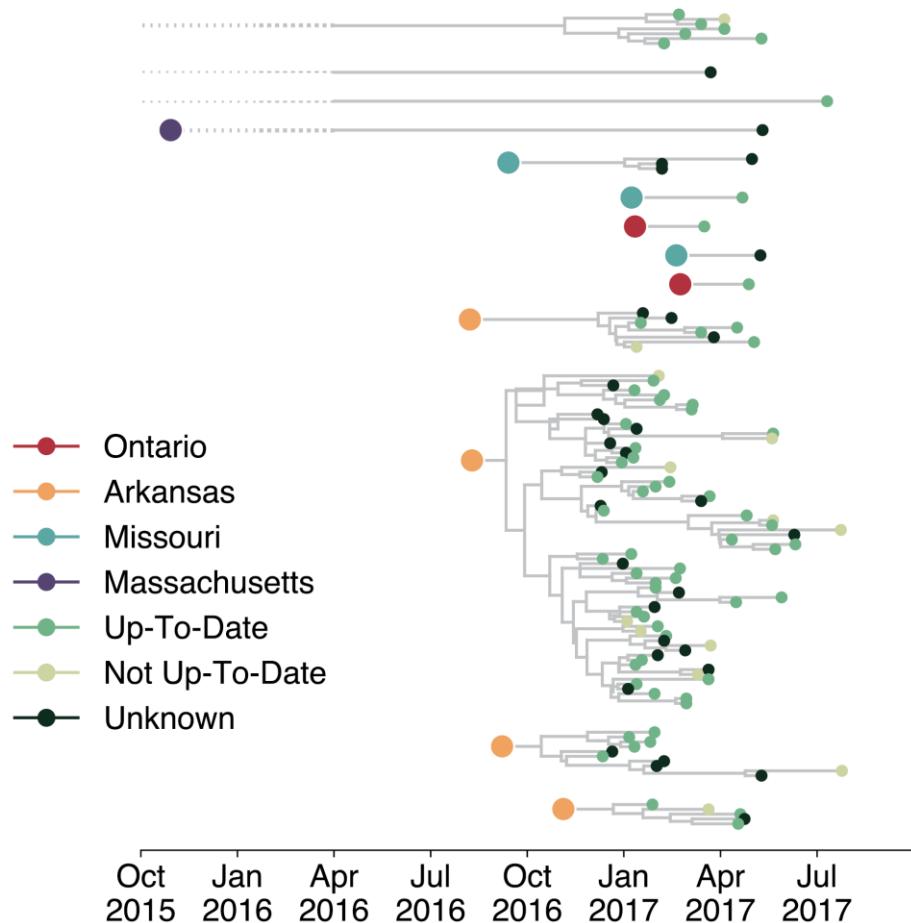
408

409 **Figure 6: Including all Washington sequences recovers majority of transmission in Marshallese**
410 To ensure that excluding non-Marshallese clusters did not skew our findings, we inferred a single tree
411 using all Washington sequences. We performed a structured coalescent analysis specifying 3 groups:
412 Marshallese, not Marshallese, and not Washington. Each internal node is colored by its most probable
413 group, with its opacity specifying the posterior probability of being in that group (fully opaque being
414 probability = 1, fully transparent being probability = 0).

415

416 **Viruses infecting individuals in different vaccination groups are genetically similar**
417 Although only 9.7% of reported mumps cases in Washington were not up-to-date for mumps
418 vaccination, infection of these individuals could have disproportionately impacted transmission

419 in the state. Emergence of an antigenically novel strain of mumps could also allow infection of
 420 previously vaccinated individuals, and result in different virus lineages infecting individuals in
 421 different vaccination categories. We colored the tips of all Washington cases in our phylogeny to
 422 represent whether they were derived from individuals who were up-to-date, not up-to-date, or
 423 whose vaccination status was unknown. Mirroring overall vaccination coverage in Washington,
 424 the vast majority of samples in our dataset were from up-to-date individuals. The not up-to-date
 425 individuals present in our dataset are dispersed throughout the phylogeny and do not cluster
 426 together (**Figure 7**), suggesting that there is no genetic difference between viruses infecting
 427 individuals with different vaccination statuses.



428
 429 **Figure 7: Individuals in different vaccination groups are infected by genetically similar viruses.**
 430 The exploded tree as shown in Figure 3a is shown, but tips are now colored by whether they represent
 431 cases from individuals who are up-to-date for mumps vaccination, not up-to-date, or cases for which

432 vaccination status was unknown. The color of the large dot represents the inferred geographic location
433 from which the Washington introduction was seeded.

434

435 Discussion

436 The resurgence of mumps in North America has ushered renewed attention towards
437 understanding post-vaccine era mumps transmission. While many studies have used
438 phylodynamic approaches to elucidate viral patterns of geographic spread (Dudas et al., 2017;
439 Gouma et al., 2016; Grubaugh et al., 2017; Stapleton et al., 2019), using genomics to
440 distinguish transmission patterns among epidemiologically distinct groups is novel. We employ a
441 phylogenetic method (Vaughan et al., 2014) traditionally applied to geography that is robust to
442 sampling bias (Dudas et al., 2018) to investigate drivers of mumps transmission in Washington.
443 We show that the Washington State outbreak was fueled by approximately 13 independent
444 introductions, primarily from Arkansas, leading to multiple co-circulating transmission chains.
445 Within Washington, transmission was more efficient within the Marshallese community.
446 Marshallese individuals were more often sampled at the beginnings of transmission chains,
447 contributed to longer transmission chains on average, and were overwhelmingly enriched on
448 internal nodes within the phylogeny. We were unable to evaluate the precise effects of age and
449 vaccination status on transmission in our outbreak. Future, larger studies will be necessary to
450 disentangle the interplay between contact patterns, waning immunity, and vaccination status
451 during mumps transmission. However, our data do suggest that social networks can be critical
452 determinants of mumps transmission. Future work exploring how social and economic
453 disparities may amplify respiratory disease transmission will be necessary for updating outbreak
454 mitigation and prevention strategies. By combining detailed metadata, novel metrics of
455 transmission in the tree, and robust controls for sampling, we provide a framework for
456 investigating source-sink dynamics that is readily applicable to other viral pathogens.

457

458 Sampling bias presents a persistent problem for phylodynamic studies that can complicate
459 inference of source-sink dynamics (De Maio et al., 2015; Dudas et al., 2018; Frost et al., 2015;
460 Kühnert et al., 2011; Lemey et al., 2020; Stack et al., 2010). Sampling bias can arise from
461 unequal case detection or from curating a dataset that poorly represents the underlying
462 outbreak. Washington State uses a passive surveillance system for mumps detection and case
463 acquisition, which is known to result in underreporting. Because the WA Department of Health
464 did not perform active mumps surveillance, it is difficult to assess whether different
465 epidemiologic groups have different likelihoods of being sampled. Marshallese individuals are
466 less likely to seek healthcare (Towne et al., 2020), which may have resulted in particularly high
467 rates of underreporting in this group. If the number of cases within the Marshallese community
468 were in fact higher than reported, this would increase the magnitude of the patterns we
469 describe, making our estimates conservative. Given a distribution of cases, composing a
470 dataset for analysis also requires sampling decisions. Uniform sampling regimes in which
471 sampling probability is equal across groups have been shown to perform well for source-sink
472 inferences (Hall et al., 2016). By selecting sequences that matched the overall attributes of the
473 outbreak, including a near 50:50 split between Marshallese and non-Marshallese cases, we
474 adhere to this recommendation. We then specifically employed structured coalescent
475 approaches which have been shown to be robust to sampling differences (Dudas et al., 2018;
476 Müller et al., 2018; Vaughan et al., 2014), rather than using other common approaches that treat
477 sampling intensity as informative of population size (Lemey et al., 2009). Within this framework,
478 we further explore the possibility that unequal sampling within Washington clades could skew
479 internal node reconstruction by forcing the sampling within each Washington clade to be equal
480 between Marshallese and non-Marshallese tips. In doing so, differences within each clade must
481 necessarily be driven by differences in transmission dynamics, rather than sampling. By
482 combining careful sample selection with overlapping approaches to evaluate sampling bias, we

483 were able to mitigate concerns that our source-sink reconstructions are driven by sampling
484 artifacts.

485

486 Our results highlight the utility of genomic data to clarify epidemiologic hypotheses. While
487 genomic data and epidemiologic investigation (including case interviews and contact follow up)
488 suggested an Arkansas introduction as the Washington outbreak's primary origin, sequence
489 data revealed repeated and ongoing introductions into Washington, similar to patterns observed
490 in Massachusetts, and the Netherlands (Gouma et al., 2016; Wohl et al., 2020). We also find
491 widespread geographic mixing across the phylogeny, consistent with investigations from the US
492 (Wohl et al., 2020), Canada (Stapleton et al., 2019), and Europe (Gavilán et al., 2018; Gouma et
493 al., 2016). Like others (Gouma et al., 2016; Wohl et al., 2020), we confirm that SH genotyping
494 alone is insufficient for fine-grained resolution of geographic transmission patterns. While CDC
495 guidelines currently recommend SH-based genotyping specifically for tracking transmission
496 pathways (Clemons et al., 2020), building public health capacity for full-genome sequencing
497 may be more useful for resolving local mumps transmission patterns.

498

499 Our finding that most introductions sparked short transmission chains suggests that mumps did
500 not transmit efficiently among the general Washington populace. We suspect that more diffuse
501 contact patterns may help explain this. Mumps has historically caused outbreaks in communities
502 with strong, interconnected contact patterns (Barskey et al., 2012; Fields et al., 2019; Nelson et
503 al., 2013), and in dense housing environments (Snijders et al., 2012), highlighted most recently
504 by outbreaks in US detention centers (Lo et al., 2021). In 2016, most outbreaks in the US were
505 associated with university settings (Albertson et al., 2016; Bonwitt et al., 2017; Donahue et al.,
506 2017; Golwalkar et al., 2018; Shah et al., 2018; Wohl et al., 2020), including a separate, smaller
507 outbreak in Washington State associated with Greek housing (Bonwitt et al., 2017). Outside of
508 university settings, other outbreaks in 2016 were reported within close-knit ethnic communities

509 (Fields et al., 2019; Marx et al., 2018). We speculate that while waning immunity may promote
510 outbreaks by increasing susceptibility among young adults, outbreaks in younger age groups
511 may be possible in sufficiently high-contact settings. Provision of an outbreak dose of mumps-
512 containing vaccine to high-risk groups may therefore be especially effective for limiting mumps
513 transmission in future outbreaks. Others have reported success in using outbreak dose mumps
514 vaccinations to reduce mumps transmission on college campuses (Cardemil et al., 2017; Shah
515 et al., 2018) and in the US army (Arday et al., 1989; Eick et al., 2008; Green, 2006; Kelley et al.,
516 1991), and the CDC currently recommends providing outbreak vaccine doses to individuals with
517 increased risk due to an outbreak (Marlow et al., 2020). Future work to quantify the interplay
518 between contact rates and vaccine-induced immunity among different age and risk groups
519 should be used to guide updated vaccine recommendations.

520

521 Recent research has focused on identifying groups at risk for mumps infection due to their age
522 (Lewnard and Grad, 2018), with less attention to other factors that may make populations
523 vulnerable. While a combination of waning immunity and dense housing settings make college
524 campuses ideal for mumps outbreaks, the Washington and Arkansas outbreaks show that
525 populations other than young adults are at risk. Soliciting feedback from the Marshallese
526 community allowed us to contextualize our genomic results with the lived experience of
527 individuals most heavily affected during the outbreak and to identify reasonable hypotheses for
528 efficient transmission. Based on these interviews and previously published studies, we
529 speculate that within the Marshallese community, a combination of factors likely led to a high
530 force of infection. The following paragraph outlines contributing factors that were brought to light
531 during our interviews with a collaborating community activist, along with corroborating citations
532 from the literature. Each of these factors were specifically cited as important and directly stem
533 from our interviews with her.

534

535 Multigenerational living is common in the Marshallese community (Fields et al., 2019), and
536 Marshallese households tend to be larger on average (average household size = 5.28 (Harris
537 and Jones, 2005), average household size for entire US populace = 2.52 (US Census Bureau,
538 n.d.)). Having more household contacts may have facilitated a greater number and higher
539 intensity of interactions among individuals, allowing the force of infection to overcome pre-
540 existing immunity. The Marshallese community is often described as close-knit, with frequent
541 and close interactions among individuals, a strong sense of community, and a broader sense of
542 family than the single-family unit typical of broader American culture (Barker, 2012; Embassy of
543 the Republic of the Marshall Islands to the United States of America, n.d.). Contacts within the
544 community could therefore be more frequent or intense, which may facilitate transmission. It is
545 also possible that infection intensity within the Marshallese community was exacerbated by low
546 rates of insurance coverage and poor access to healthcare (McElfish et al., 2017; Towne et al.,
547 2020), hesitancy to seek medical care (Williams and Hampton, 2005), and health disparities
548 stemming from US occupation, nuclear testing, and exclusion from healthcare services. As part
549 of reparations for US nuclear testing, the US signed the Compact of Free Association Treaty
550 (COFA)(Congress, 2003) with the Marshall Islands in 1989, permitting Marshallese residents to
551 live and work in the US without visas. However, eligibility for Medicaid was revoked for COFA
552 immigrants in 1996, and US-residing Marshallese remain economically disadvantaged and
553 under-insured (McElfish, 2016; McElfish et al., 2017, 2015). The passage of the Affordable Care
554 Act (ACA) has not ameliorated these issues. Interviews with US-residing Marshallese note
555 confusion among ACA staff regarding the legal status of COFA recipients, leading to drawn out
556 enrollment processes that often leave individuals uninsured, frustrated (McElfish et al., 2016),
557 and far less likely to access care (Towne et al., 2020). A study of healthcare-seeking behavior
558 among patients with diabetes showed that while multiple factors contribute to forgone care in
559 the US populace, 77% of surveyed Marshallese individuals reported recent forgone care and
560 lack of insurance was the primary reason (Towne et al., 2020). Marshallese trust in US medical

561 institutions was seriously undermined by the unconsented use of Marshallese individuals for
562 experiments on health impacts of nuclear exposure, with effects lingering today (Barker, 2012).
563 Banked historical samples confirm uptake of radioactive materials in Marshallese inhabitants of
564 affected Islands (Simon et al., 2010), but there has been limited published data on long-term
565 health impacts of nuclear exposure, and significant concern remains within the community
566 (Bordner et al., 2016). Finally, when Marshallese individuals do access care, they report
567 experiencing disdain from healthcare workers (Duke, 2017) and sub-optimal care (McElfish et
568 al., 2016). Interviews with medical workers show that blame for poor Marshallese health
569 outcomes is sometimes placed on host genetics or cultural practices (Duke, 2017), poor health
570 literacy (McElfish et al., 2018), or choosing to delay care (McElfish et al., 2018), with less
571 consideration given to how the economic and legal impacts of US occupation affect the health of
572 Marshallese individuals. These factors compound, and Marshallese individuals report hesitation
573 to seek medical care, even when sick (McElfish et al., 2016). Hesitancy to seek care could have
574 contributed to mumps transmission if sick individuals were primarily cared for at home without
575 knowledge of or the ability to implement community-isolation protocols.

576

577 Our findings highlight that social networks can be the primary risk factor for a respiratory virus
578 outbreak, even when a vaccine is effective and widely used. This finding is especially pertinent
579 as SARS-CoV2 continues to disproportionately impact populations who live and work in high-
580 risk settings, including the Marshallese (Center et al., 2020; McElfish et al., 2021), and for whom
581 vaccine licensure and distribution alone may not be a panacea. Future work should explore
582 whether nuclear exposure has impacted Marshallese immune function and susceptibility to
583 infectious disease. The passing of federal legislation remedying the exclusion of Marshallese
584 individuals from Medicaid access (Hirono, 2019) in December 2020 marks an important step
585 toward improving healthcare access. Future work to evaluate whether this change improves
586 Marshallese access to healthcare and mitigates increased disease risk will be crucial follow-up.

587 The findings of this paper demonstrate the importance of expanding our understanding of
588 populations at risk for mumps re-emergence, so that rapid and comprehensive outbreak
589 response strategies can be implemented to mitigate negative health impacts for all affected
590 communities. Finally, future work to disentangle the complex interplay between healthcare
591 access, social and economic disparity, and respiratory virus risk will be essential for mitigating
592 health impacts of mumps and other respiratory viruses.

593

594 **Materials and Methods**

Key Resources Table				
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
biological sample (Mumps virus)	110 buccal swabs from mumps positive patients in Washington	Washington State Department of Health	Sequences were deposited in Genbank under accessions MT859507-MT859672. Raw reads were deposited under SRA project number PRJNA641715	Full metadata for each sequence is available in the manuscript in Supplementary File 1a
biological sample (Mumps virus)	56 buccal swabs from mumps positive patients from other US states	Wisconsin State Lab of Hygiene	Sequences were deposited in Genbank under accessions MT859507-MT859672. Raw reads were deposited under SRA project number PRJNA641715	Full metadata for each sequence is available in the manuscript in Supplementary File 1a

biological sample (Mumps virus)	Publicly available mumps genomes	NIAID Virus Pathogen Database and Analysis Resource (ViPR) (Pickett et al., 2012)	http://www.viprbrc.org/	
sequence-based reagent	mumps_1.5k b primers	This paper	PCR primers	Full list of PCR primer sequences is available in the Methods section under “Viral RNA extraction, cDNA synthesis, and amplicon generation”
commercial assay or kit	QiAmp Viral RNA Mini Kit	Qiagen, Valencia, CA, USA	Cat #: 52904	
commercial assay or kit	Protoscript II First strand synthesis kit	New England Biolabs, Ipswich MD, USA	Cat #: E6560L	
commercial assay or kit	Q5 Hotstart DNA polymerase	New England Biolabs, Ipswich, MD, USA	Cat #: M0493L	
commercial assay or kit	Ampure XP beads	Beckman Coulter	Cat #: A63881	
commercial assay or kit	Nextera XT DNA Library Prep Kit	Illumina, San Diego, CA, USA	Cat #: FC-131-1096	

software, algorithm	Bowtie2	<u>Langmead and Salzberg, 2012</u>	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	RRID: SCR_016368
software, algorithm	MAFFT	Katoh et al., 2002	https://mafft.cbrc.jp/alignment/software/	RRID: SCR_016368
software, algorithm	TreeTime	Sagulenko et al., 2018	https://github.com/nehmerlab/treetime	
software, algorithm	BEAST (versions 1.8.4 and 2.6.2)	<u>Drummond et al., 2012</u> , <u>Lemey et al., 2009</u> , <u>Bouckaert et al., 2019</u>	https://beast.community/ and https://www.beast2.org/	RRID: SCR_010228
software, algorithm	IQTREE	Nguyen et al., 2015	http://www.iqtree.org	
software, algorithm	Github repo with protocols for generating mumps sequences from buccal swabs	This paper	https://github.com/blab/mumps-seq	This github repository contains documentation and protocols for all lab procedures and bioinformatics pipelines used to generate consensus genomes from mumps buccal swabs.
software, algorithm	Github repo with scripts used to analyze data and generate figures for this manuscript	This paper	https://github.com/blab/mumps-wa-phylodynamics	This github repository contains all of the code used to generate figures and perform the analyses described in this

manuscript. This repository also contains xml files used for input for BEAST analyses and alignments and tree files used to generate and plot phylogenetic trees.

595

596 **Data and Code Availability**

597 All code used to analyze data, input files for BEAST, and all code used to generate figures for
598 this manuscript are publicly available at <https://github.com/blab/mumps-wa-phyloynamics>. Raw
599 FASTQ files with human reads removed are available under SRA project number
600 PRJNA641715. All protocols for generating sequence data as well as the consensus genomes
601 are available at <https://github.com/blab/mumps-seq>. Consensus genomes have also been
602 deposited to Genbank under accessions MT859507-MT859672.

603

604 **Community feedback**

605 In order to ensure that this study was faithful to the experience of the Marshallese community in
606 Washington State, we sought paid consultation from a local Marshallese community health
607 advocate. We conducted video and telephone interviews to directly address the impacts of
608 mumps transmission on the Marshallese community, community healthcare goals and priorities,
609 and the impacts of the mumps outbreak on stigmatization. This feedback informed what is being
610 presented herein, provided crucial context for understanding mumps transmission, and allowed
611 us to work with the community to determine how best to discuss Marshallese involvement in the

612 outbreak.

613

614 **Mumps surveillance in Washington State**

615 Mumps is a notifiable condition in Washington State. Therefore, per the Washington
616 Administrative Code (WAC), as specified in WAC Chapter 246-101(Washington State
617 Legislature, 2014), healthcare providers, healthcare facilities, and laboratories must report
618 cases of mumps or possible mumps to the local health jurisdiction (LHJ) of the patient's
619 residence. LHJ staff initiate case investigations and facilitate optimal collection and testing of
620 diagnostic specimens. Buccal swabs and urine are acceptable specimens for real-time reverse
621 transcription polymerase chain reaction (qRT-PCR), a preferred diagnostic test for mumps. Most
622 mumps rRT-PCR tests for Washington State residents are performed at the Washington State
623 Public Health Laboratories, where all positive specimens are archived.

624

625 Individuals testing positive for mumps ribonucleic acid (RNA) by qRT-PCR are classified as
626 confirmed mumps cases if they have a clinically-compatible illness (i.e., an illness involving
627 parotitis or other salivary gland swelling lasting at least 2 days, aseptic meningitis, encephalitis,
628 hearing loss, orchitis, oophoritis, mastitis, or pancreatitis). During case investigations, case-
629 patients or their proxies are interviewed. Information about demographics, illness
630 characteristics, vaccination history, and potential for exposure to and transmission of mumps
631 are solicited from each case-patient. In concordance with CDC guidelines (Centers for Disease
632 Control and Prevention, 2019b), only vaccine doses for which there was written documentation
633 with the date of vaccine receipt were considered valid. Individuals for which such documentation
634 could not be provided were classified as having an unknown vaccination status. For individuals
635 with documented vaccine doses, they were further characterized as up-to-date or not up-to-date
636 based on their age. The Washington State Department of Health (DOH) receives, organizes,

637 performs quality control on, and analyzes data from, LHJ case reports and supports
638 investigations upon request.

639

640 **Sample collection and IRB approval**

641 This study was approved by the Fred Hutchinson Cancer Research Center (FHCRC)
642 Institutional Review Board (IR File #: 6007-944) and by the Washington State Institutional
643 Review Board, and classified as not involving human subjects. Samples were selected for
644 sequencing to maximize temporal and epidemiologic breadth and to ensure successful
645 sequencing. As such, samples were chosen based on the date of sample collection, the PCR
646 cycle threshold (Ct), case vaccination status, and community status (Marshallese or non-
647 Marshallese). Samples were selected for sequencing in 2 batches. In the first, samples were
648 selected based on covering a wide geographic range within Washington, a full range of dates
649 covering the outbreak, and having a Ct value < 36. This initial sampling regime resulted in a
650 sample set skewed slightly towards samples from Marshallese individuals. To ensure that the
651 proportion of samples in our data closely matched the distribution of cases in the outbreak, we
652 then selected a second batch of samples using the same criteria as above, but excluded
653 samples from Marshallese individuals. We then randomly sampled an additional 30 samples
654 from non-Marshallese individuals. This sampling regime resulted in a dataset that closely
655 mirrors the distribution of metadata categories in the outbreak overall. All metadata, including
656 case vaccination status, were transferred from WA DOH to FHCRC in a de-identified form.

657

658 We also sequenced an additional set of 56 samples collected in Wisconsin, Ohio, Missouri,
659 Alabama, and North Carolina provided by the Wisconsin State Laboratory of Hygiene. 10 of
660 these samples were collected in Wisconsin during the 2006/2007 Midwestern college campus
661 outbreaks, 6 samples were collected in 2014, and the rest were collected between 2016 and

662 2018. For these samples, we received metadata describing sample Ct value and date of
663 collection. All metadata were received by FHCRC in de-identified form.

664

665 **Viral RNA extraction, cDNA synthesis, and amplicon generation**

666 Viral RNA was extracted from buccal swabs using either the QiAmp Viral RNA Mini Kit (Qiagen,
667 Valencia, CA, USA) or the Roche MagNA Pure 96 DNA and viral NA small volume kit (Roche,
668 Basel, Switzerland). For samples extracted with the QiAmp Viral RNA Mini Kit, 500 µl of buccal
669 swab fluid was spun at 5000 x g for 5 minutes at 4°C to pellet host cells. The supernatant was
670 then removed and centrifuged at 14,000 rpm for 90 minutes at 4°C to pellet virions. Excess fluid
671 was discarded, and the pelleted virions were resuspended in 150-200 µl of fluid. Resuspended
672 viral particles were then used as input to the QiAmp Viral RNA Mini Kit (Qiagen, Valencia, CA,
673 USA), following manufacturer's instructions, and eluting in 30 µl of buffer AVE. For extraction
674 with the MagNA Pure, we followed manufacturer's instructions.

675

676 cDNA was generated with the Protoscript II First strand synthesis kit (New England Biolabs,
677 Ipswich MD, USA), using 8 µl of vRNA as input and priming with 2 µl of random hexamers.
678 vRNA and primers were incubated at 65°C for 5 minutes. Following this incubation, 10 µl of
679 Protoscript II reaction mix (2x) and 2 µl of Protoscript II enzyme mix (10x) were added to each
680 reaction and incubated at 25°C for 5 minutes, then 42°C for 1 hour, followed by a final
681 inactivation step at 80°C for 5 minutes. To amplify the full mumps genomes, we used Primal
682 Scheme (<http://primal.zibraproject.org/>) to design overlapping, ~1500 base pair amplicons
683 spanning the entirety of the mumps virus genome, where each tiled set of primes overlapped by
684 ~100 base pairs. Primers are listed below.

685

Primer	Primer sequence	Forward/Reverse	Primer pool

mumps_1.5kb_1F	ACCAAGGGGAAAATGAAGATGGG	Forward	pool 1
mumps_1.5kb_1R	TAACGGCTGTGCTCTAAAGTCAT	Reverse	pool 1
mumps_1.5kb_2_F	TTGTTGACAGGCTTGCAAGAGG	Forward	pool 2
mumps_1.5kb_2_R	TTGTTCAAGATGTTGCAGGCGA	Reverse	pool 2
mumps_1.5kb_3_F	TGCAACCCCATAATGCTCACCTA	Forward	pool 1
mumps_1.5kb_3_R	AGTTTGTTCCTGCCTTGCACA	Reverse	pool 1
mumps_1.5kb_4_F	AGTGAGAGCAGTCAGATGGAAGT	Forward	pool 2
mumps_1.5kb_4_R	CCCTCCATTAGACCAGGCACTTA	Reverse	pool 2
mumps_1.5kb_5_F	AACAAACAGTGTCCAGGCCACAA	Forward	pool 1
mumps_1.5kb_5_R	GGTGGCACTGTCCGATATTGTG	Reverse	pool 1
mumps_1.5kb_6_F	TGCCGTTCAATCATGAGACATAAAGA	Forward	pool 2
mumps_1.5kb_6_R	CGTAGAGGAGTTCATACGGCCA	Reverse	pool 2
mumps_1.5kb_7_F	TGTCTGTGCCTGGAATCAGATCT	Forward	pool 1
mumps_1.5kb_7_R	CGTCCTTCCAACATATCAGTGACC	Reverse	pool 1
mumps_1.5kb_8_F	CCAAAAGACAGGTGAGTTAACAGATT	Forward	pool 2
mumps_1.5kb_8_R	ACGAGCAAAGGGATGATGACT	Reverse	pool 2
mumps_1.5kb_9_F	TTTGGCACACTCCGGTTCAAAT	Forward	pool 1
mumps_1.5kb_9_R	TGACAATGGTCTCACCTCCAGT	Reverse	pool 1
mumps_1.5kb_10_F	ACTCGCACAGTATCTATTAGATCGTG	Forward	pool 2

	A		
mumps_1.5kb_10_R	GCCCAGCCAGAGTAAACAAACA	Reverse	pool 2
mumps_1.5kb_11_F	GCCAAGCAGATGGTAAACAGCA	Forward	pool 1
mumps_1.5kb_11_R	GGCTCTCTCCAACATGCTGTTTC	Reverse	pool 1
mumps_1.5kb_12_F	GCAGGGGCCTCTATGTCACTTAT	Forward	pool 2
mumps_1.5kb_12_R	CCAAGGGAGAAAGTAAAATCAAT	Reverse	pool 2

686

687 Primers were pooled into 2 pools as follows: the first contained primer pairs 1, 3, 5, 7, 9, and 11,
 688 all pooled at 10 uM. The second pool contained primer pairs 2, 4, 6, 8, 10, and 12. All primers in
 689 pool 2 were pooled at 10 uM, except for primer pair 4, which was added at a 20 uM
 690 concentration.

691

692 PCR was performed with the Q5 Hotstart DNA polymerase (New England Biolabs, Ipswich,
 693 MD, USA), using 11.75 μ l of nuclease-free water, 5 μ l of Q5 reaction buffer, 0.5 μ l of 10 mM
 694 dNTPs, 0.25 μ l, 2.5 of pooled primers, and 5 μ l of cDNA. Amplicons were generated with the
 695 following PCR cycling conditions: 98°C for 30 seconds, followed by 30 cycles of: 98°C for 15
 696 seconds, then 67°C for 5 minutes. Cycling was concluded with a 10°C hold. PCR products were
 697 run on a 1% agarose gel, and bands were cut out and purified using the QiAquick gel extraction
 698 kit (Qiagen, Valencia, CA, USA), following the manufacturer's protocol. All optional steps were
 699 performed, and the final product was eluted in 30 μ l of buffer EB. For samples extracted on the
 700 MagNA Pure, amplicons were cleaned using a 1x bead cleanup with Ampure XP beads. Final
 701 cleaned amplicons were quantified using the Qubit dsDNA HS Assay kit (Thermo Fisher,
 702 Waltham, MA, USA).

703

704 **Library preparation and sequencing**

705 For each sample, pool 1 and pool 2 amplicons were combined in equimolar concentrations to a
706 total of 0.5 ng in 2.5 μ l. Libraries were prepared using the Nextera XT DNA Library Prep Kit
707 (Illumina, San Diego, CA, USA), following manufacturer's instructions, but with reagent volumes
708 halved for each step, for the majority of samples in our dataset. For samples processed in our
709 last sequencing run, several samples had higher Ct values. We therefore chose to process
710 these samples using the standard 1x reagent volumes for the library preparation step. All
711 libraries were purified using Ampure XP beads (Beckman Coulter, Brea, CA, USA), using a 0.6x
712 cleanup, a 1x cleanup, and a final 0.7x cleanup. At each step, beads were washed twice with
713 200-400 μ l of 70% ethanol. The final product was eluted off the beads with 10 μ l of buffer EB.
714 Tagmentation products were quantified with the Qubit dsDNA HS Assay kit (Thermo Fisher,
715 Waltham, MA, USA), and run on a Tapestation with the TapeStation HighSense D5K assay
716 (Agilent, Santa Clara, CA, USA) to determine the average fragment length. All but 8 samples
717 and negatives were pooled together in 6 nM libraries and run on 300 bp x 300 bp v3 kits on the
718 Illumina MiSeq, with a 1% spike-in of PhiX. The remaining 8 samples
719 (MuVs/Washington.USA/1.17/FH77[G], MuVs/Washington.USA/12.17/FH78[G],
720 MuVs/Washington.USA/16.17/FH79[G], MuVs/Washington.USA/19.17/FH80[G],
721 MuVs/Washington.USA/20.17/FH81[G], MuVs/Washington.USA/20.17/FH82[G],
722 MuVs/Washington.USA/29.17/FH83[G], and MuVs/Washington.USA/2.17/FH84[G]) were
723 pooled to a 1.2 nM library, and run as a 50 pM library with 2% PhiX on the Illumina iSeq, with a
724 151 bp x 151 bp v3 kit.

725 **Negative controls**

726 A negative control (nuclease-free water) was run for each viral RNA extraction, reverse
727 transcription reaction, and for each pool for each PCR reaction. These negative controls were
728 carried through the library preparation process and sequenced alongside actual samples. Any

729 samples whose negative controls from any step in the process resulted in >10x mumps genome
730 coverage were re-extracted and sequenced.

731 **Bioinformatic processing of sequencing reads**

732 Human reads were removed from raw FASTQ files by mapping to the human reference genome
733 GRCH38 with bowtie2 (Langmead and Salzberg, 2012)(RRID: SCR_016368) version 2.3.2
734 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Reads that did not map to the human
735 genome were output to separate FASTQ files and used for all subsequent analyses. Illumina
736 data was analyzed using the pipeline described in detail at
737 https://github.com/lmoncla/illumina_pipeline. Briefly, raw FASTQ files were trimmed using
738 Trimmomatic (Bolger et al., 2014) (<http://www.usadellab.org/cms/?page=trimmomatic>), trimming
739 in sliding windows of 5 base pairs and requiring a minimum Q-score of 30. Reads that were
740 trimmed to a length of <100 base pairs were discarded. Trimming was performed with the
741 following command: java -jar Trimmomatic-0.36/trimmomatic-0.36.jar SE input.fastq output.fastq
742 SLIDINGWINDOW:5:30 MINLEN:100. Trimmed reads were mapped to a consensus sequence
743 from Massachusetts (Genbank accession: MF965301) using bowtie2(Langmead and Salzberg,
744 2012) version 2.3.2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>), using the following
745 command: bowtie2 -x reference_sequence.fasta -U read1.trimmed.fastq,read2.trimmed.fastq -S
746 output.sam --local. We selected this Massachusetts sequence as an initial reference sequence
747 because at the time, it represented one of the only available genomes of a genotype G mumps
748 virus that had been sampled during a US outbreak in 2016. Mapped reads were imported into
749 Geneious (<https://www.geneious.com/>) for visual inspection and consensus calling. To avoid
750 issues with mapping to an improper reference sequence, we then remapped each sample's
751 trimmed FASTQ files to its own consensus sequence. These bam files were again manually
752 inspected in Geneious, and a final consensus sequence was called, with nucleotide sites with
753 <20x coverage output as an ambiguous nucleotide ("N"). All genomes with >50% Ns were

754 discarded. In total, we generated 140 genomes with at least 80% non-N bases, and 26
755 genomes with 50-80% non-N bases. Our median completeness (percent of bases that are not
756 Ns) across the dataset is 90%. All genomes used in these analyses are available at
757 <https://github.com/blab/mumps-seq/tree/master/data>.

758

759 **Dataset curation and maximum likelihood divergence tree generation**

760 We downloaded all currently available (as of June 2020), complete mumps genomes from North
761 America, and separately from any country in the world, from the NIAID Virus Pathogen
762 Database and Analysis Resource (ViPR) (Pickett et al., 2012) through <http://www.viprbrc.org/>.
763 We also obtained mumps genomes from British Columbia, Ontario, and Arkansas. We obtained
764 written permission from sequence authors for any sequence that had not previously been
765 published on. In total, this dataset includes 437 full mumps genomes from North America.
766 Sequences and metadata were cleaned and organized using fauna, a database system that is
767 part of the Nextstrain platform. Sequences were processed using Nextstrain's augur software
768 (Hadfield et al., 2018), and filtered to include only those with at least 8,000 bases and were
769 sampled in North America in 2006 or later. Genomes were aligned with MAFFT (Katoh et al.,
770 2002)(RRID: SCR_016368), and trimmed to the reference sequence (MuV/Gabon/13/2[G],
771 GenBank accession: KM597072). We inferred a maximum likelihood phylogeny using IQTREE
772 (Nguyen et al., 2015) with a GTR nucleotide substitution model, and inferred a molecular clock
773 and temporally-resolved phylogeny using TreeTime (Sagulenko et al., 2018). Sequences with
774 an estimated clock rate that deviated from the other sequences by >4 times the interquartile
775 distance were removed from subsequent analysis. We inferred the root-to-tip distance with
776 TempEst version 1.5.1 (Rambaut et al., 2016) with the best fitting root by the heuristic residual
777 mean squared function. Trees were output in JSON format and are available at
778 <https://github.com/blab/mumps-wa-phylogenetics/blob/master/auspice>.

779

780 **Phylogenetic analysis of full North American mumps genomes**

781 Using the same set of genome sequences used for divergence tree estimation, we aligned
782 sequences with MAFFT and inferred time-resolved phylogenies in BEAST version 1.8.4
783 (Drummond et al., 2012)(RRID: SCR_010228). We used a skygrid population size prior with 100
784 bins, and a skygrid cutoff of 25 years, allowing us to estimate 4 population sizes each year. We
785 used an HKY nucleotide substitution model with 4 gamma rate categories, and a strict clock with
786 a CTMC prior. We used a discrete trait model (Lemey et al., 2009) and estimated migration
787 rates using BSSVS and ancestral states with 27 geographic locations. Here, “state” refers to the
788 inferred ancestral identity of an internal node, where the inferred identity could be any of the 27
789 geographic locations (US states and Canadian provinces) in the dataset. For the prior on non-
790 zero rates for BSSVS, we specified an exponential distribution with a mean of 26. As a prior on
791 each pairwise migration rate, we used an exponential distribution with mean 1. All other priors
792 were left at default values. We ran this analysis for 100 million steps, sampling every 10,000,
793 and removed the first 10% of sampled states as burnin. A maximum clade credibility tree was
794 summarized with TreeAnnotator, using the mean heights option. All tree plotting was performed
795 with baltic (<https://github.com/evogytis/baltic>). Input XML files and output results are available at
796 <https://github.com/blab/mumps-wa-phylodynamics/tree/master/phylogeography>.

797

798 **Quantifying transmission in divergence trees using basal and terminal tips: formulation
799 and rationale**

800 To determine whether specific groups were more likely to be part of sustained, serially sampled
801 transmission chains, we developed a statistic to quantify transmission in the tree. Our aim was
802 to develop a heuristic method that would capture patterns similar to those captured by more
803 complex structured coalescent models. In a population with high rates of transmission and high
804 sampling intensity, it is possible that sampled individuals may represent true ancestors to
805 subsequent infections (Gavryushkina et al., 2014). While this is theoretically possible in our

dataset, we expect this to be rare. Because viruses accumulate mutations at a constant rate over time, a tip's branch length should correlate with its position along the underlying transmission chain, i.e., a short branch length should indicate that the tip is closer to the true ancestral infection than a longer branch. Plotting the number of mutations on each branch vs. its estimated branch length in time units confirms that mutations and time-calibrated branch length are correlated (**Figure 8**). This suggests that on average, branches with fewer mutations also tend to represent shorter periods of time. For our dataset, most tips with an estimated branch length within the mumps serial interval of ~18 days (Vink et al., 2014) have 0 mutations (**Figure 8, circles below dashed line**). In a population with high rates of transmission in which true ancestral infections are not directly sampled, we therefore expect that tips that are genetically closer to the ancestral node should be closer to the true ancestral infection, than tips that are genetically dissimilar.

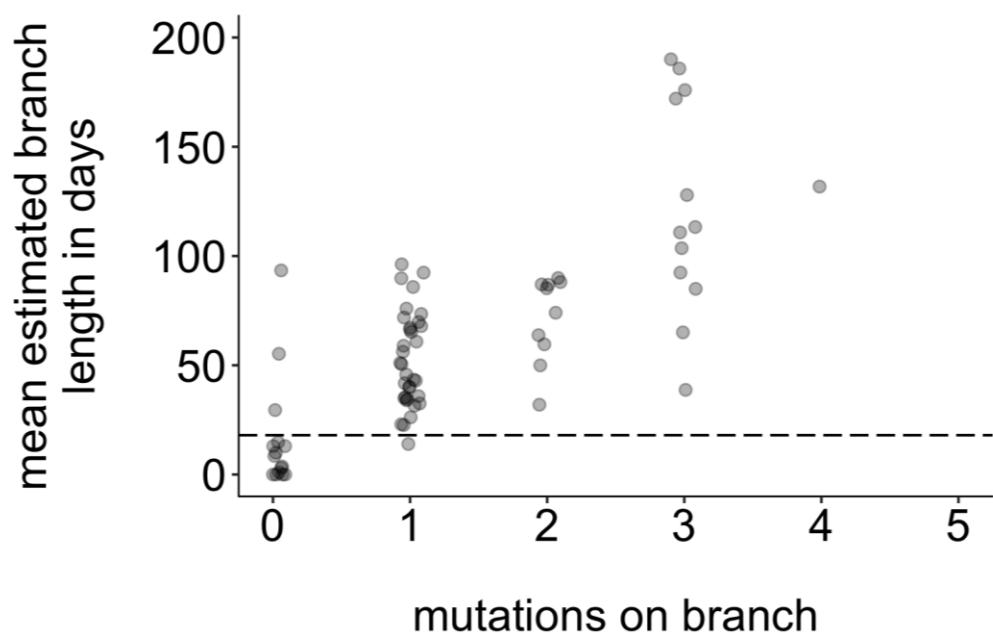


Figure 8: Mutations vs. estimated branch length in days
For each Washington tip in the full North American phylogeny with an estimated branch length in time units of ≤ 1 year, we show the number of mutations on that branch on the x-axis vs. the mean estimated branch length in days on the y-axis. The dashed line at 18 represents the mumps serial interval.

824
825 Either time or genetic divergence could be used to categorize how close tips are to their
826 parental node. Here, we have opted to use divergence for two main reasons. Mumps has a
827 relatively slow substitution rate and a long serial interval, resulting in stacks of identical
828 genomes at several points in the divergence phylogeny (**Figure 2-Figure supplement 4**). In
829 time-resolved phylogenies, branching among identical genomes is resolved by sampling date,
830 and the x-coordinate of the internal node is inferred based on time information. Because the
831 internal node location is based not on actual genetic information, there is variability in the
832 estimated placement of the internal node on the tree, resulting in branch lengths that may vary
833 among realizations of the tree. This is reflected in the 95% confidence interval of internal node
834 dates. Plotting the estimated branch length from each Washington tip to its internal node and
835 incorporating the 95% confidence interval of the internal node date shows that a wide range of
836 branch lengths are plausible for most tips (**Figure 9**). This complicates setting a simple branch
837 length threshold based on the serial interval. In contrast, divergence trees have branch lengths
838 expressed in the number of mutations arising along on that branch, which is intrinsic to the
839 sequences themselves. Secondly, higher transmission in one group could result in shorter serial
840 intervals within that group, which complicates defining a branch length cutoff based on serial
841 interval. For these reasons, we have opted to use genetic divergence as our metric of
842 “closeness” to avoid arbitrary time cutoffs and issues of uncertainty in timetree internal node
843 placements.

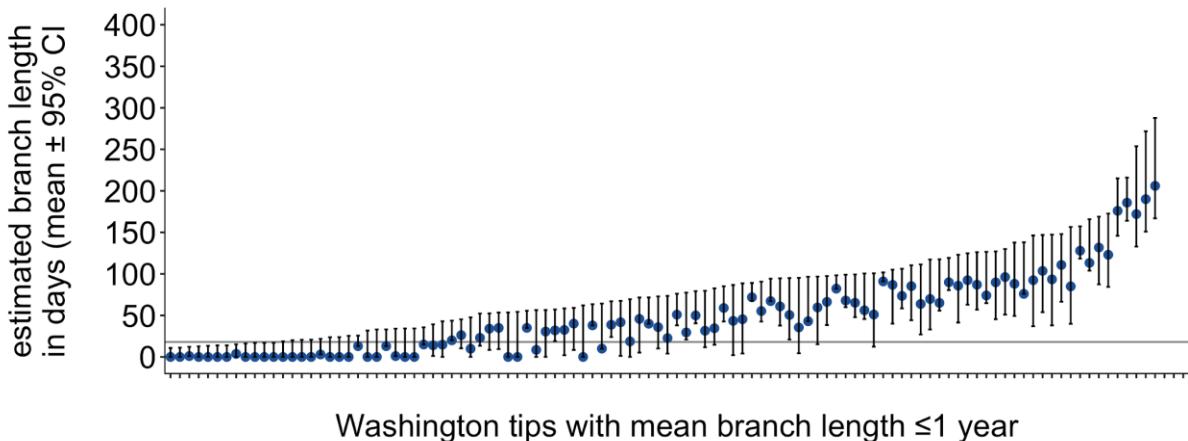


Figure 9: Washington tip branch lengths in days

For each Washington tip in the full North American phylogeny with an estimated branch length in time units of ≤ 1 year, we show the estimated branch length in days with the 95% confidence interval. The solid line at 18 represents the mump serial interval. For most tips, the estimated branch length is variable, depending on the placement of its parental internal node. This variability in internal node placement complicates setting a clear threshold for branch lengths based on time.

852

Given a divergence tree, we next categorized tips by how close they are to their parental node. To maximize the similarity between the tip and its inferred ancestral node, we classified tips as “close” to their ancestral node if no mutations occurred on the branch leading to that tip, i.e., the branch length was less than 1 divided by the alignment length. This cutoff can be set to the lower bound of the tree software (for IQTree, this cutoff is $\sim < 1 \times 10^{-16}$) with the same results. Using the JSON for the North American full genome mumps tree output from the Nextstrain pipeline (**shown in Figure 2-Figure supplement 2**) (Hadfield et al., 2018), we traversed the tree from root to tip. We collapsed very small branches (branches with no mutations) to obtain polytomies, and then classified tips as either “basal” (i.e., there were tips in the tree that descended from that internal node) or “terminal” (meaning that no sampled tips descended from that branch). Here, we define a “descendant” tip as a tip that occurs in any downstream portion of the tree, i.e., it falls within the same lineage but to the right of the parent tip. A diagram of what we classify as basal vs. terminal tips is shown in **Figure 4a**.

866
867 We expect that requiring branches to have 0 mutations should be robust regardless of mutation
868 rate and serial interval, because a branch length of 0 will always be the closest in sequence to
869 the true ancestor. However, variation in the substitution rate will impact the power of the
870 analysis for detecting associations. Because mumps has a low substitution rate, some internal
871 nodes contain stacks of identical genomes that cannot be ordered in terms of their placement
872 along the underlying transmission chain. Instead, we treat each of these nodes as basal with
873 equal probability of being upstream in the transmission chain. A higher substitution rate would
874 jitter these polytomies and increase resolution, while a lower substitution rate would further
875 reduce power. Our application is therefore conservative, but likely underpowered. Future work
876 will be necessary to define the precise interaction between mutation rate, serial interval,
877 sampling intensity and effect sizes in determining the power of this test.

878

879 **Regression model for quantifying a tip's probability of being basal**

880 For each Washington tip in the tree, we classified it as either being basal (coded as a 1) or
881 being terminal (coded as 0). For each tip, we coded its corresponding age, vaccination status,
882 and community membership as a predictor variable input into a logistic regression model. We
883 coded these attributes as follows: For community membership, non-Marshallese tips were
884 coded as 0 and Marshallese tips were coded as 1. For age, we split our dataset into adults and
885 children, with individuals aged <20 coded with a 0 and ≥20 coded with a 1. In our dataset, there
886 were 3 classifications for vaccination status: up-to-date, not up-to-date, and unknown
887 vaccination status. According to the Advisory Committee on Immunization Practices (ACIP)
888 (McLean et al., 2013), individuals aged 5-18 had to have received both recommended doses of
889 mumps-containing vaccine, children aged 15 months to 5 years required 1 dose of mumps-
890 containing vaccine, and adults over 18 had to have received at least 1 dose of mumps-
891 containing vaccine to be classified as up-to-date for mumps vaccination. Individuals under 15

892 months are considered up-to-date without any doses of mumps-containing vaccine. Not up-to-
893 date individuals are those with a known vaccination status who did not qualify under criteria to
894 be classified as up-to-date. Individuals who could not provide documentation regarding their
895 MMR vaccination history were considered to have “unknown” vaccination status. Individuals
896 with “known” vaccination status could either be fully up-to-date, undervaccinated, or
897 unvaccinated. To ensure that we measured the effect of vaccination among individuals who
898 knew their vaccination status, we coded vaccination information using two dummy variables in
899 our logistic regression, one signifying whether vaccination status was known or not, and one
900 indicating whether vaccination was up-to-date or not. We then fit a logistic regression model to
901 this data using the `glm` package in R

902 (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm>), specifying a
903 binomial model as

904

905 $\text{Pr}(\text{being basal}) \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$, where x_1 represents 0 or 1 value for member of
906 Marshallese community (Not Marshallese coded as a 0, Marshallese coded as a 1), x_2
907 represents a 0 or 1 value for age, where individuals were classified as adults (≥ 20 years of age,
908 coded as a 1) or children (< 20 years of age, coded as a 0). x_3 represents a 0 or 1 value for
909 whether vaccination status is unknown (having a known vaccination status coded as a 0, having
910 an unknown vaccination status coded as a 1) and x_4 represents 0 or 1 value for whether
911 vaccination status is up-to-date (up-to-date coded as a 0 and not up-to-date coded as a 1).

912 Under this formulation, an individual with unknown vaccination status would be coded as $x_3=1$,
913 $x_4=0$, an individual who is up-to-date would be coded as $x_3=0$, $x_4=0$, and an individual who is not
914 up-to-date is coded as $x_3=0$, $x_4=1$. This encoding allows us to evaluate the effects of having an
915 unknown vaccination status and a vaccination status that is not up-to-date.

916

917 P-values were assigned via a Wald test, and inferred coefficients were exponentiated to return
918 odds ratios. All code used to parse the divergence tree and formulate and fit the regression
919 model are available at <https://github.com/blab/mumps-wa->
920 [phylogenetics/blob/master/divergence-tree-analyses/Regression-analysis-on-descendants-in-](#)
921 [divergence-tree.ipynb](#).

922

923 **Rarefaction analysis to estimate transmission clusters**

924 Using the full set of North American mumps sequences, we designated all non-Washington
925 North American sequences as “background” sequences. We then separated Washington
926 sequences into Marshallese tips (57 total sequences) and non-Marshallese tips (52 total
927 sequences). For this analysis, we excluded the genotype K sequence in our dataset due to its
928 extreme divergence from other viruses sampled in Washington, which were all genotype G. For
929 each group (Marshallese vs. non-Marshallese), we then generated subsampled datasets
930 comprised of a random sample of 1 to n sequences, where n is the number of total sequences
931 available for that group. For each number of sequences, we performed 10 independent
932 subsampling trials. Subsampling was performed without replacement. So, for community
933 members, we generated 10 datasets in which 1 community member sequence was sampled,
934 then 10 datasets in which 2 community members sequences were sampled, etc. up to 10
935 datasets in which all 57 community members sequences were sampled. For each subsampled
936 dataset, we then combined these subsampled datasets with the background North American
937 sequences, and reran the Nextstrain pipeline. For each subsample and trial, we infer
938 geographic transmission history across the tree and enumerate the number of introductions into
939 Washington. Geographic transmission history was inferred using a discrete trait model in
940 TreeTime(Sagulenko et al., 2018). For each number of sequences tested, n , we report the
941 number of trials resulting in that number of inferred introductions, and the mean number of
942 inferred introductions across the 10 trials. Each resulting “cluster” consisted of a set of

943 sequences that are related to one another that descend from a single inferred introduction of
944 mumps into Washington.

945

946 **Inference of community transmission dynamics using a structured coalescent model**

947 To infer the rates of migration between community and non-community members and to infer
948 ancestral states of Washington internal nodes, we employed a structured coalescent model.
949 Here, “state” refers to the inferred ancestral identity of an internal node, where the identity could
950 be inferred as “Marshalllese” or “not Marshalllese”. The multitype tree model (Vaughan et al.,
951 2014) in BEAST 2 v2.6.2 (Bouckaert et al., 2019) infers the effective population sizes of each
952 deme and the migration rates between them. Because the multitype tree model requires that all
953 partitions contain all demes, we could only analyze 4 clades that circulated in Washington State
954 and included both Marshalllese and non-Marshalllese tips. We generated an XML in BEAUTi
955 v2.6.2 with 4 partitions, and linked the clock, site, and migration models. We used a strict, fixed
956 clock, set to 4.17×10^{-4} substitutions per site year, and used an HKY substitution model with 4
957 gamma-distributed rate categories. This clock rate was set based on the inferred substitutions
958 per site per year from all North American mumps genomes on nextstrain.org/mumps/na.
959 Migration rates were inferred with the prior specified as a truncated exponential distribution with
960 a mean of 1 and a maximum of 50. Effective population sizes were inferred with the prior
961 specified as a truncated exponential distribution with a mean of 1, a minimum value of 0.001,
962 and a maximum value of 10,000. All other priors were left at default values. In order to improve
963 convergence, we employed 3 heated chains using the package CoupledMCMC (Müller and
964 Bouckaert, 2019), where proposals for chains to swap were performed every 100 states. The
965 analysis was run for 100 million steps, with states sampled every 1 million steps. We ran this
966 analysis 3 independent times, and combined log and tree file output from those independent
967 runs using LogCombiner, with the first 10% (1000 states) of each run discarded as burnin. We
968 then summarized these combined output log and tree files. A maximum clade credibility tree

969 was inferred using TreeAnnotator with the mean heights option. To ensure that results were not
970 appreciably altered by the migration rate prior, we also repeated these analyses with migration
971 rates inferred with the prior specified as a truncated exponential distribution with a mean of 10
972 and a maximum of 50.

973

974 Although our complete dataset contains approximately equal numbers of sequences from
975 Marshallese and non-Marshallese cases, the 4 clusters analyzed above are enriched among
976 Marshallese tips. To assess the impact of uneven sampling within these clusters on ancestral
977 state inference, we performed a subsampling analysis. For each cluster, we subsampled down
978 the number of Marshallese tips to be equal to the number of non-Marshallese tips, and reran the
979 analysis as above. While the original analysis used 4 subclades containing both Marshallese
980 and non-Marshallese tips, one of these subclades only has 5 tips. Subsampling this particular
981 subtree would have resulted in a subtree with only 2 tips, thus we excluded this clade from the
982 subsampling analysis. For this sensitivity analysis, the 3 subsampled datasets had the following
983 tip composition: primary outbreak clade: 26 Marshallese and 26 non-Marshallese tips; 10-tip
984 introduction: 3 Marshallese and 3 non-Marshallese tips; 8-tip introduction: 4 Marshallese and 4
985 non-Marshallese tips. We generated 3 randomly subsampled datasets, and for each one ran 3
986 independent chains, with each chain run for 50 million steps, sampling every 500,000. For one
987 of the subsampled datasets, none of the chains converged after 20 days. In each of the
988 remaining 2 subsampled datasets, 2 out of 3 chains converged. We combined these converged
989 chains using LogCombiner, with the first 10% of each run discarded as burn-in. We then
990 summarized these combined output log and tree files, and inferred a maximum clade credibility
991 tree using TreeAnnotator with the mean heights option.

992

993 The analysis as described above assumes that each introduction into Washington State is an
994 independent observation of the same structured coalescent process, and that the dataset

995 represents a random sample of the underlying population. Additionally, this approach requires a
996 *priori* definition of which sequences are part of the same Washington State transmission chain.
997 Finally, the above analysis could only make use of the 4 Washington introductions with both
998 Marshallese and non-Marshallese tips, and excludes other transmission chains. Because of
999 these issues, we supplemented the above approach with an additional analysis using the
1000 approximate structured coalescent (Müller et al., 2017) in MASCOT (Müller et al., 2018). Using
1001 all of the Washington sequences, we specified three demes: Marshallese in Washington, non-
1002 Marshallese in Washington, and outside of Washington. To account for any transmission that
1003 happened outside of Washington State, the “outside of Washington” deme acted as a “ghost
1004 deme” from which we did not use any samples. The effective population size of this “outside of
1005 Washington” deme then describes the rate at which lineages between any location outside of
1006 Washington share a common ancestor. Including specific samples from outside of Washington
1007 would bias the inferred effective population size towards the coalescent rates of the sampled
1008 locations, by incorporating local transmission dynamics of other locations. We then estimated
1009 migration rates and effective population sizes for all 3 demes, but fixed the migration rates such
1010 that the unsampled deme (“outside of Washington”) could only act as a source population. This
1011 is motivated by not having observed obvious migration out of Washington State in our previous
1012 analysis here. We ran this analysis for 10 million steps, sampling every 5000, and discarded the
1013 first 10% of states as burnin.

1014

1015 **References**

- 1016 Abella MKIL, Molina MR, Nikolić-Hughes I, Hughes EW, Ruderman MA. 2019. Background
1017 gamma radiation and soil activity measurements in the northern Marshall Islands. *Proc Natl
1018 Acad Sci U S A* **116**:15425–15434. doi:10.1073/pnas.1903421116
1019 Adams WH, Fields HA, Engle JR, Hadler SC. 1986. Serologic markers for hepatitis B among
1020 Marshallese accidentally exposed to fallout radiation in 1954. *Radiat Res* **108**:74–79.
1021 Albertson JP, Clegg WJ, Reid HD, Arbise BS, Pryde J, Vaid A, Thompson-Brown R, Echols F.
1022 2016. Mumps Outbreak at a University and Recommendation for a Third Dose of Measles-

- 1023 Mumps-Rubella Vaccine - Illinois, 2015-2016. *MMWR Morb Mortal Wkly Rep* **65**:731–734.
1024 doi:10.15585/mmwr.mm6529a2
- 1025 Arday DR, Kanjarpane DD, Kelley PW. 1989. Mumps in the US Army 1980-86: should recruits
1026 be immunized? *Am J Public Health* **79**:471–474. doi:10.2105/ajph.79.4.471
- 1027 Barker HM. 2012. Bravo for the Marshallese: Regaining Control in a Post-Nuclear, Post-Colonial
1028 World. Cengage Learning.
- 1029 Barskey AE, Schulte C, Rosen JB, Handschur EF, Rausch-Phung E, Doll MK, Cummings KP,
1030 Alleyne EO, High P, Lawler J, Apostolou A, Blog D, Zimmerman CM, Montana B, Harpaz R,
1031 Hickman CJ, Rota PA, Rota JS, Bellini WJ, Gallagher KM. 2012. Mumps outbreak in
1032 Orthodox Jewish communities in the United States. *N Engl J Med* **367**:1704–1713.
1033 doi:10.1056/NEJMoa1202865
- 1034 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
1035 data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170
- 1036 Bonwitt J, Kawakami V, Wharton A, Burke RM, Murthy N, Lee A, Dell B, Kay M, Duchin J,
1037 Hickman C, McNall RJ, Rota PA, Patel M, Lindquist S, DeBolt C, Routh J. 2017. Notes from
1038 the Field: Absence of Asymptomatic Mumps Virus Shedding Among Vaccinated College
1039 Students During a Mumps Outbreak - Washington, February-June 2017. *MMWR Morb
1040 Mortal Wkly Rep* **66**:1307–1308. doi:10.15585/mmwr.mm6647a5
- 1041 Bordner AS, Crosswell DA, Katz AO, Shah JT, Zhang CR, Nikolic-Hughes I, Hughes EW,
1042 Ruderman MA. 2016. Measurement of background gamma radiation in the northern
1043 Marshall Islands. *Proc Natl Acad Sci U S A* **113**:6833–6838. doi:10.1073/pnas.1605535113
- 1044 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fournent M, Gavryushkina A, Heled
1045 J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du
1046 Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu C-H, Xie D,
1047 Zhang C, Stadler T, Drummond AJ. 2019. BEAST 2.5: An advanced software platform for
1048 Bayesian evolutionary analysis. *PLoS Comput Biol* **15**:e1006650.
1049 doi:10.1371/journal.pcbi.1006650
- 1050 Cardemil CV, Dahl RM, James L, Wannemuehler K, Gary HE, Shah M, Marin M, Riley J, Feikin
1051 DR, Patel M, Quinlisk P. 2017. Effectiveness of a Third Dose of MMR Vaccine for Mumps
1052 Outbreak Control. *N Engl J Med* **377**:947–956. doi:10.1056/NEJMoa1703309
- 1053 CDCMMWR. 2019. Notifiable Diseases and Mortality Tables.
- 1054 Center KE, Da Silva J, Hernandez AL, Vang K, Martin DW, Mazurek J, Lilo EA, Zimmerman NK,
1055 Krow-Lucal E, Campbell EM, Cowins JV, Walker C, Dominguez KL, Gallo B, Gunn JKL,
1056 McCormick D, Cochran C, Smith MR, Dillaha JA, James AE. 2020. Multidisciplinary
1057 Community-Based Investigation of a COVID-19 Outbreak Among Marshallese and
1058 Hispanic/Latino Communities - Benton and Washington Counties, Arkansas, March-June
1059 2020. *MMWR Morb Mortal Wkly Rep* **69**:1807–1811. doi:10.15585/mmwr.mm6948a2
- 1060 Centers for Disease Control and Prevention. 2019a. Health Departments: How to Optimize
1061 Mumps Testing | CDC. *Guidance for Optimizing Mumps Testing*.
1062 <https://www.cdc.gov/mumps/health-departments/optimize-testing.html>
- 1063 Centers for Disease Control and Prevention. 2019b. Epidemiology and Prevention of Vaccine-
1064 Preventable Diseases, The Pink Book: Course Textbook. Centers for Disease Control and
1065 Prevention.
- 1066 Clemmons N, Hickman C, Lee A, Mona Marin M, Patel M. 2020. Manual for the Surveillance of
1067 Vaccine-Preventable Diseases. Centers for Disease Control and Prevention.
- 1068 Congress 108th United States. 2003. Compact of Free Association Amendments Act of 2003.
- 1069 Davidkin I, Jokinen S, Broman M, Leinikki P, Peltola H. 2008. Persistence of measles, mumps,
1070 and rubella antibodies in an MMR-vaccinated cohort: a 20-year follow-up. *J Infect Dis*
1071 **197**:950–956. doi:10.1086/528993
- 1072 De Maio N, Wu CH, O'Reilly KM, Wilson D. 2015. New Routes to Phylogeography: A Bayesian
1073 Structured Coalescent Approximation. *PLoS Genet*. doi:10.1371/journal.pgen.1005421

- 1074 Donahue M, Schneider A, Ukegbu U, Shah M, Riley J, Weigel A, James L, Wittich K, Quinlisk P,
1075 Cardemil C. 2017. Notes from the Field: Complications of Mumps During a University
1076 Outbreak Among Students Who Had Received 2 Doses of Measles-Mumps-Rubella
1077 Vaccine - Iowa, July 2015-May 2016. *MMWR Morb Mortal Wkly Rep* **66**:390–391.
1078 doi:10.15585/mmwr.mm6614a4
- 1079 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and
1080 the BEAST 1.7. *Mol Biol Evol* **29**:1969–1973. doi:10.1093/molbev/mss075
- 1081 Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A,
1082 Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Di Caro A, Diclaro
1083 JW, Duraffour S, Elmore MJ, Fakoli LS, Faye O, Gilbert ML, Gevao SM, Gire S, Gladden-
1084 Young A, Gnrke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kugelman JR, Liu
1085 D, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J,
1086 Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF,
1087 Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, Stoecker K, Thorne L, Tobin EA,
1088 Vandi MA, Watson SJ, West K, Whitmer S, Wiley MR, Winnicki SM, Wohl S, Wölfel R,
1089 Yozwiak NL, Andersen KG, Blyden SO, Bolay F, Carroll MW, Dahn B, Diallo B, Formenty
1090 P, Fraser C, Gao GF, Garry RF, Goodfellow I, Günther S, Happi CT, Holmes EC, Kargbo B,
1091 Keïta S, Kellam P, Koopmans MPG, Kuhn JH, Loman NJ, Magassouba N 'faly, Naidoo D,
1092 Nichol ST, Nyenswah T, Palacios G, Pybus OG, Sabeti PC, Sall A, Ströher U, Wurie I,
1093 Suchard MA, Lemey P, Rambaut A. 2017. Virus genomes reveal factors that spread and
1094 sustained the Ebola epidemic. *Nature* **544**:309–315. doi:10.1038/nature22040
- 1095 Dudas G, Carvalho LM, Rambaut A, Bedford T. 2018. MERS-CoV spillover at the camel-human
1096 interface. *eLife*. doi:10.7554/eLife.31257
- 1097 Duke MR. 2017. Neocolonialism and Health Care Access among Marshall Islanders in the
1098 United States. *Med Anthropol Q* **31**:422–439. doi:10.1111/maq.12376
- 1099 Eick AA, Hu Z, Wang Z, Nevin RL. 2008. Incidence of mumps and immunity to measles, mumps
1100 and rubella among US military recruits, 2000-2004. *Vaccine* **26**:494–501.
1101 doi:10.1016/j.vaccine.2007.11.035
- 1102 Embassy of the Republic of the Marshall Islands to the United States of America. n.d. Embassy
1103 of the Republic of the Marshall Islands to the United States of America. *Embassy of the*
1104 *Republic of the Marshall Islands to the United States of America*.
1105 http://www.rmiembassyus.org/index.php/about_marshall-islands/culture
- 1106 Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black
1107 A, da Costa AC, Franco LC, Silva SP, Wu C-H, Raghwani J, Cauchemez S, du Plessis L,
1108 Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques
1109 CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Muñoz-
1110 Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO,
1111 Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de
1112 Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC,
1113 Dhalia R, França RF, Magalhães T, Marques ET Jr, Jaenisch T, Wallau GL, de Lima MC,
1114 Nascimento V, de Cerqueira EM, de Lima MM, Mascarenhas DL, Neto JPM, Levin AS,
1115 Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado A, Holmes
1116 EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ, Pybus OG.
1117 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas.
1118 *Nature* **546**:406–410. doi:10.1038/nature22401
- 1119 Fields VS, Safi H, Waters C, Dillaha J, Capelle L, Riklon S, Wheeler JG, Haselow DT. 2019.
1120 Mumps in a highly vaccinated Marshallese community in Arkansas, USA: an outbreak
1121 report. *Lancet Infect Dis* **19**:185–192. doi:10.1016/S1473-3099(18)30607-8
- 1122 Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. 2015. Eight challenges in
1123 phylodynamic inference. *Epidemics* **10**:88–92. doi:10.1016/j.epidem.2014.09.001
- 1124 Gavilán AM, Fernández-García A, Rueda A, Castellanos A, Masa-Calles J, López-Perea N,

- 1125 Torres de Mier MV, de Ory F, Echevarría JE. 2018. Genomic non-coding regions reveal
1126 hidden patterns of mumps virus circulation in Spain, 2005 to 2015. *Euro Surveill* **23**.
1127 doi:10.2807/1560-7917.ES.2018.23.15.17-00349
- 1128 Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled
1129 ancestor trees for epidemiology and fossil calibration. *PLoS Comput Biol* **10**:e1003919.
1130 doi:10.1371/journal.pcbi.1003919
- 1131 Golwalkar M, Pope B, Stauffer J, Snively A, Clemons N. 2018. Mumps Outbreaks at Four
1132 Universities - Indiana, 2016. *MMWR Morb Mortal Wkly Rep* **67**:793–797.
1133 doi:10.15585/mmwr.mm6729a1
- 1134 Gouma S, Cremer J, Parkkali S, Veldhuijzen I, van Binnendijk RS, Koopmans MPG. 2016.
1135 Mumps virus F gene and HN gene sequencing as a molecular tool to study mumps virus
1136 transmission. *Infect Genet Evol* **45**:145–150. doi:10.1016/j.meegid.2016.08.033
- 1137 Green CB. 2006. Updated Vaccine Guidance for Adults and Accessions-June 2006.
- 1138 Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, Wiley MR, White
1139 S, Thézé J, Magnani DM, Prieto K, Reyes D, Bingham AM, Paul LM, Robles-Sikisaka R,
1140 Oliveira G, Pronty D, Barcellona CM, Metsky HC, Baniecki ML, Barnes KG, Chak B, Freije
1141 CA, Gladden-Young A, Gnrke A, Luo C, MacInnis B, Matranga CB, Park DJ, Qu J,
1142 Schaffner SF, Tomkins-Tinch C, West KL, Winnicki SM, Wohl S, Yozwiak NL, Quick J,
1143 Fauver JR, Khan K, Brent SE, Reiner RC Jr, Lichtenberger PN, Ricciardi MJ, Bailey VK,
1144 Watkins DI, Cone MR, Kopp EW 4th, Hogan KN, Cannons AC, Jean R, Monaghan AJ,
1145 Garry RF, Loman NJ, Faria NR, Porcelli MC, Vasquez C, Nagle ER, Cummings DAT,
1146 Stanek D, Rambaut A, Sanchez-Lockhart M, Sabeti PC, Gillis LD, Michael SF, Bedford T,
1147 Pybus OG, Isern S, Palacios G, Andersen KG. 2017. Genomic epidemiology reveals
1148 multiple introductions of Zika virus into the United States. *Nature* **546**:401–405.
1149 doi:10.1038/nature22400
- 1150 Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T,
1151 Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*
1152 **34**:4121–4123. doi:10.1093/bioinformatics/bty407
- 1153 Hallgren EA, McElfish PA, Rubon-Chutaro J. 2015. Barriers and opportunities: a community-
1154 based participatory research study of health beliefs related to diabetes in a US Marshallese
1155 community. *Diabetes Educ* **41**:86–94. doi:10.1177/0145721714559131
- 1156 Hall MD, Woolhouse MEJ, Rambaut A. 2016. The effects of sampling strategy on the quality of
1157 reconstruction of viral population dynamics using Bayesian skyline family coalescent
1158 methods: A simulation study. *Virus Evol* **2**:vew003. doi:10.1093/ve/vew003
- 1159 Harris PM, Jones NA. 2005. We the People: Pacific Islanders in the United States. US
1160 Department of Commerce.
- 1161 Hirono MK. 2019. Covering our FAS Allies Act.
- 1162 Katoh K, Misawa K, Kuma KK-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple
1163 sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066.
1164 doi:10.1093/nar/gkf436
- 1165 Kelley PW, Petruccelli BP, Stehr-Green P, Erickson RL, Mason CJ. 1991. The susceptibility of
1166 young adult Americans to vaccine-preventable infections. A national serosurvey of US
1167 Army recruits. *JAMA* **266**:2724–2729. doi:10.1001/jama.1991.03470190072032
- 1168 Kühnert D, Wu C-H, Drummond AJ. 2011. Phylogenetic and epidemic modeling of rapidly
1169 evolving infectious diseases. *Infect Genet Evol* **11**:1825–1841.
1170 doi:10.1016/j.meegid.2011.08.005
- 1171 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
1172 **9**:357–359. doi:10.1038/nmeth.1923
- 1173 Lemey P, Hong SL, Hill V, Baele G, Poletto C, Colizza V, O'Toole Á, McCrone JT, Andersen
1174 KG, Worobey M, Nelson MI, Rambaut A, Suchard MA. 2020. Accommodating individual
1175 travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-

- 1176 CoV-2. *Nat Commun* **11**:5110. doi:10.1038/s41467-020-18877-9
- 1177 Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its
1178 roots. *PLoS Comput Biol*. doi:10.1371/journal.pcbi.1000520
- 1179 Lewnard JA, Grad YH. 2018. Vaccine waning and mumps re-emergence in the United States.
1180 *Sci Transl Med* **10**. doi:10.1126/scitranslmed.aaa5945
- 1181 Lo NC, Nyathi S, Chapman LAC, Rodriguez-Barraquer I, Kushel M, Bibbins-Domingo K,
1182 Lewnard JA. 2021. Influenza, Varicella, and Mumps Outbreaks in US Migrant Detention
1183 Centers. *JAMA* **325**:180–182. doi:10.1001/jama.2020.20539
- 1184 Marlow MA, Marin M, Moore K, Patel M. 2020. CDC Guidance for Use of a Third Dose of MMR
1185 Vaccine During Mumps Outbreaks. *J Public Health Manag Pract* **26**:109–115.
1186 doi:10.1097/PHH.00000000000000962
- 1187 Marx GE, Burakoff A, Barnes M, Hite D, Metz A, Miller K, Davizon ES, Chase J, McDonald C,
1188 McClean M, Miller L, Albanese BA. 2018. Mumps Outbreak in a Marshallese Community -
1189 Denver Metropolitan Area, Colorado, 2016-2017. *MMWR Morb Mortal Wkly Rep* **67**:1143–
1190 1146. doi:10.15585/mmwr.mm6741a2
- 1191 McElfish PA. 2016. Marshallese COFA Migrants in Arkansas. *J Ark Med Soc* **112**:259–60, 262.
- 1192 McElfish PA, Chughtai A, Low LK, Garner R, Purvis RS. 2018. “Just doing the best we can”:
1193 health care providers’ perceptions of barriers to providing care to Marshallese patients in
1194 Arkansas. *Ethn Health* 1–14. doi:10.1080/13557858.2018.1471670
- 1195 McElfish PA, Hallgren E, Yamada S. 2015. Effect of US health policies on health care access
1196 for Marshallese migrants. *Am J Public Health* **105**:637–643.
1197 doi:10.2105/AJPH.2014.302452
- 1198 McElfish PA, Purvis RS, Maskarinec GG, Bing WI, Jacob CJ, Ritok-Lakien M, Rubon-Chutaro J,
1199 Lang S, Mamis S, Riklon S. 2016. Interpretive policy analysis: Marshallese COFA migrants
1200 and the Affordable Care Act. *Int J Equity Health* **15**:91. doi:10.1186/s12939-016-0381-1
- 1201 McElfish PA, Purvis R, Willis DE, Riklon S. 2021. COVID-19 Disparities Among Marshallese
1202 Pacific Islanders. *Prev Chronic Dis* **18**:E02. doi:10.5888/pcd18.200407
- 1203 McElfish PA, Rowland B, Long CR, Hudson J, Piel M, Buron B, Riklon S, Bing WI, Warmack TS.
1204 2017. Diabetes and Hypertension in Marshallese Adults: Results from Faith-Based Health
1205 Screenings. *J Racial Ethn Health Disparities* **4**:1042–1050. doi:10.1007/s40615-016-0308-y
- 1206 McLean HQ, Amy Parker Fiebelkorn MSN, Temte JL, Wallace GS. 2013. Prevention of
1207 Measles, Rubella, Congenital Rubella Syndrome, and Mumps, 2013.
1208 <https://www.cdc.gov/mmwr/preview/mmwrhtml/rr6204a1.htm>
- 1209 Müller NF, Bouckaert R. 2019. Coupled MCMC in BEAST 2. *bioRxiv*. doi:10.1101/603514
- 1210 Müller NF, Rasmussen DA, Stadler T. 2017. The Structured Coalescent and Its Approximations.
1211 *Molecular Biology and Evolution*. doi:10.1093/molbev/msx186
- 1212 Müller NF, Rasmussen D, Stadler T. 2018. MASCOT: parameter and state inference under the
1213 marginal structured coalescent approximation. *Bioinformatics*.
1214 doi:10.1093/bioinformatics/bty406
- 1215 Nelson GE, Aguon A, Valencia E, Oliva R, Guerrero ML, Reyes R, Lizama A, Diras D, Mathew
1216 A, Camacho EJ, Monforte M-N, Chen T-H, Mahamud A, Kutty PK, Hickman C, Bellini WJ,
1217 Seward JF, Gallagher K, Fiebelkorn AP. 2013. Epidemiology of a mumps outbreak in a
1218 highly vaccinated island population and use of a third dose of measles-mumps-rubella
1219 vaccine for outbreak control--Guam 2009 to 2010. *Pediatr Infect Dis J* **32**:374–380.
1220 doi:10.1097/INF.0b013e318279f593
- 1221 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
1222 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–
1223 274. doi:10.1093/molbev/msu300
- 1224 Niedenthal J. 1997. A history of the people of Bikini following nuclear weapons testing in the
1225 Marshall Islands: with recollections and views of elders of Bikini Atoll. *Health Phys* **73**:28–
1226 36. doi:10.1097/00004032-199707000-00003

- 1227 Nomenclature MV, Des virus N. n.d. Weekly epidemiological record *Relevé épidémiologique*
1228 hebdomadaire.
- 1229 Palafox NA, Riklon S, Alik W, Hixon AL. 2007. Health consequences and health systems
1230 response to the Pacific U.S. Nuclear Weapons Testing Program. *Pac Health Dialog*
1231 **14**:170–178.
- 1232 Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S,
1233 Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH. 2012. ViPR: an open
1234 bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*
1235 **40**:D593–8. doi:10.1093/nar/gkr859
- 1236 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of
1237 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**:vew007.
1238 doi:10.1093/ve/vew007
- 1239 Sagulenko P, Puller V, Neher RA. 2018. TreeTime: Maximum-likelihood phylodynamic analysis.
1240 *Virus Evolution* **4**. doi:10.1093/ve/vex042
- 1241 Shah M, Quinlisk P, Weigel A, Riley J, James L, Patterson J, Hickman C, Rota PA, Stewart R,
1242 Clemmons N, Kalas N, Cardemil C, Iowa Mumps Outbreak Response Team. 2018. Mumps
1243 Outbreak in a Highly Vaccinated University-Affiliated Setting Before and After a Measles-
1244 Mumps-Rubella Vaccination Campaign-Iowa, July 2015-May 2016. *Clin Infect Dis* **66**:81–
1245 88. doi:10.1093/cid/cix718
- 1246 Simon SL. 1997. A brief history of people and events related to atomic weapons testing in the
1247 Marshall Islands. *Health Phys* **73**:5–20. doi:10.1097/00004032-199707000-00001
- 1248 Simon SL, Bouville A, Melo D, Beck HL, Weinstock RM. 2010. Acute and chronic intakes of
1249 fallout radionuclides by Marshallese from nuclear weapons testing at Bikini and Enewetak
1250 and related internal radiation doses. *Health Phys* **99**:157–200.
1251 doi:10.1097/HP.0b013e3181dc4e51
- 1252 Snijders BEP, van Lier A, van de Kassteele J, Fanoy EB, Ruijs WLM, Hulshof F, Blauwhof A,
1253 Schipper M, van Binnendijk R, Boot HJ, de Melker HE, Hahné SJM. 2012. Mumps vaccine
1254 effectiveness in primary schools and households, the Netherlands, 2008. *Vaccine* **30**:2999–
1255 3002. doi:10.1016/j.vaccine.2012.02.035
- 1256 Stack JC, Welch JD, Ferrari MJ, Shapiro BU, Grenfell BT. 2010. Protocols for sampling viral
1257 sequences to study epidemic dynamics. *J R Soc Interface* **7**:1119–1127.
1258 doi:10.1098/rsif.2009.0530
- 1259 Stapleton PJ, Eshaghi A, Seo CY, Wilson S, Harris T, Deeks SL, Bolotin S, Goneau LW,
1260 Gubbay JB, Patel SN. 2019. Evaluating the use of whole genome sequencing for the
1261 investigation of a large mumps outbreak in Ontario, Canada. *Sci Rep* **9**:12615.
1262 doi:10.1038/s41598-019-47740-1
- 1263 Takahashi T, Trott KR, Fujimori K, Simon SL, Ohtomo H, Nakashima N, Takaya K, Kimura N,
1264 Satomi S, Schoemaker MJ. 1997. An investigation into the prevalence of thyroid disease on
1265 Kwajalein Atoll, Marshall Islands. *Health Phys* **73**:199–213. doi:10.1097/00004032-
1266 199707000-00017
- 1267 Towne SD, Yeary KHK, Narcisse M-R, Long C, Bursac Z, Totaram R, Rodriguez EM, McElfish
1268 P. 2020. Inequities in Access to Medical Care Among Adults Diagnosed with Diabetes:
1269 Comparisons Between the US Population and a Sample of US-Residing Marshallese
1270 Islanders. *J Racial Ethn Health Disparities*. doi:10.1007/s40615-020-00791-x
- 1271 US Census Bureau. n.d. Historical Households Tables.
- 1272 Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ. 2014. Efficient Bayesian
1273 inference under the structured coalescent. *Bioinformatics*.
1274 doi:10.1093/bioinformatics/btu201
- 1275 Vink MA, Bootsma MCJ, Wallinga J. 2014. Serial intervals of respiratory infectious diseases: a
1276 systematic review and analysis. *Am J Epidemiol* **180**:865–875. doi:10.1093/aje/kwu209
- 1277 Volz EM. 2012. Complex population dynamics and the coalescent under neutrality. *Genetics*

1278 **190**:187–201. doi:10.1534/genetics.111.134627
1279 Washington State Legislature. 2014. Chapter 246-101 WAC. *Chapter 246-101 WAC Notifiable*
1280 *Conditions*. <https://app.leg.wa.gov/wac/default.aspx?cite=246-101>
1281 Williams DP, Hampton A. 2005. Barriers to health services perceived by Marshallese
1282 immigrants. *J Immigr Health* **7**:317–326. doi:10.1007/s10903-005-5129-8
1283 Wohl S, Metsky HC, Schaffner SF, Piantadosi A, Burns M, Lewnard JA, Chak B, Krasilnikova
1284 LA, Siddle KJ, Matranga CB, Bankamp B, Hennigan S, Sabina B, Byrne EH, McNall RJ,
1285 Shah RR, Qu J, Park DJ, Gharib S, Fitzgerald S, Barreira P, Fleming S, Lett S, Rota PA,
1286 Madoff LC, Yozwiak NL, MacInnis BL, Smole S, Grad YH, Sabeti PC. 2020. Combining
1287 genomics and epidemiology to track mumps virus transmission in the United States. *PLoS*
1288 *Biol* **18**:e3000611. doi:10.1371/journal.pbio.3000611
1289 Wong DC, Purcell RH, Rosen L. 1979. Prevalence of antibody to hepatitis A and hepatitis B
1290 viruses in selected populations of the South Pacific. *Am J Epidemiol* **110**:227–236.
1291 doi:10.1093/oxfordjournals.aje.a112807
1292 Yamada S, Dodd A, Soe T, Chen T-H, Bauman K. 2004. Diabetes mellitus prevalence in out-
1293 patient Marshallese adults on Ebeye Island, Republic of the Marshall Islands. *Hawaii Med J*
1294 **63**:45–51.

1295

1296 **Acknowledgments**

1297 We would like to extend our sincerest thank you to Jiji Jally for her help and input on the project.
1298 Jiji Jally is an advocate for affordable access to healthcare services, supportive services for the
1299 Marshallese community, and works as a translator to assist the community in Washington State.
1300 These insightful discussions were absolutely critical for contextualizing our results. We would
1301 also like to sincerely thank Kelsey Florek for locating and sharing mumps samples from
1302 Wisconsin, Ohio, Missouri, Alabama, and North Carolina, which greatly enhanced the analyses
1303 presented here. We also thank Jeff Joy for graciously sharing mumps genomes from British
1304 Columbia. Finally, we would like to thank the Fred Hutchinson Cancer Research Center
1305 sequencing core for providing excellent sequencing services and Fred Hutch Scientific
1306 Computing Infrastructure. LHM is an Open Philanthropy Project fellow of the Life Sciences
1307 Research Foundation. AB was supported by the National Science Foundation Graduate
1308 Research Fellowship Program under Grant No. DGE-1256082. TB is a Pew Biomedical Scholar
1309 and is supported by NIH R35 GM119774-01. Scientific Computing Infrastructure at Fred Hutch
1310 is supported by NIH ORIP S10OD028685.

1311

1312 **Supplemental figure legends**

1313 **Figure 2-figure supplement 1: Mumps genomes accumulate mutations linearly over time**

1314 We inferred a maximum likelihood phylogeny using IQTREE for all available complete mumps
1315 genomes of genotype G, sampled from North America between 2006 and 2018. We inferred the
1316 root-to-tip distance with TempEst and plot the root to tip divergence vs. sample collection date.
1317 Color represents geographic location (either Canadian province or US state), with colors
1318 corresponding to those in Figure 2. We infer that mumps genomes accumulate mutations at a
1319 rate of 3.75×10^{-4} substitutions per site per year.

1320

1321 **Figure 2-figure supplement 3: Placement of divergent Washington and non-genotype G**
1322 **genomes on a global phylogeny**

1323 To place the divergent Washington genomes and non-genotype G Washington and Wisconsin
1324 genomes in the context of global mumps diversity, we generated a full genome divergence
1325 phylogeny using all publicly available mumps genotypes collected from anywhere in the world.
1326 The branchpoint for genotype G viruses is marked with a black circle and annotated with text.
1327 Color indicates the geographic region from which the sample was collected and the x-axis
1328 represents substitutions per site. For ease of viewing, we have collapsed all sequences that fall
1329 within the main genotype G lineage of North American mumps that is shown in **Figure 2** into the
1330 blue triangle. This includes all sequences descending from the 2006 midwest outbreak
1331 sequences, but does not include the divergent Washington lineages shown at the top of **Figure**
1332 **2**. The genotype H and A genomes from Wisconsin are highlighted in callouts 1 and 2. The
1333 Washington genotype K sequence is shown in callout 3, while the nine divergent Washington
1334 genomes shown in **Figure 2** are highlighted here in callouts 4 and 5.

1335

1336 **Figure 2-figure supplement 4: The full genome divergence tree closely matches the time-**
1337 **resolve phylogeny**

1338 We inferred a maximum likelihood phylogeny using IQTREE for all available complete mumps
1339 genomes of genotype G, sampled from North America between 2006 and 2018. Color
1340 represents geographic location, and the x-axis displays divergence in substitutions per site per
1341 year. To reduce the number of displayed colors, we grouped US states by geography as
1342 follows: non-Washington West include California and Montana; Midwest USA includes North
1343 Dakota, Kansas, Missouri, Iowa, Wisconsin, Indiana, Michigan, Ohio, and Illinois; South USA
1344 includes North Carolina, Alabama, Virginia, Georgia, Texas, Arkansas, and Louisiana;
1345 Northeast USA includes New York, Massachusetts, Pennsylvania, New Hampshire, and New
1346 Jersey. Canadian provinces are also grouped by geographic area.

1347

1348 **Figure 2-figure supplement 5: SH gene sequences are inadequate for fine-scale**
1349 **resolution of mumps transmission**

1350 To compare whether we would recover similar tree topologies if we had only sequenced the SH
1351 gene, we downloaded all available complete mumps genomes of genotype G, sampled from
1352 North America between 2006 and 2018, and truncated our sequences to include only the coding
1353 region for SH. We then inferred a maximum likelihood phylogeny using the same procedure as
1354 in **Figure 2-figure supplement 2**. The vast majority of North American mumps sequences are
1355 identical and form a single polytomy, suggesting that SH sequencing alone provides limited
1356 resolution for inferring geographic spread. Color represents geographic location, and the x-axis
1357 displays divergence in substitutions per site per year. To reduce the number of displayed colors,
1358 we grouped US states by geography as follows: non-Washington West include California and
1359 Montana; Midwest USA includes North Dakota, Kansas, Missouri, Iowa, Wisconsin, Indiana,
1360 Michigan, Ohio, and Illinois; South USA includes North Carolina, Alabama, Virginia, Georgia,
1361 Texas, Arkansas, and Louisiana; Northeast USA includes New York, Massachusetts,

1362 Pennsylvania, New Hampshire, and New Jersey. Canadian provinces are also grouped by
1363 geographic area.

1364

1365 **Figure 4-figure supplement 1: Rarefaction results by vaccination status**

1366 We repeated the rarefaction analysis shown in **Figure 4b** for vaccination status. We separated
1367 all Washington tips and classified them by vaccination status into up-to-date, not up-to-date, or
1368 unknown vaccination status. We then performed a rarefaction analysis and plot the number of
1369 inferred Washington clusters (y-axis) as a function of the number of sequences included in the
1370 analysis (x-axis). Dark green represents unknown vaccination status, light green represents not
1371 up-to-date, and green represents up-to-date. The majority of sequences in our dataset were
1372 derived from individuals who were up-to-date for mumps vaccine. Each dot represents the
1373 number of trials in which that number of clusters was inferred, and the solid line represents the
1374 mean across trials.

1375

1376 **Figure 5-figure supplement 1: Inferences are similar under a higher migration rate prior**

1377 The results are shown for the exact same analyses displayed in **Figure 5**, except inferred under
1378 a model with a higher migration rate prior (mean of 10 instead of mean of 1). **a.** Using the 4
1379 Washington clusters that had a mixture of Marshallese and non-Marshallese cases, we inferred
1380 phylogenies using a structured coalescent model. Each group of sequences shared a clock
1381 model, migration model, and substitution model, but each topology was inferred separately,
1382 allowing us to incorporate information from all 4 clusters into the migration estimation. For each
1383 cluster, the maximum clade credibility tree is shown, where the color of each internal node
1384 represents the posterior probability that the node is Marshallese. **b.** For each internal node
1385 shown in panel **a**, we plot the posterior probability of that node being Marshallese. Across all 4
1386 clusters, almost every internal node is inferred as Marshallese with high probability. **c.** The
1387 posterior distribution of the number of “jumps” or transmission events from Marshallese to not

1388 Marshallese (light blue) and not Marshallese to Marshallese (dark blue) inferred for the primary
1389 outbreak clade.

1390

1391 **Figure 5-figure supplement 2: Structured coalescent analyses are robust to sampling
1392 differences**

1393 To ensure that our results were robust to differences in sampling of Marshallese and non-
1394 Marshallese tips within the clusters used for this analysis, we subsampled our dataset 3
1395 independent times, and ran 3 independent chains per unique subsampling. In each subsampled
1396 dataset, the number of Marshallese tips was randomly subsampled to be equal to the number of
1397 non-Marshallese tips in each of the 4 clusters. We then ran each of these subsampled datasets
1398 with the exact same model as run with the full dataset. In subsampled datasets 1 and 2, 2 out of
1399 3 chains converged, and results were combined and displayed here. In the 3rd subsampled
1400 dataset, none of the 3 chains converged, so those results are not shown. **a.** For each
1401 subsampled dataset, we plot the inferred maximum clade credibility tree from the combined tree
1402 outputs from the 2 converged chains. The color of each tip represents whether that sample was
1403 derived from a Marshallese or non-Marshallese case, and the color of the internal node
1404 represents the posterior probability of that internal node being Marshallese. **b.** For each tree
1405 shown in **a**, the posterior probability that each internal node is labelled as Marshallese is shown.
1406 The number of the subsampled dataset is shown on the x-axis and the posterior probability is
1407 shown on the y-axis.

1408

1409 **Figure 6-figure supplement 1: Posterior probabilities of internal node states**

1410 **a.** For the tree shown in **Figure 6**, each internal node is plotted. For each internal node, its color
1411 and placement on the x-axis represents its inferred most probable group (Marshallese, Not
1412 Marshallese, or Not Washington). The posterior probability of being labelled its most probable
1413 group is shown on the y-axis. We recover moderate support for a small number of non-

1414 Marshallese internal nodes, while the vast majority of internal nodes remain inferred as
1415 Marshallese. **b.** The 95% highest posterior density intervals of the inferred effective population
1416 sizes for Marshallese, non-Marshallese, and not Washington demes.

1417

1418 **Supplementary file legends**

1419 **Supplementary file 1a: Sample metadata**

1420 All genomes generated for this analysis are described above. Dates are formatted as Year-
1421 month-day. Vaccination status, Ct, and sample collection type are all available for the
1422 Washington samples. Genome coverage represents the total proportion of bases in the genome
1423 with at least 20x coverage for which we were able to call a base. Sites with <20x coverage were
1424 labelled as Ns. Only samples with at least 50% non-N bases were included in the analysis.

1425

1426 **Source data file legends**

1427 **Figure 1-source data 1: Washington State mumps case counts in 2016-2017**

1428

1429 **Figure 1-source data 2: Metadata for sequences generated in this manuscript with**
1430 **collection dates**

1431

1432 **Figure 2-source data 1: XML file to run discrete trait phylogeographic analysis of North**
1433 **American mumps transmission shown in Figure 2, with combined mcc tree and output**
1434 **log files**

1435

1436 **Figure 2-source data 2: Divergence trees with metadata for divergence trees shown in**
1437 **Figure 2-figure supplement 4 and Figure2-figure supplement 5**

1438

1439 **Figure 3-source data 1: Inferred introductions into Washington State across posterior**
1440 **distribution**

1441

1442 **Figure 4-source data 1: Rarefaction results for community status analysis shown in**
1443 **Figure 4b**

1444

1445 **Figure 4-source data 2: Rarefaction results for vaccination status analysis shown in**
1446 **Figure 4-figure supplement 1**

1447

1448 **Figure 5-source data 1: XML file to run structured coalescent analysis and combined**
1449 **output log and tree files with a migration rate prior of 1 (shown in Figure 5, identifiable**
1450 **metadata have been removed)**

1451

1452 **Figure 5-source data 2: XML file to run structured coalescent analysis and combined**
1453 **output log and tree files with a migration rate prior of 10 (shown in Figure 5-figure**
1454 **supplement 1, identifiable metadata have been removed)**

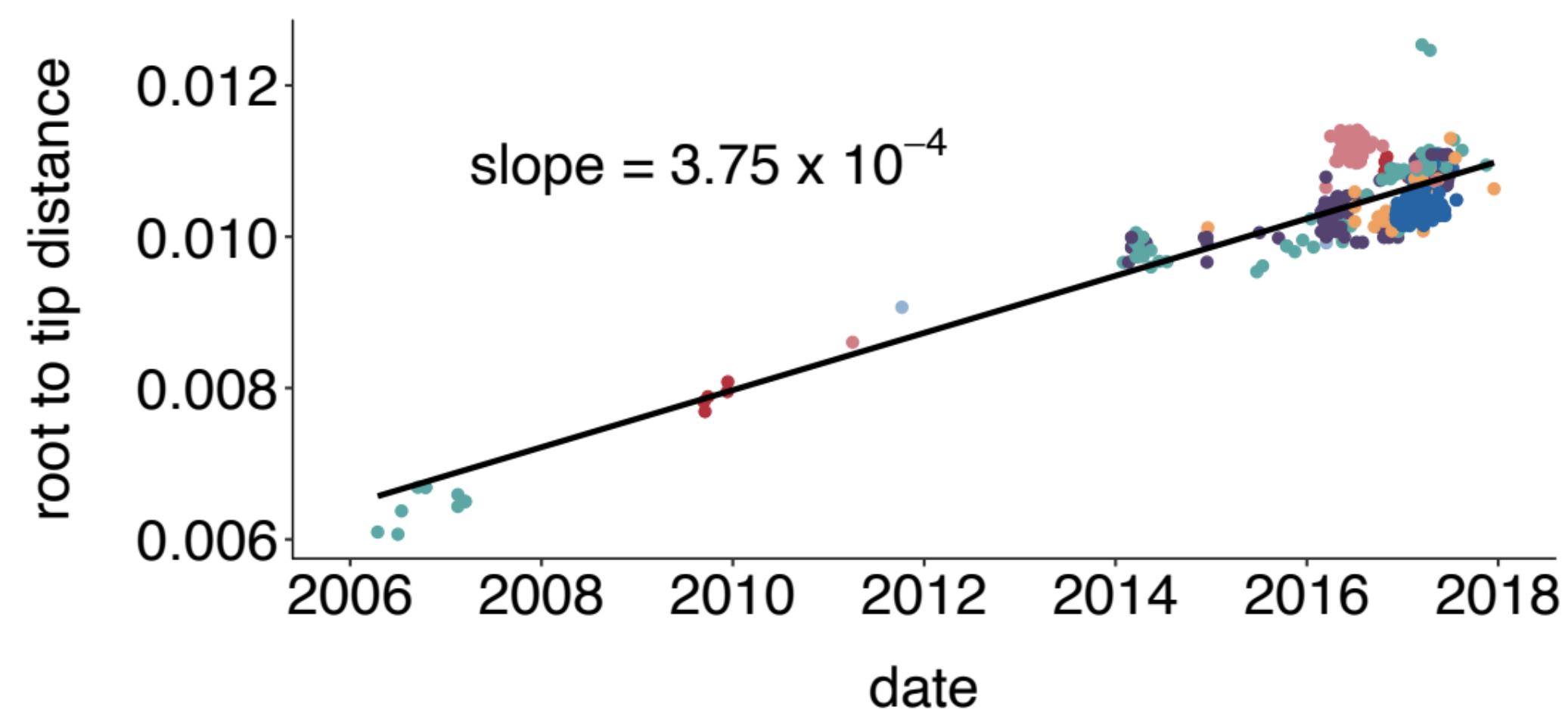
1455

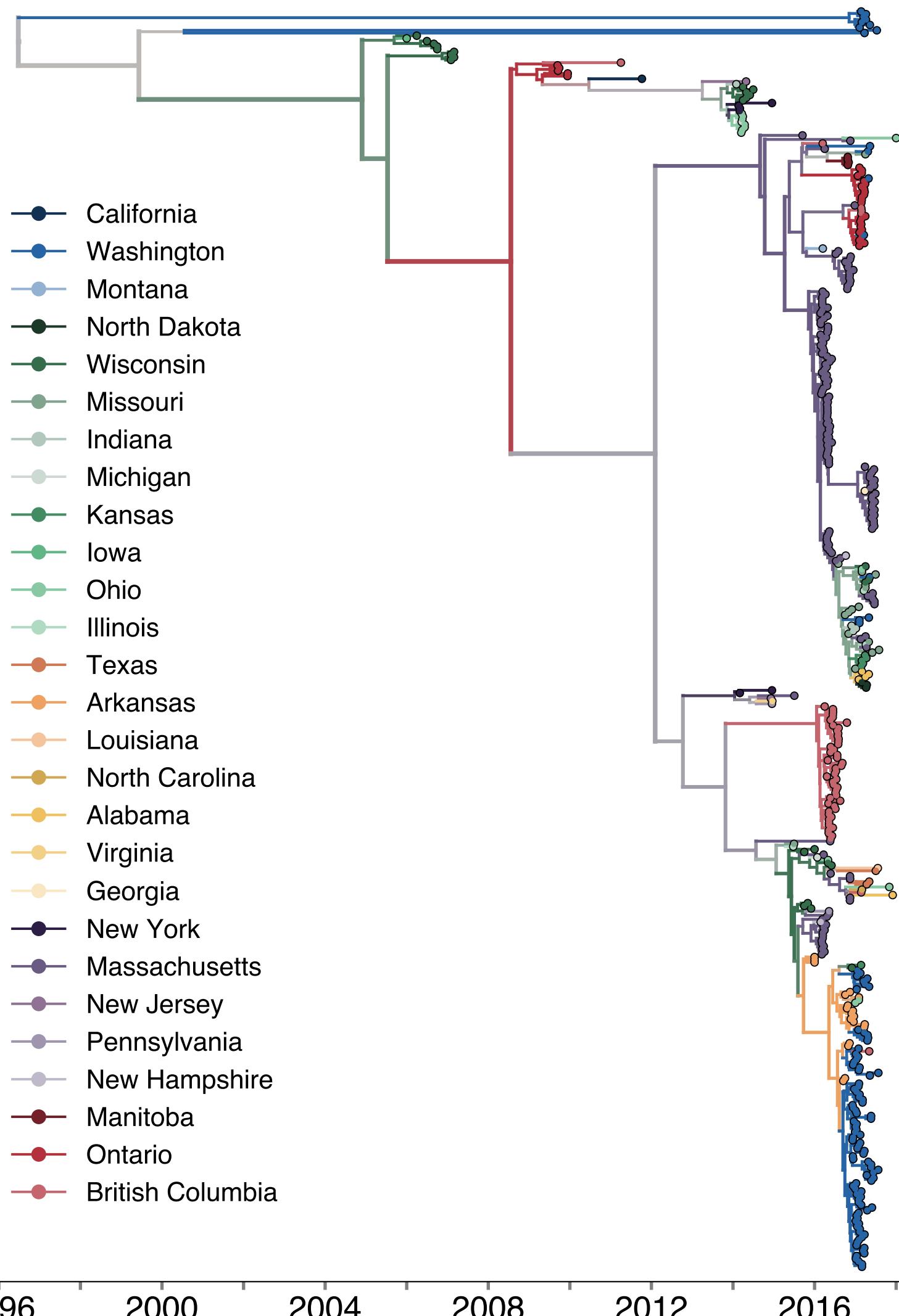
1456 **Figure 5-source data 3: XML files and combined output files to run structured coalescent**
1457 **analysis where clades were subsampled to have equal numbers of Marshallese and non-**
1458 **Marshallese tips. (shown in Figure 5-figure supplement 2, identifiable metadata have**
1459 **been removed)**

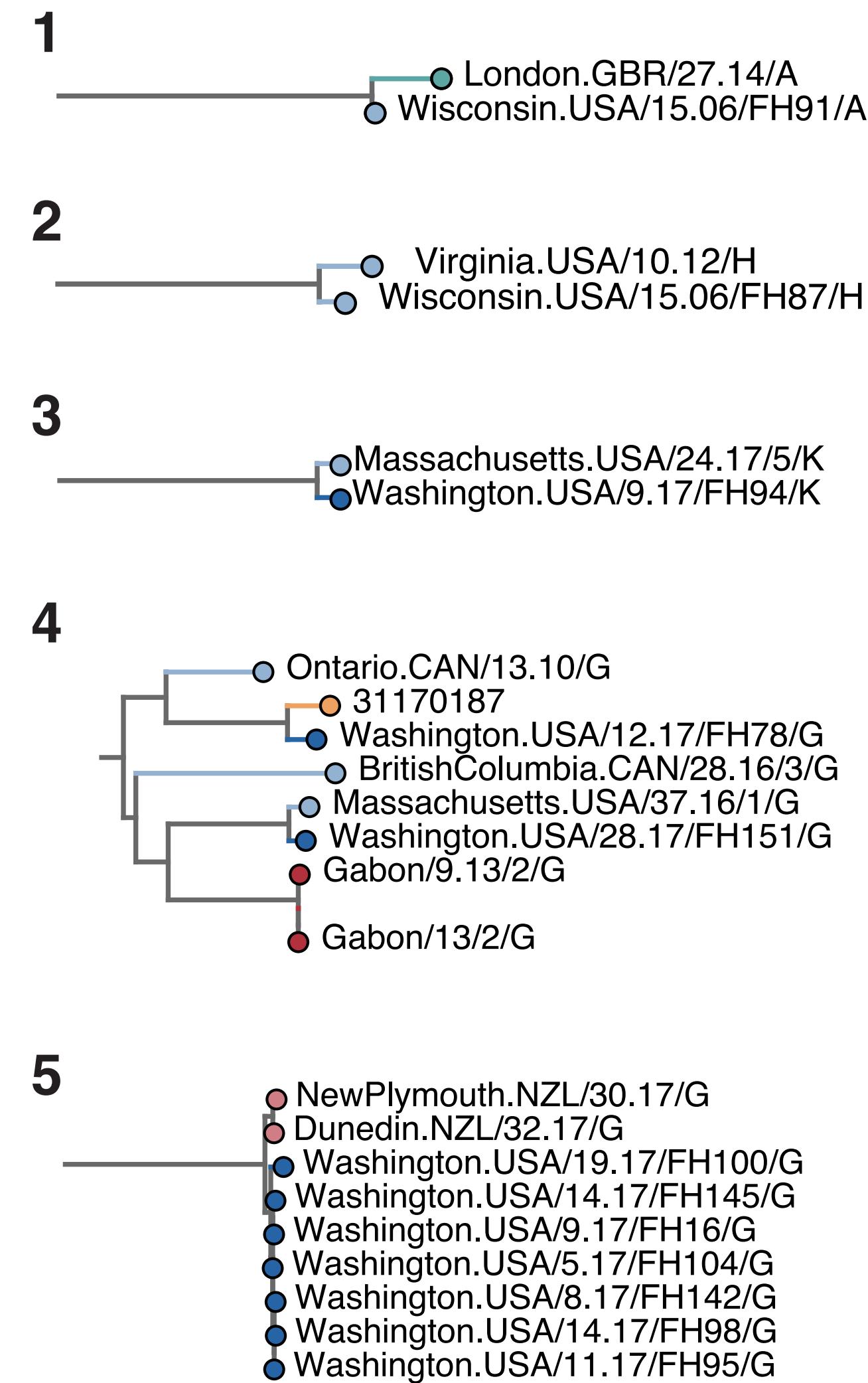
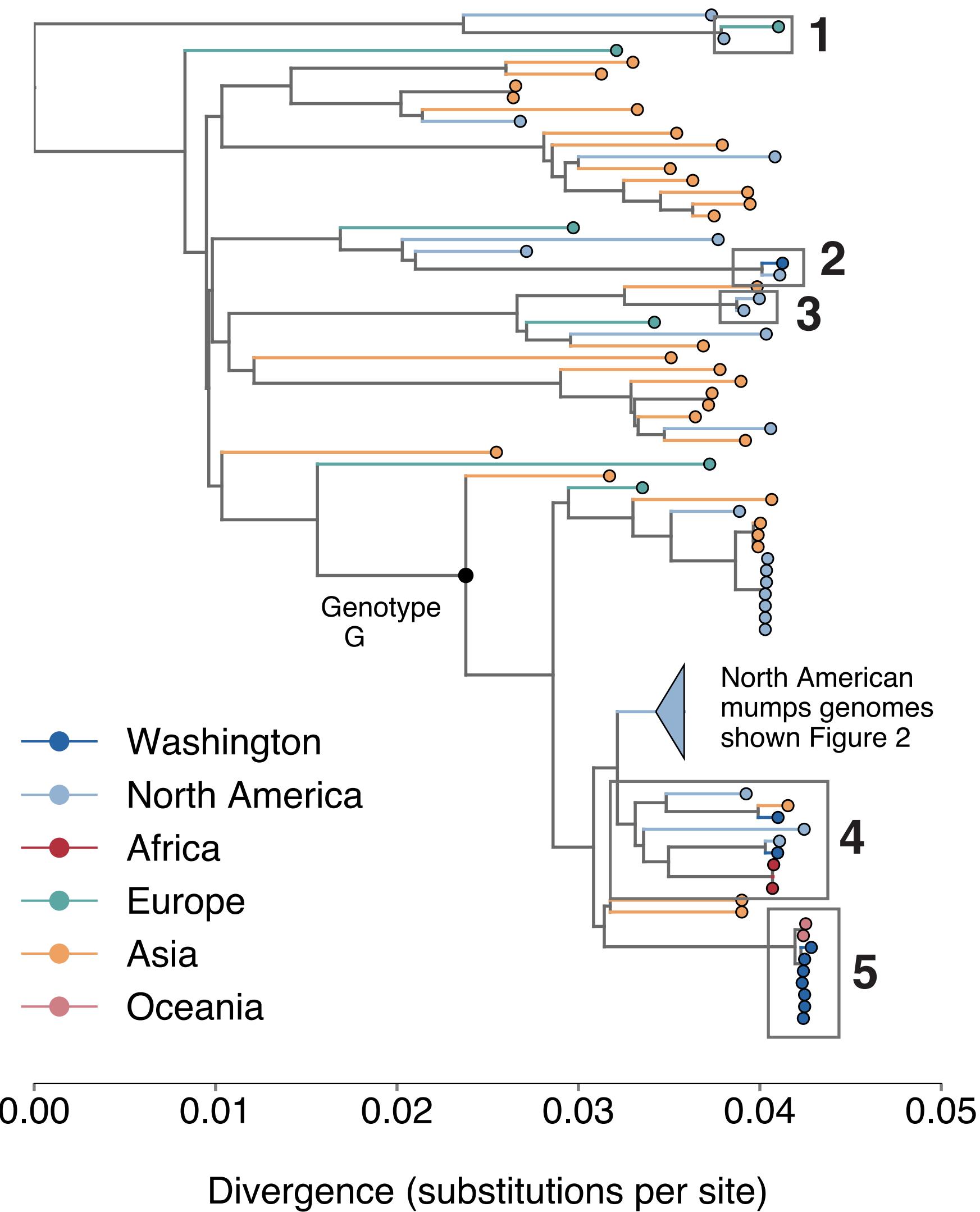
1460

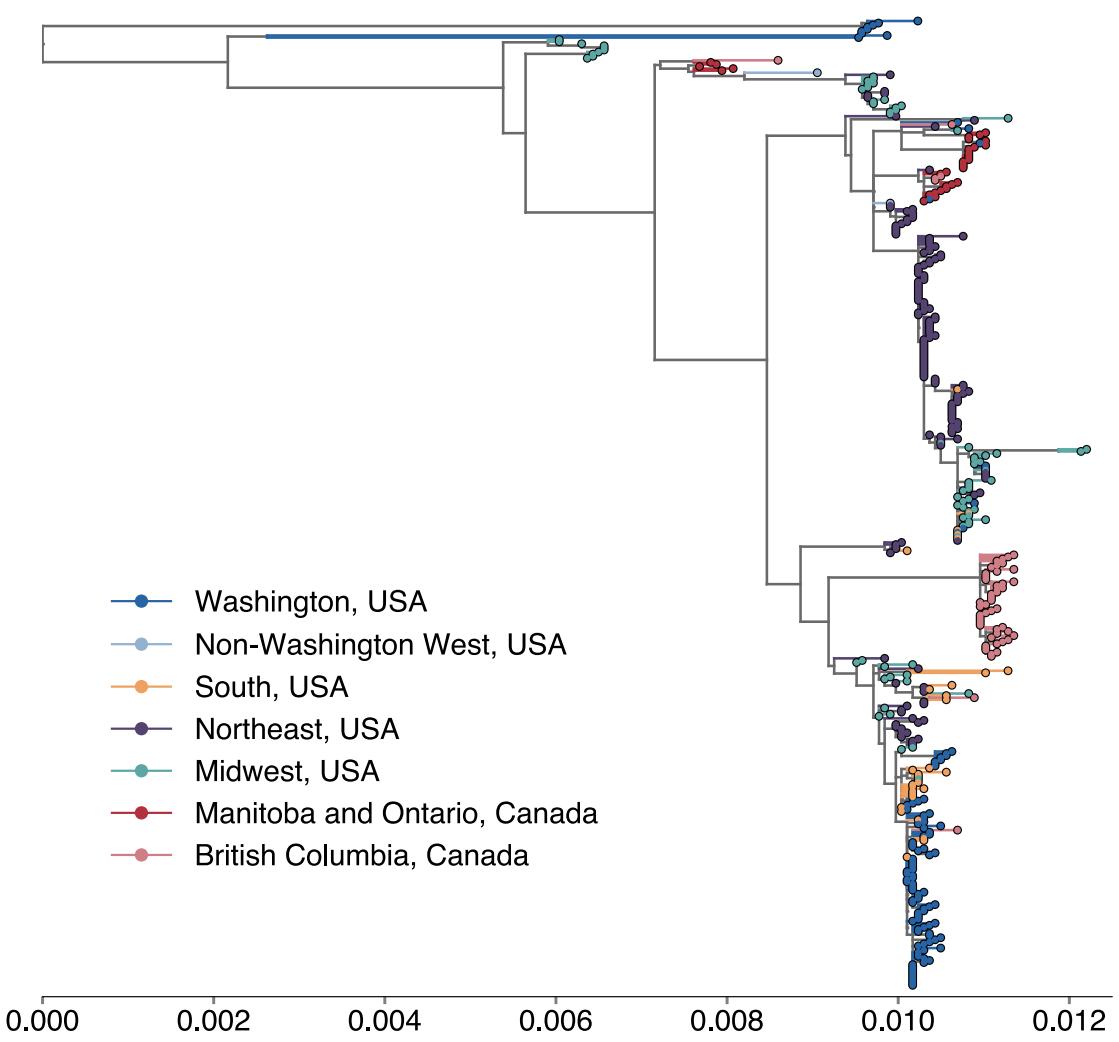
1461 **Figure 6-source data 1: XML file and output files to run structured coalescent analysis**
1462 **with unsampled “ghost” deme shown in Figure 6 (identifiable metadata have been**
1463 **removed)**

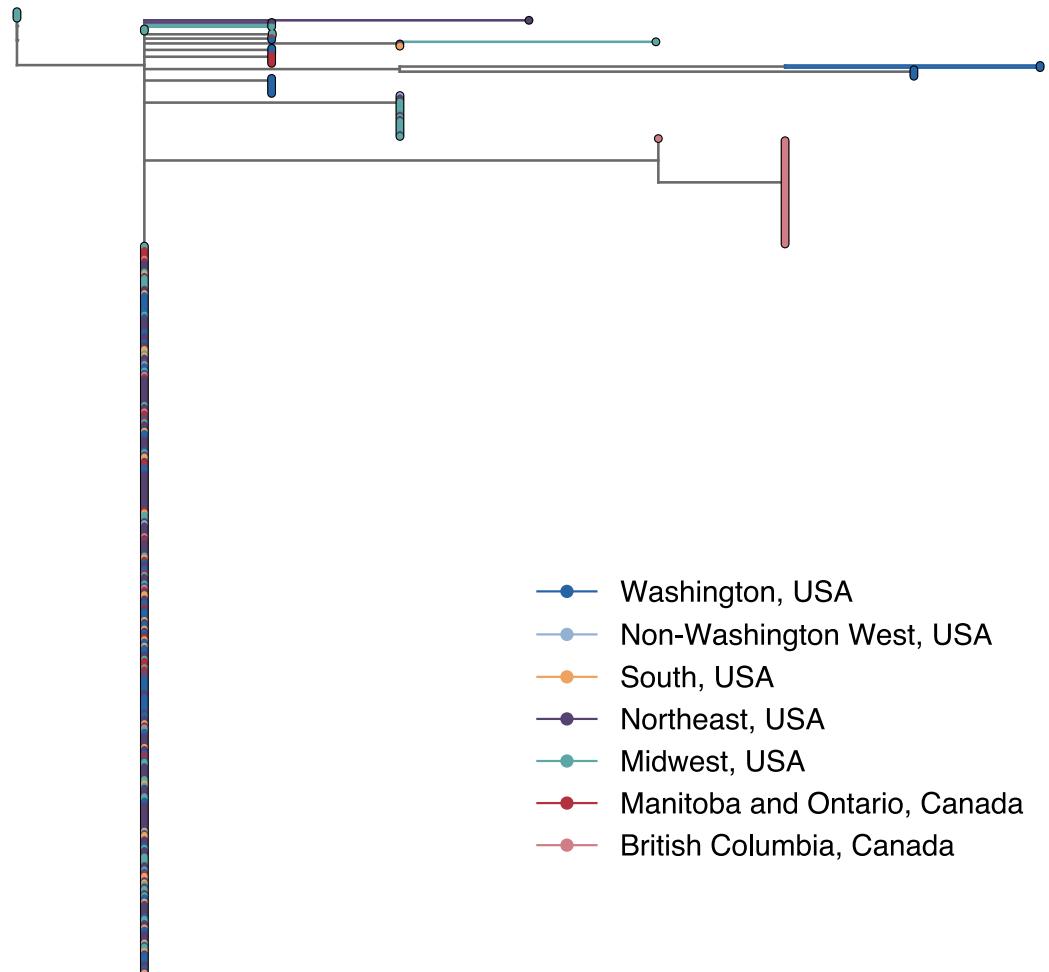
1464

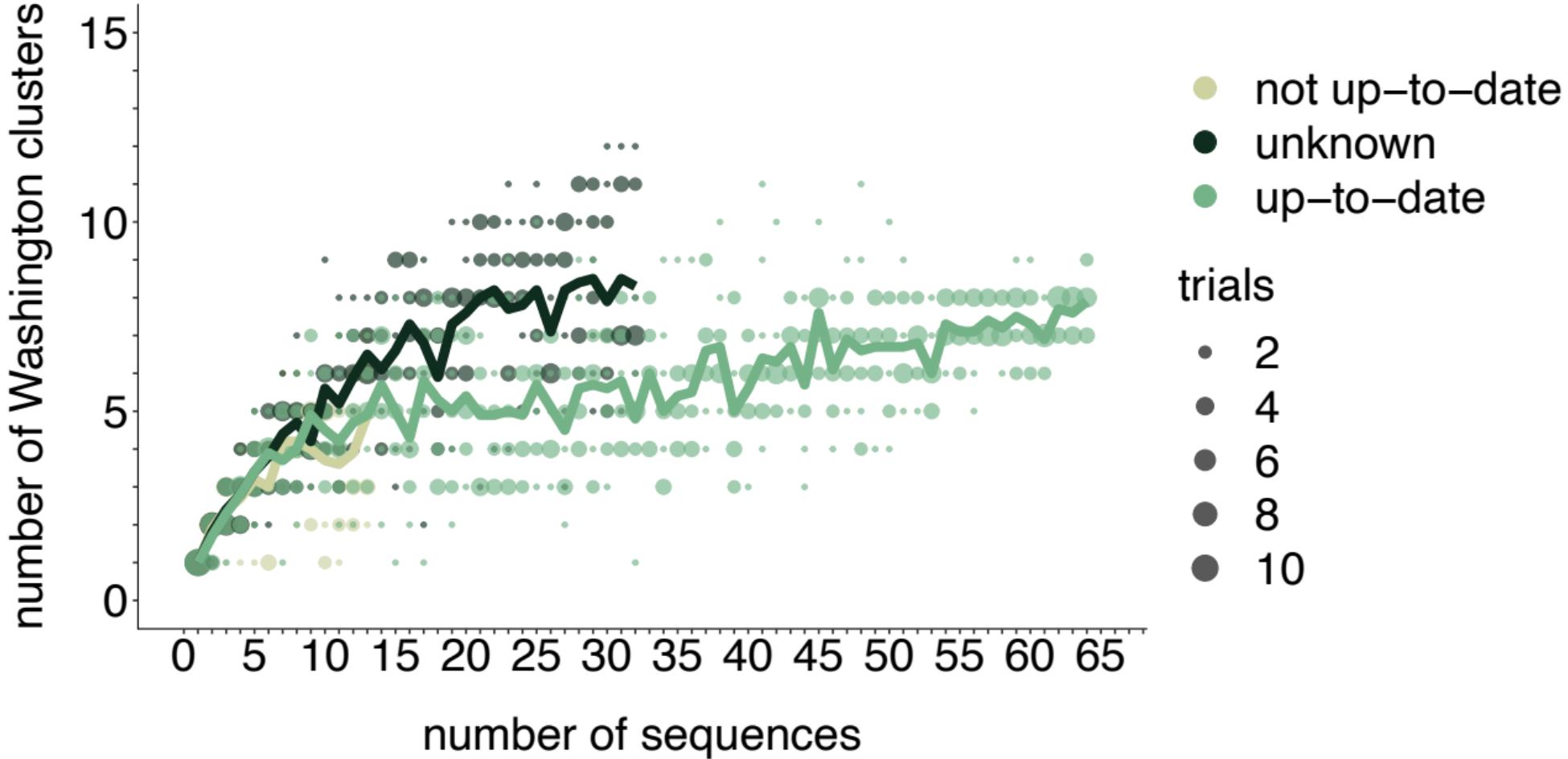


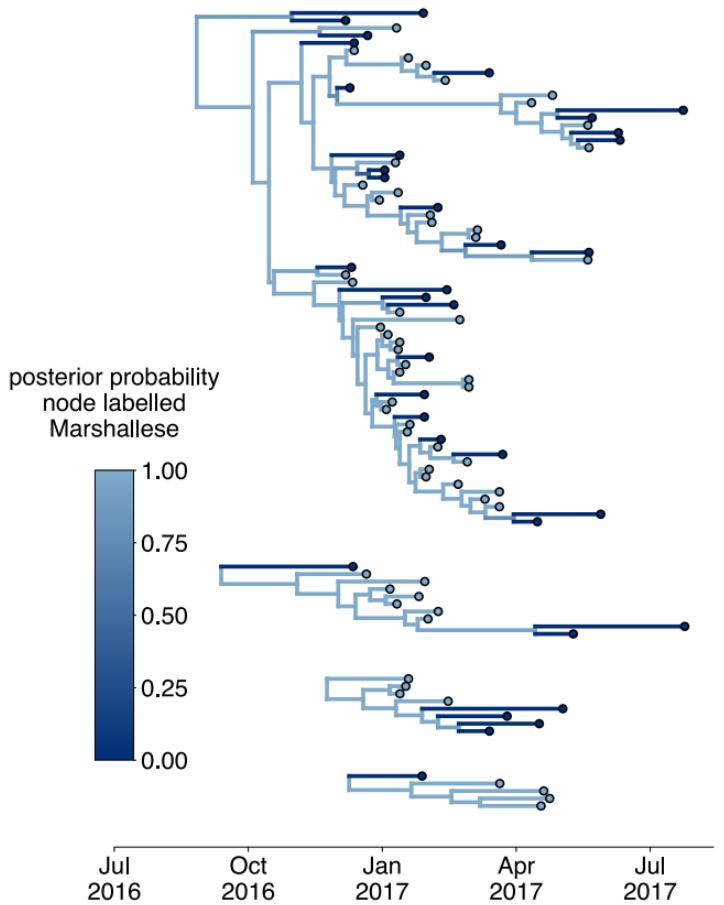
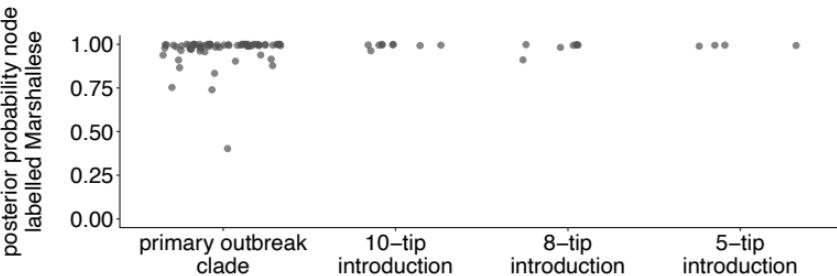
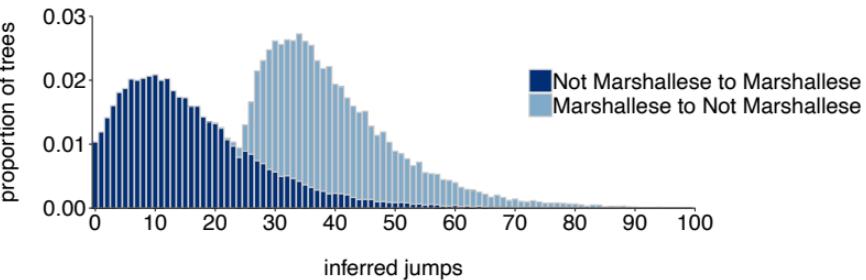


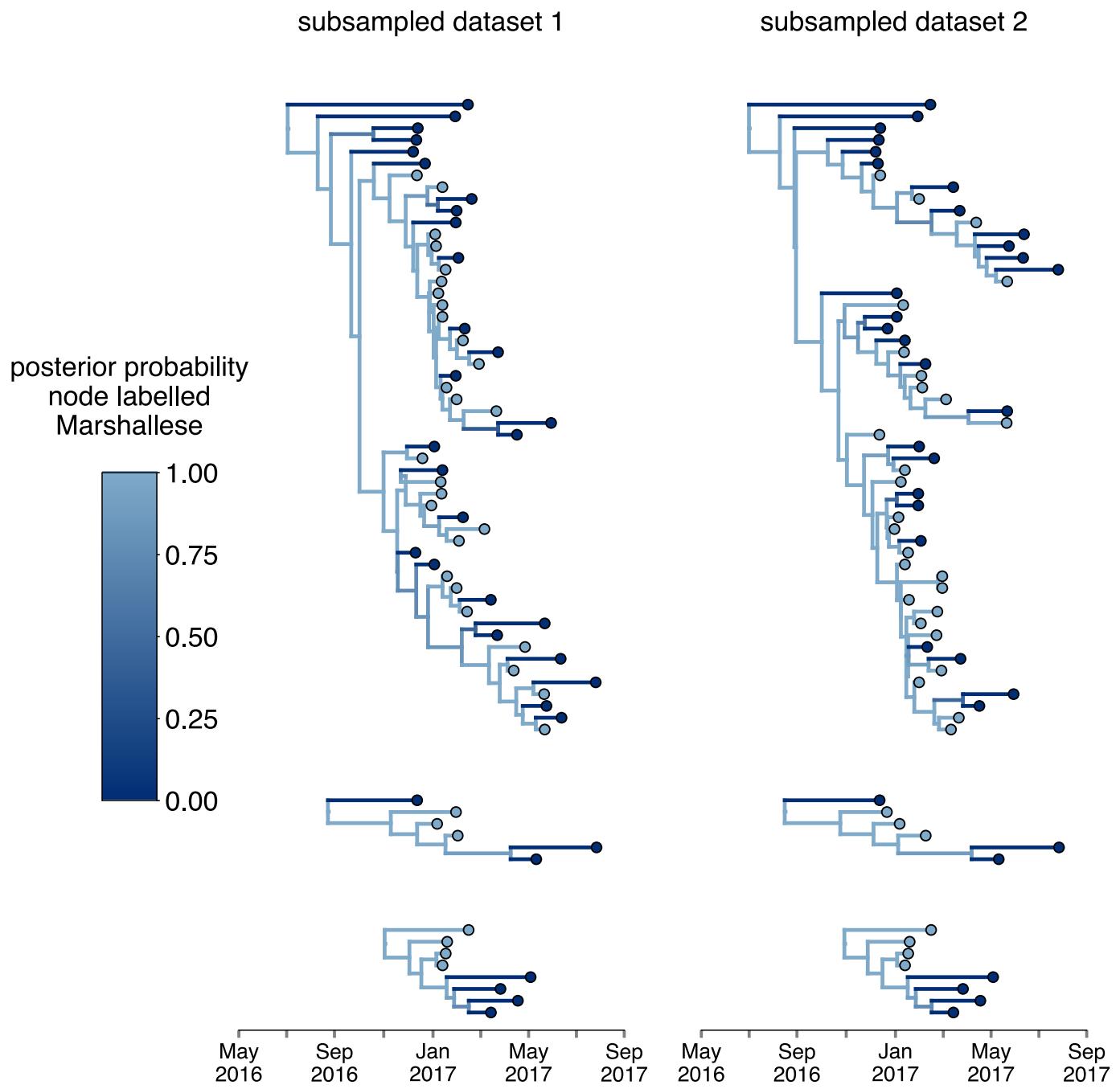
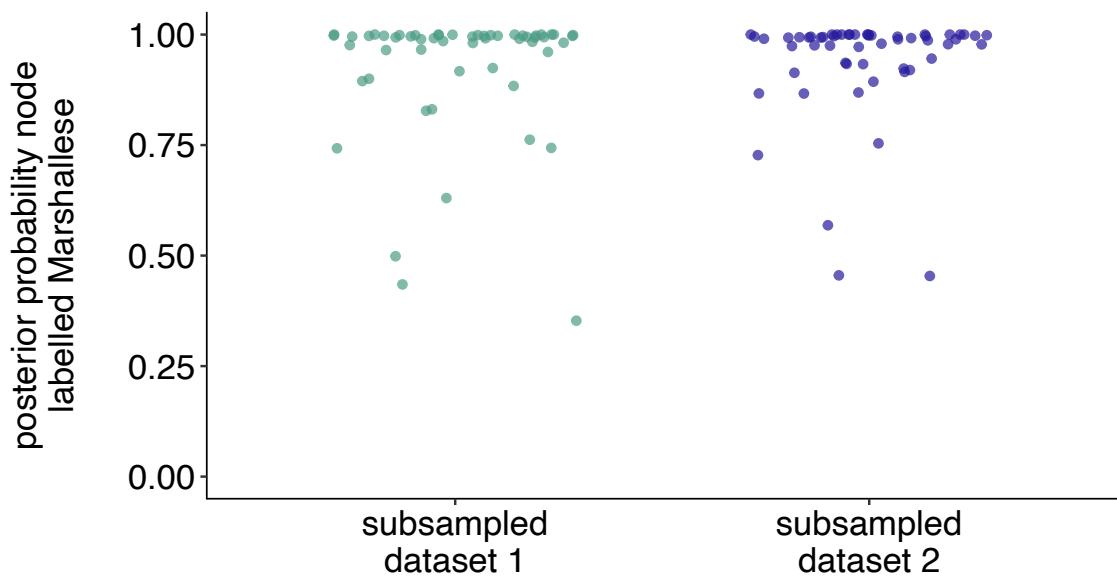


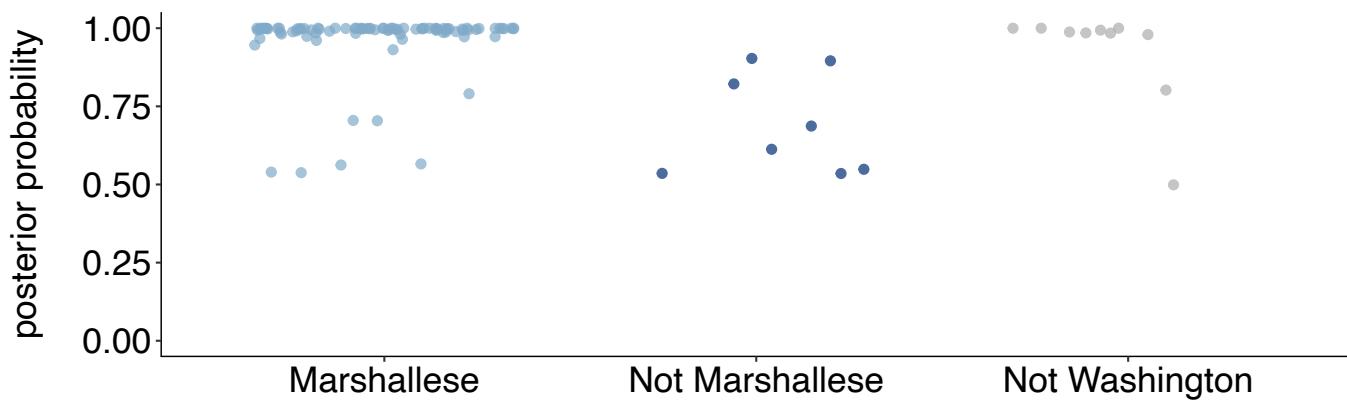






a**b****c**

a**b**

a**b**