

1  
2 **Spatial-temporal and phylogenetic analyses of epidemiologic data to help**  
3 **understand the modes of transmission of endemic typhoid fever in Samoa**  
4

5 **Short title:** Spatial-temporal and genomic epidemiology of typhoid fever in Samoa  
6

7 Michael J. Sikorski<sup>1,2,3</sup>, Jianguo Ma<sup>4</sup>, Tracy H. Hazen<sup>1,3</sup>, Sachin N. Desai<sup>2</sup>, Siasosi  
8 Tupua<sup>5</sup>, Susana Nimarota-Brown<sup>5</sup>, Michelle Sialeipata<sup>5</sup>, Savitra Rambocus<sup>6</sup>, Susan A.  
9 Ballard<sup>6</sup>, Mary Valcanis<sup>6</sup>, Robert E. Thomsen<sup>5</sup>, Roy M. Robins-Browne<sup>6</sup>, Benjamin P.  
10 Howden<sup>6</sup>, Take K. Naseri<sup>5</sup>, Myron M. Levine<sup>1,2,3</sup>, David A. Rasko<sup>1,3\*</sup>  
11

12 <sup>1</sup> Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore,  
13 Maryland, United States

14 <sup>2</sup> Center for Vaccine Development and Global Health, University of Maryland School of  
15 Medicine, Baltimore, Maryland, United States

16 <sup>3</sup> Department of Microbiology and Immunology, University of Maryland School of  
17 Medicine, Baltimore, Maryland, United States

18 <sup>4</sup> Department of Geographical Sciences, University of Maryland, College Park,  
19 Maryland, United States

20 <sup>5</sup> Ministry of Health, Government of Samoa, Apia, Samoa

21 <sup>6</sup> Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology  
22 and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection  
23 and Immunity, Melbourne, Victoria, Australia

24 \*Corresponding Author

25 E-mail: [drasko@som.umaryland.edu](mailto:drasko@som.umaryland.edu)

26 Alternative Corresponding Author

27 E-mail: [mlevine@som.umaryland.edu](mailto:mlevine@som.umaryland.edu)

## 28 **ABSTRACT**

29 *Salmonella enterica* serovar Typhi (S. Typhi) is either widely distributed or proximally  
30 transmitted via fecally-contaminated food or water to cause typhoid fever. In Samoa,  
31 where endemic typhoid fever has persisted over decades despite water quality and  
32 sanitation improvements, the local patterns of S. Typhi circulation remain  
33 undistinguished. From April 2018-June 2020, epidemiologic data and GPS coordinates  
34 were collected during household investigations of 260 acute cases of typhoid fever, and  
35 27 asymptomatic shedders of S. Typhi were detected among household contacts.  
36 Spatial and temporal distributions of cases were examined using Average Nearest  
37 Neighbor and space-time hotspot analyses. In rural regions, infections occurred in  
38 sporadic, focal clusters contrasting with persistent, less clustered cases in the Apia  
39 Urban Area. Restrictions to population movement during nationwide lockdowns in 2019-  
40 2020 were associated with marked reductions of cases. Phylogenetic analyses of  
41 isolates with whole genome sequences (n=186) revealed one dominant genotype 3.5.4  
42 (n=181/186) that contains three Samoa-exclusive sub-lineages: 3.5.4.1, 3.5.4.2, and  
43 3.5.4.3. Variables of patient sex, age, and geographic region were examined by  
44 phylogenetic groupings, and significant differences (p<0.05) associated genetically-  
45 similar isolates in urban areas with working ages (20-49 year olds), and in rural areas  
46 with age groups typically at home (<5, 50+). Isolates from asymptomatic shedders were  
47 among all three sub-lineages. Whole genome sequencing also corroborated bacterial  
48 genetic similarity in 10/12 putative epidemiologic linkages among cases and  
49 asymptomatic shedders as well as 3/3 repeat positives (presumed relapses), with a  
50 median of one single nucleotide polymorphism difference. These findings highlight

51 various patterns of typhoid transmission in Samoa that differ between urban and rural  
52 regions as well as genomic subtypes. Asymptomatic shedders, detectable only through  
53 household investigations, are likely an important reservoir and mobile agent of infection.  
54 This study advances a “Samoaan *S. Typhi* framework” that supports current and future  
55 typhoid surveillance and control efforts in Samoa.

56

## 57 **AUTHOR SUMMARY**

58 Many typhoid endemic countries have evident transmission contributions from widely  
59 distributed contaminated water supplies and/or asymptomatic *S. Typhi* carriers who  
60 intermittently cause sporadic outbreaks. However, these patterns have not yet been  
61 examined in the island nation of Samoa, where typhoid has remained endemic for  
62 decades. In this study, we incorporated the discerning powers of spatial-temporal  
63 cluster analyses as well as phylogenetics of whole genome sequences (WGS) of *S.*  
64 *Typhi* isolates to examine detailed epidemiologic data collected through household  
65 investigations of culture-confirmed cases of typhoid fever occurring in Samoa from April  
66 2018 through June 2020. We detected patterns consistent with both modes of  
67 transmission, varying between urban and rural regions, and we provided evidence of  
68 intra-household transmission of genetically similar isolates, thereby supporting a  
69 majority of putative epidemiologic linkages made during household investigations, and  
70 identifying important roles for asymptomatic shedders of *S. Typhi*. These findings  
71 advance our understanding of persistently endemic typhoid fever in Samoa and directly  
72 support the efforts of the Samoa Typhoid Fever Control Program of the Ministry of  
73 Health of Samoa.

## 74 INTRODUCTION

75 Typhoid fever is a potentially fatal human host-restricted infection caused by ingestion  
76 of *Salmonella enterica* serovar Typhi (*S. Typhi*) via contaminated food or water. Typhoid  
77 incidence, reported as cases per 100,000 people per year, is stratified into low (<10),  
78 moderate (10-100), high (>100–<500), and very high (≥500), with different transmission  
79 mechanisms capable of maintaining typhoid fever at different incidence levels [1].  
80 Widespread distribution of fecally-contaminated indirect sources (called “vehicles”),  
81 such as water supplies [2,3], crops irrigated with raw sewage [4,5], or processed foods  
82 [6,7], results in amplified, high and very high incidence transmission of *S. Typhi*.  
83 Sporadic, focal outbreaks of typhoid fever are typically associated with asymptomatic  
84 shedding of *S. Typhi* from a temporary (weeks to months) [8] or chronic (lifelong)  
85 carrier, such as “Typhoid Mary” [9]. These asymptomatic shedders inadvertently  
86 contaminate food that is consumed by contacts in close proximity, such as household  
87 members or attendees at gatherings (e.g. picnics and celebrations) [10].

88 Samoa, an island nation in the Polynesian sub-region of Oceania (the continental  
89 grouping of thousands of islands comprising 14 countries and 9 dependencies in the  
90 Central and South Pacific Ocean), has endured persistent endemic typhoid fever for  
91 decades despite improvements in drinking water quality, sanitation infrastructure, and  
92 economic development [11–14]. We have previously described endemic typhoid in  
93 Samoa by person, place, and time using blood culture surveillance data from 2008-2019  
94 [14] and characterized the population structure and evolutionary origins of a collection of  
95 *S. Typhi* genomes from Samoa spanning 1983-2020 [15]. However, it remains unclear  
96 to what extent endemic typhoid in Samoa is sustained by widespread, continuous

97 distribution of an unknown fecally-contaminated vehicle (e.g., municipal water supplies)  
98 or focal, sporadic clusters caused by ingestion of food contaminated by temporary or  
99 chronic shedders. Differing in spatial and temporal distribution, these transmission  
100 cycles can be examined using geospatial point-pattern analyses [16,17]. Additionally,  
101 when paired with data from epidemiologic investigations, whole genome sequencing  
102 (WGS) analyses can correlate epidemiology with phylogeny, infer patterns of infection  
103 by genetic subtype, and quantify relatedness between isolates to evaluate suspected  
104 transmission events [18].

105 In Samoa, teams from the Samoa Typhoid Fever Control Program (STFCP) of  
106 the Ministry of Health (MoH) perform epidemiological investigations of the households of  
107 every culture-confirmed typhoid case to ascertain clinical status, environmental risks,  
108 prevalence of *S. Typhi* among household contacts, and evidence of epidemiologic  
109 association to other cases. In this study, we combined epidemiologic data collected  
110 through household investigations from April 2018 through June 2020 with geospatial  
111 point-pattern statistical tools and the high resolution discernment of WGS. We examined  
112 the spatial and temporal distribution of typhoid cases, as well as the sex, age, and  
113 geographic occurrences of infections by phylogenetic groupings. Putative epidemiologic  
114 linkages identified through case investigations and rare repeat positive isolates from the  
115 same individual were examined for genomic similarity and provided primary data to  
116 inform a phylogenetic threshold metric of relatedness among endemic *S. Typhi* strains.  
117 Our findings can help elucidate how *S. Typhi* circulates in Samoa and support the  
118 further integration of WGS into the epidemiologic surveillance and control efforts of  
119 typhoid control programs.

120

## 121 **METHODS**

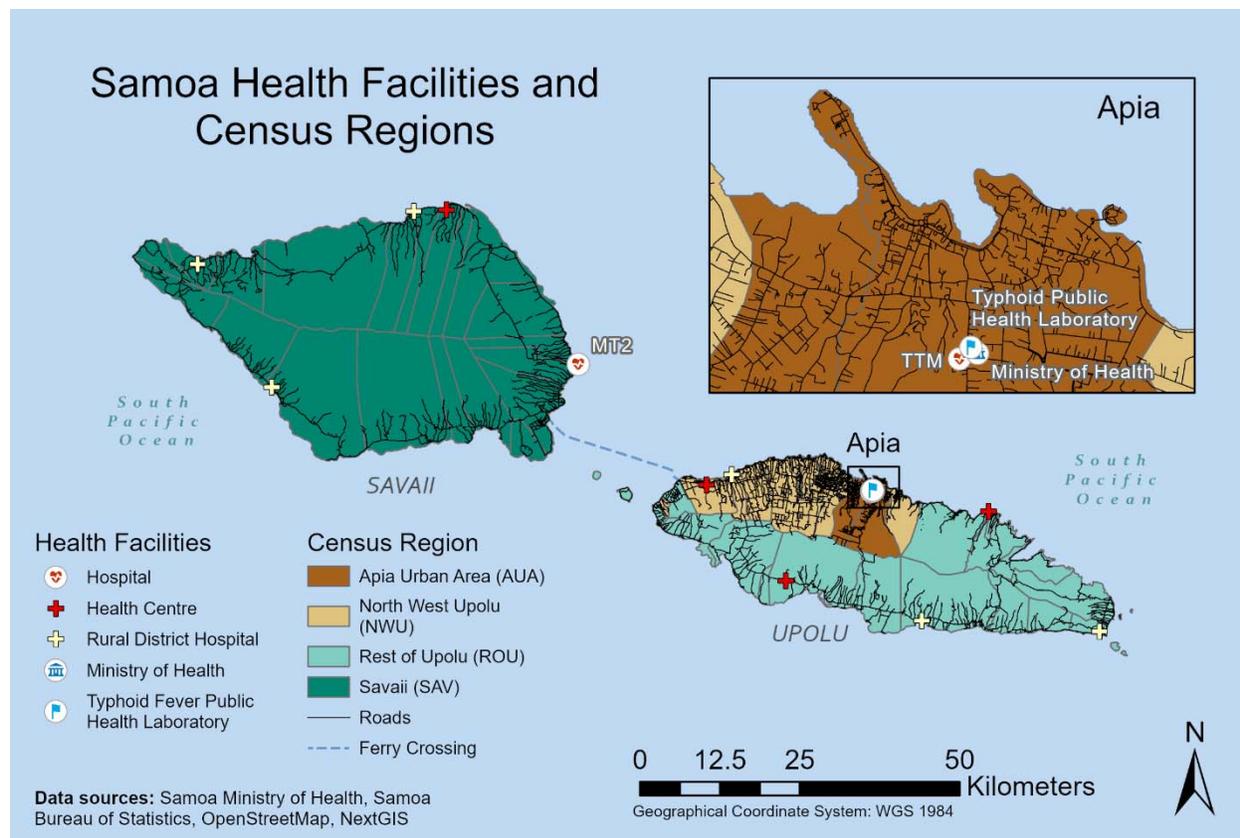
122 This study received ethical clearance from the Health Research Committee of the MoH  
123 of Samoa, as well as The University of Maryland, Baltimore (UMB) Institutional Review  
124 Board protocol HP-00087489.

125

### 126 **Study setting**

127 This study took place from 2018-2020 in Samoa, a Pacific Island nation of area 2,842  
128 km<sup>2</sup> (1,097 sq. mi) and ~200,000 population. Samoa comprises two populated islands,  
129 Upolu and Savaii, each with a central clinical laboratory and several peripheral health  
130 facilities (Fig 1). Typhoid fever is a notifiable infectious disease in Samoa that triggers  
131 epidemiologic investigation of any and all culture-confirmed cases [19,20]. Since 2018,  
132 the STFCP has strengthened epidemiologic surveillance for typhoid fever, improved  
133 clinical and environmental microbiology laboratory capacity, and initiated WGS at a  
134 collaborating regional reference laboratory in Melbourne, Australia. Typhoid Fever  
135 Epidemiologic “SWAT” Teams (equipped with specialized epidemiologic tools and  
136 tactics), one on each island, were trained to perform epidemiologic investigations of the  
137 household (and/or workplace or school) of every bacteriologically confirmed case of  
138 typhoid fever occurring anywhere in Samoa.

139



140

141 **Fig 1. Map of Samoa with census regions, roads, and health facilities.** Samoa is  
142 comprised of two primary islands, Upolu to the east and Savaii to the west. About three  
143 quarters of the population and the capital town of Apia are located on Upolu. The black  
144 lines indicate roads associated with populated areas, which are predominantly located  
145 around the periphery of the islands except for the urban areas where inland settlements  
146 are increasing. The islands are subdivided into four statistical census regions, 43  
147 political districts (outlined in light grey), and >300 villages (not depicted). The four  
148 colored geographic areas differentiate the four census regions: Apia Urban Area (AUA),  
149 North West Upolu (NWU), Rest of Upolu (ROU), and Savaii (SAV). AUA is most dense  
150 with ~25% of the total population. Apia is the capital seat of government and includes  
151 the Ministry of Health headquarters, Samoa Typhoid Fever Public Health Laboratory,  
152 and Tupua Tamasese Meaole (TTM) national hospital, which processes the blood  
153 cultures for Upolu (insert). NWU is mixed peri-urban and rural in population density and  
154 separates Apia from the inter-island ferry (navy dashed line) and adjacent international  
155 airport (not shown). The Malietoa Tanumafili II (MT2) hospital processes blood cultures  
156 for Savaii. The health facility icons indicate the two central hospitals with clinical  
157 microbiology laboratories, TTM and MT2, and the 6 district hospitals and 4 peripheral  
158 health centers.

## 159 **Case ascertainment**

160 Febrile patients who presented to an MoH health facility with clinical signs and  
161 symptoms of typhoid fever [21] triggered the collection of a blood culture. Patients from  
162 whom *S. Typhi* was isolated and confirmed by standard methods [22,23] were  
163 considered acute blood culture-confirmed cases of typhoid fever. Every such case was  
164 reported to the Typhoid SWAT Team on the incident island for expeditious investigation.  
165 On three occasions, the same individual presented to a health facility with culture-  
166 confirmed typhoid fever within ~1 month of a prior confirmed typhoid infection. These  
167 three repeat positive cases (presumed relapses) were counted as distinct cases but  
168 considered epidemiologically linked by person.

169

## 170 **Case investigations and epidemiologic linkages**

171 During household visits, the Typhoid SWAT teams evaluated environmental risk factors  
172 for typhoid transmission and ascertained the clinical wellbeing of the index case and  
173 their household contacts. Global positioning system (GPS) coordinates of the home  
174 were collected using Garmin GPSMAP 64sc handheld navigators. The Typhoid SWAT  
175 teams also sought evidence of epidemiologic links to other recent typhoid cases  
176 occurring within two months in the same home/venue, in neighboring homes, or in  
177 contacts living elsewhere in Samoa. To detect temporary or chronic asymptomatic  
178 shedders of *S. Typhi* among household contacts, up to three rectal swabs collected 1-2  
179 days apart from each contact were transported in Cary-Blair transport medium to the  
180 Typhoid Fever Public Health Laboratory in Apia and examined using standard  
181 bacteriologic methods [22,23]. Asymptomatic contacts from whom *S. Typhi* was isolated

182 from stool cultures were referred for treatment [24]. These asymptomatic shedders were  
183 considered epidemiologically linked to the index case by person, place, and time.

184

## 185 **Spatial and temporal point pattern analyses**

186 Using the spatial statistics toolbox in Esri ArcGIS Pro v2.9 [25], we performed two point  
187 pattern cluster analyses of case households (n=260) to ascertain significant ( $p < 0.05$ )  
188 spatial and spatial-temporal clusters. First, we used the Average Nearest Neighbor  
189 (ANN) analysis to determine if the spatial distribution of points (i.e., case households)  
190 was clustered based on the average distances between them [26,27]. The resulting  
191 nearest neighbor ratio ( $R_n$ ) ranged from zero (significant clustering) to 2.15 (uniform  
192 equidistant dispersion) and was supported by critical values (Z-score and p-value); an  
193  $R_n$  equal to 1 indicated randomness or no detectable spatial pattern (S1 and S2 Figs).  
194 The ANN analysis was performed on household coordinates in aggregate and then by  
195 each of the four census regions (Fig 1, S3A Fig), which separate the two Samoan  
196 islands into the Apia Urban Area (AUA; urban; ~625 persons per km<sup>2</sup>), North West  
197 Upolu (NWU; mixed peri-urban and rural; ~275 persons per km<sup>2</sup>), Rest of Upolu (ROU;  
198 rural; ~60 persons per km<sup>2</sup>) and Savaii (SAV; rural; ~26 persons per km<sup>2</sup>) [28].

199 We also used the Space-Time Hotspot analysis tool to detect significant  
200 clustering of acute case households over time. Case household coordinates were  
201 binned into 1 km<sup>2</sup> x 2-month space-time cubes and visualized in three dimensions with  
202 red shading to indicate space-time clustering confidence levels (Fig 2). Detailed  
203 descriptions and methodologies for the ANN and Space-Time Hotspot tools are  
204 included in the Supplementary Information.

## 205 **Whole genome sequencing and phylogenetic analysis**

206 All available *S. Typhi* blood isolates from acute cases and stool isolates from  
207 asymptomatic shedders were shipped to the Microbiological Diagnostic Unit Public  
208 Health Laboratory (University of Melbourne, Australia) for whole genome sequencing,  
209 as previously described [15]. Briefly, extracted DNA was sequenced on the Illumina  
210 NextSeq 500 with 150-cycle paired end chemistry. Genotypes were determined from  
211 reads mapped to the reference strain CT18 (GCA\_000195995.1) using GenoTyphi  
212 v1.9.1 [29,30]. Genotype 3.5.4.1, 3.5.4.2, and 3.5.4.3 sub-lineages unique to Samoa  
213 were assigned based on canonical single nucleotide polymorphisms (SNPs) against  
214 CT18, as previously described [15]. To generate a locally referenced maximum-  
215 likelihood (ML) phylogeny, reads were assembled and core genome SNPs were  
216 identified relative to the 2012 Samoan *S. Typhi* reference strain H12ESR00394-001A  
217 (GCA\_001118185.2). After removal of recombination sites, a ML phylogeny was  
218 inferred using the generalized time-reversible (GTR) site substitution model with a  
219 Gamma rate distribution and the Lewis ascertainment bias correction  
220 (ASC\_GTRGAMMA) and 100 bootstrap pseudo-replicates in RAxML v8.2.10 [31].  
221 Pairwise SNP distances between genomes were calculated using snp-dists v0.8.2 [32]  
222 and compared among epidemiologically linked isolates. Detailed protocols are found in  
223 the Supplementary Information, with accession numbers listed in S1 Table and  
224 assemblies available on Figshare (doi: 10.6084/m9.figshare.18665686).  
225

## 226 **Epidemiologic analysis of genotypes/sub-lineages**

227 Patient characteristics (sex, age group, and mean age) were assessed for differences  
228 between all cases and the subset of sequenced isolates using Pearson's chi-squared  
229 test of independence for the categorical variables and unpaired two-tailed T test of  
230 means for mean age. Phylogenetic groupings of *S. Typhi* were examined for  
231 associations with epidemiologic variables (infection status, sex, age group, census  
232 region, year) using Fisher's exact test. The ANN statistic was performed in ArcGIS Pro  
233 v2.9 on groupings of household coordinates by each genotype 3.5.4 sub-lineage (S3B-  
234 D Fig). The absolute and relative counts of phylogenetic groupings within each census  
235 region were plotted over 2-month time intervals using Microsoft Excel v16.58.

236

## 237 **Statistical analyses.**

238 Spatial statistics were performed in ArcGIS Pro v2.9 using default tools. Epidemiologic  
239 comparisons were performed using the biostatistics package RVAideMemoire v0.9-81  
240 [33] in R v4.1.1 [34]. Statistical significance was reported at  $p < 0.05$  and indicated in the  
241 text, figure, and/or table legends, accordingly.

242

## 243 **RESULTS**

### 244 **Characteristics of the dataset**

245 From April 27, 2018 to June 9, 2020, the STFCP identified 260 blood culture-confirmed  
246 acute typhoid fever cases through a strengthened surveillance system and detected 27  
247 asymptomatic *S. Typhi* fecal shedders through household investigations. We examined

248 case household coordinates for spatial and spatial-temporal clustering, and then  
 249 incorporated phylogenetic analyses of the sequenced subset of 186 Samoan *S. Typhi*  
 250 isolates – 172 from acute cases and 14 from asymptomatic shedders – among  
 251 household contacts (S1 Table). Among the 186 whole genome sequenced isolates,  
 252 three *S. Typhi* genotypes were identified: genotype 3.5.4 (181/186, 97.3%), genotype  
 253 4.1 (4/186, 2.2%), and genotype 3.5.3 (1/186; 0.5%). All but one genotype 3.5.4 isolates  
 254 could be further categorized into three distinct sub-lineages denoted 3.5.4.1 (n=49),  
 255 3.5.4.2 (n=51), and 3.5.4.3 (n=80) and differentiated from all other global *S. Typhi*  
 256 genomes by single canonical SNPs (S1 Table), as previously described [15]. The  
 257 sequenced subset of acute cases showed no significant difference in patient  
 258 characteristics (patient sex, mean age, or age group distribution) when compared to all  
 259 reported cases (Table 1).

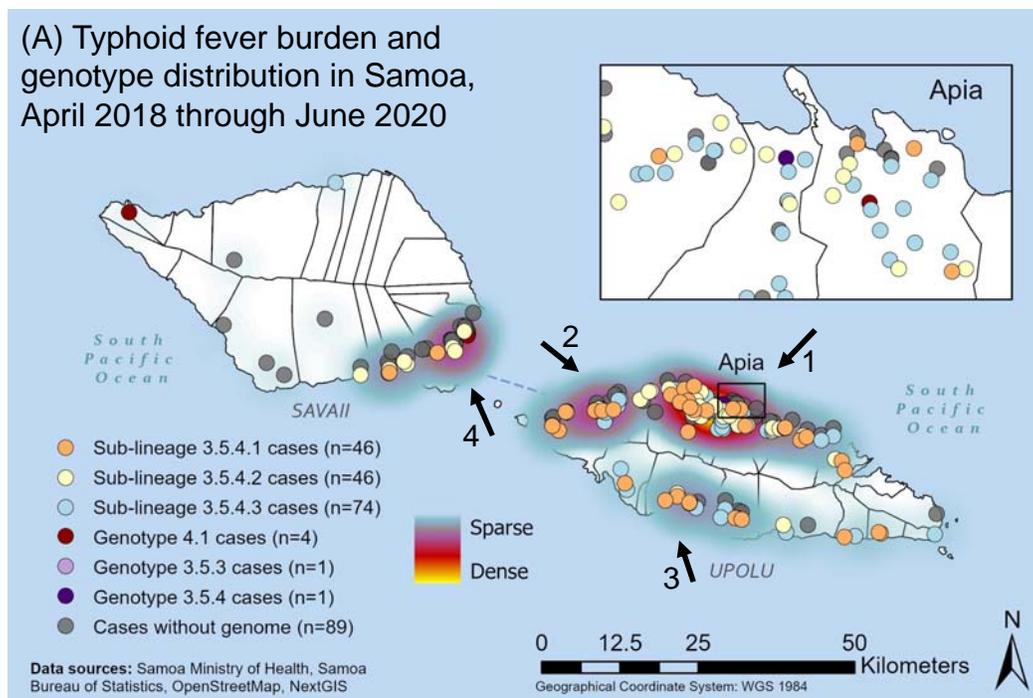
**Table 1.** Comparison of age and sex variables of acute cases associated with an *S. Typhi* whole genome sequence (WGS) versus acute cases missing WGS data

Variable		All cases	Cases, WGS available	Cases, WGS missing	Significance
<b>Total</b>		<b>260</b>	<b>172</b>	<b>88</b>	-
Sex	Female, N (%)	112 (43.1%)	72 (41.9%)	40 (45.5%)	p=0.560
	Male, N (%)	147 (56.5%)	99 (57.6%)	48 (54.5%)	
Age	Mean age (SD)	24.4 (16.7%)	24.1 (16.5%)	25.0 (17.3%)	p=0.709
Age groups	0-4 years old, N (%)	21 (8.1%)	15 (8.7%)	6 (6.8%)	p=0.864
	5-19 years old, N (%)	103 (39.6%)	66 (38.4%)	37 (42.0%)	
	20-59 years old, N (%)	111 (42.7%)	75 (43.6%)	36 (40.9%)	
	50+ years old, N (%)	25 (9.6%)	16 (9.3%)	9 (10.2%)	
Categorical variables (sex and age group) tested with Pearson's Chi-squared test of independence. Continuous variable (mean age) tested with unpaired two-tailed T test of means. Statistics were performed using the RVAideMemoire v0.9-81 [33] packages in R v4.1.1 [34].					

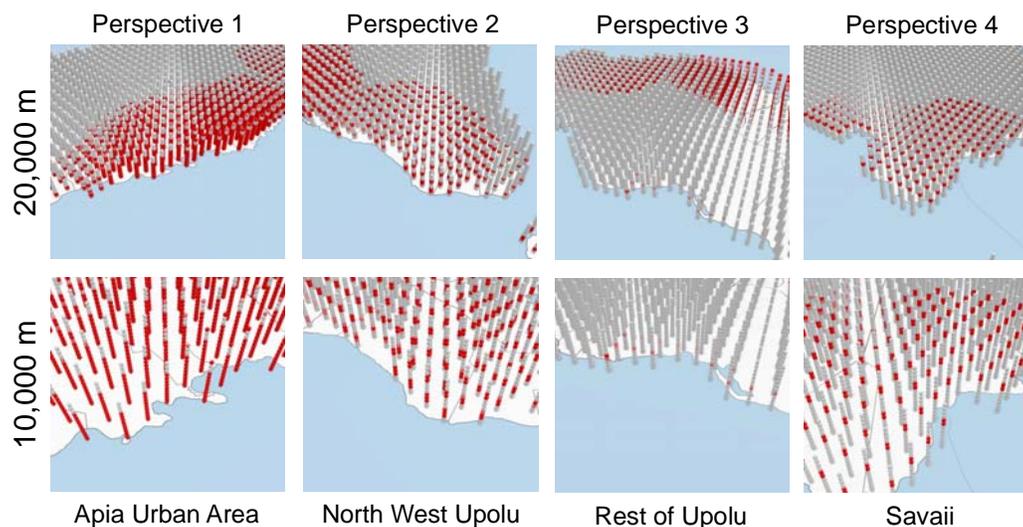
260

261 **Spatial clustering and spatial-temporal hotspot analyses of**  
262 **acute cases**

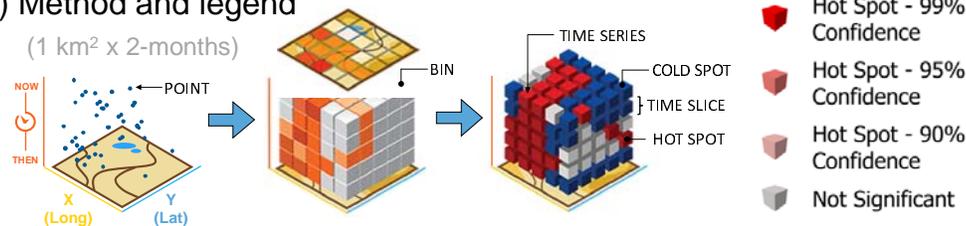
263 When acute case households (n=260) were mapped, areas with many typhoid cases  
264 (higher burden) – Apia, northwestern Upolu, southeastern Savaii, and Upolu’s southern  
265 coast – were prominent among the four census regions (arrows in Fig 2A). We used the  
266 Average Nearest Neighbor (ANN) and Space-Time Hotspot analyses in ArcGIS Pro  
267 v2.9 to assess spatial and spatial-temporal clustering, respectively. As expected for a  
268 human host-restricted communicable disease, case households (n=260) demonstrated  
269 significant clustering by ANN analysis ( $R_n < 1$ ,  $p < 0.005$ ). However, when grouped by  
270 census region (S3A Fig, Table 2), local differences emerged in the degree of clustering,  
271 i.e.,  $R_n$  closer to 0, indicating more tightly grouped clusters. Case households in the  
272 AUA (n=42), where population density is greatest and where residents receive  
273 putatively treated drinking water through a reticulated system, were the least clustered  
274 of the census regions ( $R_n = 0.62$ ). In contrast, case households in the rural ROU (n=55),  
275 where districts each have independent (often untreated) water schemes, were the most  
276 clustered ( $R_n = 0.25$ ) (Table 2). Notably, SAV (n=53), which is more rural than ROU,  
277 demonstrated a similar degree of clustering ( $R_n = 0.60$ ) to AUA (Table 2); however, there  
278 were ~8 case households remotely located in the north and west of Savaii far from any  
279 nearest neighbor (S3A Fig).



(B) Space-time hotspot analysis visualized in 3D at two altitudes



(C) Method and legend



280  
281 **Fig 2. Spatial-temporal and genomic analysis of typhoid fever cases (burden) in**  
282 **Samoa from April 2018 through June 2020. (A)** Using ArcGIS Pro v2.9, household  
283 GPS coordinates of acute cases of *S. Typhi* (n=260) from April 27, 2018 to June 9, 2020  
284 are plotted on a map of Samoa and colored according to *S. Typhi* genotype/sub-

285 lineage. A point density heatmap is inferred from case coordinates. The four prominent  
 286 hotspots are identified with arrows labeled 1-4, which also indicate aerial perspectives  
 287 shown in the reconstructions in panel B. **(B)** Coordinates were binned into space-time  
 288 cubes of 1km<sup>2</sup> x 2-months (see panel C and Supplementary Information for more  
 289 details) and tested for significant clustering by the emerging hotspot analysis  
 290 implemented in ArcGIS Pro. The resulting 3D space-time hotspots were visualized from  
 291 an aerial view of 20,000 meters and 10,000 meters focusing on the four typhoid  
 292 hotspots identified in panel A. In the 3D renderings, each vertical column on the map  
 293 represents a stack of cubes over 1 km<sup>2</sup> area and with a height representing 2 months of  
 294 time, totaling 13 time steps, or all 26 months of this study. Aggregation of points in both  
 295 space and time generate a space-time hotspot, visible as a red cube or band of several  
 296 adjacent cubes. The intensity of red shading indicates statistical confidence. **(C)** To  
 297 generate space-time cubes, points were graphed in a 3-dimensional plane based on  
 298 coordinates and time, and then binned into cubes of space and time dimensions.  
 299 Counts within each cube were compared to nearby cubes to extrapolate hot and cold  
 300 spots with statistical support (99%, 95%, 90%, not significant). Hotspots in the x-y plane  
 301 occur within a time slice. Hotspots inferred from the z-direction fall within a time series.  
 302 Adapted with permission from Esri (Redlands, CA: Environmental Systems Research  
 303 Institute, Inc.).  
 304

**Table 2.** Results of the Average Nearest Neighbor (ANN) statistic performed on geographic and phylogenetic groupings of *S. Typhi* acute infections.

Group <sup>a</sup>	No. of points <sup>a</sup>	$\bar{O}$ distance (meters)	$\bar{E}$ distance (meters)	Nearest Neighbor Ratio ( $R_N$ )	ANN Z-score <sup>b</sup>	ANN P-value <sup>c</sup>	S3 Fig. Panel
All cases	260	676.07	2312.04	0.29	-21.83	p<0.005	A
Cases AUA	42	204.87	331.34	0.62	-4.73	p<0.005	A
Cases NWU	110	351.30	858.30	0.41	-11.85	p<0.005	A
Cases ROU	55	607.40	2406.22	0.25	-10.61	p<0.005	A
Cases SAV	53	1800.82	3007.99	0.60	-5.59	p<0.005	A
Cases 3.5.4.1	46	1589.89	3230.86	0.49	-6.59	p<0.005	B
Cases 3.5.4.2	46	1211.20	2990.57	0.41	-7.64	p<0.005	C
Cases 3.5.4.3	74	1710.15	3304.99	0.52	-7.94	p<0.005	D

ANN is a statistical comparison of the observed mean ( $\bar{O}$ ) and expected mean ( $\bar{E}$ ) distances between points yields a nearest neighbor ratio ( $R_N$ ) further described by a probability (p-value) that the observed spatial pattern was created by a random process and standard deviations (z-score). The  $R_N$  ranges from 0 (indicating discrete clustering of points) to ~2.15 (indicating uniform dispersion of points equidistant from one another). An  $R_N$  value of 1 indicates complete spatial randomness.

<sup>a</sup> Only household coordinates of acute cases were included to assess circulating *S. Typhi* detected through routine passive surveillance and to avoid biasing the point-pattern analyses to households where asymptomatic shedders were localized by active investigation.

<sup>b</sup> All z-scores are small (<-1.96) corresponding to a confidence level above 95%.

<sup>c</sup> All p-values are small (<0.005) which indicate statistical significance.

305

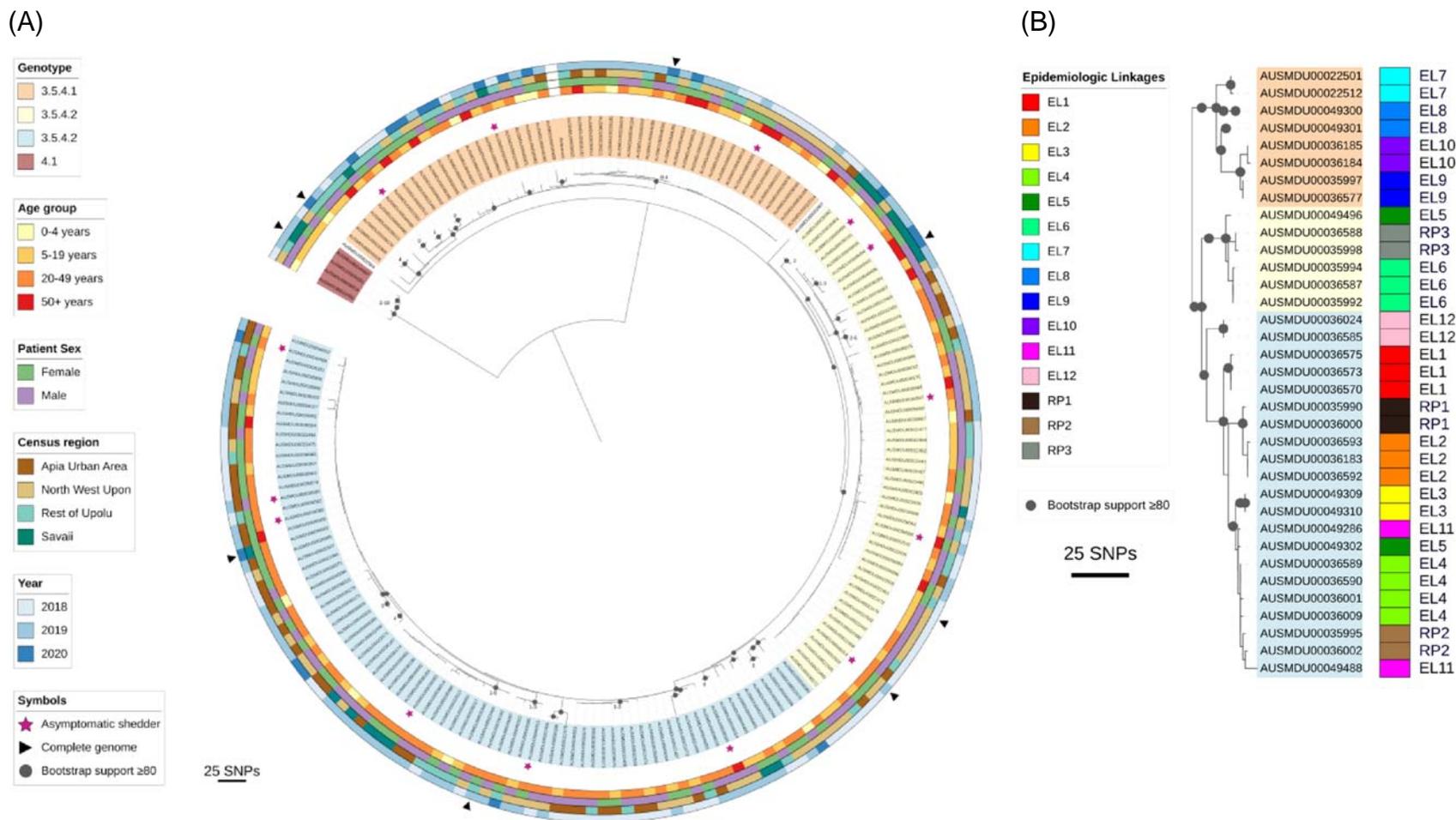
306 To assess for temporal clustering by census region, a Space-Time Hotspot  
307 analysis was performed using a window of 1 km<sup>2</sup> x 2-months to bin case household  
308 coordinates (n=260) and visualized in three-dimensions at two altitudes, 20,000 meters  
309 and 10,000 meters, to aid in visualization of each time series (Fig 2B). Temporal  
310 persistence of typhoid burden was detectable in AUA as sequential spatial-temporal  
311 hotspots, visible as contiguous red columns, ( $p<0.01$ ) associated with that geographic  
312 region (Fig 2B-C). This was in contrast with sporadic (short-lived) hotspots of infection  
313 evident in the rest of the country as identified by the discrete red bands ( $p<0.01$ )  
314 between grey bars in those geographic regions (Fig 2B-C).

315

## 316 **Phylogenetic analysis by genotype, person, place, and time**

317 Subtyping of *S. Typhi* isolates facilitates elucidation of patterns of transmission based  
318 on the assumption that similar isolates form a network or chain of infection via a  
319 common vehicle and/or source. The presence of three genotype 3.5.4 sub-lineages in  
320 Samoa (Fig 3A) allowed the demarcation of useful groupings based on genetic similarity  
321 for subsequent, more detailed epidemiologic analyses of associations. The 14  
322 sequenced isolates from asymptomatic shedders detected through household  
323 investigations were genotype 3.5.4 and showed no association with any particular sub-  
324 lineage (Table 3, Fig 3A). Sub-lineage 3.5.4.1 was significantly ( $p<0.05$ ) associated with  
325 (i) age group (greater-than-expected frequency [GTEF] among ages 0-4 years and 50+;  
326 less-than-expected frequency [LTEF] in 5-49 year-olds), (ii) census region (GTEF in  
327 ROU, LTEF in AUA and SAV), and (iii) year (GTEF in 2019) (Table 3, Fig 3A). Sub-  
328 lineage 3.5.4.2 was significantly ( $p<0.05$ ) associated with the patient's sex, being

329 isolated from males more frequently than females (Table 3, Fig 3A). Sub-lineage 3.5.4.3  
330 was significantly ( $p < 0.05$ ) associated with 20-49 year-olds and LTEF in the 5-19 and  
331 50+ year old age groups, as well as census region (common to the four census regions  
332 but GTEF in the AUA and ROU) (Table 3, Fig 3A).  
333



334

335 **Fig 3. Maximum-likelihood phylogeny, associated epidemiologic variables, and epidemiologic linkages among *S.***  
 336 ***Typhi* from Samoa from April 2018 through June 2020. (A)** Maximum-likelihood phylogeny from *de novo* assembled  
 337 whole genome sequences of Samoan *S. Typhi* genotype 3.5.4 isolates from April 27, 2018 through June 9, 2020 (N=186,  
 338 S1 Table). Core genome single nucleotide polymorphisms (SNPs) were identified relative to the 2012 Samoa Reference  
 339 genome H12ESR00394-001A (GCA\_001118185.2). Genotype/sub-lineage is indicated by isolate label shading. The  
 340 labels of the three genotype 3.5.4 sub-lineages 1-3 are identified with orange, yellow, and blue label shading, respectively.

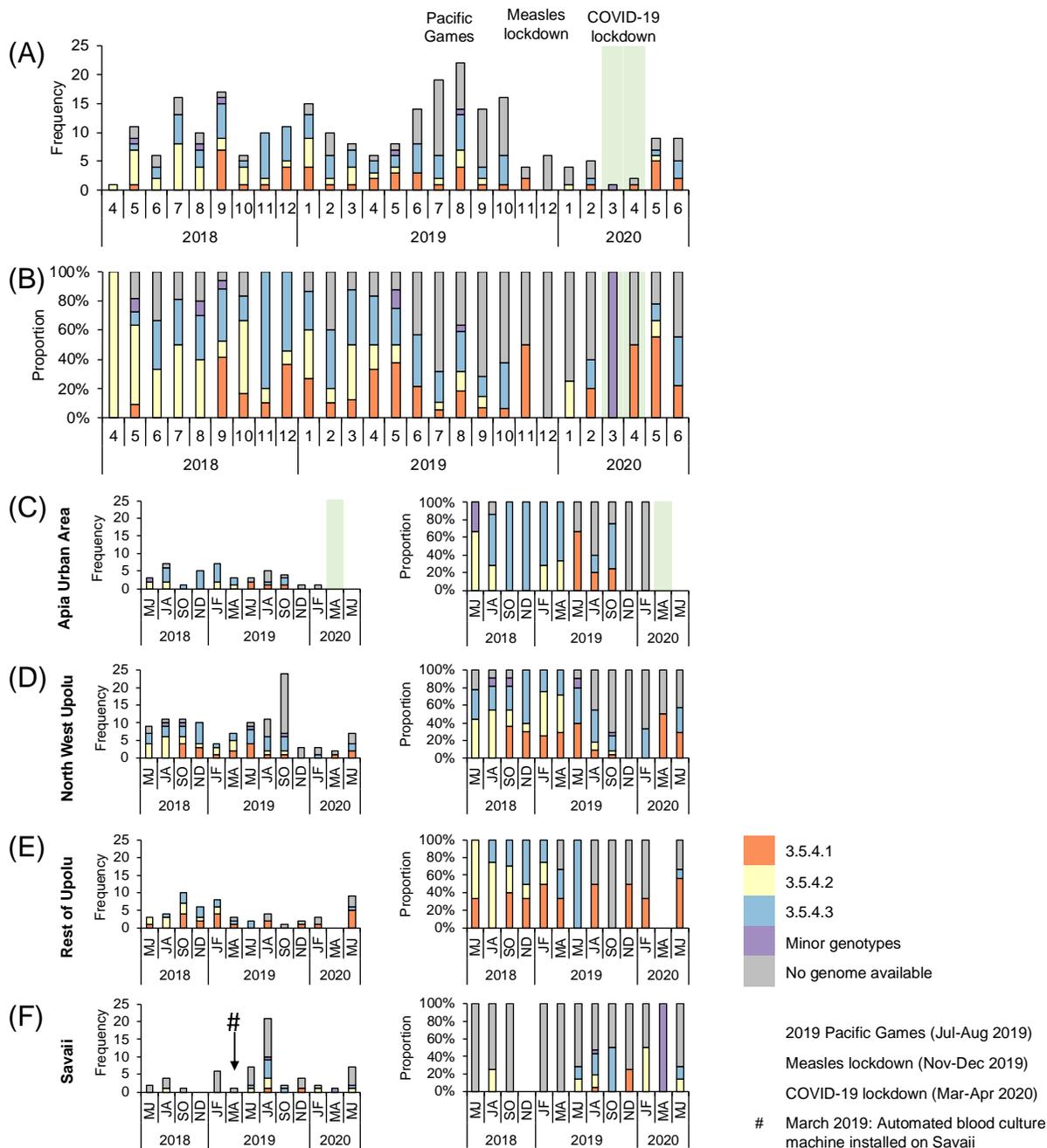
341 Genotype 4.1 isolates are shaded in red. Two outlier isolates are not shaded. Asymptomatic shedders of *S. Typhi* are  
342 denoted with a pink star. The patient age group, patient sex, census region of residence, and year of isolation are  
343 annotated in the circumscribing colored rings. **(B)** Epidemiologic linkages (EL) identified through case household  
344 investigations are compared against genomic relatedness of the isolates based on core genome phylogeny. The visible  
345 isolates were extracted from the tree in Panel A. In total 2/12 epidemiologic linkages are not supported by the genomic  
346 variation observed in the isolates from those samples (EL5/dark green and EL11/pink groups are not similar). Three  
347 repeat positive (RP) blood cultures isolated ~1 month apart are also compared, showing close genetic similarity.  
348

**Table 3.** Comparison of epidemiologic variables relating to person, place, and time for each Samoa dominant S. Typhi genotype/sub-lineage.

	Genotype/Sub-lineage						TOTAL
	3.5.4.1	3.5.4.2	3.5.4.3	4.1	3.5.4	3.5.3	
<b>Total</b>	49	51	80	4	1	1	186
<b>Status</b>							
Acute Case	46 (93.9%)	46 (90.2%)	74 (92.5%)	4 (100%)	1 (100%)	1 (100%)	172
Asymptomatic Shedder	3 (6.1%)	5 (9.8%)	6 (7.5%)	0 (0%)	0 (0%)	0 (0%)	14
Significance	p=1	p=0.535	p=1	p=1	p=1	p=1	
<b>Sex</b>							
Female	24 (49%)	16 (31.4%)	40 (50%)	1 (25%)	0 (0%)	0 (0%)	81
Male	25 (51%)	35 (68.6%)	40 (50%)	3 (75%)	1 (100%)	1 (100%)	105
Significance	p=0.404	<b>p=0.0470</b>	p=0.137	p=0.634	p=1	p=1	
<b>Age group</b>							
0-4 years	7 (14.3%)	3 (5.9%)	4 (5.0%)	1 (25%)	0 (0%)	0 (0%)	15
5-19 years	17 (34.7%)	23 (45.1%)	23 (28.8%)	3 (75%)	0 (0%)	1 (100%)	67
20-49 years	15 (30.6%)	19 (37.3%)	50 (62.5%)	0 (0%)	1 (100%)	0 (0%)	85
50+ years	10 (20.4%)	6 (11.8%)	3 (3.8%)	0 (0%)	0 (0%)	0 (0%)	19
Significance	<b>p=0.004</b>	p=0.363	<b>p=0.0004</b>	p=0.093	p=1	p=0.543	
<b>Census region</b>							
AUA	4 (8.2%)	7 (14%)	23 (28.8%)	0 (0%)	1 (100%)	0 (0%)	35
NWU	20 (40.8%)	23 (44%)	35 (43.8%)	2 (50%)	0 (0%)	1 (100%)	81
ROU	22 (44.9%)	12 (24%)	13 (16.3%)	0 (0%)	0 (0%)	0 (0%)	47
SAV	3 (6.1%)	9 (18%)	9 (11.3%)	2 (50%)	0 (0%)	0 (0%)	23
Significance	<b>p=0.0013</b>	p=0.452	<b>p=0.0072</b>	p=0.127	p=0.312	p=1	
<b>Year</b>							
2018	15 (30.6%)	28 (54.9%)	32 (40.0%)	1 (25%)	1 (100%)	1 (100%)	78
2019	23 (46.9%)	20 (39.2%)	42 (52.5%)	3 (75%)	0 (0%)	0 (0%)	88
2020	11 (22.4%)	3 (5.9%)	6 (7.5%)	0 (0%)	0 (0%)	0 (0%)	20
Significance	<b>p=0.0083</b>	p=0.081	p=0.328	p=0.762	p=0.527	p=0.527	

Numbers in parentheses indicate the percentages of the number of times the genotype/sub-lineage was detected for each variable relative to the total number of the genotypes; P-values were calculated using Fisher's exact test using the RVAideMemoire v0.9-81 [33] packages in R v4.1.1 [34]. Bolded p-values are significant at p<0.05.

350 Geospatially, case households associated with each genotype/sub-lineage were  
351 found in the four census regions (Table 3) and were overlapping (Fig 2A), indicating  
352 widespread circulation among the islands. Spatial clustering of each 3.5.4 sub-lineage  
353 was significant ( $p < 0.005$ ) and intermediate ( $0.4 < R_n < 0.6$ ) among the groupings tested by  
354 ANN (Table 2). Within each census region, we identified variation in the frequency as  
355 well as transient expansions and contracts in the proportions of each 3.5.4 sub-lineage  
356 over time (Fig 4). Total case counts peaked in July-August 2019 and then sharply  
357 declined in November-December 2019 and again in March-April 2020 (highlighted  
358 background colors in Fig 4), coinciding with three major national events.  
359



360

361 **Fig 4. Bar graphs depicting the number and proportion of blood culture-**  
 362 **confirmed cases of typhoid fever in Samoa over time from April 27, 2018 through**  
 363 **June 9, 2020 by genotype/sub-lineage. (A-B)** The frequency and proportion of acute  
 364 cases (N=260) occurring each month are colored by genotype/sub-lineage (n=172) or  
 365 greyed if no genome was available (n=88). Genotype 3.5.4 sub-lineages 1-3,  
 366 comprising a total of 166/172 sequenced isolates from acute cases, are identified as  
 367 orange, yellow, and blue bars, respectively. Minor genotypes (purple bar) include  
 368 genotype 4.1 (4/186), genotype 3.5.3 (1/186), and one genotype 3.5.4 outlier (1/186)  
 369 that did not fall into one of the three sub-lineages. **(C-F)** The frequency and proportion

370 of acute cases occurring over 2-month intervals in each census region. Three major  
371 events impacting population mobility – increases during the 2019 Pacific Games in Nov-  
372 Dec 2019, and restrictions during the measles lockdown from Nov-Dec 2019 [35] and  
373 COVID-19 lockdown from Mar-Apr 2020 [36] – are labeled as light red, light purple, and  
374 light green background colors, respectively, over the two-month time intervals affected.

375

## 376 **Genetic distances between epidemiologic linkages**

377 WGS analyses can also be used to verify epidemiologic linkages, thereby providing  
378 evidence of transmission based on the relatedness of *S. Typhi* isolates. We defined  
379 relatedness based on the number of core genome SNP differences between isolates,  
380 with fewer differences indicating closer relatedness. During household investigations,  
381 the Typhoid SWAT teams attempted to trace infections and link index cases by person,  
382 place, and/or time. In 12 instances, genomes were available for putatively linked  
383 infections in relation to person, place, and/or time (Fig 3B, Table 4). In six of these, an  
384 asymptomatic shedder of *S. Typhi* was identified. In nine epidemiologic linkages, the  
385 genotype sub-lineages matched and SNP distances between isolates were two or  
386 fewer, indicating close relatedness (Fig 3B, Table 4). In two linkages, EL8 and EL11,  
387 the genotype sub-lineages matched, but the SNP distances between isolates were  
388 greater, with nine and 14 SNP differences, respectively (Table 2). In another (EL5), the  
389 isolates were of completely different *S. Typhi* sub-lineages and 28 SNPs apart (Table  
390 4). Additionally, microbiological surveillance during the study period identified three  
391 instances in which two positive *S. Typhi* cultures derived from the same individual were  
392 collected ~1 month apart (Table 4). In each instance, the genotype sub-lineages  
393 matched (Fig 3B) and the SNP distances between the first and second isolation were 1-  
394 3 SNPs, indicating the isolates were closely related (Table 4).

**Table 4.** Suspected epidemiologic linkages detected through active surveillance and associated genotype and pairwise SNP distance between isolates

Linkage ID <sup>a</sup>	Year	Month	Sex	Age	Status <sup>b</sup>	Relation to Index Case <sup>c</sup>	Epidemiologic Association	Strain ID	Genotype <sup>d</sup>	SNP Dist <sup>e</sup>
EL1	2018	12	F	10-14	Acute Case	Index	Index	AUSMDU00036575	3.5.4.3	-
	2018	12	M	35-39	Acute Case	Family; same HH	Person, place, time	AUSMDU00036570	3.5.4.3	1
	2018	12	F	20-24	Acute Case	Family; same HH	Person, place, time	AUSMDU00036573	3.5.4.3	1
EL2	2019	8	F	20-24	Acute Case	Index	Index	AUSMDU00036183	3.5.4.3	-
	2019	8	F	30-34	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036593	3.5.4.3	1
	2019	8	F	25-29	Acute Case	Family; different HH	Person, time	AUSMDU00036592	3.5.4.3	0
EL3	2020	6	M	40-44	Acute Case	Index	Index	AUSMDU00049310	3.5.4.3	-
	2020	6	M	10-14	Acute Case	Family; same HH	Person, place, time	AUSMDU00049309	3.5.4.3	0
EL4	2019	2	F	10-14	Acute Case	Index	Index	AUSMDU00036001	3.5.4.3	-
	2019	2	F	25-29	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036589	3.5.4.3	1
	2019	3	F	10-14	Acute Case	Family; same HH	Person, place, time	AUSMDU00036009	3.5.4.3	2
	2019	3	M	30-34	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036590	3.5.4.3	1
EL5	2020	5	M	65-69	Acute Case	Index	Index	AUSMDU00049496	3.5.4.2	-
	2020	5	F	20-24	Acute Case	Family; different HH	Person, time	AUSMDU00049302	3.5.4.3	28
EL6	2019	1	M	10-14	Acute Case	Index	Index	AUSMDU00035992	3.5.4.2	-
	2019	1	M	65-69	Acute Case	Family; same HH	Person, place, time	AUSMDU00035994	3.5.4.2	1
	2019	1	M	35-39	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036587	3.5.4.2	0
EL7	2018	9	M	5-9	Acute Case	Index	Index	AUSMDU00022501	3.5.4.1	-
	2018	9	F	50-54	Acute Case	Family; same HH	Person, place, time	AUSMDU00022512	3.5.4.1	1

EL8	2020	5	F	60-64	Acute Case	Index	Index	AUSMDU00049300	3.5.4.1	-
	2020	5	M	25-29	Acute Case	Caregiver; different HH	Person, place, time	AUSMDU00049301	3.5.4.1	14
EL9	2018	12	M	10-14	Acute Case	Index	Index	AUSMDU00036577	3.5.4.1	-
	2019	1	M	20-24	Acute Case	Gathering (funeral); diff HH	Person, place, time	AUSMDU00035997	3.5.4.1	0
EL10	2019	8	M	10-14	Acute Case	Index	Index	AUSMDU00036184	3.5.4.1	-
	2019	8	F	0-4	Acute Case	Family; different HH	Person, time	AUSMDU00036185	3.5.4.1	1
EL11	2019	10	F	0-4	Acute Case	Index	Index	AUSMDU00049286	3.5.4.3	-
	2020	4	F	25-29	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00049488	3.5.4.3	9
EL12	2019	6	M	15-19	Acute Case	Index	Index	AUSMDU00036024	3.5.4.3	-
	2019	6	M	65-69	Asymptomatic Shedder	Family; same HH	Person, place, time	AUSMDU00036585	3.5.4.3	0
RP1	2019	1	F	10-14	Acute Case	Index	Index	AUSMDU00035990	3.5.4.3	-
	2019	2	F	10-14	Acute Case	Self	Person	AUSMDU00036000	3.5.4.3	3
RP2	2019	1	M	15-19	Acute Case	Index	Index	AUSMDU00035995	3.5.4.3	-
	2019	2	M	15-19	Acute Case	Self	Person	AUSMDU00036002	3.5.4.3	3
RP3	2019	1	F	60-64	Acute Case	Index	Index	AUSMDU00035998	3.5.4.2	-
	2019	2	F	60-64	Asymptomatic Shedder	Self	Person	AUSMDU00036588	3.5.4.2	1

<sup>a</sup> EL = Epidemiologic linkage; RP = repeat positive (culture); numbers are ascending based on date of isolate from index case  
<sup>b</sup> Acute cases are blood isolates; asymptomatic shedders are rectal swab stool isolates  
<sup>c</sup> The index case is the first culture-confirmed infection in the linkage; HH = Household  
<sup>d</sup> Genotypes were determined using GenoTyphi v1.9.1; sub-lineages were determined as previously described [15]  
<sup>e</sup> Single Nucleotide Polymorphism (SNP) pairwise distances calculated using snp-dists v0.8.2 [32].

## 396 **DISCUSSION**

397 Identifying the modes of transmission that maintain the endemicity of typhoid in Samoa  
398 has been a subject of interest for many decades [12,14]. Some insights were derived  
399 from our review of the Samoa typhoid burden from 2008-2019 [14]: 53-193 cases of  
400 blood culture-confirmed typhoid fever occurred annually without obvious seasonality,  
401 equating to a persistently moderate annual incidence. The greatest incidence and  
402 numbers of cases was observed among school age children and working age adults in  
403 the contiguous regions of NWU and AUA where piped treated water reaches most  
404 households, indicating that their daily life patterns put them at enhanced risk. Herein, we  
405 relate results of applying the potential discerning power of spatial-temporal cluster and  
406 phylogenetic analyses to typhoid infections in Samoa from April 2018 through June  
407 2020.

408       The burden of typhoid fever was most concentrated in four locations (hotspots;  
409 arrows in Fig 2A) determined by point densities: Apia, northwestern Upolu, southeastern  
410 Savaii near the ferry terminal, and Upolu's southern coast. We hypothesized that spatial  
411 and temporal differences among the underlying census regions would support either  
412 widespread, continuous typhoid circulation or focal, sporadic case clusters around  
413 households. Case households in both urban and rural regions showed significant  
414 clustering ( $p < 0.005$ ,  $R_n < 1$ ); however, the distances between cases were much smaller  
415 in rural ROU (lower  $R_n$ ), indicating stronger foci of spatial clusters compared with more  
416 dispersed case clustering (greater  $R_n$ ) in AUA (Table 2). In Savaii, the  $R_n$  was closer to  
417 the value for AUA, likely due to skewing from ~8 individual cases occurring in remote  
418 isolation distant from nearest neighbors (S3A Fig). Spatial-temporal hotspot analysis

419 revealed consecutive, significant ( $p < 0.01$ ) temporal hotspots in AUA and the adjacent  
420 peri-urban regions of NWU (Fig 4), which could represent a locus of persistent,  
421 amplified transmission and also a source of infection for persons from other parts of the  
422 islands who work, visit, or consume food prepared in and around Apia.

423 Contextualization of cases over time amid major national events that either  
424 greatly expanded or abruptly halted population mobility highlighted an association  
425 between population mobility and case frequency. Typhoid cases, all of which occurred  
426 in Samoan people, rose dramatically during the 2019 Pacific Games in July-August  
427 2019 (light red background in Fig 4A) when there was an influx of athletes and  
428 supporters. Notably, we did not detect any non-Samoan genotypes during or after this  
429 period among the sequenced isolates (Fig 4A), suggesting that dissemination of an  
430 imported strain was less likely responsible for the observed increase in cases. Months  
431 later, gatherings and population mobility were limited during an intensive nationwide  
432 measles lockdown in November-December 2019 (light purple background in Fig 4A)  
433 [35], and during COVID-19 lockdowns in March-April 2020 (light green background in  
434 Fig 4A) [36]. During these periods of reduced population mobility, the burden of typhoid  
435 diminished markedly (Fig 4A). These data highlight a potential role for community  
436 interactions and population movement in transmission.

437 Sub-typing of clinically indistinguishable infections can help to ascertain unique  
438 networks of infection. Our previous WGS analyses of Samoan *S. Typhi* from 1983-2020  
439 revealed that dominant Samoa-exclusive genotypes 3.5.4 and 3.5.3 can be divided into  
440 sub-lineages based on canonical SNPs [15]. Herein, we analyzed the 2018-2020 *S.*  
441 *Typhi* isolates by sub-lineages in relation to variables of person, place, and time (Table

442 3). Infections with sub-lineage 3.5.4.1 were more common in ROU and NWU among  
443 very young children and older adults who are typically home. Whereas, infections with  
444 sub-lineage 3.5.4.3 were more common in urban regions among 20-49 year olds, who  
445 represent the dominant workforce [28] that and may be exposed to *S. Typhi* outside the  
446 home. Infections with sub-lineage 3.5.4.2 were more common among men than women.  
447 These data suggest different groups of people in different regions have variable  
448 exposures to *S. Typhi*. Regrettably, because of the disruptions of COVID-19 and  
449 reallocation of resources, it was not possible to undertake a case/control study to verify  
450 whether adults and school age children who commute from NWU and ROU to work or  
451 school in Apia are at significantly increased risk for developing typhoid versus age- and  
452 village-matched persons who remain near home.

453       Spatial clustering by genotype 3.5.4 sub-lineages revealed variable degrees of  
454 clustering (Table 2) and considerable overlap among the household clusters where  
455 each sub-lineage was detected (Fig 2A), indicating that the sub-lineages are not  
456 geographically restricted to any area of Samoa. In fact, fluctuations and lapses in the  
457 absolute and relative proportions of the sub-lineages in each census region over time  
458 (Fig 4) indicated exchanges among the regions through a mobile agent (either human  
459 carrier or food/water vehicle) and/or a heterogeneous reservoir emitting distinct  
460 genotypes at different times. The water supplies are not interconnected between  
461 regions, making contaminated products or mobile human carriers more likely to be  
462 responsible. Evidence to support an important role for carriers are the 27 household  
463 contacts of acute cases that were asymptotically shedding *S. Typhi*. The subset  
464 available for WGS analysis (n=14) demonstrated similar phylogenetic distribution of the

465 isolates among each of the three 3.5.4 sub-lineages (Table 3, Fig 3) indicating roughly  
466 even contribution to the dominant networks of infection. Our experience suggests that  
467 asymptomatic shedders are an important and underestimated reservoir of *S. Typhi* in  
468 Samoa, detectable only through active surveillance and detailed case investigations.

469 We also utilized WGS to verify epidemiologic linkages implicated through the  
470 Typhoid SWAT team investigations. Related isolates based on limited genomic SNP  
471 differences would support these linkages. In our analysis, nearly all linkages were  
472 strongly supported by genetic similarity, with a median pairwise SNP distance of 1 SNP,  
473 and demonstrated likely within-household transmission events and contributions by  
474 asymptomatic shedders. These findings are consistent with previous typhoid studies in  
475 endemic regions where WGS was paired with household investigations in Jakarta,  
476 Kathmandu, and Kolkata [37–39] to support the occurrence of within-household  
477 transmission events.

478 There is not a consensus threshold for SNP differences that would associate *S.*  
479 *Typhi* infections in genomic analyses alone. We detected from 0 to 28 pairwise SNP  
480 differences among the 12 putative epidemiologic linkages and 3 instances of repeat  
481 positive cultures from the same individual collected ~1 month apart (Table 4).  
482 Relatedness criteria of 10 SNPs or less [40] would encompass 13/15 (86.7%) linkages,  
483 and 12/13 were within  $\leq 4$  SNPs, a suggested cutoff for *Salmonella enterica* [41]. Future  
484 studies to characterize *S. Typhi* population diversity in acute infections and  
485 asymptomatic carriers by sequencing and comparing multiple colonies obtained from  
486 stool cultures collected over time, as has been completed for other enteric pathogens

487 [42], could aid in establishing appropriate thresholds for SNP-based relatedness among  
488 *S. Typhi* isolates for cluster identification.

489         This study has several limitations. Some *S. Typhi* isolates from cases and  
490 asymptomatic shedders were unavailable for sequencing, resulting in lost genomic data.  
491 Fortunately, all culture-confirmed cases had epidemiologic investigations that included  
492 recording the household GPS coordinates and permitting comparisons of the  
493 sequenced subset with the total collection of cases; no significant difference in patient  
494 sex, mean age, or age groupings was observed (Table 1). Another limitation is that we  
495 generally assumed household coordinates were a reasonable indicator of where an  
496 individual was infected, a necessary assumption for consistency in the spatial analyses.  
497 However, particularly for AUA and NWU cases, there were few epidemiologic  
498 investigations of workplaces of adults or schools of schoolchildren, venues where  
499 transmission may have occurred.

500         Sustained surveillance and institution of routine determination of WGS in Samoa  
501 could help guide interventions during the Consolidation Phase of the STFCP,  
502 which will focus on detection of asymptomatic carriers of *S. Typhi* and possible linked  
503 acute cases. Following completion of the mass vaccination of all Samoans 1-45 years of  
504 age [43], the genotype/sub-lineages of remaining sporadic cases can be compared to  
505 pre-vaccination campaign genotypes/sub-lineages. Findings in the current study build  
506 upon a functional “*Samoan S. Typhi* framework” that can inform public health decision  
507 making and future genomic epidemiology analyses in Samoa as a paradigm for the  
508 understudied region of Oceania.

509

## 510 **ACKNOWLEDGEMENTS**

511 The authors wish to acknowledge the staff at the Ministry of Health of Samoa for their  
512 administrative and technical support of the Samoa Typhoid Fever Control Program, the  
513 staff at the Microbiological Diagnostic Unit Public Health Laboratory, Jane Han for her  
514 logistical support of the Samoa Typhoid Fever Surveillance Initiative, and Yuanyuan  
515 Liang of the Center for Vaccine Development and Global Health for her discussions of  
516 statistical methods.

517

## 518 REFERENCES

- 519 1. Crump JA. Progress in typhoid fever epidemiology. *Clin Infect Dis*. 2019;68: S4–  
520 S9. doi:10.1093/cid/ciy846
- 521 2. Mermin JH, Villar R, Carpenter J, Roberts L, Samariddin A, Gasanova L, et al. A  
522 massive epidemic of multidrug-resistant typhoid fever in Tajikistan associated with  
523 consumption of municipal water. *J Infect Dis*. 1999;179: 1416–1422.  
524 doi:10.1086/314766
- 525 3. Egoz N, Shihab S, Leitner L, Lucian M. An outbreak of typhoid fever due to  
526 contamination of the municipal water supply in northern Israel. *Isr J Med Sci*.  
527 1988;24: 640–3.
- 528 4. Sears SD, Ferreccio C, Levine MM, Cordano AM, Monreal J, Black RE, et al. The  
529 use of Moore swabs for isolation of *Salmonella typhi* from irrigation water in  
530 Santiago, Chile. *J Infect Dis*. 1984;149: 640–642. doi:10.1093/infdis/149.4.640
- 531 5. Shuval HI. Investigation of typhoid fever and cholera transmission by raw  
532 wastewater irrigation in Santiago, Chile. *Water Sci Technol*. 1993;27: 167–174.  
533 doi:10.2166/wst.1993.0341
- 534 6. Walker W. The Aberdeen typhoid outbreak of 1964. *Scott Med J*. 1965;10: 466–  
535 479. doi:10.1177/003693306501001203
- 536 7. Taylor A, Santiago A, Gonzalez-Cortes A, Gangarosa EJ. Outbreak of typhoid  
537 fever in Trinidad in 1971 traced to a commercial ice cream product. *Am J*  
538 *Epidemiol*. 1974;100: 150–157. doi:10.1093/oxfordjournals.aje.a112017
- 539 8. Srinivasan M, Sindhu KN, Giri S, Kumar N, Mohan VR, Grassly NC, et al.  
540 *Salmonella Typhi* shedding and household transmission by children with blood  
541 culture-confirmed typhoid fever in Vellore, South India. *J Infect Dis*. 2021;224:  
542 S593–S600.
- 543 9. Soper GA. The curious career of Typhoid Mary. *Bull N Y Acad Med*. 1939;15:  
544 698–712.
- 545 10. Olsen SJ, Bleasdale SC, Magnano AR, Landrigan C, Holland BH, Tauxe R V., et  
546 al. Outbreaks of typhoid fever in the United States, 1960-99. *Epidemiol Infect*.  
547 2003;130: 13–21. doi:10.1017/s0950268802007598
- 548 11. World Health Organization. Typhoid and paratyphoid fevers from 1950 to 1954.  
549 *Bull World Health Organ*. 1955;13: 173–191.
- 550 12. Cvjetanović B, Grab B, Uemura K. Epidemiological model of typhoid fever and its  
551 use in the planning and evaluation of antityphoid immunization and sanitation  
552 programmes. *Bull World Health Organ*. 1971;45: 53–75.

- 553 13. Ministry of Natural Resources and Environment. National water and sanitation  
554 baseline survey. Apia, Samoa; 2015.
- 555 14. Sikorski MJ, Desai SN, Tupua S, Thomsen RE, Han J, Rambocus S, et al.  
556 Tenacious endemic typhoid fever in Samoa. *Clin Infect Dis*. 2020;71: S120–S126.  
557 doi:10.1093/cid/ciaa314
- 558 15. Sikorski MJ, Hazen TH, Desai SN, Nimarota-Brown S, Tupua S, Sialeipata M, et  
559 al. Persistence of rare *Salmonella* Typhi genotypes susceptible to first-line  
560 antibiotics in the remote islands of Samoa. 2022; In submission.
- 561 16. Gu H, Fan W, Liu K, Qin S, Li X, Jiang J, et al. Spatio-temporal variations of  
562 typhoid and paratyphoid fevers in Zhejiang Province, China from 2005 to 2015.  
563 *Sci Rep*. 2017;7: 1–11. doi:10.1038/s41598-017-05928-3
- 564 17. Ismail K, Maiga G, Ssebuggwaawo D, Nabende P, Mansourian A. Spatio-  
565 temporal trends and distribution patterns of typhoid disease in Uganda from 2012  
566 to 2017. *Geospat Health*. 2020;15: 326–336. doi:10.4081/GH.2020.860
- 567 18. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation  
568 sequencing technologies and their application to the study and control of bacterial  
569 infections. *Clin Microbiol Infect*. 2018;24: 335–341. doi:10.1016/j.cmi.2017.10.013
- 570 19. Parliament of Samoa. Consolidated Acts of Samoa 2010. Health Ordinance 1959,  
571 No.19. Pacific Islands Legal Information Institute; 2010 [cited 15 Jan 2022].  
572 Available: [http://www.pacii.org/ws/legis/num\\_act/haa1969149](http://www.pacii.org/ws/legis/num_act/haa1969149)
- 573 20. Parliament of Samoa. Health Amendment Act 1969, No. 23. Pacific Islands Legal  
574 Information Institute; 1969 [cited 15 Jan 2022]. Available:  
575 [http://www.pacii.org/ws/legis/consol\\_act/ho1959122](http://www.pacii.org/ws/legis/consol_act/ho1959122)
- 576 21. Basnyat B, Qamar FN, Rupali P, Ahmed T, Parry CM. Enteric fever. *BMJ*.  
577 2021;372: n437. doi:10.1136/bmj.n437
- 578 22. Perilla MJ, Bopp C, Wells J. *Salmonella* serotype Typhi: Identification and  
579 antimicrobial susceptibility testing. In: Manual for the laboratory identification and  
580 antimicrobial susceptibility testing of bacterial pathogens of public health  
581 importance in the developing world. Geneva: World Health Organization; 2003.  
582 pp. 103–162.
- 583 23. Strockbine NA, Bopp CA, Fields PI, Kaper JB, Nataro JP. *Escherichia*, *Shigella*,  
584 and *Salmonella*. In: Jorgensen JH, Pfaller MA, Carroll KC, editors. Manual of  
585 clinical microbiology. 11th ed. 11th ed. Washington, DC, USA: ASM Press; 2015.  
586 pp. 685–713. doi:10.1128/9781555817381.ch37
- 587 24. Kalra SP, Naithani N, Mehta SR, Swamy AJ. Current trends in the management  
588 of typhoid fever. *Med J Armed Forces India*. 2003;59: 130–135.  
589 doi:10.1016/S0377-1237(03)80060-6

- 590 25. Esri. ArcGIS Pro v2.9. Redlands, CA: Environmental Systems Research Institute,  
591 Inc.; 2021.
- 592 26. Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial  
593 relationships in populations. *Ecology*. 1954;35: 445–453. doi:10.2307/1931034
- 594 27. Pinder DA, Witherick ME. The principles, practice and pitfalls of nearest-  
595 neighbour analysis. *Geography*. 1972;57: 277–288.
- 596 28. Samoa Bureau of Statistics. 2016 Census Briefs Nos. 1-4. Apia; 2016.
- 597 29. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended  
598 genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human  
599 typhoid. *Nat Commun*. 2016;7: 12827. doi:10.1038/ncomms12827
- 600 30. Dyson ZA, Holt KE. Five years of GenoTyphi: Updates to the global *Salmonella*  
601 Typhi genotyping framework. *J Infect Dis*. 2021;224: S775–S780.  
602 doi:10.1093/infdis/jiab414
- 603 31. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-  
604 analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313.  
605 doi:10.1093/bioinformatics/btu033
- 606 32. Seeman T. Snp-dists v0.8.2. GitHub; 2021. Available:  
607 <https://github.com/tseemann/snp-dists>
- 608 33. Hervé M. Package 'RVAideMemoire' v0.9-81. 2021 [cited 14 Feb 2022].  
609 Available: <https://cran.r-project.org/package=RVAideMemoire>
- 610 34. R Core Team. R: A language and environment for statistical computing. Vienna,  
611 Austria: R Foundation for Statistical Computing; 2022.
- 612 35. Craig AT, Heywood AE, Worth H. Measles epidemic in Samoa and other Pacific  
613 islands. *Lancet Infect Dis*. 2020;20: 273–275. doi:10.1016/S1473-3099(20)30053-  
614 0
- 615 36. UN.ESCAP. Asia-Pacific countries' COVID-19 policy responses: Samoa. 2021.
- 616 37. Baker S, Holt K, Van De Vosse E, Roumagnac P, Whitehead S, King E, et al.  
617 High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing  
618 geographical assignment of haplotypes and pathotypes within an urban district of  
619 Jakarta, Indonesia. *J Clin Microbiol*. 2008;46: 1741–1746.  
620 doi:10.1128/JCM.02249-07
- 621 38. Baker S, Holt KE, Clements ACA, Karkey A, Arjyal A, Boni MF, et al. Combined  
622 high-resolution genotyping and geospatial analysis reveals modes of endemic  
623 urban typhoid fever transmission. *Open Biol*. 2011;1: 110008.  
624 doi:10.1098/rsob.110008

- 625 39. Holt KE, Dutta S, Manna B, Bhattacharya SK, Bhaduri B, Pickard DJ, et al. High-  
626 resolution genotyping of the endemic *Salmonella* Typhi population during a Vi  
627 (typhoid) vaccination trial in Kolkata. PLoS Negl Trop Dis. 2012;6.  
628 doi:10.1371/journal.pntd.0001490
- 629 40. Burnsed LJ, Kovar LD, Angelo KM, Trees EK, Concepción-Acevedo J, McDermott  
630 MD, et al. Use of whole genome sequencing to complement characterisation of a  
631 typhoid fever outbreak among a Marshallese community: Oklahoma, 2015.  
632 Epidemiol Infect. 2019;147: e11. doi:10.1017/S0950268818002601
- 633 41. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome  
634 sequencing options for bacterial strain typing and epidemiologic analysis based  
635 on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin  
636 Microbiol Infect. 2018;24: 350–354. doi:10.1016/j.cmi.2017.12.016
- 637 42. Sahl JW, Sistrunk JR, Fraser CM, Hine E, Baby N, Begum Y, et al. Examination  
638 of the enterotoxigenic *Escherichia coli* population structure during human  
639 infection. Gilmore MS, editor. MBio. 2015;6: 1–6. doi:10.1128/mBio.00501-15
- 640 43. Pauulu U. New typhoid fever vaccine roll out. 2021 Aug 25 [cited 25 Aug 2022].  
641 In: Samoa Observer [Newspaper]. Vaitele, Samoa. Available:  
642 <https://www.samoaobserver.ws/category/samoa/90045>

643

644 **Funding:** This work was supported by a grant [OPP1194582 / INV-000049] from the Bill  
645 & Melinda Gates Foundation (M.M.L.). Under the grant conditions of the Foundation, a  
646 Creative Commons Attribution 4.0 Generic License has already been assigned to the  
647 Author Accepted Manuscript version that might arise from this submission. Research  
648 support was in part from federal funds from National Institutes of Health under National  
649 Institute of Allergy and Infectious Diseases grants F30AI156973 (M.J.S.) and  
650 U19AI110820 (D.A.R.), as well as National Institute of Diabetes and Digestive and  
651 Kidney Diseases training grant T32DK067872 (M.J.S.). M.M.L. is supported in part by  
652 the Simon and Bessie Grollman Distinguished Professorship at the University of  
653 Maryland School of Medicine. The funders had no role in the design and conduct of the  
654 study; collection, management, analysis, and interpretation of the data; preparation,

655 review, or approval of the manuscript; and decision to submit the manuscript for  
656 publication.

657 **Data availability:** Raw sequence data have been submitted to GenBank and accession  
658 numbers of individual isolates are listed in S1 Table. Assembled genomes are available  
659 on Figshare (doi: 10.6084/m9.figshare.18665686).

660 **Competing interests:** M.M.L. reports grants from Bill & Melinda Gates Foundation  
661 during the conduct of the study; in addition, M.M.L. has a patent entitled “Broad  
662 spectrum vaccine against typhoidal and nontyphoidal *Salmonella* disease” (US  
663 9,011,871 B2) issued. R.M.R.-B. reports grants from Bill & Melinda Gates Foundation,  
664 and non-financial support from the Government of Samoa during the conduct of the  
665 study. All other authors report no potential conflicts or competing interests.

666

667 **Supplementary Information:**

668

669 **DETAILED METHODS**

670 Whole genome sequencing, assembly, and mapping

671 Maximum-likelihood phylogeny

672 Spatial and temporal point pattern analyses

673 Average nearest neighbor (ANN) analysis

674 Space-time hotspot analysis

675

676 **SUPPLEMENTARY FIGURES**

677 S1 Fig. Range and interpretation of average nearest neighbor ratio ( $R_n$ ) values

678 S2 Fig. Average nearest neighbor output summary diagram

679 S3 Fig. Maps of average nearest neighbor datasets

680

681 **SUPPLEMENTARY TABLES**

682 S1 Table. Samoa *S. Typhi* isolates analyzed in this study

683