

1 Persistence of rare *Salmonella* Typhi genotypes susceptible to first-line antibiotics in the  
2 remote islands of Samoa

3

4 Michael J. Sikorski<sup>a,b,c</sup>, Tracy H. Hazen<sup>a,c</sup>, Sachin N. Desai<sup>b</sup>, Susana Nimarota-Brown<sup>d</sup>,  
5 Siasoi Tupua<sup>d</sup>, Michelle Sialeipata<sup>d</sup>, Savitra Rambocus<sup>e</sup>, Danielle J. Ingle<sup>f</sup>, Sebastian  
6 Duchene<sup>f</sup>, Susan A. Ballard<sup>e</sup>, Mary Valcanis<sup>e</sup>, Sara Zufan<sup>e,f</sup>, Jianguo Ma<sup>g</sup>, Jason W.  
7 Sahl<sup>h</sup>, Mailis Maes<sup>i</sup>, Gordon Dougan<sup>i</sup>, Robert E. Thomsen<sup>d</sup>, Roy M. Robins-Browne<sup>f,j</sup>,  
8 Benjamin P. Howden<sup>e,f</sup>, Take K. Naseri<sup>d</sup>, Myron M. Levine<sup>b,c</sup>, David A. Rasko<sup>a,c#</sup>

9

10 <sup>a</sup>Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore,  
11 Maryland, United States of America

12 <sup>b</sup>Center for Vaccine Development and Global Health, University of Maryland School of  
13 Medicine, Baltimore, Maryland, United States of America

14 <sup>c</sup>Department of Microbiology and Immunology, University of Maryland School of  
15 Medicine, Baltimore, Maryland, United States of America

16 <sup>d</sup>Ministry of Health, Government of Samoa, Apia, Samoa

17 <sup>e</sup>Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne  
18 at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

19 <sup>f</sup>Department of Microbiology and Immunology at the Peter Doherty Institute for Infection  
20 and Immunity, The University of Melbourne, Melbourne, Victoria, Australia

21 <sup>g</sup>Department of Geographical Sciences, University of Maryland, College Park, Maryland,  
22 United States of America

23 <sup>h</sup>Pathogen & Microbiome Institute, Northern Arizona University, Flagstaff, Arizona,  
24 United States of America

25 <sup>i</sup>Cambridge Institute of Therapeutic Immunology & Infectious Disease, Department of  
26 Medicine, University of Cambridge, Cambridge, United Kingdom

27 <sup>j</sup>Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria,  
28 Australia

29

30 Running Head: Genomics of *Salmonella* Typhi in Samoa

31

32 #Corresponding Author: David A. Rasko, Ph.D.

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

34

35 Alternate corresponding author: Myron M. Levine, M.D., D.T.P.H.

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

37

38 **Abstract word count:** 248

39 **Importance word count:** 110

40 **Text word count:** 5262

41 **Keywords:** typhoid fever; *Salmonella* Typhi; genomics; epidemiology; Oceania; Samoa

42

43 **ABSTRACT**

44 For decades, the remote island nation of Samoa (pop. ~200,000) has faced endemic  
45 typhoid fever despite improvements in water quality, sanitation, and economic  
46 development. We recently described the epidemiology of typhoid fever in Samoa from  
47 2008-2019 by person, place, and time; however, the local *Salmonella enterica* serovar  
48 Typhi (*S. Typhi*) population structure, evolutionary origins, and genomic features  
49 remained unknown. Herein, we report whole genome sequence analyses of 306 *S.*  
50 Typhi isolates from Samoa collected between 1983 and 2020. Phylogenetics revealed a  
51 dominant population of rare genotypes 3.5.4 and 3.5.3, together comprising 292/306  
52 (95.4%) of Samoan versus 2/4934 (0.04%) global *S. Typhi* isolates. Three distinct 3.5.4  
53 genomic sub-lineages were identified and their defining polymorphisms were  
54 determined. These dominant Samoan genotypes, which likely emerged in the 1970s,  
55 share ancestry with other clade 3.5 isolates from South America, Southeast Asia, and  
56 Oceania. Additionally, a 106-kb pHCM2 phenotypically-cryptic plasmid, detected earliest  
57 in a 1992 Samoan *S. Typhi* isolate, was identified in 106/306 (34.6%) of Samoan  
58 isolates; this is more than double the observed proportion of pHCM2-containing isolates  
59 in the global collection. In stark contrast with global *S. Typhi* trends, resistance-  
60 conferring polymorphisms were detected in only 15/306 (4.9%) of Samoan *S. Typhi*,  
61 indicating overwhelming susceptibility to antibiotics that are no longer effective in most  
62 of South and Southeast Asia. This country-level genomic framework can help local  
63 health authorities in their ongoing typhoid surveillance and control efforts, as well as to  
64 fill a critical knowledge gap in *S. Typhi* genomic data from Oceania.

## 65 **IMPORTANCE**

66 In this study we used whole genome sequencing and comparative genomics analyses  
67 to characterize the population structure, evolutionary origins, and genomic features of *S.*  
68 *Typhi* associated with decades of endemic typhoid fever in Samoa. Our analyses of  
69 Samoan isolates from 1983 to 2020 identified a rare *S. Typhi* population in Samoa that  
70 likely emerged around the early 1970s and evolved into sub-lineages that presently  
71 dominate. The dominance and persistence of these endemic genotypes in Samoa are  
72 not readily explained by any apparent genomic competitive advantage or widespread  
73 acquisition of antimicrobial resistance. These data establish the necessary framework  
74 for future genomic surveillance of *S. Typhi* in Samoa for public health benefit.

75

## 76 **INTRODUCTION**

77 Whole genome sequencing (WGS) has become increasingly informative for global  
78 surveillance for *Salmonella enterica* serovar *Typhi* (*S. Typhi*), the causative agent of  
79 typhoid fever (1). Following the completion of the first full genome sequence of *S. Typhi*  
80 in 2001 (2), WGS has proved integral in exploring *S. Typhi* isolate relationships and  
81 relatedness (3), spatial distribution (4), antimicrobial resistance (AMR) (5), virulence (6),  
82 and outbreaks (7). Diminishing sequencing costs has permitted large-scale analysis of  
83 >1800 *S. Typhi* genomes published in 2015 (8, 9). This facilitated large genomic  
84 epidemiology studies of *S. Typhi* collected from endemic countries (10–18) and  
85 returning travelers to non-endemic countries (19–23). Genomics has also been used to  
86 identify and describe the emergence and international spread of multidrug and  
87 extensively drug resistant typhoid fever of great public health concern (11, 24–28), as

88 well as international exchanges of *S. Typhi* (29, 30) primarily among the interconnected  
89 hyper-endemic global sub-regions of South Asia, Southeast Asia, and sub-Saharan  
90 Africa (31).

91 Oceania was recently recognized as the fourth global sub-region with a relatively  
92 high incidence of typhoid fever (31, 32), while reports spanning multiple decades  
93 document the duration of endemic typhoid fever in the region (33–38). Additionally,  
94 Oceania has not been a focus of previous large-scale genomic analyses. Excluding  
95 Australia and New Zealand, Oceania by nature includes a plethora of small remote  
96 island nations scattered across the Pacific Ocean isolated from one another by  
97 hundreds to thousands of kilometers of open water. These island nations, especially in  
98 the sub-region of Polynesia, are primarily rural, have small capital towns, and generally  
99 lack sprawling densely populated metropolitan regions that characteristically sustain  
100 high-incidence endemic typhoid through amplified waterborne transmission (39). It has  
101 been observed that typhoid fever may behave differently amongst these isolated  
102 regions and populations. For example, in a randomized, controlled field trial in Tonga,  
103 acetone-inactivated parenteral typhoid vaccine was only modestly protective (34),  
104 whereas that same vaccine provided markedly higher efficacy in other field sites,  
105 particularly in British Guiana (Guyana) (40), suggesting differences in the host-pathogen  
106 relationship by geography.

107 Samoa, a small island nation (land area ~2,840 km<sup>2</sup>, population ~200,000)  
108 located in the Polynesian sub-region of Oceania, has experienced decades of endemic  
109 typhoid fever of moderate incidence, despite steady improvements in water supplies  
110 and sanitation (41). It is unclear how *S. Typhi* has sustained itself in this remote island

111 population. To explore the population structure and evolutionary origins of *S. Typhi* in  
112 Samoa, we characterized the genomic content of isolates collected from 1983 to 2020  
113 in the context of global and ancestral collections of *S. Typhi*. Herein we identify and  
114 characterize the dominant *S. Typhi* genotype 3.5.4 and nested genotype 3.5.3, which  
115 are exclusive to Samoa and have remained susceptible to antimicrobial agents despite  
116 decades of endemicity and liberal prescribing practices. As with other *S. Typhi* sub-  
117 lineages we can distinguish these three circulating 3.5.4 sub-lineages by single SNPs.  
118 We also estimated the time of emergence of these genotypes to the 1970s and  
119 demonstrated their evolutionary relationship to clade 3.5 isolates from Oceania, South  
120 America, and Southeast Asia.

121

## 122 **METHODS**

123 This study received ethical clearance from the Health Research Committee of the  
124 Ministry of Health (MoH) of Samoa. The University of Maryland, Baltimore (UMB)  
125 Institutional Review Board determined this project, protocol HP-00087489, to be  
126 exempt.

127

### 128 **Study setting**

129 Samoa is comprised of two main populated islands, Upolu and Savaii. The MoH of  
130 Samoa operates two Clinical Microbiology Laboratories, one in Tupua Tamasese  
131 Meaole (TTM) Hospital on Upolu and the other in Malietoa Tanumafili II (MT2) Hospital  
132 on Savaii. Typhoid fever is a notifiable communicable disease in Samoa that can trigger  
133 intense epidemiologic investigation of even individual cases (42, 43). In 2018, the

134 Samoa Typhoid Fever Control Program (TFCP) of the MoH established two Typhoid  
135 Fever Epidemiologic “SWAT” Teams (i.e., equipped with specializes epidemiologic tools  
136 and tactics), one on each island, trained to perform epidemiological investigations of the  
137 household (and/or workplace or school) of every bacteriologically confirmed case of  
138 typhoid fever occurring anywhere in Samoa. The Typhoid SWAT teams ascertain the  
139 clinical status of all contacts in the case household (or workplace, or school), obtain  
140 stool cultures from contacts to detect asymptomatic excretors of *S. Typhi*, assess  
141 environmental risk factors for typhoid transmission, and seek evidence of links to other  
142 typhoid cases.

143

#### 144 **Bacterial isolates sequenced in this study**

145 This study includes contemporary *S. Typhi* isolates collected in Samoa by the MoH  
146 Clinical Microbiology Laboratories and from the Typhoid Fever Public Health Laboratory  
147 (from stool cultures obtained by the Typhoid SWAT Teams). It also includes historical *S.*  
148 *Typhi* isolates collected from travelers who acquired typhoid in Samoa (but were  
149 diagnosed upon return to their home country); these travel-associated isolates were  
150 either made available for sequencing by collaborators or had been previously  
151 characterized (Tables S1 and S2).

152 From April 2018 through June 2020, *S. Typhi* isolates from blood cultures of  
153 acutely febrile patients were bacteriologically confirmed by standard methods (41, 44–  
154 46), tested for antimicrobial susceptibility (AST), and stored at -70°C. Cases of typhoid  
155 fever refer to patients with clinical signs and symptoms of enteric fever (47) that  
156 triggered the collection of a blood culture from which *S. Typhi* was isolated and

157 confirmed. Under the routine public health surveillance activities of the MoH of Samoa,  
158 every case of typhoid fever was rapidly reported to one of the MoH Typhoid SWAT  
159 Teams. During household (or other venue) visits, up to three rectal swabs were  
160 collected ~1-2 days apart from asymptomatic contacts, transported in Cary-Blair  
161 transport media to the Typhoid Fever Public Health Laboratory and examined using  
162 standard bacteriology methods (44, 45). These asymptomatic, subclinical *S. Typhi*  
163 infections detected by stool culture in household contacts of a confirmed acute clinical  
164 case are referred to as asymptomatic shedders.

165 All stored *S. Typhi* isolates from cases and asymptomatic shedders were shipped  
166 from Samoa to the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL)  
167 at The Peter Doherty Institute for Infection and Immunity in Melbourne, Australia for  
168 whole genome sequencing (Table S1). Additionally, from the MDU PHL historical  
169 freezer collection, all 14 previously uncharacterized historical isolates of *S. Typhi*  
170 originating from travelers from Samoa spanning the years 1983-2011 were included in  
171 this study (Table S1).

172

### 173 **Phenotypic antimicrobial susceptibility testing (AST)**

174 At the TTM and MT2 Clinical Microbiology Laboratories, *S. Typhi* blood isolates from  
175 2018 to 2020 were tested for phenotypic AST by the Kirby-Bauer Disc Diffusion method  
176 (48) following the 2009 Clinical and Laboratory Standards Institute (CLSI) breakpoints  
177 (49) until 2019, when a switch was made to the more cost-effective European  
178 Committee for Antimicrobial Susceptibility Testing (EUCAST) guidelines (50).

179



## 180 **DNA extraction and whole genome sequencing (WGS)**

181 DNA extraction and WGS were performed at the MDU PHL following standard  
182 procedures (51). Unique dual-indexed libraries were prepared using the Nextera XT  
183 DNA sample preparation kit (Illumina). Libraries were sequenced on the Illumina  
184 NextSeq 500 with 150-cycle paired-end chemistry. Illumina sequence reads underwent  
185 preliminary inspection using the MDU PHL bioinformatics pipeline, Nullarbor (52).

186

## 187 **Draft genome assembly**

188 Illumina reads were quality filtered and trimmed using Trimmomatic v0.38 (53) and  
189 assembled *de novo* using SPAdes v3.14.1 (54) in Python v2.7.14. The final draft  
190 assemblies were filtered to contain only contigs  $\geq 500$  bp in length and with  $\geq 5x$  k-mer  
191 coverage, as previously described (55). One isolate (AUSMDU00049311) not meeting  
192 quality metrics was excluded from subsequent analyses.

193

## 194 **Read alignment and genotyping**

195 Genotypes were assigned according to GenoTyphi (9, 56). Paired-end Illumina reads  
196 were mapped to the reference strain CT18 (GenBank Accession AL513382.1), using  
197 Snippy v4.6.0 (57). GenoTyphi v1.9.1 was then used to analyze the Snippy output vcf  
198 files. GenoTyphi uses a set of core genome SNPs relative to CT18 to sequentially  
199 classify *S. Typhi* genomes into primary clusters/clades, clades, subclades, and sub-  
200 lineages represented by a numerical sequence of one, two, three, or four digits,  
201 respectively.

202

## 203 **Complete genome sequencing and assembly**

204 Fifteen Samoan *S. Typhi* strains were also selected to undergo long-read sequencing to  
205 obtain a complete genome. Within each genotype of *S. Typhi* presently circulating in  
206 Samoa, the oldest and most recent isolates were selected for genome closure at the  
207 MDU PHL using established pipelines (58, 59). Long-read sequencing libraries were  
208 prepared using the Oxford Nanopore Technologies ligation sequencing kit (SQK-  
209 LSK109) and sequenced on the GridION system using R9.4.1 flow cells. Final  
210 closed/completed assemblies were polished using the full long-read set with Medaka  
211 v1.0.3 (60) and with the short-read set using Pilon v1.23 (61).

212

## 213 ***S. Typhi* genomes from other studies**

214 High-throughput short-read Illumina data for all Samoan *S. Typhi* genomes (n=117) and  
215 all non-Samoan clade 3.5 genomes (n=12) previously published by Wong et al. (8, 9)  
216 were downloaded from the European Nucleotide Archive (ENA) BioProject PRJEB3215.  
217 Duplicate Samoan isolates were removed. Illumina reads for all clade 3.5 genomes from  
218 Colombia (n=20) and Chile (n=68) were downloaded from ENA BioProjects  
219 PRJEB42858 (62) and PRJEB20778 (63), respectively. These *S. Typhi* sequences from  
220 other studies and their accession numbers are listed in Table S2. All short read Illumina  
221 data from these studies were assembled using SPAdes v3.14.1 (54), as described  
222 above.

223

224 **Global collection and *in silico* typing**

225 Pathogenwatch is an online bacterial genome surveillance tool for *S. Typhi* (64, 65) with  
226 *in silico* typing modules for uploaded data and a global collection of published *S. Typhi*  
227 assemblies for comparison (65). All genome assemblies from this study (Tables S1- S3)  
228 were uploaded to Pathogenwatch for plasmid replicons detection using the  
229 PlasmidFinder (66) Enterobacteriaceae database implemented as “IncTyper” in  
230 Pathogenwatch. Genomic predictions of AMR were determined using the  
231 Pathogenwatch curated AMR library, which includes both genes and point mutations  
232 known to confer phenotypic resistances (64).

233

234 **Maximum-likelihood phylogenies**

235 A global phylogeny based on the single nucleotide polymorphisms (SNPs) of the core  
236 sequences of assembled genomes was generated following similar configurations to  
237 Wong et al. (9). For all assemblies in this study and global assemblies downloaded from  
238 Pathogenwatch, SNPs were determined against the reference strain CT18 using NASP  
239 v1.2.0 (67). To exclude regions of recombination, the resultant alignment was analyzed  
240 with Gubbins v2.4.1 (68). RAxML v8.2.10 (69) was run on the PHYLIP format alignment  
241 of filtered polymorphic sites using the generalized time-reversible (GTR) site substitution  
242 model with discrete Gamma ( $\Gamma$ ) distributed rate variation and the Lewis ascertainment  
243 bias correction (ASC\_GTRGAMMA), and 100 bootstrap pseudo-replicates. The  
244 resulting phylogeny was rooted at genotype 0.0 and visualized in iTOL (70).

245 For a clade-specific phylogeny, these steps were repeated with clade 3.5  
246 assembled genomes aligned to a local 2012 Samoan *S. Typhi* reference strain

247 H12ESR00394-001A (GenBank Accession LT904890.1). Pairwise SNP distances  
248 between genomes were calculated using snp-dists v0.8.2 (71). The resulting phylogeny  
249 was midpoint rooted and visualized in iTOL (70). This was repeated with strain CT18,  
250 and unique SNPs defining the Samoan 3.5.4 sub-lineages were manually identified from  
251 global and clade 3.5 NASP matrices using the criteria established in GenoTyphi for the  
252 identification and characterization of canonical SNPs (9).

253

### 254 **Whole genome and pHCM2 gene content analyses**

255 To assess whether any unique or lost genes may impart a competitive advantage to the  
256 Samoa dominant genotypes, a Large Scale BLAST Score Ratio (LS-BSR) analysis (72,  
257 73) was performed on draft and complete genomes (Tables S1-S2) including a diverse  
258 collection of complete reference genomes from GenBank spanning 1916 to 2019,  
259 derived from typhoid endemic regions, and representing most major *S. Typhi*  
260 phylogenetic groups (Table S3). The LS-BSR output matrix (Table S4) encodes a  
261 measure of homology of all predicted protein-coding sequences (genes) across all  
262 genomes in the comparison to examine total gene content. As the *S. Typhi* organism is  
263 genomically conserved and monophyletic (74, 75), a stringent threshold to define a  
264 putative gene as present was set to BSR  $\geq 0.9$ . Less than 0.9 was considered absent.  
265 Scoary v1.6.16 (76) was then used to examine this matrix against multiple variables  
266 (Table S5). Genes with a p-value  $< 0.05$  adjusted with Bonferroni's adjustment method  
267 for multiple comparisons were considered significant. The predicted protein function of  
268 each gene was determined using an in-house annotation pipeline (77).

269 To confirm the presence/absence of the full-length pHCM2 identified by the  
270 IncFIB(pHCM2) replicon in Pathogenwatch, LS-BSR analysis was completed on all  
271 Samoan *S. Typhi* draft genomes using the 106,710 bp pHCM2 molecule from the  
272 complete genome of the 1992 Samoan strain AUSMDU00051359 (GenBank Accession  
273 CP090227; Table S1) annotated in RAST (78–80) as input (Table S7). To assess  
274 homology of the available complete pHCM2 molecules, all four circularized pHCM2-like  
275 molecules present in the completed Samoan *S. Typhi* genomes (this study) were  
276 compared against all five pHCM2-like molecules present in the GenBank reference  
277 collection.

278

### 279 **Temporal and evolutionary analyses**

280 For temporal and evolutionary phylogenetic analyses, raw reads for all clade 3.5  
281 isolates were first aligned to the local Samoan reference strain H12ESR00394-001A  
282 and SNPs were determined using Snippy v4.6.0 (57), with filtering of phage regions  
283 identified using PHASTER (81). The final core SNPs were extracted using Snippy core  
284 v4.6.0 (57) and used to construct a maximum-likelihood (ML) phylogeny in MEGAX  
285 v10.1.8 (82, 83) using the GTR+ $\Gamma$  substitution model and 100 bootstrap pseudo-  
286 replications. The resulting NEWICK tree was examined in TempEst v1.5.3 (84). Tree  
287 tips were assigned their sampling time and a regression analysis of the root-to-tip  
288 branch distances as a function of sampling times was conducted with the best-fitting  
289 root selected according to heuristical residual mean squared, similar to previous studies  
290 (13, 85).

291 To estimate a date of divergence of a most recent common ancestor (MRCA),  
292 the same core alignment was analyzed in BEAST v1.10.4 (86). The molecular clock  
293 model was calibrated using dates of isolation and specified with the GTR+ $\Gamma$  nucleotide  
294 substitution model, uncorrelated relaxed (UCR) clock model, and exponential growth  
295 coalescent tree priors. The number of constant nucleotide patterns were specified as  
296 the number of specific nucleotides. Three independent runs performed using a Markov  
297 chain Monte Carlo length of 200 million steps, sampling every 20,000 steps, were then  
298 combined using LogCombiner v1.10.4 (87) and assessed in Tracer v1.7.1 (88) to verify  
299 that the effective sample size (EES) of all key parameters was at least 200. The  
300 resultant summary maximum clade credibility (MCC) tree was constructed using  
301 TreeAnnotator v1.10.4 (87) after discarding the first 10% of iterations as burn-in and  
302 using median node-height values. The MCC tree was decorated in R v4.1.1 using the  
303 ggtree package (89–91). All above steps were also repeated using the classical  
304 reference genome for *S. Typhi* strain CT18.

305

### 306 **Data availability**

307 All raw sequences and complete assemblies generated in this study have been  
308 deposited on GenBank BioProject PRJNA319593 with the accessions listed in Table  
309 S1. All assemblies analyzed have been deposited on Figshare (doi:  
310 10.6084/m9.figshare.18665686).

311

## 312 **RESULTS**

### 313 **Number and sources of Samoan *S. Typhi* sequences**

314 During the study period of April 2018 to June 2021, the MoH of Samoa stored and  
315 shipped 173 blood isolates from culture-confirmed acute cases of typhoid fever and 15  
316 rectal swab isolates from asymptomatic shedders of *S. Typhi* (Table S1). One 2020  
317 Samoan sequence (AUSMDU00049311) from an acute case was removed due to poor  
318 assembly metrics. Together with the 14 Samoan *S. Typhi* isolates from the MDU PHL  
319 historical freezer collections sequenced in this study (Table S1) and the 105 non-  
320 duplicate Samoan *S. Typhi* previously published by Wong et al. (8) (Table S2), we  
321 examined a final collection of 306 Samoan *S. Typhi*. Most isolates were from 2012 and  
322 2018 to 2020 (Fig. 1).

323

### 324 **Phylogenetic structure of *S. Typhi* from Samoa**

325 The population structure of *S. Typhi* from Samoa from 1983 to 2020 was dominated by  
326 isolates in parent clade 3.5 (Fig. 2A), namely genotypes 3.5.4 (285/306; 93.1%) and  
327 3.5.3 (7/306; 2.3%). Genotypes 3.5.4/3.5.3 appeared to be exclusively associated with  
328 Samoa as they have not yet been reported elsewhere in the world, apart from two  
329 travel-related isolations of unknown origin recovered in Australia in 2011 (8) (Table S2).  
330 Within the Samoa-exclusive genotype 3.5.4, we identified three distinct sub-lineages  
331 numbered 1-3 (genotype designations 3.5.4.1, 3.5.4.2, and 3.5.4.3) (Fig. 2B). These  
332 sub-lineages can be exclusively delineated from other *S. Typhi* by the SNPs listed in  
333 Table 1.

334 A small cluster of Samoan *S. Typhi* isolated from 1996 to 2020 were genotype  
335 4.1 (11/306; 3.6%), and single isolates of genotypes 2.2.1, 2.3.2, and 3.5 were detected  
336 in 2012, 1983, and 1985, respectively (Fig. 2A). During the period of intensified typhoid  
337 surveillance, i.e., 2018 to 2020, only genotypes 3.5.3, 3.5.4, and 4.1 were identified as  
338 circulating in Samoa (Fig. 1).

339 Clade 3.5 (i.e., genotypes 3.5, 3.5.1, 3.5.2, 3.5.3, and 3.5.4) includes isolates  
340 sampled over the past 40 years from Samoa (n=293), Chile (n=68), Colombia (n=20),  
341 Laos (n=5), Fiji (n=3), Indonesia (n=1), Vietnam (n=1) and Australia from unknown  
342 travel (n=2) (Tables S1-S2). A ML phylogeny inferred from all clade 3.5 genomes  
343 demonstrated that the Samoan genotypes 3.5.4 and 3.5.3 do not intermingle with  
344 isolates from these other countries, despite sharing a common core genome indicating  
345 shared evolutionary ancestry (Fig. 2B). The only Samoan isolate of genotype 3.5  
346 (AUSMDU00021621) was isolated in 1983 and is phylogenetically most similar to  
347 isolates from Chile in the 1980s (Fig. 2B). Collectively, these data demonstrate that the  
348 genotypes 3.5.4 and 3.5.3 are exclusive to Samoa and have been present for decades  
349 with only sporadic isolations of alternative genotypes.

350  
351 **Clock-like behavior and date of emergence of common ancestor to Samoa-**  
352 **exclusive *S. Typhi* genotypes**

353 Root-to-tip regression analysis as a function of sampling time was performed on the  
354 clade 3.5 ML tree using TempEst (84) with the best-fitting root selected (Fig. 3A). We  
355 observed a positive temporal association between year isolated and root-to-tip



356 evolutionary distance, with an  $R^2$  value of 0.302. This regression coefficient is  
357 comparable to those reported in other studies of *S. Typhi* (13, 29, 85).

358 To infer a date of emergence of the MRCA of genotypes 3.5.4 and 3.5.3, we  
359 analyzed the summary statistics and MCC tree output of our clade 3.5 BEAST analysis  
360 (87) (Fig. 2B). The date of emergence of the MRCA to the dominant Samoa genotypes  
361 (3.5.4 and 3.5.3) was estimated to be around 1974 (95% Highest Posterior Density  
362 [HPD]: 1968, 1980). The substitution rate (meanRate parameter in BEAST) was  
363  $5.82 \times 10^{-8}$  (95% HPD:  $5.0 \times 10^{-8}$ ,  $6.6 \times 10^{-8}$ ) genome-wide substitutions per site per year, or  
364  $\sim 0.27$  (95% HPD: 0.24, 0.32) substitutions per genome per year. Compared with a rate  
365 of  $\sim 0.8$  SNPs per genome per year calculated for genotype 4.3.1 (13, 85) (previously  
366 designated the H58 haplotype (8)), these estimates represent a reduced rate of SNP  
367 acquisition across the core genome in the dominant Samoan genotypes compared to  
368 genotypes in other geographic locations.

369

### 370 **Unique gene-content analysis**

371 To determine if any genetic characteristics might explain the persistence and endemic  
372 nature of the *S. Typhi* in Samoa, an all-versus-all gene-by-gene comparative analysis  
373 was performed using LS-BSR (72, 73) including all Samoan and clade 3.5 isolates  
374 (Tables S1-2) versus a diverse collection of 47 complete isolates from GenBank (Table  
375 S3). The number and annotation of significantly differential genes (uniquely present or  
376 absent with  $p < 0.05$  adjusted with Bonferroni's method for multiple comparisons  
377 correction) were identified based on metadata groupings. There were  $>500$  differential  
378 genes among *S. Typhi* of Samoan origin versus non-Samoan isolates, as well as in the

379 Samoan genotypes 3.5.4/3.5.3 versus other genotypes. Each 3.5.4 sub-lineage  
380 contained between 200-270 differential genes distinguishing them from the other sub-  
381 lineages and genotypes. With respect to time, ~200 differential genes were encoded in  
382 *S. Typhi* from the 1980s compared to only <15 across more recent decades. These  
383 differential genes by each trait are detailed in Table S5. Within each trait, roughly 30-  
384 60% of the significant differential genes were hypothetical or putative in their functional  
385 annotation, which is not uncommon. Of the genes uniquely present in Samoa or  
386 genotypes 3.5.4/3.5.3, which are most descriptive of Samoan *S. Typhi*, the functional  
387 annotations predominantly included regulatory genes, metabolic factors, and mobile  
388 elements (Table S5). No genes known to confer enhanced virulence or antimicrobial  
389 resistance to first-line antibiotics were detected.

390

### 391 **Antimicrobial resistance**

392 Among the Samoan *S. Typhi* isolates, 15/306 (4.9%) contained a point mutation in DNA  
393 gyrase subunit A, a quinolone resistance determining region (QRDR) conferring  
394 predicted intermediate resistance to fluoroquinolones (64), which are commonly  
395 prescribed in the treatment of typhoid fever (47) (Fig. 1B, Table S6). These were *gyrA*-  
396 S83F (n=10/15), *gyrA*-D87G (n=4/15), and *gyrA*-D87N (n=1/15). Of these 15 isolates  
397 with predicted resistance, ten were isolated in Samoa in the Clinical Microbiology  
398 Laboratories and phenotypic AST assays for ciprofloxacin were completed and  
399 available for eight of ten isolates. All eight isolates were reported susceptible to  
400 ciprofloxacin (Table S6) indicating disagreement with the predicted intermediate  
401 resistance and the genotype. Azithromycin is an oral macrolide drug also commonly

402 prescribed in the treatment of typhoid fever (47). An additional point mutation in  
403 acriflavin resistance protein B gene, *acrB*-R717Q, known to limit susceptibility to  
404 azithromycin (15), was also identified in the isolate containing the *gyrA*-D87N point  
405 mutation. Unfortunately, we did not have access to this isolate to perform AST because  
406 it was collected in New Zealand from a traveler from Samoa.

407

#### 408 **Comparative analysis of pHCM2 plasmid**

409 Using IncTyper in Pathogenwatch (64), the plasmid replicon corresponding to  
410 IncFIB(pHCM2) was detected in 106/306 (34.6%) of Samoan *S. Typhi* isolates (Figs. 1B  
411 and 4A). This replicon is significantly ( $p < 0.01$ , chi-square test of independence)  
412 associated with *S. Typhi* from Samoa, compared with 618/4934 (12.5%) of non-Samoan  
413 global *S. Typhi*. The pHCM2 sequence from the 1994 Samoan isolate,  
414 AUSMDU00051359, was completed and circularized and is 106,710 bp in length  
415 (GenBank Accession CP090228; Table S1). The plasmid contains one tRNA gene  
416 sequence and 129 genes, of which 89/130 (68.5%) encode a hypothetical protein with  
417 no known function (Table S7). The remaining 41 genes include functions for replication,  
418 central metabolism or mobile elements. No genes for antimicrobial resistance or  
419 virulence were detected. The full-length pHCM2 was confirmed in three other complete  
420 Samoan genome sequences from 2008, 2019, and 2020 selected for closure (Table  
421 S1). All four completed circularized plasmids were identical in length (Table S1) and  
422 gene content (Fig. 4), despite more than 25 years between the oldest and most recent  
423 isolation.

424 A focused LS-BSR analysis of the 130 pHCM2 genes against all 306 Samoan *S.*  
425 *Typhi* strains (Fig. 4A) revealed the presence of the pHCM2 genes in 106/306 (34.6%)  
426 of the isolates. In two isolates, 10071\_8\_49\_Sam0049\_2012 and AUSMDU00049494,  
427 LS-BSR detected the majority of the gene content of the pHCM2, but the >90%  
428 coverage of the replicon region required for IncTyper (64, 65) was lacking from the draft  
429 assemblies, resulting in omission from identification via incompatibility typing. In  
430 contrast, in the 1983 Samoan isolate AUSMDU00021620, IncTyper identified the IncFIB  
431 replicon, but the pHCM2 gene content was absent from the analyzed assembly. The  
432 remaining replicon determination resulted in an agreement of the replicon and the  
433 presence of remainder of the pHCM2 gene content.

434 Comparison of plasmid gene content between the four complete Samoan  
435 pHCM2 molecules and five GenBank complete reference sequences identified a near  
436 complete plasmid homology (Fig. 4B). Only two hypothetical genes, 220341.113.peg.18  
437 and 220341.113.peg.2 (Table S7) demonstrated divergence in pHCM2 from strain CT18  
438 compared with the pHCM2 molecules of four Samoan strains, indicating a significant  
439 level of cryptic plasmid conservation.

440

## 441 **DISCUSSION**

442 The phylogenetic analyses in this study demonstrate that 95.4% (292/306) of *S. Typhi*  
443 isolates from Samoa, spanning 1983 to 2020, are of rare genotypes 3.5.4 and 3.5.3  
444 (Fig. 1-2, Table S1). Among non-Samoan *S. Typhi*, genotypes 3.5.4 and 3.5.3 were  
445 recorded through passive surveillance in only two Australian isolates of unknown travel  
446 origin (Fig. 2A). Considering that 41/306 (13.4%) of the Samoan *S. Typhi* were collected

447 in Australia from travelers to Samoa (Tables S1 and S2) and the frequent migration and  
448 travel between Samoans and their extended families living in Australia, it is plausible  
449 that these two unknown isolates of genotypes 3.5.4 and 3.5.3 may have been acquired  
450 in Samoa. Additionally, the intermingling of these imported isolates of unknown origin  
451 amongst the Samoan *S. Typhi* further supports this putative association (Fig. 2B).

452 Therefore, we consider the genotypes 3.5.4 and 3.5.3 as globally exclusive to Samoa.

453 Within these dominant genotypes, we defined three discrete genotype 3.5.4 sub-  
454 lineages circulating in Samoa via core genomic variation and canonical SNPs (Fig. 2B;  
455 Table 1). These sub-lineages represent distinct core genome evolution detectable by  
456 comparative genomic analysis. Including the relatively infrequent genotypes 3.5.3 and  
457 4.1 in this Samoan *S. Typhi* collection, a total of five genotypes/sub-lineages of Samoan  
458 *S. Typhi* distinguishable by single SNPs form a functional WGS framework with public  
459 health utility. Specifically, future isolates may be sequenced, rapidly categorized by  
460 genotype/sub-lineage, and utilized to infer putative linkages with other genomically  
461 related isolates for targeted investigation by the Samoan Typhoid SWAT teams (92).

462 Phylogenetic ancestry of genotypes 3.5.4/3.5.3 is shared with clade 3.5 *S. Typhi*  
463 isolates from countries in Oceania, South America, and Southeast Asia (Fig. 2B).

464 Genotypes 3.5.4/3.5.3 are predicted to have emerged from the clade 3.5 genomic  
465 background in ~1974 (95% HPD: 1968-1980) (Fig. 3B). However, *S. Typhi* has been  
466 present in Samoa decades prior. Endemic typhoid fever in Samoa was recorded in  
467 World Health Organization surveillance data in the 1950s (93) and in epidemiological  
468 modeling data collected in the 1960s (33). These records corroborate a persistently  
469 endemic situation predating the estimated emergence of genotypes 3.5.4/3.5.3. It is not

470 known whether *S. Typhi* in Samoa in the 1950s and 1960s were predecessors of the  
471 3.5.4/3.5.3 Samoan *S. Typhi* or other genotypes that were eventually displaced by the  
472 3.5.4/3.5.3 Samoan *S. Typhi*.

473 The estimated core genome substitution rate of clade 3.5 isolates is ~0.27 SNPs  
474 per genome per year, which is approximately 3-fold less than previous *S. Typhi* studies  
475 focusing on the rapidly spreading multiple drug resistant H58/genotype 4.3.1 (13, 85).  
476 These results may indicate different environmental pressures in Samoa or an  
477 intrinsically more slowly evolving *S. Typhi* population in a geographically remote island  
478 setting. Additionally, it is remarkable that while there have been sporadic isolations of  
479 non-3.5.4/3.5.3 in Samoa since 1983, and most likely before, there has not been the  
480 replacement of the 3.5.4/3.5.3 genotypes by these incoming external genotypes. This  
481 suggests that there may be other factors, such as host genetics (94, 95), microbiome  
482 (96), or environmental factors (97), as well as unknown mechanisms of transmission,  
483 that support the maintenance of the 3.5.4/3.5.3 genotypes in Samoa and their  
484 endurance, heretofore, despite classical typhoid control efforts.

485 We have assessed Samoan *S. Typhi* for intrinsic factors that may provide such a  
486 competitive or survival advantage. While many putative protein-coding sequences  
487 missed through mapping analyses were uniquely present in the Samoan *S. Typhi*  
488 isolates and absent from the diverse set of reference genomes as well as vice-versa,  
489 the functions of most of these genes are currently unknown. An ~106-kb phenotypically-  
490 cryptic pHCM2 molecule that does not encode resistance or virulence factors (98) was  
491 detected in Samoan *S. Typhi* isolates as early as 1994 and has retained identical  
492 plasmid length and gene content 25 years later (Fig. 4B). The proportion of Samoan *S.*

493 Typhi bearing IncFIB(pHCM2) was more than double the observed proportion of  
494 pHCM2-containing isolates in the global collection. This significant ( $p < 0.01$ ) enrichment  
495 of this conserved plasmid in the Samoan *S. Typhi* population may be due to enhanced  
496 plasmid acquisition and stability, an intrinsic advantage for its maintenance, or an  
497 artifact of the local *S. Typhi* population. Our genomic content analyses of pHCM2 from  
498 Samoa, India, Bangladesh, and Vietnam (Fig. 4B) revealed numerous genes encoding  
499 hypothetical proteins, and a limited number of proteins associated with gene regulation,  
500 central metabolism or mobile elements, but none are predicted to impart an advantage  
501 based on the current functional annotations (Table S7). This remarkable gene  
502 conservation across significant geographic distances and time suggests that some  
503 functions beneficial to their persistence or to the *S. Typhi* host may be encoded on this  
504 plasmid, but these are yet to be elucidated.

505 Finally, we observed that there has been limited development of AMR in Samoa,  
506 as mutations associated with AMR were detected in only 15/306 (4.9%) of Samoan  
507 isolates (Fig. 2, Table S1-S2). Pan-susceptibility to first-line antibiotics was identified in  
508 2018 to 2020 by the clinical microbiology laboratories in Samoa (Table S1) and our  
509 examination of genomic features associated with antibiotic resistance largely  
510 corroborates this observation. There was incongruence between genotype and  
511 phenotype observed; however, the reason for this incongruence is presently unknown.  
512 This finding should give pause to assigning phenotypic resistance and to making clinical  
513 or regulatory decisions based solely on WGS data. A limited study of antibiotic  
514 prescriptions and usage in 2004 in Samoa revealed higher levels of prescribing (66.4%  
515 percent of prescriptions included an antibiotic) and a greater reliance on penicillins (63%

516 of defined daily doses per 1000 inhabitants per day) compared with other low- and  
517 middle-income countries (99). Unmeasured over-the-counter sales of antibiotics were  
518 also reported (99). There appears to be less AMR development among the endemic  
519 Samoan strains of *S. Typhi* when compared to other countries in spite of liberal  
520 antibiotic usage. It is currently unclear if this is associated with the dominant and stable  
521 3.5.4/3.5.3 genotypes or the patterns of antibiotic usage.

522         Limitations of this study include inconsistent sampling over time, which is an  
523 issue with many genomic studies in resource-restricted locations. Therefore, we relied  
524 on historical travel-associated isolates collected in Australia and New Zealand where  
525 detailed epidemiologic data was not always available. For example, singular historical  
526 isolations of genotypes 2.2.1, 2.3.2, and 3.5 in 2012, 1983, and 1985, respectively, were  
527 detected in individuals in New Zealand and Australia who had recent travel history to  
528 Samoa. These could indicate under-sampled genotypes that were once common in  
529 Samoa. However, these genotypes are identified globally and could have been  
530 mischaracterized as Samoa-associated importations if other travel linkages were not  
531 disclosed. Since 2018, the Samoa TFCP has paired WGS with improved blood culture  
532 surveillance and household investigations to mitigate this limitation and provide  
533 contiguous years of genome sequencing. Future investment in local sequencing  
534 capacity and implementation of rapid bioinformatics pipelines of public health utility  
535 would greatly strengthen local and regional infectious disease surveillance with  
536 complementary WGS data.

537         In sum, we have determined that typhoid fever in Samoa is caused by a Samoa-  
538 restricted *S. Typhi* population comprised of rare but stable genotypes 3.5.4 and 3.5.3



539 that have not developed antimicrobial resistance over decades of endemicity and  
540 suboptimal antibiotic prescribing practices. However, this relatively simple situation,  
541 where first-line drugs remain effective in treating typhoid fever in Samoa, could  
542 suddenly change. *S. Typhi* belonging to H58/genotype 4.3.1 encoding drug resistance,  
543 as well as enhanced virulence (8), have emerged and spread internationally (24, 25,  
544 100). While genotype 4.3.1 was not detected in this collection of Samoan *S. Typhi*,  
545 Samoa has endured persistently endemic typhoid fever for decades and could be  
546 vulnerable to the development of antimicrobial resistance in an endemic isolate or the  
547 importation and dissemination of a resistant clone similar to H58/genotype 4.3.1

548 Through this study, we have established a Samoa-specific *S. Typhi* genomic  
549 framework that will allow for enhanced typhoid surveillance. For example, pairing these  
550 WGS data with traditional epidemiology and geospatial clustering analyses can provide  
551 insights into the transmission mechanics of *S. Typhi* in Samoa, which may be applicable  
552 to other islands in Oceania. Such data can inform the setting of interventions to be  
553 carried out by the Samoa MoH and can help monitor their progress, such as assessing  
554 the *S. Typhi* genomic population structure before and after mass vaccination with Vi  
555 conjugate vaccine, which commenced in Samoa in August 2021 (101). Should AMR *S.*  
556 *Typhi* emerge in Samoa, timely WGS analysis can likely provide a genomic examination  
557 of the phenotypic AMR, which may implicate the geographic origin (locally or  
558 internationally). Finally, once the incidence of acute cases is drastically reduced in high  
559 burden age groups following the current mass vaccination program, WGS data can  
560 assist the MoH during the Consolidation Phase of the Samoa TFPCP, which is intended

561 to actively seek out silent chronic typhoid carriers who will constitute the long-term  
562 human reservoir of *S. Typhi*.

563

## 564 **ACKNOWLEDGEMENTS**

565 The authors wish to acknowledge the staff at the Ministry of Health of Samoa for their  
566 administrative and technical support of the Samoa Typhoid Fever Control Program, the  
567 staff at the Microbiological Diagnostic Unit Public Health Laboratory, Jane Han for her  
568 logistical support of the Samoa Typhoid Fever Surveillance Initiative, and Corin Yeats  
569 for his technical support with Pathogenwatch.

570 This work was supported by the Bill & Melinda Gates Foundation OPP1194582  
571 (INV-000049) (PI: Prof. Myron M. Levine). Under the grant conditions of the Foundation,  
572 a Creative Commons Attribution 4.0 Generic License has already been assigned to the  
573 Author Accepted Manuscript version that might arise from this submission. M. J. S.  
574 received research support in part by federal funds from National Institutes of Health  
575 under National Institute of Allergy and Infectious Diseases grants F30AI156973 (PI:  
576 Michael J. Sikorski) and U19AI110820 (PI: David A. Rasko), as well as National Institute  
577 of Diabetes and Digestive and Kidney Diseases training grant T32DK067872 (PI: Jean-  
578 Pierre Raufman). M. M. L. is supported in part by the Simon and Bessie Grollman  
579 Distinguished Professorship at the University of Maryland School of Medicine. The  
580 funders had no role in the design and conduct of the study; collection, management,  
581 analysis, and interpretation of the data; preparation, review, or approval of the  
582 manuscript; and decision to submit the manuscript for publication.

583 M. M. L. reports grants from Bill & Melinda Gates Foundation during the conduct  
584 of the study; in addition, M. M. L. has a patent entitled “Broad spectrum vaccine against  
585 typhoidal and nontyphoidal *Salmonella* disease” (US 9,011,871 B2) issued. R. M. R.-B.  
586 reports grants from Bill & Melinda Gates Foundation, and non-financial support from the  
587 Government of Samoa during the conduct of the study. All other authors report no  
588 potential conflicts.  
589

590 **REFERENCES**

- 591 1. Dougan G, Baker S. 2014. *Salmonella enterica* serovar Typhi and the  
592 pathogenesis of typhoid fever. *Annu Rev Microbiol* 68:317–336.
- 593 2. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C,  
594 Mungall KL, Bentley SD, Holden MTGG, Sebaihia M, Baker S, Basham D, Brooks  
595 K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N,  
596 Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A,  
597 Larsen TS, Leather S, Moule S, Ó'Gaora P, Parry C, Quail M, Rutherford K,  
598 Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Complete  
599 genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi  
600 CT18. *Nature* 413:848–852.
- 601 3. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, Rance R,  
602 Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G. 2008. High-  
603 throughput sequencing provides insights into genome variation and evolution in  
604 *Salmonella* Typhi. *Nat Genet* 40:987–993.
- 605 4. Baker S, Holt KE, Clements ACA, Karkey A, Arjyal A, Boni MF, Dongol S,  
606 Hammond N, Koirala S, Duy PT, Nga TVT, Campbell JI, Dolecek C, Basnyat B,  
607 Dougan G, Farrar JJ. 2011. Combined high-resolution genotyping and geospatial  
608 analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol*  
609 1:110008.
- 610 5. Kariuki S, Revathi G, Kiiru J, Mengo DM, Mwituria J, Muyodi J, Munyalo A, Teo  
611 YY, Holt KE, Kingsley RA, Dougan G. 2010. Typhoid in Kenya is associated with  
612 a dominant multidrug-resistant *Salmonella enterica* serovar Typhi haplotype that

- 613 is also widespread in Southeast Asia. J Clin Microbiol 48:2171–2176.
- 614 6. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming  
615 clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601–  
616 612.
- 617 7. Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL. 2015. Whole-genome  
618 sequencing in outbreak analysis. Clin Microbiol Rev 28:541–563.
- 619 8. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, Kingsley RA,  
620 Thomson NR, Keane JA, Weill FX, Edwards DJ, Hawkey J, Harris SR, Mather  
621 AE, Cain AK, Hadfield J, Hart PJ, Thieu NTV, Klemm EJ, Glinos DA, Breiman RF,  
622 Watson CH, Kariuki S, Gordon MA, Heyderman RS, Okoro C, Jacobs J, Lunguya  
623 O, Edmunds WJ, Msefula C, Chabalgoity JA, Kama M, Jenkins K, Dutta S, Marks  
624 F, Campos J, Thompson C, Obaro S, Maclennan CA, Dolecek C, Keddy KH,  
625 Smith AM, Parry CM, Karkey A, Mulholland EK, Campbell JI, Dongol S, Basnyat  
626 B, Dufour M, Bandaranayake D, Naseri TT, Singh SP, Hatta M, Newton P,  
627 Onsare RS, Isaia L, Dance D, Davong V, Thwaites G, Wijedoru L, Crump JA, De  
628 Pinna E, Nair S, Nilles EJ, Thanh DP, Turner P, Soeng S, Valcanis M, Powling J,  
629 Dimovski K, Hogg G, Farrar J, Holt KE, Dougan G. 2015. Phylogeographical  
630 analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi  
631 identifies inter- and intracontinental transmission events. Nat Genet 47:632–639.
- 632 9. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, Murphy N, Holliman  
633 R, Sefton A, Millar M, Dyson ZA, Dougan G, Holt KE. 2016. An extended  
634 genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human  
635 typhoid. Nat Commun 7:12827.

- 636 10. Pham Thanh D, Thompson CN, Rabaa MA, Sona S, Sopheary S, Kumar V,  
637 Moore C, Tran Vu Thieu N, Wijedoru L, Holt KE, Wong V, Pickard D, Thwaites  
638 GE, Day N, Dougan G, Turner P, Parry CM, Baker S. 2016. The molecular and  
639 spatial epidemiology of typhoid fever in rural Cambodia. *PLoS Negl Trop Dis*  
640 10:e0004785.
- 641 11. Klemm EJ, Shakoor S, Page AJ, Qamar FN, Judge K, Saeed DK, Wong VK,  
642 Dallman TJ, Nair S, Baker S, Shaheen G, Qureshi S, Yousafzai MT, Saleem MK,  
643 Hasan Z, Dougan G, Hasan R. 2018. Emergence of an extensively drug-resistant  
644 *Salmonella enterica* serovar Typhi clone harboring a promiscuous plasmid  
645 encoding resistance to fluoroquinolones and third-generation cephalosporins.  
646 *MBio* 9:e00105-18.
- 647 12. Liaquat S, Sarwar Y, Ali A, Haque A, Farooq M, Martinez-Ballesteros I, Laorden  
648 L, Garaizar J, Bikandi J. 2018. Virulotyping of *Salmonella enterica* serovar Typhi  
649 isolates from Pakistan: Absence of complete SPI-10 in Vi negative isolates. *PLoS*  
650 *Negl Trop Dis* 12:e0006839.
- 651 13. Britto CD, Dyson ZA, Duchene S, Carter MJ, Gurung M, Kelly DF, Murdoch DR,  
652 Ansari I, Thorson S, Shrestha S, Adhikari N, Dougan G, Holt KE, Pollard AJ.  
653 2018. Laboratory and molecular surveillance of paediatric typhoidal *Salmonella* in  
654 Nepal: Antimicrobial resistance and implications for vaccine policy. *PLoS Negl*  
655 *Trop Dis* 12:e0006408.
- 656 14. Tanmoy AM, Westeel E, De Bruyne K, Goris J, Rajoharison A, Sajib MSI, Van  
657 Belkum A, Saha SK, Komurian-Pradel F, Endtz HP, Bruyne K De, Goris J,  
658 Rajoharison A, Sajib MSI, Belkum A van, Saha SK, Komurian-Pradel F, Endtz

- 659 HP. 2018. *Salmonella enterica* serovar Typhi in Bangladesh: Exploration of  
660 genomic diversity and antimicrobial resistance. MBio 9:e02112-18.
- 661 15. Hooda Y, Sajib MSI, Rahman H, Luby SP, Bondy-Denomy J, Santosham M,  
662 Andrews JR, Saha SK, Saha S. 2019. Molecular mechanism of azithromycin  
663 resistance among typhoidal *Salmonella* strains in Bangladesh identified through  
664 passive pediatric surveillance. PLoS Negl Trop Dis 13:e0007868.
- 665 16. Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE, Chowdhury EK, Dougan  
666 G, Qadri F. 2020. Population structure and antimicrobial resistance patterns of  
667 *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. PLoS  
668 Negl Trop Dis 14:e0008036.
- 669 17. Wang K-Y, Lee D-J, Shie S-S, Chen C-J. 2019. Population structure and  
670 transmission modes of indigenous typhoid in Taiwan. BMC Med Genomics  
671 12:126.
- 672 18. Britto CD, Dyson ZA, Mathias S, Bosco A, Dougan G, Jose S, Nagaraj S, Holt KE,  
673 Pollard AJ. 2019. Persistent circulation of a fluoroquinolone-resistant *Salmonella*  
674 *enterica* Typhi clone in the Indian subcontinent. J Antimicrob Chemother 75:337–  
675 341.
- 676 19. Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA, Day M, Freedman J,  
677 Chattaway MA, Holt KE, Dallman TJ. 2019. Informal genomic surveillance of  
678 regional distribution of *Salmonella* Typhi genotypes and antimicrobial resistance  
679 via returning travellers. PLoS Negl Trop Dis 13:e0007620.
- 680 20. Wong W, Rawahi H Al, Patel S, Yau Y, Eshaghi A, Zittermann S, Tattum L, Morris  
681 SK. 2019. The first Canadian pediatric case of extensively drug-resistant

- 682 *Salmonella* Typhi originating from an outbreak in Pakistan and its implication for  
683 empiric antimicrobial choices. IDCases 15:e00492.
- 684 21. Engsbro AL, Riis Jespersen HS, Goldschmidt MI, Mollerup S, Worning P,  
685 Pedersen MS, Westh H, Schneider UV. 2019. Ceftriaxone-resistant *Salmonella*  
686 *enterica* serotype Typhi in a pregnant traveller returning from Karachi, Pakistan to  
687 Denmark, 2019. Eurosurveillance 24:1–5.
- 688 22. Meyer Sauter PM, Stevens MJA, Paioni P, Wüthrich D, Egli A, Stephan R,  
689 Berger C, Bloemberg G V. 2019. Siblings with typhoid fever: An investigation of  
690 intrafamilial transmission, clonality, and antibiotic susceptibility. Travel Med Infect  
691 Dis 101498.
- 692 23. Procaccianti M, Motta A, Giordani S, Riscassi S, Guidi B, Ruffini M, Maffini V,  
693 Esposito S, Dodi I. 2020. First case of typhoid fever due to extensively drug-  
694 resistant *Salmonella enterica* serovar Typhi in Italy. Pathogens 9:151.
- 695 24. Levine MM, Simon R. 2018. The gathering storm: Is untreatable typhoid fever on  
696 the way? MBio 9:e00482-18.
- 697 25. Dyson ZA, Klemm EJ, Palmer S, Dougan G. 2019. Antibiotic resistance and  
698 typhoid. Clin Infect Dis 68:S165–S170.
- 699 26. Duy PT, Dongol S, Giri A, Nguyen To NT, Dan Thanh HN, Nhu Quynh NP, Duc  
700 Trung P, Thwaites GE, Basnyat B, Baker S, Rabaa MA, Karkey A. 2020. The  
701 emergence of azithromycin-resistant *Salmonella* Typhi in Nepal. JAC Antimicrob  
702 Resist 2:1–4.
- 703 27. Shin E, Park J, Jeong HJ, Park AK, Na K, Lee H, Chun J, Hwang KJ, Kim C-J,  
704 Kim J. 2021. Emerging high-level ciprofloxacin-resistant *Salmonella enterica*



- 705 serovar Typhi haplotype H58 in travelers returning to the Republic of Korea from  
706 India. PLoS Negl Trop Dis 15:e0009170.
- 707 28. Carey ME, Jain R, Yousuf M, Maes M, Dyson ZA, Thu TNH, Nguyen Thi Nguyen  
708 T, Ho Ngoc Dan T, Nhu Pham Nguyen Q, Mahindroo J, Thanh Pham D, Sandha  
709 KS, Baker S, Taneja N. 2021. Spontaneous emergence of azithromycin  
710 resistance in independent lineages of *Salmonella* Typhi in northern India. Clin  
711 Infect Dis 72:e120–e127.
- 712 29. Park SE, Pham DT, Boinett C, Wong VK, Pak GD, Panzner U, Espinoza LMC,  
713 von Kalckreuth V, Im J, Schütt-Gerowitt H, Crump JA, Breiman RF, Adu-Sarkodie  
714 Y, Owusu-Dabo E, Rakotozandrindrainy R, Soura AB, Aseffa A, Gasmelseed N,  
715 Keddy KH, May J, Sow AG, Aaby P, Biggs HM, Hertz JT, Montgomery JM,  
716 Cosmas L, Olack B, Fields B, Sarpong N, Razafindrabe TJL, Raminosoa TM,  
717 Kabore LP, Sampo E, Teferi M, Yeshitela B, El Tayeb MA, Sooka A, Meyer CG,  
718 Krumkamp R, Dekker DM, Jaeger A, Poppert S, Tall A, Niang A, Bjerregaard-  
719 Andersen M, Løfberg SV, Seo HJ, Jeon HJ, Deerin JF, Park J, Konings F, Ali M,  
720 Clemens JD, Hughes P, Sendagala JN, Vudriko T, Downing R, Ikumapayi UN,  
721 Mackenzie GA, Obaro S, Argimon S, Aanensen DM, Page A, Keane JA, Duchene  
722 S, Dyson Z, Holt KE, Dougan G, Marks F, Baker S. 2018. The phylogeography  
723 and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. Nat  
724 Commun 9:5094.
- 725 30. da Silva KE, Tanmoy AM, Pragasam AK, Iqbal J, Islam Sajib MS, Mutreja A,  
726 Veeraraghavan B, Tamrakar D, Qamar FN, Dougan G, Bogoch I, Seidman JC,  
727 Shakya J, Vaidya K, Carey ME, Shrestha R, Irfan S, Baker S, Luby SP, Cao Y,

- 728 Dyson ZA, Garrett DO, John J, Kang G, Hooda Y, Saha SK, Saha S, Andrews JR.  
729 2021. The international and intercontinental spread and expansion of  
730 antimicrobial-resistant *Salmonella* Typhi. medRxiv 2021.09.03.21262852.
- 731 31. GBD 2017 Typhoid and Paratyphoid Collaborators. 2019. The global burden of  
732 typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of  
733 Disease Study 2017. Lancet Infect Dis 19:369–381.
- 734 32. World Health Organization. 2018. Typhoid vaccines: WHO position paper – March  
735 2018. Wkly Epidemiol Rec 93:153–172.
- 736 33. Cvjetanović B, Grab B, Uemura K. 1971. Epidemiological model of typhoid fever  
737 and its use in the planning and evaluation of antityphoid immunization and  
738 sanitation programmes. Bull World Health Organ 45:53–75.
- 739 34. Tapa S, Cvjetanović B. 1975. Controlled field trial on the effectiveness of one and  
740 two doses of acetone-inactivated and dried typhoid vaccine. Bull World Health  
741 Organ 52:75–80.
- 742 35. Hirshman JH. 1976. Communicable disease in the South Pacific islands. Med J  
743 Aust 2:758–760.
- 744 36. Tangitau A. 1997. Typhoid in a Tongan village. Pac Health Dialog 3:208–210.
- 745 37. Lutui T, Ofanoa M, Finau SA, Maika MK. 1999. Typhoid fever in Tonga. Pac  
746 Health Dialog 6:240–244.
- 747 38. Olsen SJ, Kafoa B, Win NSS, Jose M, Bibb W, Luby S, Waidubu G, O’Leary M,  
748 Mintz E. 2001. Restaurant-associated outbreak of *Salmonella enterica* in Nauru:  
749 An epidemiological and cost analysis. Epidemiol Infect 127:405–412.
- 750 39. Luby SP. 2018. Urban slums: A supportive ecosystem for typhoidal Salmonellae.

- 751 J Infect Dis 218:S250–S254.
- 752 40. Ashcroft MT, Singh B, Nicholson CC, Ritchie JM, Sorryan E, Williams F. 1967. A  
753 seven-year field trial of two typhoid vaccines in Guyana. *Lancet* 290:1056–1059.
- 754 41. Sikorski MJ, Desai SN, Tupua S, Thomsen RE, Han J, Rambocus S, Nimarota-  
755 Brown S, Punimata L, Tusitala S, Sialeipata M, Hoffman SA, Tracy JK, Higginson  
756 EE, Tennant SM, Gauld JS, Klein DJ, Ballard SA, Robins-Browne RM, Dougan G,  
757 Nilles EJ, Howden BP, Crump JA, Naseri TK, Levine MM. 2020. Tenacious  
758 endemic typhoid fever in Samoa. *Clin Infect Dis* 71:S120–S126.
- 759 42. Samoa. Health Ordinance 1959, No.19. Available at:  
760 [http://www.paclii.org/ws/legis/num\\_act/haa1969149/](http://www.paclii.org/ws/legis/num_act/haa1969149/).
- 761 43. Samoa. Health Amendment Act 1969, No. 23. Available at:  
762 [http://www.paclii.org/ws/legis/consol\\_act/ho1959122/](http://www.paclii.org/ws/legis/consol_act/ho1959122/).
- 763 44. Perilla MJ, Bopp C. 2003. *Salmonella* serotype Typhi: Identification and  
764 Antimicrobial Susceptibility Testing, p. 103–162. *In* Manual for the Laboratory  
765 Identification and Antimicrobial Susceptibility Testing of Bacterial Pathogens of  
766 Public Health Importance in the Developing World. World Health Organization,  
767 Geneva.
- 768 45. Strockbine NA, Bopp CA, Fields PI, Kaper JB, Nataro JP. 2015. *Escherichia*,  
769 *Shigella*, and *Salmonella*, p. 685–713. *In* Jorgensen, JH, Pfaller, MA, Carroll, KC  
770 (eds.), *Manual of Clinical Microbiology*, 11th ed. ASM Press, Washington, DC,  
771 USA.
- 772 46. Ministry of Health Laboratory Division. 2016. *Microbiology Manual*. Moto'otua,  
773 Apia, Samoa.

- 774 47. Basnyat B, Qamar FN, Rupali P, Ahmed T, Parry CM. 2021. Enteric fever. *BMJ*  
775 372:n437.
- 776 48. Bauer AW, Kirby WMM, Sherris JC, Turck M. 1966. Antibiotic susceptibility testing  
777 by a standardized single disk method. *Am J Clin Pathol* 45:493–496.
- 778 49. CLSI. Performance standards for antimicrobial disc susceptibility tests; 10th ed.  
779 CLSI standard M02-A10. Wayne, PA: Clinical and Laboratory Standards Institute;  
780 2009.
- 781 50. The European Committee on Antimicrobial Susceptibility Testing. 2019.  
782 Breakpoint tables for interpretation of MICs and zone diameters. v9.0,  
783 <http://www.eucast.org>.
- 784 51. Ingle DJ, Andersson P, Valcanis M, Wilmot M, Easton M, Lane C, Barden J,  
785 Gonçalves da Silva A, Seemann T, Horan K, Ballard SA, Sherry NL, Williamson  
786 DA, Howden BP. 2021. Genomic epidemiology and antimicrobial resistance  
787 mechanisms of imported typhoid in Australia. *Antimicrob Agents Chemother*  
788 65:e0120021.
- 789 52. Seemann T, Goncalves da Silva A, Bulach D, Schultz M, Kwong J, Howden B.  
790 2018. Nullarbor (v1.41), GitHub <https://github.com/tseemann/nullarbor>.
- 791 53. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina  
792 sequence data. *Bioinformatics* 30:2114–2120.
- 793 54. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,  
794 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler  
795 G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly  
796 algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–

- 797 477.
- 798 55. Hazen TH, Nagaraj S, Sen S, Permala-Booth J, Del Canto F, Vidal R, Barry EM,  
799 Bitoun JP, Chen WH, Tennant SM, Rasko DA. 2019. Genome and functional  
800 characterization of colonization factor antigen I- and CS6-encoding heat-stable  
801 enterotoxin-only enterotoxigenic *Escherichia coli* reveals lineage and geographic  
802 variation. *mSystems* 4:e00329-18.
- 803 56. Dyson ZA, Holt KE. 2021. Five years of GenoTyphi: Updates to the global  
804 *Salmonella* Typhi genotyping framework. *J Infect Dis* 224:S775–S780.
- 805 57. Seemann T. 2020. Snippy (v4.6.0), GitHub <https://github.com/tseemann/snippy>.
- 806 58. Baines SL, da Silva AG, Carter GP, Jennison A, Rathnayake I, Graham RM,  
807 Sintchenko V, Wang Q, Rockett RJ, Timms VJ, Martinez E, Ballard S, Tomita T,  
808 Isles N, Horan KA, Pitchers W, Stinear TP, Williamson DA, Howden BP,  
809 Seemann T, Communicable Diseases Genomics Network (CDGN). 2020.  
810 Complete microbial genomes for public health in Australia and the Southwest  
811 Pacific. *Microb Genomics* 6:1–12.
- 812 59. Ingle DJ, Ambrose RL, Baines SL, Duchene S, Gonçalves da Silva A, Lee DYJ,  
813 Jones M, Valcanis M, Taiaroa G, Ballard SA, Kirk MD, Howden BP, Pearson JS,  
814 Williamson DA. 2021. Evolutionary dynamics of multidrug resistant *Salmonella*  
815 *enterica* serovar 4,[5],12:i:- in Australia. *Nat Commun* 12:4786.
- 816 60. Oxford Nanopore Technologies. 2020. Medaka (v1.0.3), GitHub  
817 <https://github.com/nanoporetech/medaka>.
- 818 61. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA,  
819 Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for

- 820 comprehensive microbial variant detection and genome assembly improvement.  
821 PLoS One 9:e112963.
- 822 62. Guevara PD, Maes M, Thanh DP, Duarte C, Rodriguez EC, Montaña LA, Dan  
823 THN, Nguyen TNT, Carey ME, Campos J, Chinen I, Perez E, Baker S. 2021. A  
824 genomic snapshot of *Salmonella enterica* serovar Typhi in Colombia. PLoS Negl  
825 Trop Dis 15:e0009755.
- 826 63. Maes M, Sikorski MJ, Carey ME, Higginson EE, Dyson ZA, Fernandez A, Araya  
827 P, Tennant SM, Baker S, Lagos R, Hormazabal JC, Levine MM, Dougan G. 2022.  
828 Whole genome sequence analysis of *Salmonella* Typhi provides evidence of  
829 phylogenetic linkage between cases of typhoid fever in Santiago, Chile in the  
830 1980s and 2010-2016. medRxiv 2022.01.19.22269577.
- 831 64. Argimón S, Yeats CA, Goater RJ, Abudahab K, Taylor B, Underwood A, Sánchez-  
832 Busó L, Wong VK, Dyson ZA, Nair S, Park SE, Marks F, Page AJ, Keane JA,  
833 Baker S, Holt KE, Dougan G, Aanensen DM. 2021. A global resource for genomic  
834 predictions of antimicrobial resistance and surveillance of *Salmonella* Typhi at  
835 Pathogenwatch. Nat Commun 12:2879.
- 836 65. Centre for Genomic Pathogen Surveillance (CGPS). Pathogenwatch Technical  
837 Descriptions. <https://cgps.gitbook.io/pathogenwatch/technical-descriptions>.  
838 Accessed 10 December 2021.
- 839 66. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L,  
840 Møller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids  
841 using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents  
842 Chemother 58:3895–3903.

- 843 67. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, Driebe EM, Drees  
844 KP, Hicks ND, Williamson CHD, Hepp CM, Smith DE, Roe C, Engelthaler DM,  
845 Wagner DM, Keim P. 2016. NASP: an accurate, rapid method for the identification  
846 of SNPs in WGS datasets that supports flexible input and output formats. *Microb*  
847 *Genomics* 2:e000074.
- 848 68. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill  
849 J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant  
850 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15.
- 851 69. Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-  
852 analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- 853 70. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and  
854 new developments. *Nucleic Acids Res* 47:W256–W259.
- 855 71. Seeman T. 2021. Snp-dists (v0.8.2), GitHub [https://github.com/tseemann/snp-](https://github.com/tseemann/snp-dists)  
856 [dists](https://github.com/tseemann/snp-dists).
- 857 72. Rasko DA, Myers GSA, Ravel J. 2005. Visualization of comparative genomic  
858 analyses by BLAST score ratio. *BMC Bioinformatics* 6:2.
- 859 73. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale blast score  
860 ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between  
861 bacterial genomes. *PeerJ* 2:e332.
- 862 74. Kidgell C, Reichard U, Wain J, Linz B, Torpdahl M, Dougan G, Achtman M. 2002.  
863 *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000  
864 years old. *Infect Genet Evol* 2:39–45.
- 865 75. Achtman M. 2008. Evolution, population structure, and phylogeography of

- 866 genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62:53–70.
- 867 76. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in  
868 microbial pan-genome-wide association studies with Scoary. *Genome Biol*  
869 17:238.
- 870 77. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J,  
871 Mahurkar A, Giglio MG. 2011. The IGS standard operating procedure for  
872 automated prokaryotic annotation. *Stand Genomic Sci* 4:244–251.
- 873 78. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K,  
874 Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL,  
875 Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich  
876 C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST  
877 server: Rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- 878 79. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA,  
879 Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014.  
880 The SEED and the Rapid Annotation of microbial genomes using Subsystems  
881 Technology (RAST). *Nucleic Acids Res* 42:D206–D214.
- 882 80. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek  
883 R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V,  
884 Wattam AR, Xia F. 2015. RASTtk: A modular and extensible implementation of  
885 the RAST algorithm for building custom annotation pipelines and annotating  
886 batches of genomes. *Sci Rep* 5:8365.
- 887 81. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016.  
888 PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic*



- 889           Acids Res 44:W16–W21.
- 890   82.   Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular  
891           evolutionary genetics analysis across computing platforms. *Mol Biol Evol*  
892           35:1547–1549.
- 893   83.   Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis  
894           (MEGA) for macOS. *Mol Biol Evol* 37:1237–1239.
- 895   84.   Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal  
896           structure of heterochronous sequences using TempEst (formerly Path-O-Gen).  
897           *Virus Evol* 2:vew007.
- 898   85.   Kariuki S, Dyson ZA, Mbae C, Ngetich R, Kawai SM, Wairimu C, Anyona S, Gitau  
899           N, Onsare RS, Ongandi B, Duchene S, Ali M, Clemens JD, Holt KE, Dougan G.  
900           2021. Multiple introductions of multidrug-resistant typhoid associated with acute  
901           infection and asymptomatic carriage, Kenya. *Elife* 10:e67852.
- 902   86.   Drummond AJ, Bouckaert RR. 2015. Bayesian Evolutionary Analysis with BEAST.  
903           Cambridge University Press, Cambridge.
- 904   87.   Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018.  
905           Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.  
906           *Virus Evol* 4:vey016.
- 907   88.   Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior  
908           summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904.
- 909   89.   Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. GGTREE : an R package for  
910           visualization and annotation of phylogenetic trees with their covariates and other  
911           associated data. *Methods Ecol Evol* 8:28–36.

- 912 90. Yu G, Lam TT-Y, Zhu H, Guan Y. 2018. Two methods for mapping and visualizing  
913 associated data on phylogeny using ggtree. *Mol Biol Evol* 35:3041–3043.
- 914 91. Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Curr Protoc*  
915 *Bioinforma* 69:e96.
- 916 92. Ruan Z, Yu Y, Feng Y. 2020. The global dissemination of bacterial infections  
917 necessitates the study of reverse genomic epidemiology. *Brief Bioinform* 21:741–  
918 750.
- 919 93. World Health Organization. 1955. Typhoid and paratyphoid fevers from 1950 to  
920 1954. *Bull World Health Organ* 13:173–191.
- 921 94. Dunstan SJ, Stephens HA, Blackwell JM, Duc CM, Lanh MN, Dudbridge F,  
922 Phuong CXT, Luxemburger C, Wain J, Ho VA, Hien TT, Farrar J, Dougan G.  
923 2001. Genes of the class II and class III major histocompatibility complex are  
924 associated with typhoid fever in Vietnam. *J Infect Dis* 183:261–268.
- 925 95. Dunstan SJ, Hue NT, Han B, Li Z, Tram TTB, Sim KS, Parry CM, Chinh NT, Vinh  
926 H, Lan NPH, Thieu NTV, Vinh PV, Koirala S, Dongol S, Arjyal A, Karkey A,  
927 Shilpakar O, Dolecek C, Foo JN, Phuong LT, Lanh MN, Do T, Aung T, Hon DN,  
928 Teo YY, Hibberd ML, Anders KL, Okada Y, Raychaudhuri S, Simmons CP, Baker  
929 S, De Bakker PIW, Basnyat B, Hien TT, Farrar JJ, Khor CC. 2014. Variation at  
930 HLA-DRB1 is associated with resistance to enteric fever. *Nat Genet* 46:1333–  
931 1336.
- 932 96. Zhang Y, Brady A, Jones C, Song Y, Darton TC, Jones C, Blohmke CJ, Pollard  
933 AJ, Magder LS, Fasano A, Szein MB, Fraser CM. 2018. Compositional and  
934 functional differences in the human gut microbiome correlate with clinical outcome

- 935 following infection with wild-type *Salmonella enterica* serovar Typhi. MBio  
936 9:e00686-18.
- 937 97. Kingsley RA, Langridge G, Smith SE, Makendi C, Fookes M, Wileman TM, El  
938 Ghany MA, Keith Turner A, Dyson ZA, Sridhar S, Pickard D, Kay S, Feasey N,  
939 Wong V, Barquist L, Dougan G. 2018. Functional analysis of *Salmonella* Typhi  
940 adaptation to survival in water. *Environ Microbiol* 20:4079–4090.
- 941 98. Kidgell C, Pickard D, Wain J, James K, Diem Nga LT, Diep TS, Levine MM,  
942 O’Gaora P, Prentice MB, Parkhill J, Day N, Farrar J, Dougan G. 2002.  
943 Characterisation and distribution of a cryptic *Salmonella typhi* plasmid pHCM2.  
944 *Plasmid* 47:159–171.
- 945 99. Norris P, Nguyen HA. 2007. Consumption of antibiotics in a small Pacific island  
946 nation: Samoa. *Pharm Pract (Granada)* 5:36–41.
- 947 100. Kirchhelle C, Dyson ZA, Dougan G. 2019. A biohistorical perspective of typhoid  
948 and antimicrobial resistance. *Clin Infect Dis* 69:S388–S394.
- 949 101. Pauulu U. 2021. New Typhoid Fever vaccine roll out. Samoa Observer  
950 (newspaper). Vaitele, Samoa.  
951 <https://www.samoobserver.ws/category/samoa/90045>. Retrieved 25 August  
952 2021. Vaitele, Samoa.  
953

**Table 1.** Candidate canonical SNPs delineating Samoan genotype 3.5.4 sub-lineages

Genotype	n/N*	Ref. allele	Alt. allele	CT18 position	Gene	Protein
3.5.4.1	96/96	G	A <sup>2</sup>	3960783	STY4102	putative L-lactate dehydrogenase
		G	A <sup>2</sup>	4794120	<i>gpmB</i>	probable phosphoglycerate mutase 2
3.5.4.2	77/77	C	T <sup>2</sup>	1731534	STY1812	exodeoxyribonuclease III
		C	T <sup>2</sup>	1945934	STY2094	glucose 6-phosphate dehydrogenase
3.5.4.3	106/107 <sup>1</sup>	C	T <sup>3</sup>	295362	STY0278	putative drug efflux protein

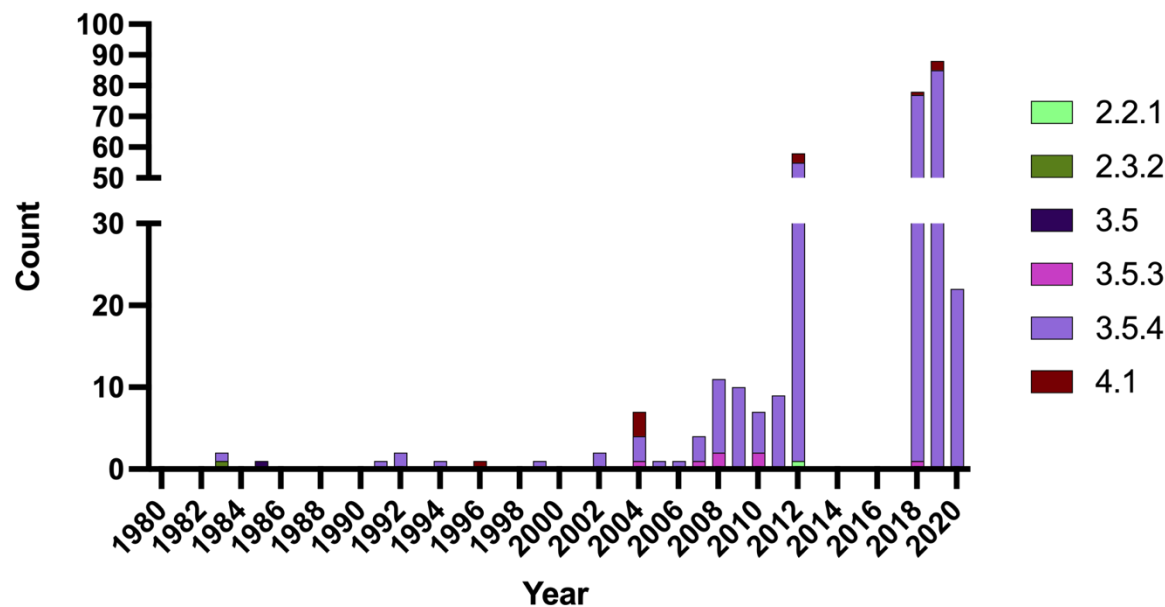
Genotype = Suggested sub-lineage designation; n/N = number of *S. Typhi* isolates in Samoa over total number of *S. Typhi* isolates in the sub-lineage; Ref. Allele = CT18 allele; Alt. allele = alternative SNP candidates in the genotype sub-lineage.

\*All SNPs were detected in 100% of their designated genotype groupings (N) and absent from 100% of out-grouped isolates (5,240 global *S. Typhi* genomes were analyzed)

<sup>1</sup> One 3.5.4.3 isolate is of unknown origin.

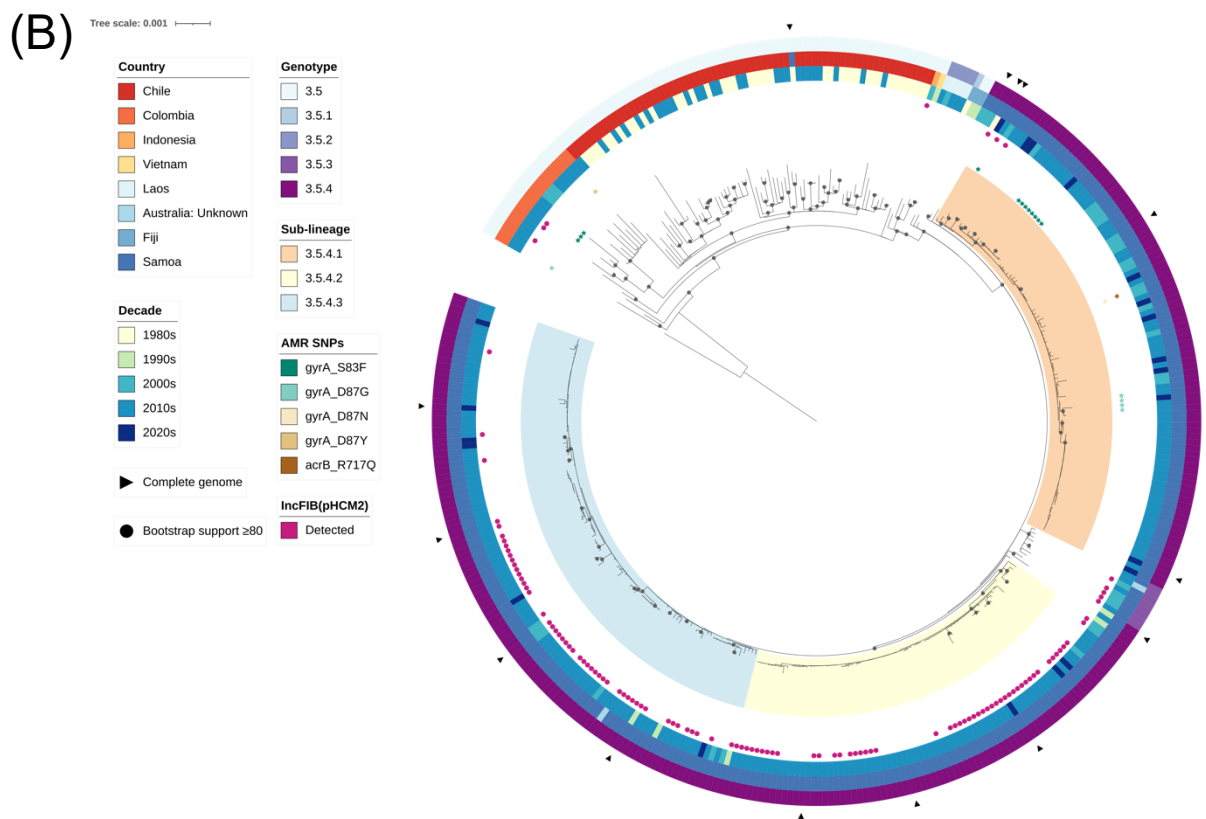
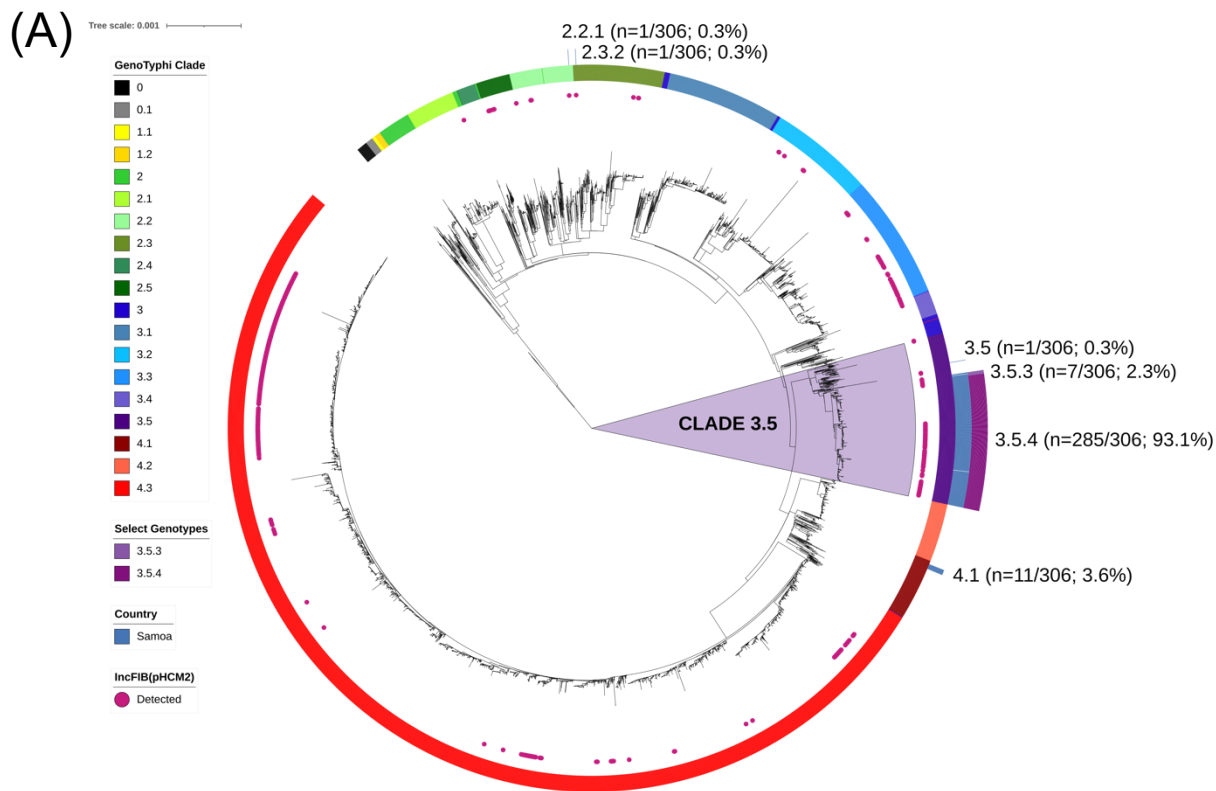
<sup>2</sup> Synonymous intragenic silent mutation

<sup>3</sup> Non-synonymous intragenic missense mutation



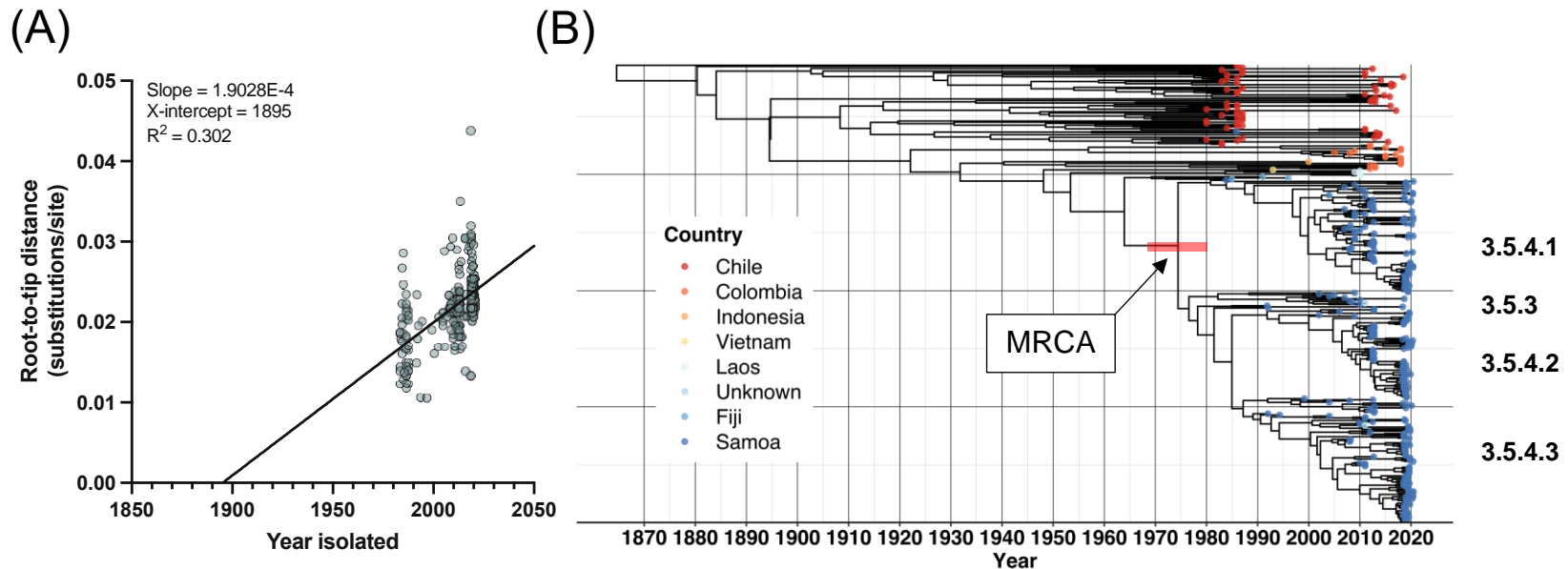
**Fig. 1. Frequency and genotype distribution of Samoan *S. Typhi* sequenced isolates by year of collection.**

Samoan isolates of *S. Typhi* (N=306) analyzed in this study span the years 1983-2020. Isolates from 1983-2011 were all collected from travelers from Samoa in Australia and New Zealand. Most isolates in 2012 and all isolates in 2018-2020 were collected in Samoa by the Ministry of Health. Genotype designations were determined using GenoTyphi (9, 56) on paired-end Illumina reads mapped to strain CT18.



**Fig. 2. Phylogenetic analysis of Samoan *S. Typhi* genomes in a global and dominant clade-specific context.**

**(A)** Maximum-likelihood tree of 5,240 *S. Typhi* genome assemblies using reference strain CT18 (GenBank Accession AL513382.1) including 306 from Samoa with GenoTyphi clade (ring 1) [innermost ring] (9, 56), country (ring 2) [Samoa], and genotypes 3.5.4 and 3.5.3 (ring 3) labeled. The distribution of the IncFIB(pHCM2) replicon detected using with PlasmidFinder (66) is indicated with pink shaded circles. Clade 3.5 genomes are highlighted with a light purple background. **(B)** Maximum-likelihood phylogeny of assembled clade 3.5 genomes (N=394) using a local Samoan reference strain H12ESR00394-001 (GenBank Accession LT904890.1). Country of origin, decade of isolation, antimicrobial resistance point mutations and the detection of the IncFIB(pHCM2) replicon are displayed.



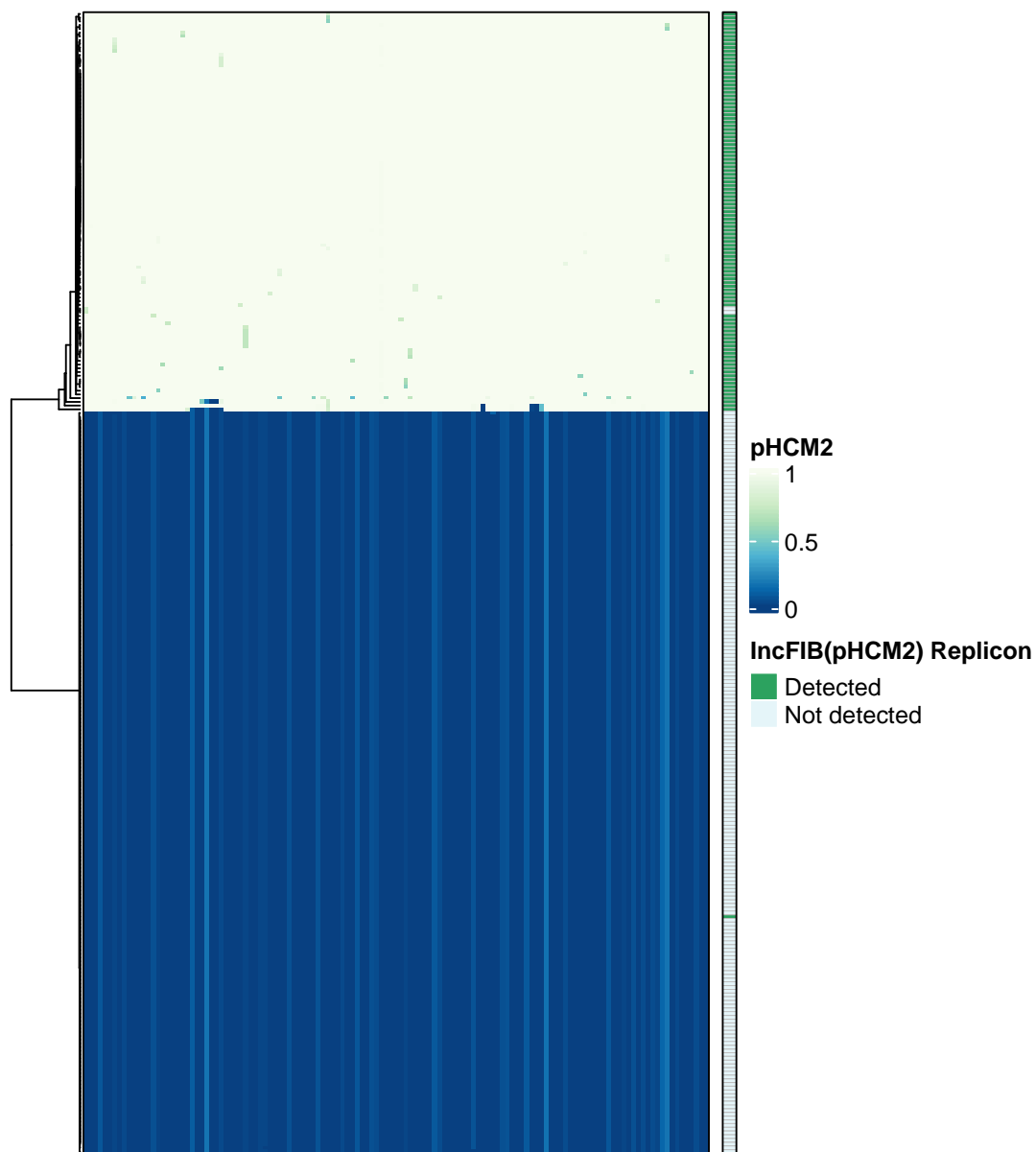
**Fig. 3. Temporal and evolutionary analysis of *S. Typhi* clade 3.5.**

**(A)** Temporal signal analysis of clade 3.5 *S. Typhi*. TempEst v1.5.3 (84) root-to-tip regression analysis as a function of sampling time is shown for the core sequence alignment against the local Samoan reference strain H12ESR00394-001A (GenBank Accession LT904890.1) of all 394 *S. Typhi* belonging to clade 3.5 (isolated between 1983 and 2020 with variable date precision) with the best-fitting root selected according to heuristical residual mean squared. **(B)** Bayesian evolutionary analysis of the same core sequence alignment showing the maximum-clade credibility (MCC) phylogenetic tree of Samoan genotypes 3.5.4 and 3.5.3 isolates in the context of global clade 3.5 *S. Typhi* isolates. The age of the node of the most recent common ancestor (MRCA) to the Samoan genotypes 3.5.3 and 3.5.4 isolates (arrow) had a mean

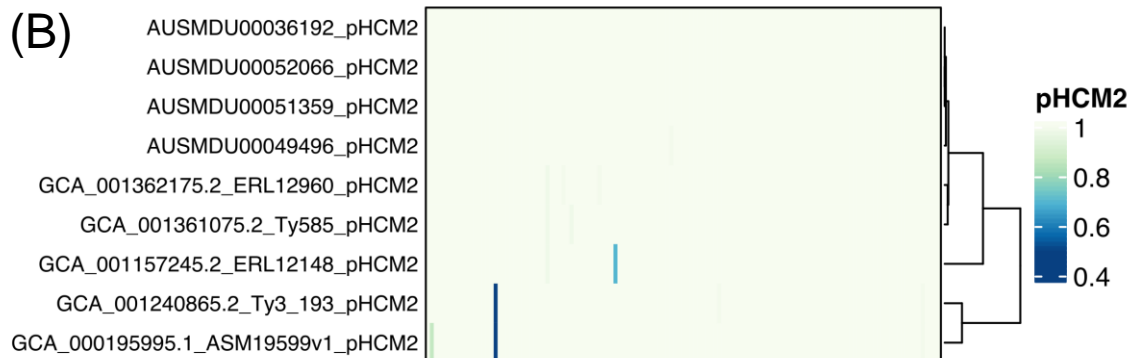


of 1974 and the red horizontal bar represents the 95% Highest Posterior Density (HPD: 1968, 1981) of the node. Tip colors indicate country of origin for each isolate. A time scale is shown along the x-axis is relative to the most recent sampling date of 2020-06-09. The genotype/sub-lineage designations are labeled to the right. The tree was constructed with BEAST v1.10.4 (87) and visualized using the ggtree package in R (89–91).

(A)



(B)



**Fig. 4. LS-BSR heatmap identifying pHCM2 genes in Samoan *S. Typhi*.**

**(A)** The 130 sequential coding regions in the circularized complete pHCM2 plasmid of 1994 Samoan strain AUSMDU00051359 was compared all Samoan *S. Typhi* (n=306) draft genomes for gene homology. The heatmap depicts gene absence (dark blue) and presence (light yellow) as a BLAST score ratio from 0 to 1, respectively. The bar on the right indicates the presence (green) or absence (light blue) of the IncFIB(pHCM2) replicon as detected by IncTyper (64). **(B)** This comparison was repeated for all four complete circularized pHCM2 molecules in Samoan *S. Typhi* and the five GenBank complete reference genomes that contain a complete circularized pHCM2 molecule. Sequential gene homology compared with the pHCM2 molecule from strain AUSMDU00051359 is depicted, with homology ranging from low (dark blue) to high (light yellow). Strain ERL12960 is genotype 4.3.1.1 isolated in India in 2012. Strain Ty585 is genotype 4.3.1.1 isolated in Bangladesh in 2012. Strain ERL 12148 is genotype 4.3.1.1 isolated in India in 2012. Strain Ty3 is genotype 4.3.1.1 isolated in Vietnam in 1997. Strain ASM19599v1 is strain CT18, a genotype 3.2.1 isolate from Vietnam in 1993. The GenBank accession numbers are listed in Table S3.