

1 **Whole genome sequence analysis of *Salmonella* Typhi in Papua New Guinea reveals an**  
2 **established population of genotype 2.1.7 sensitive to antimicrobials.**

3

4 **Short title:** Genetic epidemiology of *Salmonella* Typhi in Papua New Guinea

5

6 Zoe A. Dyson<sup>1,2,3,4,\*</sup>, Elisheba Malau<sup>5</sup>, Paul F. Horwood<sup>6</sup>, Rebecca Ford<sup>7</sup>, Valentine Siba<sup>7</sup>,

7 Mition Yoannes<sup>7</sup>, William Pomat<sup>7</sup>, Megan Passey<sup>8</sup>, Louise M. Judd<sup>2</sup>, Danielle J. Ingle<sup>9</sup>,

8 Deborah A. Williamson<sup>9,10</sup>, Gordon Dougan<sup>3</sup>, Andrew R. Greenhill<sup>5,\*</sup>, Kathryn E. Holt<sup>1,2,¶</sup>

9

10

11 1. London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

12 2. Department of Infectious Diseases, Central Clinical School, Monash University,

13 Melbourne, Victoria 3004, Australia

14 3. Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID),

15 Department of Medicine, University of Cambridge, Cambridge, CB2 0AW, UK.

16 4. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1RQ,

17 UK.

18 5. School of Science, Psychology and Sport, Federation University, Australia

19 6. College of Public Health, Medical and Veterinary Sciences, James Cook University,

20 Townsville, Australia.

21 7. Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

22 8. University Centre for Rural Health, Faculty of Medicine and Health, University of Sydney,

23 Lismore, NSW, Australia

24 9. Department of Microbiology and Immunology, The University of Melbourne at The Peter

25 Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

26 10. Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology  
27 and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection  
28 and Immunity, Melbourne, VIC, Australia

29

30 \* Indicates corresponding author(s)

31 †These authors contributed equally as senior authors

32

33 Correspondence:

34 Dr Zoe Anne Dyson [zoe.dyson@lshtm.ac.uk](mailto:zoe.dyson@lshtm.ac.uk)

35 Dr Andrew Greenhill [andrew.greenhill@federation.edu.au](mailto:andrew.greenhill@federation.edu.au)

36

## 37 **Abstract**

### 38 ***Background***

39 Typhoid fever, a systemic infection caused by *Salmonella enterica* serovar Typhi, remains a  
40 considerable public health threat in impoverished regions within many low- and middle-  
41 income settings. However, we still lack a detailed understanding of the emergence,  
42 population structure, molecular mechanisms of antimicrobial resistance (AMR), and  
43 transmission dynamics of *S. Typhi* across many settings, particularly throughout the Asia-  
44 Pacific islands. Here we present a comprehensive whole genome sequence (WGS) based  
45 overview of *S. Typhi* populations circulating in Papua New Guinea (PNG) over 30 years.

46

### 47 ***Principle findings***

48 Bioinformatic analysis of 86 *S. Typhi* isolates collected between 1980-2010 demonstrated  
49 that the population structure of PNG is dominated by a single genotype (2.1.7) that appears to  
50 have emerged in the Indonesian archipelago in the mid-twentieth century with very limited  
51 evidence of inter-country transmission. Genotypic and phenotypic data demonstrated that the

52 PNG *S. Typhi* population appears to be susceptible to former first line drugs for treating  
53 typhoid fever (chloramphenicol, ampicillin and co-trimoxazole), as well as fluoroquinolones,  
54 third generation cephalosporins, and macrolides. PNG genotype 2.1.7 was genetically  
55 conserved, with very few deletions, and no evidence of plasmid or prophage acquisition.  
56 Genetic variation among this population was attributed to either single point mutations, or  
57 homologous recombination adjacent to repetitive ribosomal RNA operons.

58

### 59 ***Significance***

60 Antimicrobials remain an effective option for the treatment of typhoid fever in PNG, along  
61 with other intervention strategies including improvements to water, sanitation and hygiene  
62 (WaSH) related infrastructure and potentially the introduction of Vi-conjugate vaccines.  
63 However, continued genomic surveillance is warranted to monitor for the emergence of AMR  
64 within local populations, or the introduction of AMR associated genotypes of *S. Typhi* in this  
65 setting.

66

67

### 68 **Author Summary**

69 Typhoid fever, caused by *Salmonella enterica* serovar Typhi, is a systemic infection common  
70 to many low- to middle-income settings. While the population structure of *S. Typhi* has been  
71 genetically characterised using whole genome sequencing in many endemic countries  
72 throughout Sub-Saharan Africa and South Asia, we are still lacking a detailed understanding  
73 for many regions including those among the Asia-Pacific islands. Genomic surveillance of  
74 isolates spanning 30 years demonstrated a population structure of *S. Typhi* in Papua New  
75 Guinea (PNG) dominated by a single genotype (2.1.7) that emerged in the mid-twentieth  
76 century, is genetically homogeneous, and sensitive to a wide range of antibiotics commonly

77 used in the treatment of typhoid. There was little evidence of inter-country transmission and  
78 the setting appeared free of *S. Typhi* genotypes commonly associated with AMR e.g. H58  
79 (genotype 4.3.1). These data suggest that former first line drugs (chloramphenicol, ampicillin  
80 and co-trimoxazole), fluoroquinolones, third generation cephalosporins and macrolides all  
81 remain viable options for controlling typhoid in addition to the introduction of Vi-conjugate  
82 vaccines and improvements to water, sanitation and hygiene (WaSH) related infrastructure.  
83 Routine molecular surveillance is necessary to monitor for introduced or emerging AMR to  
84 inform treatment guidelines and intervention strategies.

85

86

## 87 **Introduction**

88 Typhoid fever is a systemic infection of the bacterium *Salmonella enterica* serovar Typhi[1].  
89 Each year more than 10,000,000 cases of typhoid fever occur worldwide, of which more than  
90 100,000 result in death[2]. *S. Typhi* is transmitted faeco-orally, usually through contaminated  
91 food and water and therefore constitutes a major public health threat in many low- to middle-  
92 income countries (LMIC), particularly in impoverished regions where hygiene and sanitation  
93 infrastructure is limited[3].

94

95 Antimicrobial chemotherapy has become a mainstay in the treatment of typhoid fever, with  
96 death rates declining from 20-30% to 1% following the introduction of antimicrobials[4,5].  
97 However, antimicrobial resistance (AMR) has arisen multiple times in *S. Typhi*, with  
98 multidrug-resistance (MDR; resistance to former first line antimicrobials chloramphenicol,  
99 co-trimoxazole, and ampicillin) emerging from the 1970s onwards[6], followed by resistance  
100 to fluoroquinolones commonly mediated by mutations in the quinolone resistance  
101 determining region (QRDR) of genes *gyrA*, *gyrB*, and *parC*[7–9]. More recently, extensively

102 drug-resistant (XDR) *S. Typhi* populations resistant to all orally administered antimicrobials  
103 except for azithromycin have emerged in Pakistan outbreaks[10,11], and mutations mediating  
104 azithromycin resistance have been observed in non-XDR *S. Typhi* populations[12–15].

105

106 Papua New Guinea (PNG) has been recognised as a high burden setting for typhoid since the  
107 mid-1990s when the incidence rate was reported as 1,208 cases per 100,000 population,  
108 amongst the highest rates in the world at that time[16]. Although there is a paucity of recent  
109 data, one diagnostic study demonstrated that typhoid remains a common diagnosis in febrile  
110 patients in the PNG highlands[17]. Vaccines for typhoid fever are not currently part of  
111 the routine schedule in PNG, nor have they been in the past.

112

113 Early molecular studies [18–20] revealed that several ribotypes of *S. Typhi* appeared endemic  
114 to PNG, with many appearing unique to the country, and some also observed in Southeast  
115 Asian countries. Further, 80.6% of isolates from 1986-1989 belonged to a single phage type  
116 (D2) [18,20], which was also observed at lower frequency in other countries including  
117 Malaysia and Thailand. Previous reports[18], have also noted that *S. Typhi* from PNG  
118 appeared free of both detectable AMR genes and plasmids, and have suggested that genome  
119 plasticity might be common. However, despite these early observations, only two locally  
120 collected *S. Typhi* sequences from PNG have been reported on in detail[21]. A further 45  
121 isolates collected from travellers returning to Australia from PNG were sequenced as a part  
122 of a global overview study[9], but this study did not report on the specific dynamics and  
123 characteristics of the PNG pathogen population. Subsequently, there is currently no  
124 comprehensive whole genome sequence-based overview of the population structure,  
125 transmission dynamics, or evolutionary history of *S. Typhi* in PNG. Here we combine data on  
126 41 novel isolates collected locally in PNG between 1992 and 2010, with those data sequenced

127 previously from return travellers, to provide a 30-year overview (1980-2010) of *S. Typhi* in  
128 PNG.

129

130

## 131 **Methods**

### 132 *Ethics statement*

133 Ethical approval for this study was granted by the PNG Institute of Medical Research  
134 Institutional Review Board (1609), the PNG Medical Research Advisory Board (MRAC  
135 16.43) and the Federation University Human Research Ethics Committee (A17-074).

136

### 137 *Study Setting*

138 PNG is a tropical country situated approximately 6° south of the equator in the Western  
139 Pacific region. It is a LMIC, with a large proportion of its ~9 million inhabitants living in  
140 regional and remote areas of the country. Health and development indicators are low by  
141 regional and global standards, with ~40% of people having access to improved water supply  
142 and <20% with access to adequate sanitation[22,23]. In 2019, PNG's human development  
143 index of 0.55 ranked it 155 out of 189 countries and territories[24].

144

145 The PNG Institute of Medical Research (PNGIMR) is headquartered in Goroka, the capital of  
146 Eastern Highlands Province. The institute has been responsible for the majority of studies  
147 that have been conducted on typhoid fever in PNG.

148

### 149 *Isolate collection and culture processing*

150 Isolates of *S. Typhi* that were sequenced as part of this study (n=41) were obtained from the  
151 PNGIMR culture collection. All isolates had been preserved by freeze-drying in glass

152 ampules. The content of each ampule was resuspended in nutrient broth, and inoculated onto  
153 nutrient agar (incubated for 18 – 24 hours at 35°C). PCR was conducted on isolates using  
154 previously described methods [17] to confirm their identity as *S. Typhi*.

155

#### 156 ***Antimicrobial susceptibility testing and phage typing***

157 Antimicrobial susceptibility testing (AST) was carried out for n=40 (88.9%) of travel-  
158 associated cases for chloramphenicol, ampicillin, sulfamethoxazole, trimethoprim, nalidixic  
159 acid, tetracycline, kanamycin, gentamycin, and spectinomycin as detailed in **S1 Table**. AST  
160 was performed on *S. Typhi* isolates using agar breakpoint dilution and interpreted using  
161 Clinical and Laboratory Standards Institute (CLSI) breakpoints as described previously[25].  
162 Phage types were extracted from the National Enteric Pathogens Surveillance Scheme  
163 (NEPSS), a dedicated surveillance system for human and nonhuman enteric pathogens  
164 (including *Salmonella*) that has been operated by the Microbiological Diagnostic Unit Public  
165 Health Laboratory (MDU PHL)[25].

166

#### 167 ***DNA extraction and whole genome sequencing***

168 Isolated colonies (1 – 3) were suspended in sterile water to 0.5 McFarland, and 100 µl of the  
169 suspension was used for DNA extraction with the DNeasy Blood and Tissue Kit (Qiagen,  
170 Hilden, Germany), following the manufacturer’s instructions. Extracted genomic DNA  
171 (gDNA) was then subjected to indexed whole genome sequencing on an Illumina HiSeq 2500  
172 platform at the Wellcome Sanger Institute to generate paired end reads of 150 bp in length as  
173 described previously[9].

174

175 Three isolates were selected for additional long-read sequencing using the Nanopore MinION  
176 R9 device which was carried out as described previously[26]. To generate high molecular

177 weight gDNA (>60 kb) suited to Oxford Nanopore sequencing without further size  
178 selection the following protocol was used; firstly, isolates were grown overnight at 37°C on  
179 Luria-Bertani (LB) plates, before single colonies were picked for overnight culture at 37°C in  
180 LB broth medium. Bacterial cell pellets from 3.0 ml LB culture were then generated by  
181 centrifugation at 15,000 g for 5 minutes. gDNA was extracted from these pellets using  
182 Agencourt GenFind V2 (Beckman Coulter) with modifications as follows. Cell pellets were  
183 resuspended in 400 µl lysis buffer containing 9 µl Proteinase K (96 mg/ml; Beckman  
184 Coulter) and 1 µl RNase A (100 mg/ml; Sigma Aldrich) by tip mixing. Samples were lysed  
185 at 37°C for 30 minutes, and gDNA was extracted from the lysed samples by completing the  
186 remaining steps of the GenFind V2 for 200 µl of blood/serum from the binding step  
187 onwards as per the manufacturer's instructions.

188

### 189 ***Read mapping, Single Nucleotide Variant (SNV) and phylogenomic analyses***

190 For Single Nucleotide Variant (SNV) analysis, paired end reads for n=41 *S. Typhi* isolates  
191 collected in this study together with n=45 from a previous study[27] (**S1 Table**), were  
192 aligned to the CT18 reference genome (GenBank accession AL513382) [28] using the  
193 RedDog mapping pipeline (V1.beta10.3; available at <https://github.com/katholt/reddog>).  
194 RedDog maps reads to the reference genome with Bowtie2 (v2.2.9)[29], before using  
195 SAMtools (v1.3.1) [30] to identify high quality single nucleotide variant (SNV) calls as  
196 previously described[8]. A single sequence (sample MDUST355) was found to be a mixture  
197 of *S. Typhi* genotypes (~80% 2.1.7.1 and ~20% 2.1.7.2) and was subsequently removed from  
198 further phylogenomic analyses. A core SNV alignment was generated by concatenating  
199 alleles with high-quality consensus base calls (phred scores >20), for all SNV sites meeting  
200 this criteria in >95% of genomes (representing the 95% 'soft' core of the *S. Typhi* genome).  
201 This alignment was filtered to exclude SNV called in previously defined [9,31,32] prophage



202 regions and repetitive sequences (354 kb; ~7.4% of bases) in the CT18 reference  
203 chromosome, and Paratyphi A sequence AKU\_12601 (GenBank accession FM200053)  
204 alleles were also included in the alignment for outgroup rooting of the phylogenetic tree. The  
205 resultant 657 bp SNV alignment then was used to infer a whole genome pseudoalignment  
206 using the CT18 reference sequence, and any remaining recombination was removed using  
207 Gubbins (v2.3.2)[33]. The resultant 620 bp recombination-filtered alignment was used as  
208 input for RAxML (v8.2.8)[34], which was used to infer a maximum likelihood (ML)  
209 phylogeny with a generalized time-reversible model and a Gamma distribution to model site-  
210 specific rate variation (GTR +  $\Gamma$  substitution model; GTRGAMMA in RAxML), with 100  
211 bootstrap pseudo-replicates used to assess branch support (**S1 Fig**).

212

213 GenoTyphi (v1.9.1; available at <https://github.com/katholt/genotypphi/>) was used together  
214 with RedDog Bowtie2 alignment (BAM) files to infer genotypes for all sequences analysed  
215 in this study using SNVs that were defined in the extended *S. Typhi* genotyping  
216 framework[27,35]. These data revealed that our sampling of the PNG *S. Typhi* population  
217 was dominated by a single genotype (2.1.7, n=84, 98%; **S1 Fig**), necessitating the completion  
218 of a reference genome for the PNG genotype 2.1.7 population to allow for high resolution  
219 phylogenomic analyses. Three PNG genotype 2.1.7 genomes were selected for Nanopore  
220 long read sequencing and hybrid assembly (see **S1 Table**), with the oldest of these  
221 (MDUST348 from 1985), selected as a reference sequence for lineage-focused SNV analyses  
222 and phylogenetic inference using the mapping-based approach described above. For this  
223 newly completed PNG genotype 2.1.7 reference genome (MDUST348, deposited under  
224 accession TBA, repetitive regions were identified using the `ncumer` command in MUMmer  
225 (v3.23) [36] and integrated prophage sequences detected using the PHAge Search Tool  
226 Enhanced Release (PHASTER; available at <https://phaster.ca/>)[37]. These regions (detailed

227 in **S2 Table**) were then filtered from the inferred PNG genotype 2.1.7 alignment, with any  
228 further recombination removed with Gubbins as described above. The resultant  
229 recombination filtered alignment of PNG genotype 2.1.7 *S. Typhi* sequences (n=83) also  
230 included alleles from representative non-genotype 2.1.7 sequences as outgroups for  
231 phylogenetic tree rooting (detailed in **S3 Table**), and this final alignment (258 bp; 143  
232 sequences) was then used to infer the final phylogenies shown in **Fig 1 and S2 Fig**. A global  
233 context phylogeny including an additional n=50 sequences of clade 2.1 (n=133 total; **S1**  
234 **Table**) from previous studies[27,38] was inferred in the same manner with an alignment of  
235 671 bp in length (**Fig 2**; 184 sequences including outgroups). An interactive version of the  
236 resulting global context clade 2.1 phylogeny is available at:  
237 [https://microreact.org/project/icGPbF\\_O8](https://microreact.org/project/icGPbF_O8).

238  
239 Unique SNVs defining PNG genotype 2.1.7 *S. Typhi* sublineages were manually extracted  
240 from RedDog allele tables using R. SNVs causing non-synonymous mutations in highly  
241 conserved genes were prioritized for sublineage definitions. SNV distances were calculated  
242 from the core genome SNV alignment using snp-dists (v0.7.0; available at  
243 <https://github.com/tseemann/snp-dists>).

244  
245 Assessments of the gene conservation across PNG genotype 2.1.7 *S. Typhi* (**S4 Fig**) were  
246 derived from the gene presence/absence matrix output from the RedDog mapping pipeline  
247 using the MDUST348 PNG 2.1.7 *S. Typhi* reference genome completed in this study.

248

#### 249 ***Temporal analysis***

250 To investigate the temporal signal and emergence of genotype 2.1.7 *S. Typhi* in PNG, we  
251 used multiple methods. Initially we used TempEst (v1.5.1) [39] to assess the temporal

252 structure of the data (i.e. signals of clock-like behaviour) by conducting a regression analysis  
253 of root-to-tip branch distances for the PNG genotype 2.1.7 ML tree inferred using the  
254 MDUST348 reference genome (**S2 Fig**) as a function of sampling time. The heuristic residual  
255 mean squared method was used for the regression analysis with the best-fitting root option  
256 selected. The resultant data from these analyses were visualised in R.

257

258 In order to estimate divergence times we analysed the sequence data with BEAST2 [40]  
259 (v2.4.7), using the GTR +  $\Gamma$  substitution model and sampling times (tip dates) defined by the  
260 year of isolation to calibrate the molecular clock. For n=4 sequences (YB 352, YB 366, YB  
261 559, and YB 611; **S1 Table**), the precise year of isolation was not available, however, as they  
262 were known to have been collected between 2009-2010 their date of isolation was estimated  
263 with BEAST2 using a tip date sampling prior specifying a uniform distribution between  
264 2009-2010. Models were fit using both constant-coalescent population size and Bayesian  
265 skyline tree priors, together with either a strict clock model or a relaxed (uncorrelated  
266 lognormal distribution) clock model, to identify the model combination that best fit the data.  
267 For all model and tree prior combinations, a chain length of 10,000,000 steps sampling every  
268 5,000 steps was selected[41]. Preliminary analyses highlighted the constant-coalescent  
269 population size and relaxed (uncorrelated lognormal) clock models to best fit the data.  
270 However, this model combination yielded an implausible divergence date (~1914, 95% HPD  
271 = 1835-1969), compared to those inferred from other model combinations (~1953, 95% HPD  
272 = 1979-1917), and analyses of n=43 genotype 2.1.7 from the previous study alone [27]  
273 suggested the Bayesian skyline tree prior was better suited to the data. Therefore, the analyses  
274 presented here are those using the coalescent Bayesian skyline model together with a relaxed  
275 (uncorrelated log normal) clock model. To test the temporal signal we conducted a date-  
276 randomisation test whereby sampling times were randomly shuffled, and the analysis re-run

277 20 times[41,42]. These tests (**S3 Fig**) indicated that the data displayed temporal structure  
278 [41].

279

280 For the final analysis reported here, a single independent run conducted with a chain length  
281 of 10,000,000 states, sampling every 5,000 iterations, was subjected to removal of the first  
282 10% of steps as burn-in with LogCombiner (v2.4.7)[40]. Maximum-clade credibility (MCC)  
283 trees were then generated with ‘common ancestor heights’ specified for node heights using  
284 TreeAnnotator (v2.4.6)[40]. Effective sample sizes were estimated to be >200 for all  
285 parameters reported. All phylogenies were visualised using the R package ggtree (v2.2.4)  
286 [43] and the online tool Microreact [44]; an online interactive dated phylogeny is available at  
287 <https://microreact.org/project/CYcTnSmjT>.

288

### 289 ***Reference genome assembly and annotation***

290 Hybrid Illumina-Nanopore genome assemblies were generated with Unicycler (v0.4.7) [45]  
291 for three PNG genotype 2.1.7 isolates MDUST305, MDUST348, and MDUST255 (see **S2**  
292 **Fig and S1 Table**). The resultant *de Bruijn* assembly graphs were then resolved into single  
293 contigs where necessary through manual curation with Bandage (v0.8.1)[46], and annotated  
294 with PROKKA (v1.14.0)[47]. Nucleotide pairwise comparisons of the completed genomes  
295 were carried out using Mauve [48] (vSnapshot\_2015-02-25 1) and visualised using the  
296 genoPlotR (v0.8.11) package for R[49].

297

298 For analysis of prophage regions detected with PHASTER [37] in hybrid assembled  
299 sequences, these were aligned with their closest genetic relative in CT18 (prophage ST35)  
300 [50] using Mauve to confirm the prophage region boundaries and insertion site. Prophage  
301 regions were then extracted manually from the bacterial chromosome using seqret from the

302 EMBOSS toolkit[51]. Annotations were then transferred to our PNG sequences from the  
303 CT18 reference sequence for prophage ST35 using the Rapid Annotation Transfer Tool  
304 (RATT)[52], and then manually curated together with PROKKA derived annotations.  
305 BLASTn comparisons were then carried out, and visualized using genoPlotR[49].

306

### 307 ***AMR detection***

308 The mapping based allele typer SRST2 [53] (v0.1.8) was used to detect the presence of  
309 known acquired AMR genes and plasmid replicons using both the ARG-annot [54] and  
310 PlasmidFinder [55] databases, respectively. SNVs in the Quinolone Resistance Determining  
311 Regions (QRDR) of genes *gyrA*, *gyrB*, and *parC* that are known to reduce susceptibility to  
312 fluoroquinolones in *S. Typhi* [7], and *acrB* mutations associated with azithromycin resistance  
313 in typhoidal *Salmonella*[12], were detected using GenoTyphi (**described above**).

314

### 315 ***Nucleotide sequence accession numbers***

316 Raw read sequences for the 41 novel isolates from PNG have been deposited in the European  
317 Nucleotide Archive under project PRJEB20541; and individual accession numbers are listed  
318 in **S1 Table**. Genome assemblies for completed sequences of traveller isolates MDUST348,  
319 MDUST255, and MDUST305 were deposited in GenBank under accession numbers TBA,  
320 TBA, and TBA respectively (see **S1 Table** for details).

321

322

## 323 **Results**

### 324 ***Population structure and antimicrobial resistance of *S. Typhi* in PNG***

325 We combined sequence data from n=41 *S. Typhi* collected locally in PNG and sequenced for  
326 this study, with n=45 sequences from a previous study [27] that originate from returned

327 travellers who acquired their infections in PNG. Together these data span 30 years (1980-  
328 2010). Travel-associated and locally-collected isolate sequences were intermingled in the  
329 inferred phylogeny (**S1 Fig**), and revealed that the population structure of *S. Typhi* during  
330 this period was dominated by a single genotype, 2.1.7 (n=84, 98%; isolated 1985 to 2010)  
331 which has been present for at least 25 years. Only one other genotype was detected; clade 4.1  
332 (n=2, 2.3%; isolated 1980 and 1986) (**S1 Fig**).

333

334 The population structure of genotype 2.1.7 in PNG can be subdivided into two main lineages  
335 (**S2 Fig**), members of which are separated by a median pairwise distance of ~33 SNVs. We  
336 designated these as lineage I (genotype 2.1.7.1) and lineage II (2.1.7.2). In our data, genotype  
337 2.1.7.1 was more common with n=66 (78.6%) members, compared to n=18 (21.4%) 2.1.7.2  
338 members. Interestingly, 2.1.7.2 was not observed after 1998. Genotype 2.1.7.1 can be  
339 identified by the presence of a synonymous marker SNV C1503T in gene STY4417 (position  
340 4286788 in CT18), and genotype 2.1.7.2 can be identified by the presence of a synonymous  
341 marker SNV C540T in gene STY4106 (position 3967063 in CT18). These genotypes have  
342 been added to the GenoTyphi scheme available at  
343 <http://github.com/katholt/genotyphi>[27,35,56].

344

345 No known acquired AMR genes, nor mutations associated with either decreased  
346 ciprofloxacin susceptibility or azithromycin resistance (see **Methods**) were detected among  
347 any of the *S. Typhi* sequences from PNG, indicating this population is sensitive to commonly  
348 used antimicrobials. Susceptibility testing data was available for n=40 (88.9%) traveller  
349 isolates, all of which tested susceptible to all drugs tested (see **Methods** and **S1 table**).

350

351 *Emergence and transmission dynamics of PNG genotype 2.1.7 sublineages*

352 We applied Bayesian phylodynamic analysis to the set of PNG 2.1.7 sequences to estimate  
353 the dates of emergence for genotype 2.1.7, and its sublineages 2.1.7.1 and 2.1.7.2 (**Fig 1**,  
354 interactive phylogeny available at: <https://microreact.org/project/CYcTnSmjT>). This analysis  
355 yielded a local substitution rate of 0.49 SNVs per genome per year (95% HPD 0.32-0.66) or  
356  $1.1 \times 10^{-7}$  genome-wide substitutions per site per year (95% HPD  $7.3 \times 10^{-8}$ - $1.5 \times 10^{-7}$ ). These  
357 data showed temporal structure [41] to support these results (see **Methods** and **S3 Fig**),  
358 which were slightly lower than published estimates for *S. Typhi* for genotype 4.3.1 in Nepal  
359 ( $1.7 \times 10^{-7}$  genome-wide substitutions per site per year, 95% HPD  $1.1 \times 10^{-7}$ - $2.4 \times 10^{-7}$ )[8],  
360 Kenya ( $1.9 \times 10^{-7}$ , 95% HPD  $1.5 \times 10^{-7}$ - $2.2 \times 10^{-7}$ )[57], and globally ( $1.42 \times 10^{-7}$ , 95% HPD  
361  $1.0 \times 10^{-7}$  to  $1.8 \times 10^{-7}$ )[9]. We estimate that the most recent common ancestor (MRCA) for all  
362 PNG genotype 2.1.7 existed circa ~1953 (95% HPD = 1917-1979) as shown in **Fig 1**. PNG  
363 2.1.7 sublineages appear to have emerged contemporaneously in PNG in the mid-late 1970s,  
364 with the MRCA of 2.1.7.1 estimated circa ~1976 (95% HPD 1963-1985) and the MRCA of  
365 2.1.7.2 around the same time or potentially slightly later, circa ~1979 (95% HPD 1971-1984).

366

367

368 **Fig 1. Maximum clade credibility tree inferred from all available (n=83) PNG 2.1.7 *S.***  
369 ***Typhi* using BEAST2.** Tip colours indicate isolate source as per inset legend, branches are  
370 labelled by *S. Typhi* genotype. Square nodes indicate the reference sequence and other  
371 completed genomes. Divergence dates for PNG genotype 2.1.7 and sublineages are labelled  
372 (95% highest posterior density in brackets). Interactive phylogeny available at:

373 <https://microreact.org/project/CYcTnSmjT>.

374

375

376 In order to better understand the relationship of PNG genotype 2.1.7 *S. Typhi* to its nearest  
377 relatives in the pathogen population, we inferred a phylogeny of all available genome  
378 sequences belonging to its parent clade 2.1 (**Fig 2**, interactive phylogeny available at:  
379 [https://microreact.org/project/icGPbF\\_O8](https://microreact.org/project/icGPbF_O8)). Genotype 2.1.7 isolates were mostly from PNG  
380 (n=84, 93.3%). Within genotype 2.1.7, PNG sequences formed a distinct cluster (further  
381 subdivided into lineages 2.1.7.1 and 2.1.7.2), with a small number of distant relatives in India  
382 (three genotype 2.1.7 isolates separated from PNG isolates by a median pairwise distance of  
383 ~64 SNVs) and Vietnam (one sequence separated by median ~41 SNVs). The only non-PNG  
384 isolate clustering within the PNG sublineages was a single 2.1.7.1 sequence from India  
385 (MDUST301 from 1988; **S1 Table and Fig 2**), which was intermingled in the phylogeny  
386 with the PNG isolates. This sequence was separated by a distance of ~3 SNV from its closest  
387 PNG relative suggestive of transmission of this genotype from PNG to India on at least one  
388 occasion. The other (i.e. non-2.1.7) isolates from clade 2.1 originated almost exclusively  
389 from other islands in the Indonesian archipelago (total n=40 isolates, 90.1%; see **Fig 2**),  
390 suggesting that the PNG 2.1.7 group emerged from a local pathogen population that had been  
391 established in the region for some time.

392

393

394 **Fig 2. Maximum likelihood phylogeny inferred from all available (n=135) clade 2.1 *S.***  
395 ***Typhi*.** Tips are coloured by location of origin as per inset legend (for traveller isolates, this  
396 is country of travel rather than country of isolation), branches are labelled by *S. Typhi*  
397 genotype. Square nodes indicate the position of the reference sequence and other completed  
398 genomes. Phylogeny is outgroup rooted with a selection of non-clade 2.1 taxa (see **Methods**).  
399 Interactive phylogeny available at: [https://microreact.org/project/icGPbF\\_O8](https://microreact.org/project/icGPbF_O8).

400



401

402 ***Genetic variation among PNG genotype 2.1.7 S. Typhi***

403 As described in **Materials and Methods**, we completed three PNG genotype 2.1.7 genomes  
404 to facilitate high resolution analysis of this clone. Three 2.1.7 genomes from PNG *S. Typhi*  
405 were selected for completion; two were the oldest sequences from each sublineage available  
406 (genotype 2.1.7.1 sequence MDUST305 isolated in 1990, and genotype 2.1.7.2 sequence  
407 MDUST348 isolated in 1985). We also completed a recent sequence of the commonest  
408 genotype 2.1.7.1 (sequence MDUST255 isolated in 2010). The three completed genomes  
409 were markedly similar in size, differing by as little as 4.4 kbp (**Fig 3a**). Further, pairwise  
410 nucleotide sequence comparisons revealed high levels of genetic similarity (99-100%  
411 coverage, 100% identity), and evidence of genome plasticity with multiple large-scale  
412 inversions observed adjacent to repetitive ribosomal RNA (*rrn*) operons (**Fig 3b**).

413

414

415 **Fig 3. Genome plasticity among PNG *S. Typhi*. (A) Summary of genetic features of**  
416 **completed PNG 2.1.7 *S. Typhi* genomes. (B) Pairwise nucleotide alignments of PNG**  
417 **2.1.7 *S. Typhi*. Shaded regions indicate nucleotide sequence homology, red blocks indicate**  
418 **ribosomal RNA (rRNA) operons, and blue blocks indicate the ST35-like prophage as per the**  
419 **inset legend. (C) Pairwise alignment of prophage ST35-like elements in PNG *S. Typhi***  
420 **and the CT18 reference sequence. Shaded regions indicate nucleotide sequence homology,**  
421 **and genes are coloured by putative function as per the inset legend.**

422

423

424 Gene content was highly conserved among both PNG 2.1.7 sublineages (96-100% coverage;  
425 **S4 Fig**). Very few deletions were observed, and were not restricted to a particular sublineage.

426 Three sequences, two of genotype 2.1.7.1 (MDUST329 and UJ 212G) and one of 2.1.7. 2  
427 (MDUST332) appeared to have lost the *Salmonella* pathogenicity island 7 (SPI7) locus  
428 known to encode genes for Vi polysaccharide capsule formation, type IVB pilli, and  
429 bacteriophage SopE. Such deletions are known to occur in the laboratory and did not appear  
430 to affect the phage types previously determined for the PNG 2.1.7 traveller isolates as all  
431 were of phage type D2. Besides the deletion of prophage SopE within SPI7, no other  
432 prophage deletions were apparent.

433

434 Examination of detected prophage elements amongst the three completed prophage  
435 sequences revealed a smaller ~1.8 kbp inversion within an ST35-like prophage in the genome  
436 sequence of MDUST255 (**Fig 3c**). The inversion was located in the prophage tail gene  
437 module carrying homologues of prophage ST35 genes (from reference sequence CT18)  
438 encoding for a tail fibre (STY3691), tail fibre assembly chaperone (STY3692), alternate tail  
439 fibre tips (STY3693 and STY3694), a site-specific invertase (STY3695), and major tail tube  
440 filament structural gene (STY3696)[50]. The inversion, likely mediated by the site-specific  
441 invertase (STY3695), appears to have resulted in the formation of an alternative tail fibre  
442 protein; a fusion of the N-terminal region of tail fibre protein STY3691 with alternative tail  
443 fibre tip protein STY4694. Examination of alignments of read data to reference sequence  
444 MDUST348 suggested that this locus was present in both orientations across the PNG 2.1.7  
445 population in line with previous reports[50].

446

447

## 448 **Discussion**

449 These findings demonstrate that the vast majority of typhoid fever cases in PNG are caused  
450 by a single lineage of *S. Typhi* (genotype 2.1.7) which emerged in the Indonesian archipelago

451 in the mid-twentieth century before differentiating into two sublineages in the mid-late 1970s,  
452 with genotype 2.1.7.2 no longer appearing common after 1998. These findings correlate well  
453 with the broad epidemiology of typhoid fever in PNG. Typhoid fever was sporadically  
454 reported in PNG in the 1960s, becoming more commonly diagnosed in the 1970s and then  
455 becoming endemic in the PNG highlands in the 1980s [16,58]. Thus, the divergence dates for  
456 PNG *S. Typhi* 2.1.7 sublineages correspond to the approximate time when typhoid fever  
457 became endemic. *S. Typhi* genotype 2.1.7 appears to have persisted within PNG since this  
458 time, and was dominant among both returning travellers and locally detected cases over the  
459 course of 30 years of genomic surveillance. We observed little evidence of genetic variation  
460 among the genotype 2.1.7 population, with infrequent gene deletions (**S4 Fig**), no obvious  
461 evidence of prophage or plasmid acquisitions (**Fig 3** and **S4 Fig**), and a lower substitution  
462 rate compared to the widespread H58 (4.3.1) genotypes associated with intercontinental  
463 transmission and AMR[8,9,57]. Notably, our data suggest that AMR has not emerged among  
464 the *S. Typhi* population in this setting, with both genotype and phenotype data revealing all  
465 isolates to be sensitive to former first line drugs (chloramphenicol, ampicillin and co-  
466 trimoxazole), as well as fluoroquinolones, third generation cephalosporins, and macrolides.  
467 The AMR sensitivity of the PNG 2.1.7 genotype and the long term stability of this population  
468 is in sharp contrast to many other endemic settings, e.g. South Asia where infections are  
469 commonly driven by H58 genotypes[9], both MDR and reduced susceptibility to  
470 fluoroquinolones are common [7,8,59,60], and resistance to third generation  
471 cephalosporins[10,11], and azithromycin [12–15,61] have also emerged.

472

473 We observed limited evidence of inter-country transmission, with no obvious recent  
474 introductions into PNG from outside. A small number of genotype 2.1.7 isolates were found  
475 outside PNG in India and Vietnam, but these were mostly distantly related to the PNG

476 lineages and we observed only one possible case of transmission of PNG genotype 2.1.7 *S.*  
477 Typhi into India (travel-associated case of lineage 2.1.7.1 isolated in 1988 **Fig 2**). It is  
478 possible, however, that this individual may also have visited PNG as well as India in their  
479 travels with only the later recorded. The limited international movement of *S. Typhi* in and  
480 out of PNG may partially explain why AMR has not been observed in this setting, in the form  
481 of the arrival of H58 or other drug resistant genotypes of *S. Typhi*.

482

483 Among PNG 2.1.7 populations, the lack of emergence of AMR may be influenced by a lack  
484 of local selective pressures related to antimicrobial usage. Tracking antimicrobial usage is  
485 challenging in PNG, but evidence of sub-optimal administration of antibiotics has been  
486 reported[62]. Approximately 73% of outpatients with non-malarial febrile illness are  
487 administered antibiotics, commonly contrary to the relevant prescription guidelines.  
488 Amoxicillin, trimethoprim-sulfamethoxazole and benzylpenicillin collectively accounted for  
489 ~75% of antibiotic administrations, demonstrating the high dependence on  $\beta$ -lactams in PNG,  
490 but little exposure to cephalosporins. Data pertaining to other Gram-negative pathogens in  
491 PNG are limited. A recent investigation of *Shigella* from PNG and the Pacific region revealed  
492 this pathogen to have resistance to ampicillin, chloramphenicol, tetracycline, and  
493 trimethoprim-sulfamethoxazole, with resistance more common in bacteria isolated since  
494 2010. Resistance to cephalosporins (ceftriaxone) and ciprofloxacin was not detected in  
495 *Shigella* from PNG[63]. On the basis of observations in *Shigella*, resistance to first-line  
496 antibiotics could emerge in other gastrointestinal pathogens in PNG, but exposure, and  
497 therefore resistance to, cephalosporins and second-generation quinolones (e.g. ciprofloxacin)  
498 is unlikely.

499

500 Genome variation reported previously in the PNG *S. Typhi* population [18] appears to be  
501 primarily mediated by large-scale re-arrangements adjacent to repetitive *rrn* operons (**Fig**  
502 **3b**), rather than a diverse range of genotypes or the acquisition of mobile genetic elements  
503 such as prophages or plasmids consistent with previous reports[18,19]. Rearrangements  
504 between the *rrn* operons have been described previously in *S. Typhi* and are driven by  
505 homologous recombination [64,65]. This has previously been postulated to cause sequence  
506 variation (e.g. diverse ribotypes) among PNG *S. Typhi*[19,20], and our data support this  
507 hypothesis. Recent long-read data suggests that rearrangements around the *rrn* operons are  
508 relatively common in *S. Typhi*[66], and appear to be associated with host persistence and  
509 carriage in the gallbladder[67]. We also observed inversions in the tail morphogenesis gene  
510 module of the ST35-like prophage. These inversions are also well described and are likely a  
511 mechanism for generating variation in tail fibre genes that are known to be involved with the  
512 attachment stage of viral replication[68]. Such variation is known to cause alteration of phage  
513 host ranges[50,69].

514

515 This study is not without limitations, primarily relating to the challenges associated with  
516 obtaining clinical isolates of *S. Typhi*. There is no ongoing routine surveillance conducted in  
517 PNG, nor is there routine blood culture for diagnosis of febrile illness. Consequently, locally  
518 collected isolates are obtained primarily through targeted studies. Most isolates analysed in  
519 this study originated from Eastern Highlands Province, where the PNG Institute of Medical  
520 Research is headquartered and is thus the site of most studies. Eastern Highlands Province is  
521 one of 22 provinces in the country (albeit the second most populated province), thus there is a  
522 lack of geographical distribution of isolates. Given most isolates originate from targeted  
523 studies, the temporal distribution of isolates is intermittent. Moreover, the lack of ongoing  
524 surveillance precluded contemporary isolates (from the past 10 year) being collected and thus

525 included in this study. However, as locally collected and travel associated cases both  
526 demonstrated a lack of AMR and similar genotype frequencies, in line with limited  
527 previously published sequences[21], these combined data appear representative of major  
528 trends among the *S. Typhi* population in this setting.

529

530 Overall, our data highlight that unlike other recently examined settings in Kenya, Zimbabwe  
531 and other regions of Sub-Saharan Africa, as well as South Asian settings in India, Nepal,  
532 Bangladesh and Pakistan [8,10,57,59,60,70–73] the *S. Typhi* populations in PNG appear  
533 sensitive to a diverse range of antimicrobials with former first line drugs, fluoroquinolones,  
534 third generation cephalosporins and macrolides all remaining viable treatment options prior  
535 to the introduction of Vi-conjugate vaccines and improvements to water, sanitation and  
536 hygiene (WaSH) related infrastructure. However, despite this, further ongoing surveillance is  
537 required to monitor for the potential emergence or introduction of AMR *S. Typhi* in this  
538 setting.

539

540

#### 541 **Acknowledgements**

542 The authors wish to thank Sebastian Duchene for useful discussions regarding phylodynamic  
543 analyses.

544

#### 545 **Financial Disclosure Statement**

546 ZAD was supported by a grant funded by the Wellcome Trust (STRATAA; 106158/Z/14/Z),  
547 and received funding from the European Union’s Horizon 2020 research and innovation  
548 programme under the Marie Skłodowska-Curie grant agreement TyphiNET (#845681). DJI  
549 was supported by a National Health and Medical Research Council (NHMRC) Emerging

550 Leadership Fellowship (GNT1195210). DAW is supported by an NHMRC Emerging  
551 Leadership Fellowship (GNT1174555). KEH was supported by a Senior Medical Research  
552 Fellowship from the Viertel Foundation of Australia, and the Bill and Melinda Gates  
553 Foundation, Seattle (grant #OPP1175797). This work was supported, in whole or in part, by  
554 the Bill & Melinda Gates Foundation [OPP1175797]. Under the grant conditions of the  
555 Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned  
556 to the Author Accepted Manuscript version that might arise from this submission. The  
557 funders had no role in study design, data collection and analysis, decision to publish, or  
558 preparation of the manuscript.

559

560

#### 561 **Data Availability**

562 Genomic data are submitted to the European Nucleotide Archive (ENA) and accession  
563 numbers are available in the supplementary tables.

564

565

#### 566 **References**

567 1. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. *N Engl J Med*.

568 2002;347: 1770–1782. doi:10.1056/nejmra020201

569 2. Collaborators G 2017 T and P. The global burden of typhoid and paratyphoid fevers: a

570 systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious*

571 *Diseases*. 2019;19: 369–381. doi:10.1016/s1473-3099(18)30685-6

572 3. Connor BA, Schwartz E. Typhoid and paratyphoid fever in travellers. *The Lancet*

573 *Infectious Diseases*. 2005;5: 623–628. doi:10.1016/s1473-3099(05)70239-5

- 574 4. Butler T, Knight J, Nath SK, Speelman P, Roy SK, Azad MA. Typhoid fever complicated  
575 by intestinal perforation: a persisting fatal disease requiring surgical management. *Rev Infect*  
576 *Dis.* 1985;7: 244–256. doi:10.1093/clinids/7.2.244
- 577 5. Crump JA. Progress in Typhoid Fever Epidemiology. *Clin Infect Dis.* 2019;68: S4–S9.  
578 doi:10.1093/cid/ciy846
- 579 6. Dyson ZA, Klemm EJ, Palmer S, Dougan G. Antibiotic Resistance and Typhoid. *Clin*  
580 *Infect Dis.* 2019;68: S165–S170. doi:10.1093/cid/ciy1111
- 581 7. Thanh DP, Karkey A, Dongol S, Thi NH, Thompson CN, Rabaa MA, et al. A novel  
582 ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone  
583 treatment failure. *Elife.* 2016;5: e14003. doi:10.7554/elife.14003
- 584 8. Britto CD, Dyson ZA, Duchene S, Carter MJ, Gurung M, Kelly DF, et al. Laboratory and  
585 molecular surveillance of paediatric typhoidal Salmonella in Nepal: Antimicrobial resistance  
586 and implications for vaccine policy. *PLoS Negl Trop Dis.* 2018;12: e0006408.  
587 doi:10.1371/journal.pntd.0006408
- 588 9. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al. Phylogeographical  
589 analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter-  
590 and intracontinental transmission events. *Nat Genet.* 2015;47: 632–639. doi:10.1038/ng.3281
- 591 10. Klemm EJ, Shakoor S, Page AJ, Qamar FN, Judge K, Saeed DK, et al. Emergence of an  
592 Extensively Drug-Resistant Salmonella enterica Serovar Typhi Clone Harboring a  
593 Promiscuous Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation  
594 Cephalosporins. *mBio.* 2018;9. doi:10.1128/mbio.00105-18



- 595 11. Rasheed F, Saeed M, Alikhan N-F, Baker D, Khurshid M, Ainsworth EV, et al.  
596 Emergence of Resistance to Fluoroquinolones and Third-Generation Cephalosporins in  
597 Salmonella Typhi in Lahore, Pakistan. *Microorganisms*. 2020;8.  
598 doi:10.3390/microorganisms8091336
- 599 12. Hooda Y, Sajib MSI, Rahman H, Luby SP, Bondy-Denomy J, Santosham M, et al.  
600 Molecular mechanism of azithromycin resistance among typhoidal Salmonella strains in  
601 Bangladesh identified through passive pediatric surveillance. *PLoS Negl Trop Dis*. 2019;13:  
602 e0007868. doi:10.1371/journal.pntd.0007868
- 603 13. Carey ME, Jain R, Yousuf M, Maes M, Dyson ZA, Thu TNH, et al. Spontaneous  
604 Emergence of Azithromycin Resistance in Independent Lineages of Salmonella Typhi in  
605 Northern India. *Clin Infect Dis*. 2021;72: e120–e127. doi:10.1093/cid/ciaa1773
- 606 14. Duy PT, Dongol S, Giri A, Nguyen TNT, Thanh HND, Nguyen QNP, et al. The  
607 emergence of azithromycin-resistant Salmonella Typhi in Nepal. *JAC-Antimicrobial*  
608 *Resistance*. 2020;2. doi:10.1101/2020.08.07.20166389
- 609 15. Iqbal J, Dehraj IF, Carey ME, Dyson ZA, Garrett D, Seidman JC, et al. A Race against  
610 Time: Reduced Azithromycin Susceptibility in Salmonella enterica Serovar Typhi in  
611 Pakistan. *mSphere*. 2020;5. doi:10.1128/msphere.00215-20
- 612 16. Passey M. The new problem of typhoid fever in Papua New Guinea: how do we deal with  
613 it? *1995*;38: 300–304. Available: <http://europepmc.org/abstract/MED/9522872>
- 614 17. Siba V, Horwood PF, Vanuga K, Wapling J, Sehuko R, Siba PM, et al. Evaluation of  
615 serological diagnostic tests for typhoid fever in Papua New Guinea using a composite  
616 reference standard. *Clin Vaccine Immunol*. 2012;19: 1833–1837. doi:10.1128/cvi.00380-12

- 617 18. Combs BG, Passey M, Michael A, Pang T, Lightfoot D, Alpers MP. Ribotyping of  
618 *Salmonella enterica* serovar Typhi isolates from Papua New Guinea over the period 1977 to  
619 1996. *Papua and New Guinea medical journal*. 2005;48: 158–167. Available:  
620 [http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17212062&retmo](http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17212062&retmode=ref&cmd=prlinks)  
621 [de=ref&cmd=prlinks](http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17212062&retmode=ref&cmd=prlinks)
- 622 19. Thong K-L, Goh Y-L, Yasin RM, Lau MG, Passey M, Winston G, et al. Increasing  
623 genetic diversity of *Salmonella enterica* serovar typhi isolates from papua new guinea over  
624 the period from 1992 to 1999. *Journal of Clinical Microbiology*. 2002;40: 4156–4160.  
625 doi:10.1128/jcm.40.11.4156-4160.2002
- 626 20. Thong KL, Passey M, Clegg A, Combs BG, Yassin RM, Pang T. Molecular analysis of  
627 isolates of *Salmonella typhi* obtained from patients with fatal and nonfatal typhoid fever.  
628 *Journal of Clinical Microbiology*. 1996;34: 1029–1033. doi:10.1128/jcm.34.4.1029-  
629 1033.1996
- 630 21. Baddam R, Thong K-L, Avasthi TS, Shaik S, Yap K-P, Teh CSJ, et al. Whole-genome  
631 sequences and comparative genomics of *Salmonella enterica* serovar Typhi isolates from  
632 patients with fatal and nonfatal typhoid fever in Papua New Guinea. *Journal of Bacteriology*.  
633 2012;194: 5122–5123. doi:10.1128/jb.01051-12
- 634 22. UNICEF. At a glance: Papua New Guinea. Port Moresby. Available:  
635 <https://www.unicef.org/png>
- 636 23. UNICEF. Papua New Guinea. Available: [https://www.unicef.org/png/what-we-do/water-](https://www.unicef.org/png/what-we-do/water-sanitation-and-hygiene)  
637 [sanitation-and-hygiene](https://www.unicef.org/png/what-we-do/water-sanitation-and-hygiene)

- 638 24. UNDP. Papua New Guinea. Available:  
639 [http://hdr.undp.org/sites/all/themes/hdr\\_theme/country-notes/PNG.pdf](http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/PNG.pdf)
- 640 25. Williamson DA, Lane CR, Easton M, Valcanis M, Strachan J, Veitch MG, et al.  
641 Increasing Antimicrobial Resistance in Nontyphoidal Salmonella Isolates in Australia from  
642 1979 to 2015. *Antimicrob Agents Chemother.* 2018;62. doi:10.1128/aac.02012-17
- 643 26. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with  
644 multiplex MinION sequencing. *Microbial Genomics.* 2017;3: e000132.  
645 doi:10.1099/mgen.0.000132
- 646 27. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended  
647 genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid.  
648 *Nat Commun.* 2016;7: 12827. doi:10.1038/ncomms12827
- 649 28. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete  
650 genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.  
651 *Nature.* 2001;413: 848–852. doi:10.1038/35101607
- 652 29. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth.* 2012;9:  
653 357–359. doi:10.1038/nmeth.1923
- 654 30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence  
655 Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25: 2078–2079.  
656 doi:10.1093/bioinformatics/btp352
- 657 31. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, et al. High-  
658 throughput sequencing provides insights into genome variation and evolution in Salmonella  
659 Typhi. *Nat Genet.* 2008;40: 987–993. doi:10.1038/ng.195

- 660 32. Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA, Day M, et al. Informal genomic  
661 surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial  
662 resistance via returning travellers. Akullian A, editor. PLoS Negl Trop Dis. 2019;13:  
663 e0007620. doi:10.1371/journal.pntd.0007620
- 664 33. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid  
665 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences  
666 using Gubbins. Nucleic Acids Research. 2015;43: e15–e15. doi:10.1093/nar/gku1196
- 667 34. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
668 large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033
- 669 35. Dyson ZA, Holt KE. Five years of GenoTyphi: updates to the global Salmonella Typhi  
670 genotyping framework. The Journal of Infectious Diseases. 2021. doi:10.1093/infdis/jiab414
- 671 36. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in  
672 large sequence sets. Curr Protoc Bioinformatics. 2003;Chapter 10: Unit 10.3.  
673 doi:10.1002/0471250953.bi1003s00
- 674 37. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster  
675 version of the PHAST phage search tool. Nucleic Acids Research. 2016;44: W16-21.  
676 doi:10.1093/nar/gkw387
- 677 38. Consortium IT, Wong VK, Holt KE, Okoro C, Baker S, Pickard DJ, et al. Molecular  
678 Surveillance Identifies Multiple Transmissions of Typhoid in West Africa. PLoS Negl Trop  
679 Dis. 2016;10: e0004781. doi:10.1371/journal.pntd.0004781

- 680 39. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of  
681 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016;2:  
682 vew007. doi:10.1093/ve/vew007
- 683 40. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a  
684 software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10:  
685 e1003537. doi:10.1371/journal.pcbi.1003537
- 686 41. Duchene S, Holt KE, Weill F-X, Hello SL, Hawkey J, Edwards DJ, et al. Genome-scale  
687 rates of evolutionary change in bacteria. *Microbial Genomics.* 2016;2: e000094.  
688 doi:10.1099/mgen.0.000094
- 689 42. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-  
690 structured data to estimate evolutionary rates of double-stranded DNA viruses. *Molecular*  
691 *Biology and Evolution.* 2010;27: 2038–2051. doi:10.1093/molbev/msq088
- 692 43. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an rpackage for visualization and  
693 annotation of phylogenetic trees with their covariates and other associated data. McInerny G,  
694 editor. *Methods Ecol Evol.* 2016;8: 28–36. doi:10.1111/2041-210x.12628
- 695 44. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact:  
696 visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial*  
697 *Genomics.* 2016;2: e000093. doi:10.1099/mgen.0.000093
- 698 45. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome  
699 assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13: e1005595.  
700 doi:10.1371/journal.pcbi.1005595

- 701 46. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo  
702 genome assemblies. *Bioinformatics*. 2015;31: 3350–3352. doi:10.1093/bioinformatics/btv383
- 703 47. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:  
704 2068–2069. doi:10.1093/bioinformatics/btu153
- 705 48. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved  
706 genomic sequence with rearrangements. *Genome Research*. 2004;14: 1394–1403.  
707 doi:10.1101/gr.2289704
- 708 49. Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome  
709 visualization in R. *Bioinformatics*. 2010;26: 2334–2335. doi:10.1093/bioinformatics/btq413
- 710 50. Thomson N, Baker S, Pickard D, Fookes M, Anjum M, Hamlin N, et al. The role of  
711 prophage-like elements in the diversity of *Salmonella enterica* serovars. *J Mol Biol*.  
712 2004;339: 279–300. doi:10.1016/j.jmb.2004.03.058
- 713 51. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open  
714 Software Suite. *Trends Genet*. 2000;16: 276–277. doi:10.1016/s0168-9525(00)02024-2
- 715 52. Otto TD, Dillon GP, Degraeve WS, Berriman M. RATT: Rapid Annotation Transfer Tool.  
716 *Nucleic Acids Research*. 2011;39: e57. doi:10.1093/nar/gkq1268
- 717 53. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid  
718 genomic surveillance for public health and hospital microbiology labs. *Genome Med*. 2014;6:  
719 90. doi:10.1186/s13073-014-0090-6

- 720 54. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al.  
721 ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial  
722 genomes. *Antimicrob Agents Chemother.* 2014;58: 212–220. doi:10.1128/aac.01310-13
- 723 55. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, et al. In silico  
724 detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence  
725 typing. *Antimicrob Agents Chemother.* 2014;58: 3895–3903. doi:10.1128/aac.02412-14
- 726 56. Dyson ZA, Holt KE. *katholt/genotyphi: GenoTyphi v1.9.1*. Zenodo. 2021.  
727 doi:10.5281/zenodo.4707614
- 728 57. Kariuki S, Dyson ZA, Mbae C, Ngetich R, Kawai SM, Wairimu C, et al. Multiple  
729 introductions of multidrug-resistant typhoid associated with acute infection and  
730 asymptomatic carriage, Kenya. [“Lewnard, Joseph”], editors. *Elife.* 2021;10: e67852.  
731 doi:10.7554/elife.67852
- 732 58. Richens J. Typhoid in the highlands of Papua New Guinea 1984-1990: a hospital-based  
733 perspective. *Papua and New Guinea medical journal.* 1995;38: 305–314. Available:  
734 <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9522873&retmod>  
735 [e=ref&cmd=prlinks](http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9522873&retmod=e=ref&cmd=prlinks)
- 736 59. Britto CD, Dyson ZA, Mathias S, Bosco A, Dougan G, Jose S, et al. Persistent circulation  
737 of a fluoroquinolone-resistant *Salmonella enterica* Typhi clone in the Indian subcontinent. *J*  
738 *Antimicrob Chemother.* 2020;75: 337–341. doi:10.1093/jac/dkz435
- 739 60. Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE, Chowdhury EK, et al.  
740 Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in

- 741 urban Dhaka, Bangladesh from 2004 to 2016. *PLoS Negl Trop Dis*. 2020;14: e0008036.  
742 doi:10.1371/journal.pntd.0008036
- 743 61. Sajib MSI, Tanmoy AM, Hooda Y, Rahman H, Andrews JR, Garrett DO, et al. Tracking  
744 the Emergence of Azithromycin Resistance in Multiple Genotypes of Typhoidal Salmonella.  
745 *mBio*. 2021;12. doi:10.1128/mbio.03481-20
- 746 62. Saweri OPM, Hetzel MW, Mueller I, Siba PM, Pulford J. The treatment of non-malarial  
747 febrile illness in Papua New Guinea: findings from cross sectional and longitudinal studies of  
748 health worker practice. *BMC Health Serv Res*. 2017;17: 10–10. doi:10.1186/s12913-016-  
749 1965-6
- 750 63. Malau E, Ford R, Valcanis M, Jennison AV, Mosse J, Bean D, et al. Antimicrobial  
751 sensitivity trends and virulence genes in *Shigella* spp. from the Oceania region. *Infect Genet*  
752 *Evol*. 2018;64: 52–56. doi:10.1016/j.meegid.2018.06.015
- 753 64. Brüßow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens:  
754 from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*. 2004;68:  
755 560-602-table of contents. doi:10.1128/mubr.68.3.560-602.2004
- 756 65. Sanderson KE, Liu SL. Chromosomal rearrangements in enteric bacteria. *Electrophoresis*.  
757 1998;19: 569–572. doi:10.1002/elps.1150190417
- 758 66. Page AJ, Ainsworth EV, Langridge GC. socru: typing of genome-level order and  
759 orientation around ribosomal operons in bacteria. *Microbial Genomics*. 2020;6.  
760 doi:10.1099/mgen.0.000396



- 761 67. Matthews TD, Rabsch W, Maloy S. Chromosomal rearrangements in *Salmonella enterica*  
762 serovar Typhi strains isolated from asymptomatic human carriers. *mBio*. 2011;2: e00060-11.  
763 doi:10.1128/mbio.00060-11
- 764 68. Guttman B, Raya R, Kutter E. Basic Phage Biology. In: Kutter E, Sulakvelidze A, editors.  
765 Bacteriophages Biology and Applications . Bacteriophages Biology and Applications; 2005.
- 766 69. Kutter E, Raya R, Carlson K. Molecular mechanisms of phage infection. In: Kutter E,  
767 Sulakvelidze A, editors. Bacteriophages Biology and Applications . Bacteriophages Biology  
768 and Applications; 2005. pp. 165–222.
- 769 70. Mashe T, Leekitcharoenphon P, Mtapuri-Zinyowera S, Kingsley RA, Robertson V,  
770 Tarupiwa A, et al. *Salmonella enterica* serovar Typhi H58 clone has been endemic in  
771 Zimbabwe from 2012 to 2019. *J Antimicrob Chemother*. 2020. doi:10.1093/jac/dkaa519
- 772 71. Pragasam AK, Pickard D, Wong V, Dougan G, Kang G, Thompson A, et al. Phylogenetic  
773 Analysis Indicates a Longer Term Presence of the Globally Distributed H58 Haplotype of  
774 *Salmonella Typhi* in Southern India. *Clin Infect Dis*. 2020;71: 1856–1863.  
775 doi:10.1093/cid/ciz1112
- 776 72. Park SE, Pham DT, Boinett C, Wong VK, Pak GD, Panzner U, et al. The phylogeography  
777 and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. *Nat Commun*.  
778 2018;9: 5094. doi:10.1038/s41467-018-07370-z
- 779 73. Silva KE da, Tanmoy AM, Pragasam AK, Iqbal J, Sajib MSI, Mutreja A, et al. The  
780 international and intercontinental spread and expansion of antimicrobial-resistant  
781 &em&g;Salmonella&lt;/em&g; Typhi. *medRxiv*. 2021; 2021.09.03.21262852. Available:  
782 <http://medrxiv.org/content/early/2021/09/07/2021.09.03.21262852.abstract>

783

784

785 **Supporting Information Captions**

786 **S1 Fig. Maximum likelihood phylogeny for all available *S. Typhi* originating in PNG.**

787 Tip colours indicate the source of the sequence as per the inset legend. Branches are labelled

788 by *S. Typhi* genotype. Square nodes indicate the position of the reference sequence and other

789 completed genomes.

790

791 **S2 Fig. Maximum likelihood phylogeny for all available PNG genotype 2.1.7 *S. Typhi*.**

792 Tip colours indicate the source of the sequence as per the inset legend, and branches are

793 labelled by genotype. Square nodes indicate the position of the reference sequence and other

794 completed genomes. A Bayesian dated tree inferred using the same alignment and year of

795 isolation is shown in **Fig 1**.

796

797 **S3 Fig. Temporal Analysis of PNG genotype 2.1.7 *S. Typhi*. (A) Tempest regression of**

798 **root-to-tip distances as a function of time**, with the root of the tree selected using heuristic

799 residual mean squared. Each point represents a tip in the maximum likelihood phylogenetic

800 tree shown in **S2 Fig**. The slope is a crude estimate of the annual substitution rate for the

801 SNV alignment, the x-intercept corresponds to the age of the root node, and  $R^2$  is a measure

802 of clock-like behaviour among the data. **(B) Date randomisation test results** with the right

803 most box plot showing the posterior substitution rate estimate from the SNV alignment of the

804 data with the correct sampling times, and the remaining 20 boxplots showing the posterior

805 distributions of the rate estimate from replicate runs where the dates were subjected to

806 randomisation. The data are considered to have strong temporal structure if the estimates

807 using the correct dates do not overlap with those where the dates were randomised.

808

809 **S4 Fig. Genetic conservation among PNG 2.1.7 *S. Typhi*.** Branches are labelled by  
810 genotype. Coloured square nodes indicate the position of the reference sequence and other  
811 completed genomes. Heatmap shows the presence or absence of all annotated genes in the  
812 reference sequence as per the inset legend. Features of interest are highlighted with red boxes  
813 and labels.

814

815 **S1 Table. Accession numbers and data for *S. Typhi* sequences used in this study (Excel**  
816 **file)**

817

818 **S2 Table. Excluded repeat and phage regions in PNG MDUST348 2.1.7.2 completed**  
819 **reference sequence (Excel file)**

820

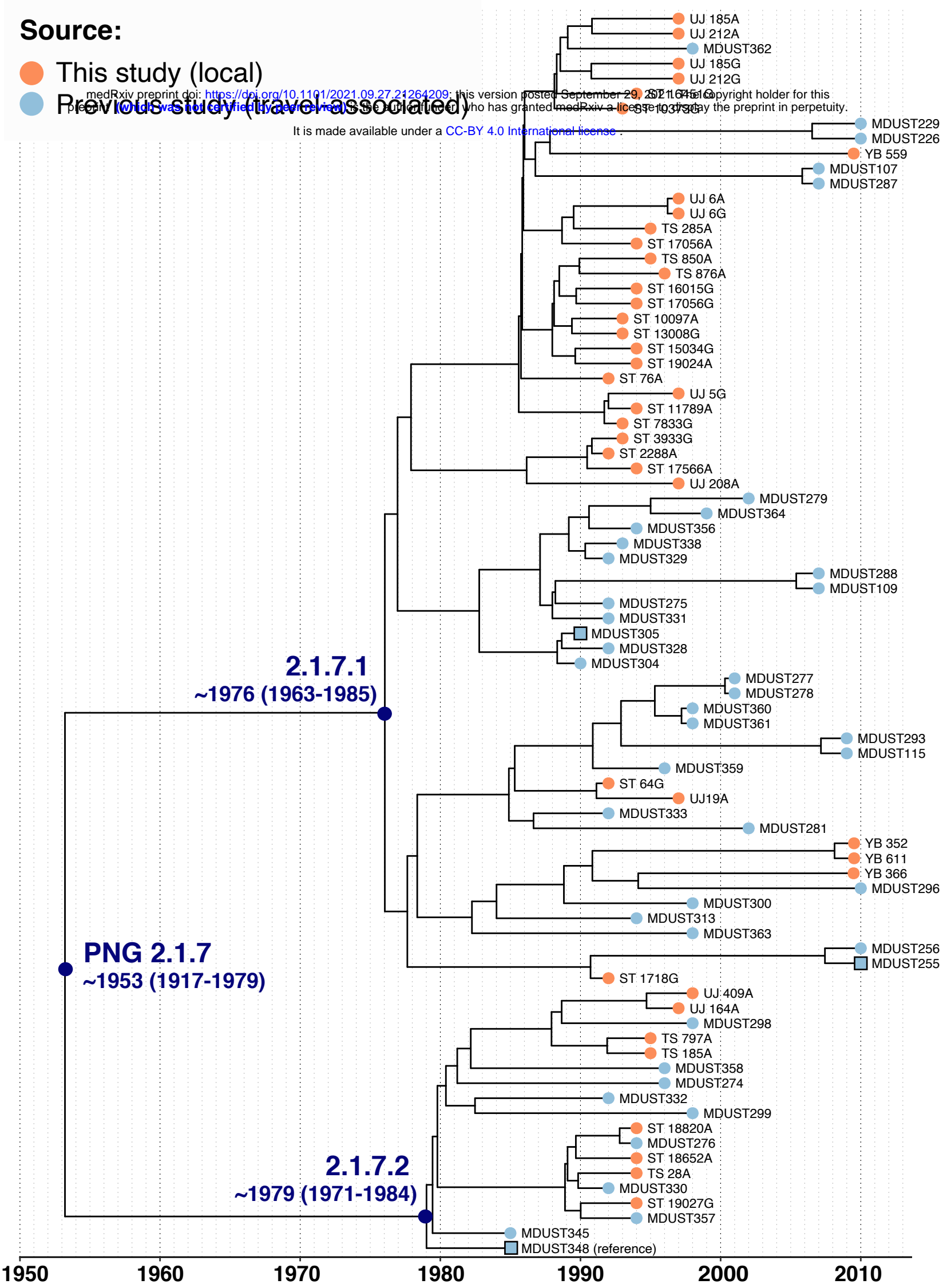
821 **S3 Table. Outgroups used for phylogenetic tree rooting (Excel file)**

# Source:

● This study (local)

● Previous study (travel-associated)

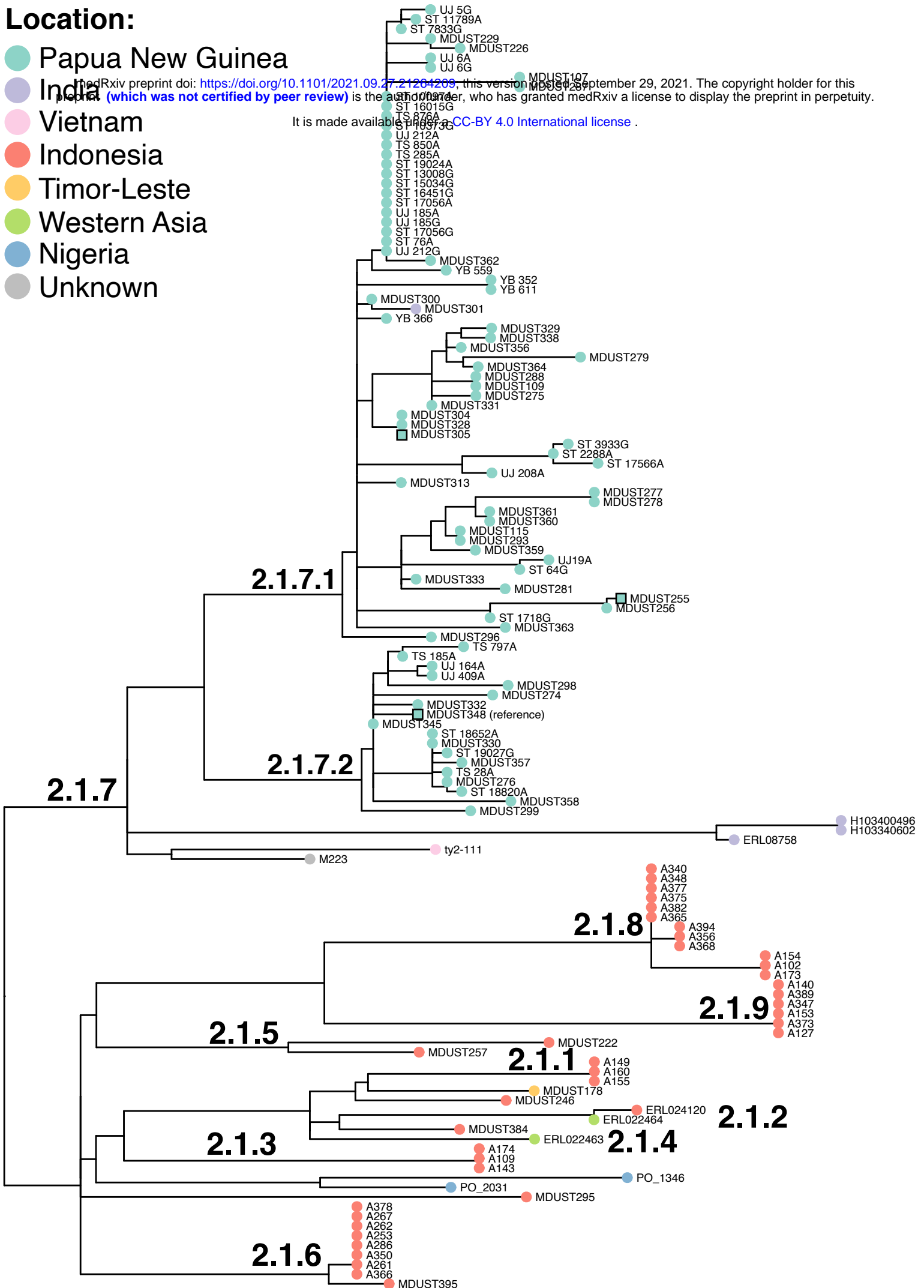
medRxiv preprint doi: <https://doi.org/10.1101/2021.09.27.21264209>; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



# Location:

- Papua New Guinea
- India
- Vietnam
- Indonesia
- Timor-Leste
- Western Asia
- Nigeria
- Unknown

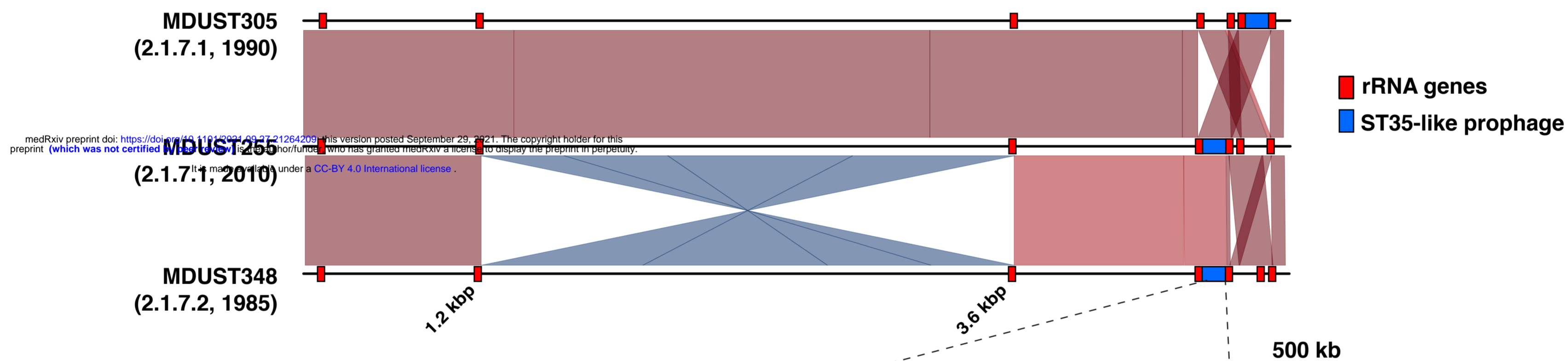
medRxiv preprint doi: <https://doi.org/10.1101/2021.09.27.21264209>; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



0.008

**A**

Sequence	Year	Sample type	Genotype	Genome size (bp)	G+C mol %	Prophage regions	Phage type
MDUST305	1990	Blood	2.1.7.1	4,796,602	52.04	6	D2
MDUST255	2010	Blood	2.1.7.1	4,797,669	52.04	6	D2
MDUST348	1985	Unknown	2.1.7.2	4,793,333	52.04	6	D2

**B****C**